Springer Undergraduate Texts in Mathematics and Technology

More information about this series at http://www.springer.com/series/7438

Jay S. Treiman

Calculus with Vectors

Springer

Jay S. Treiman
Department of Mathematics
Western Michigan University
Kalamazoo, MI, USA

Additional material to this book can be downloaded from http://extras.springer.com

ISSN 1867-5506 ISSN 1867-5514 (electronic)
ISBN 978-3-319-35424-8 ISBN 978-3-319-09438-0 (eBook)
DOI 10.1007/978-3-319-09438-0
Springer Cham Heidelberg New York Dordrecht London

Mathematics Subject Classifications (2010): 97I40, 97I50, 97I30

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book grew out of a need for a text to use in a calculus class that is intended for students going into science, engineering, and mathematics. The course includes early vectors and early transcendentals to help prepare students for their calculus-based physics classes. There are a number of books that have tried to fill this need. Unfortunately, none of the books written for this audience have made it to a second edition, even though they all need significant revisions. In addition, some of my colleagues were lamenting the number of typographical errors in the text we were using. Since the only book that was being updated for this type of course was a version of a regular calculus text with a small amount of material on vectors and vector-valued functions tacked onto a few early chapters, I felt that a new text was needed that could be revised as appropriate.

There were a few things that I felt were important to help our students prepare for their first physics course that are often not in a first calculus course. These include vectors, a rigorous but not overly formal approach, examples and applications that they will see again, the early introduction of transcendental functions and using notation that is correct, but that does not always match what they have seen.

It has been an adventure thinking about how to approach some things so that they may be easier for students to get some basic understanding. One of these things is a change from the common ε-δ approach to limits to a sequence approach to limits. This change is partially motivated by the fact that the intuition given in most calculus texts coincides more closely with the sequence definition of limits of functions as opposed to the ε-δ definition of the limit of a function. The use of sequences for limits is also meant to prepare students for the chapter on series where everything depends on the limits of partial sums. Among the other things that do not follow the "standard" approach are the derivation of the derivatives of sine and cosine, leaving explicit discussion about the connection between increasing and decreasing functions and derivatives until the mean value theorem is available, rethinking the use of "tables" of integration and reordering the techniques of integration to allow the use of partial fractions for trigonometric integrals.

Including the use of technology has been a challenge since I use technology, a CAS calculator or Maple, as a normal part of my teaching. Since the tools available vary and may change rapidly, I did not want too much information that is specific to any software or hardware platform. On the other hand, there are large numbers of people teaching calculus who do not want to use technology and large numbers of people who are wedded to a specific technology. The science and engineering students will be using technology in almost all of their later classes. Given these facts, I do not have many examples of the specific use of technology. The major exception is in the section on integration tables. There is an example of how different CAS systems may give differing results. In the text, there are a good number of problems and examples where technology is required and the materials do not avoid questions involving technology.

One consequence of the availability of technology is that students can easily find quantities such as sums of vectors, dot and cross products, derivatives, integrals, tangent lines, equations of planes, and volumes of revolution using current computer algebra systems. Since students now have the tools to find correct answers to many of the problems in this book, including answers to “drill” problems does not seem important. With access to the Internet, I am uncertain if there are many problems for which one cannot find a solution to an almost identical problem on some web site. I am fairly certain that solutions to any “new” problem in a calculus book will appear online within a few years. Because of this access to answers and solutions for calculus problems, I see very little need to include answers or solutions for exercises in this book.

In the United States, many of our students have gaps in the background expected for calculus. These missing pieces include a lack of familiarity with trigonometric functions and poor algebra skills. Several sections in the first appendix are included to help with some of the pieces, including some basic algebra review and some basic trigonometry.

A large number of proofs are not presented in the text, or are only presented in part. Many of the results are partially justified or the proofs are done as derivations rather than formal proofs. For many mathematics teachers and students, this is not enough. Since there are times when I would like to include certain formal proofs but I do not feel that they will help the readability of this book, some formal proofs are included materials that will be available online.

Some additional material for the text, including Maple worksheets with rotatable 3D graphs that can be viewed using Maple Player from Maplesoft, can be found at http://extras.springer.com. I will always appreciate being notified of all errata at jay.treiman@wmich.edu. As I collect more proofs, problems, and other material, I will post them on my homepage, http://homepages.wmich.edu/~treiman.

I wish to thank all those people who have helped me learn to teach and those who have helped with this book. Among those who have given me examples of mathematics teaching to strive toward are G. D. Chakerian, Virginia M. Warfield, Isaac Namioka, and R. T. Rockafellar. My thanks go to all of them. I also wish to thank those who have helped me with this book. They include Yuri Ledyaev, who made it clear that there are different ways to approach topics in calculus; Daniela Hernandez, who found numerous typos and other problems in the text; Christine Horsmon; Dennis Pence; and all of the students who have suffered through my errors. I am also indebted to Jonathan Borwein for years of encouragement and help, including his help getting this book published.

The reviewers made many useful suggestions and comments that have made this book better and deserve my thanks. Ms. Elizabeth Loew has been a helpful editor and I wish to thank her.

Finally, I would like to thank my wife, Janice Selden, for her support.

Kalamazoo, USA Jay S. Treiman

Contents

Chapter 1
Points and Vectors

Almost all physical situations involve more than one quantity that depends on time or on position. One physical system that has been widely studied for many years is our solar system. Consider only the sun, the earth, and Jupiter and assume the sun position is fixed. Since the orbits of the earth and Jupiter are not in the same plane, the positions of the sun, the earth and Jupiter cannot be put into a single plane. This means the positions of the two planets must be considered in three dimensions. For two movable objects that gives six position variables that depend on time. We will write these six variables as a 6-tuple of numbers $(x_1, x_2, x_3, x_4, x_5, x_6)$ and call these ordered sequences of numbers vectors. In this course vectors are used to work with ordered lists of variables.

The purpose of this chapter is to introduce vectors, their properties, and computations with vectors. Even though most of the calculus you will learn in the first two semesters of calculus can be done in terms of functions with one input and one output, almost all of the applications of calculus involve multiple variables. A simple example is giving the position of a ship on the ocean. This requires two variables, latitude and longitude. If one wants to give the position of a ship on a trip from Hong Kong to London in terms of time, there is one input, time, and there are two outputs, latitude and longitude. We will call such a function a vector-valued function. This type of function is used throughout these notes.

1.1 Points

We start with how to specify position in terms of rectangular coordinates. You should have already seen this when you plotted points for a scatter plot of data or plotted points for the graph of a function.

To plot a point in a plane one chooses a point, the *origin* and two perpendicular lines that intersect at the origin. These lines are usually horizontal, the x-axis, and vertical, the y-axis, see Fig. 1.1a on page 2.

We assign the coordinates $(0,0)$ to the origin. The axes are labeled with signed distances from the origin. The x values increase toward the right and the y values increase toward the top. A point $\mathbf{x}$ in the plane is designated by its *coordinates*, an x value and a y value. To get to the point $(1,-2)$ one goes one unit in the positive x direction, right, and then two units in the

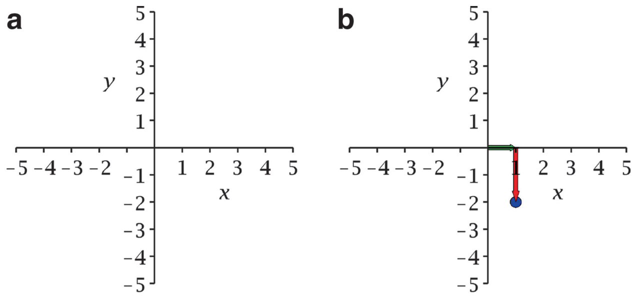
---

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_1) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_1

negative y direction, down. See Fig. 1.1b on page 2. We can also get to the point $(1,-2)$ by first going down two units and then to the right one unit.

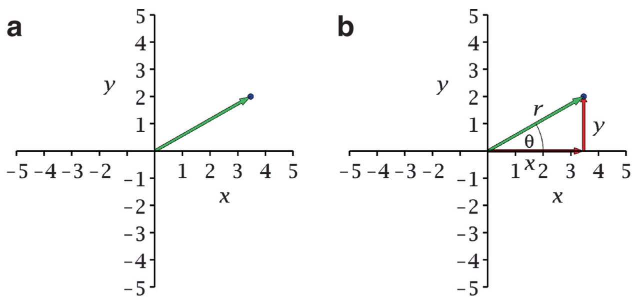
There is another way to describe a point in the plane. We can start at the origin and move in a given direction a given distance. For example, we could move four units in the direction $\pi/6$, (or $30°$), up from the x-axis, see Fig. 1.2a on page 2.



Axes in the plane

A point in the plane

Fig. 1.1



The polar representation of a point

The polar and rectangular representations of a point

Fig. 1.2

The convention for describing points this way is to use an ordered pair of numbers (r,θ) where r is the distance from the origin to the point and θ is the counter clockwise angle from the positive x-axis to the desired direction. These are called the *polar coordinates* of the point. When it may be unclear that a point is specified in polar coordinates or rectangular, we will use the notation $\mathbf{x}_P$ to indicate the point is in polar coordinates and $\mathbf{x}_R$ to indicate the point is in rectangular coordinates. Since almost all points, and vectors, will be in rectangular coordinates, we will almost always use the notation $\mathbf{x}$ to denote a point or a vector in rectangular coordinates.

Using right triangle geometry, it is easy to translate from polar coordinates to rectangular coordinates. In the Fig. 1.2b on page 2, one can see that the x distance traveled to get to the point is $x = r\cos(\theta)$ and the y distance is $y = r\sin(\theta)$.

Example 1. The translation of the point $\mathbf{x}_P = (4,-\pi/3)$ from polar coordinates to rectangular coordinates is given by

$$\mathbf{x}_R = (4\cos(-\pi/3), 4\sin(-\pi/3)) = (2,-2\sqrt{3}). \tag{1.1}$$

The *distance* r between two points (x_1,y_1) and (x_2,y_2) in the plane is found using the Pythagorean Theorem and rectangular coordinates. We use the idea from Fig. 1.3 on page 3.

Since r is the hypotenuse of the right triangle, our distance formula is

$$r = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}. \tag{1.2}$$

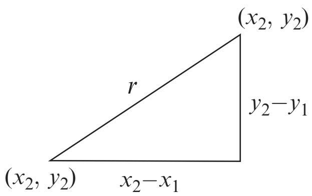


Fig. 1.3 The distance between points in the plane

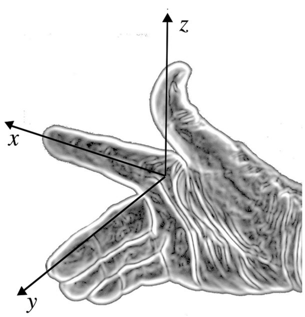


Fig. 1.4 How to visualize a right-hand system

Example 2. The distance from the point $(-1,2)$ to the point $(3,5)$ is

$$r = \sqrt{(3-(-1))^2+(5-2)^2} = \sqrt{4^2+3^2} = \sqrt{25} = 5. \tag{1.3}$$

Working with points in three dimensions using rectangular coordinates is similar. We use an ordered triple of numbers to represent the three coordinates. In a manner similar to what is done in two dimensions, we choose a point as the origin and takes three lines that meet at the origin and make right angles with each other as axes. The origin is again labeled with 0 for all coordinates, $\mathbf{0} = (0,0,0)$.

The axes are labeled to form a *right-handed system*. This means that if one has a flat right hand, the fingers point in the first direction. Bending all the fingers except the index finger 90° toward the palm of ones hand points these fingers in the second direction. Extending the thumb out from the hand at 90° from the index finger and the other fingers gives the third direction. It is common to label the directions of the fingers in the above positions as the positive x direction, the positive y direction, and the positive z direction respectively, see Fig. 1.4 on page 3.

Getting to a point is now done in the same way as getting to a point in the plane. To get to the point (a,b,c), one goes a units along the x axis from the origin, then b units parallel to the y axis, and finally c units parallel to the z axis. As with points in the plane, the order of the directions parallel to the axes does not matter.

Example 3. Figure 1.5 on page 4 shows how to plot the point $(2,-1,3)$.

The distance between two points in three dimensions is found using the Pythagorean Theorem. The example below gives a specific example to illustrate the reasoning. To find the distance from (x_1,y_1,z_1) to(x_2,y_2,z_2) we first find the distance from (x_1,y_1,z_1) to (x_2,y_2,z_1). These two

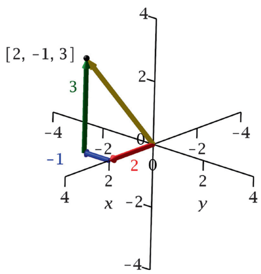
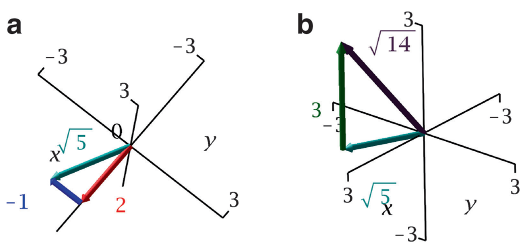


Fig. 1.5 A point in $\mathbb{R}^3$



First application of the Pythagorean Theorem

Second application of the Pythagorean Theorem

Fig. 1.6

points are in a plane parallel to the xy-plane. This means that the distance between them, since the z coordinate does not change, is

$$r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1.4}$$

The line segment from (x_2,y_2,z_1) to (x_2,y_2,z_2) is perpendicular to the line segment from (x_1,y_1,z_1) to (x_2,y_2,z_1) and has length $|z_2 - z_1|$. Since the distance from (x_1,y_1,z_1) to (x_2,y_2,z_2) is the length of the line segment between the points and that line segment is the hypotenuse of the right triangle with sides (x_2,y_2,z_1) to (x_2,y_2,z_2) and (x_1,y_1,z_1) to (x_2,y_2,z_1), the distance is

$$s = \sqrt{r^2 + (z_2 - z_1)^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \tag{1.5}$$

Example 4. Consider the distance from the origin, $(0,0,0)$ to $(2,-1,3)$. One could first find the distance from $(0,0,0)$ to $(2,-1,0)$ using the Pythagorean Theorem, see Fig. 1.6b on page 4. The distance is $r = \sqrt{2^2 + (-1)^2} = \sqrt{5}$.

Then one calculates the distance from the origin to $(2,-1,3)$ using the Pythagorean Theorem again, see Fig. 1.6b on page 4.

The distance is $s = \sqrt{\left(\sqrt{5}\right)^2 + 3^2} = \sqrt{14}$ or $s = \sqrt{2^2 + (-1)^2 + 3^2} = \sqrt{14}$.

There are two common generalizations of polar coordinates to three dimensions. The first simply adds a z coordinate to the polar r and θ, this system is called *cylindrical coordinates*. The second generalization of polar coordinates to three dimensions called *spherical coordinates* uses two angles and a distance, radius. Many students find these, coordinate systems more difficult to use than polar coordinates. Therefore, since they are not needed at this point, we leave them for later.

Exercises

1. Plot the following points in the plane.

(a)	$(2,3)$	(c)	$(-5,1)$	(e)	$(-2,-1)$
(b)	$(-3,-5)$	(d)	$(3,-6)$		

2. Plot the following points in $\mathbb{R}^3$.

(a)	$(1,2,3)$	(c)	$(-5,-2,1)$	(e)	$(-2,-1,-3)$
(b)	$(-3,-5,3)$	(d)	$(3,-6,3)$		

3. Find the distances between each of the pairs of points.

(a)	$(2,3)$, $(0,0)$	(f)	$(1,2,3)$, $(0,0,0)$
(b)	$(-3,-5)$, $(2,1)$	(g)	$(-3,-5,3)$, $(1,0,2)$
(c)	$(-5,1)$, $(0,3)$	(h)	$(-5,-2,1)$, $(3,-2,1)$
(d)	$(3,-6)$, $(3,-5)$	(i)	$(3,-6,3)$, $(5,-1,4)$
(e)	$(-2,-1)$, $(-2,-1)$	(j)	$(-2,-1,-3)$, $(2,1,3)$

4. Which of the following are right hand systems? The notation here is that $\hat{\imath}$ represents the positive x-axis, $\hat{\jmath}$ represents the positive y-axis, and $\hat{k}$ represents the positive z-axis. Also, $-\hat{\imath}$ represents the negative x-axis.

(a)	$\hat{\imath}$, $\hat{k}$, $\hat{\jmath}$	(c)	$\hat{\jmath}$, $\hat{k}$, $\hat{\imath}$
(b)	$\hat{\imath}$, $-\hat{k}$, $\hat{\jmath}$	(d)	$\hat{k}$, $\hat{\jmath}$, $\hat{\imath}$

5. You are 20 km away from a mountain with height 8,000 m. If your elevation is 3,000 m, how far are you from the top of the mountain?

6. Your target is 500 m north, 300 m west, and 500 m below you. Is the target within 1,000 m of you?

7. One car is 3 mi northeast of an intersection and a second car is 5 mi south of the same intersection. What is the distance between the two cars?

8. A door that is 2.5 ft by 7 ft starts with its bottom hinged corner at the origin, the top hinged corner along the z-axis and its bottom free corner along the y-axis. Find the position of the top free corner of after the door is rotated 30° toward the x-axis, see Fig. 1.7 on page 6.

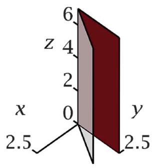


Fig. 1.7 A door rotated on its hinges

1.2 Vectors

The idea of a vector used in this class comes from the physical ideas of change in position, velocity, and acceleration. The velocity of a car on a flat surface has two components, the direction of travel and the speed of the car. In simple terms, the velocity has a direction and a

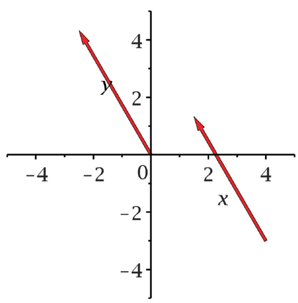


Fig. 1.8 Two representations of a single vector

magnitude. It does not have a position since a car can be going west at 100 km/h in New York or in New Delhi. Our definition of a vector is an object with direction and length.

When reading this book it is important to realize that real numbers can be taken as vectors. The magnitude of a real number x is its absolute value, $|x|$ and its direction is its sign, $\mathrm{sgn}(x)$. Since 0 has no sign and 0 length, it can have either direction. Given this approach, all of the theorems stated with vectors also hold for scalars, real numbers.

In the plane one often specifies directions with the counter clockwise angle from the positive x-axis. Adding a length gives a specification in polar coordinates. For example, one can move in the direction $2\pi/3$, (or $120°$), counter clockwise from the x-axis for 5 units. If one starts at the origin, this puts one at $(-5/2, 5\sqrt{3}/2)$. If one starts at $(4,-3)$, the vector puts one at $(4-5/2, -3+5\sqrt{3}/2)$, see Fig. 1.8 on page 6.

When we start a vector at a point $\mathbf{a}$ and end it at a point $\mathbf{b}$, $\mathbf{a}$ is called the *tail* of the vector and $\mathbf{b}$ is called the *head* of the vector.

Vectors can also be represented in rectangular coordinates. We can take a vector with direction θ and a length r and write it as $\big(r\cos(\theta), r\sin(\theta)\big)$. This is the form we will use most often in this course since calculus is easier to do in rectangular coordinates.

Example 5. We can write the vector, in polar coordinates, $\mathbf{x}_P(r,\theta) = (5, 3\pi/4)$ as

$$\mathbf{x}_R = \big(5\cos(3\pi/4), 5\sin(3\pi/4)\big) = (-5\sqrt{2}/2, 5\sqrt{2}/2). \tag{1.6}$$

in rectangular coordinates. In this case the vector can be thought of as moving 5 units in the direction $3\pi/4$ counter clockwise from the x-axis. The rectangular coordinates version can be thought of as moving $5\sqrt{2}/2$ in the negative x direction and $5\sqrt{2}/2$ in the positive y direction.

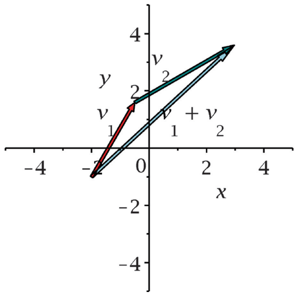


Fig. 1.9 The sum of two vectors

Remark 1. You have probably seen polar coordinates before. Recall that one translates a point (x,y) from rectangular to polar coordinates using

$$r = \sqrt{x^2+y^2} \text{ and}$$
$$\theta = \arctan\left(\frac{y}{x}\right).$$

One also needs to add or subtract π from θ if (x,y) is in the second or third quadrant since the range of $\arctan(\theta)$ is $(-\pi/2,\pi/2)$.

Example 6. Consider the point $(-4,5)$ in rectangular coordinates. In polar coordinates it becomes $(\sqrt{41}, \arctan(-5/4)+\pi)$.

The *length*, norm or magnitude, of a vector $\mathbf{x} = (x,y)$ in rectangular coordinates is defined as $\|\mathbf{x}\| = \sqrt{x^2+y^2}$. This is the distance from the tail of a vector to its tip. It is also the distance from the origin to the point (x,y). In a similar manner we take the length of a vector $\mathbf{x} = (x,y,z)$ in $\mathbb{R}^3$ to be the distance from $(0,0,0)$ to (x,y,z), $\|\mathbf{x}\| = \sqrt{x^2+y^2+z^2}$.

Addition of vectors is now easy to define. The model we use is the idea of moving in direction θ_1 a distance r_1 units and then moving in direction θ_2 a distance r_2 units (Fig. 1.9 on page 7).

One can view this as starting at a point, going in the direction of $\mathbf{v}_1$ the length of $\mathbf{v}_1$, to the tip of $\mathbf{v}_1$. Then one moves from the tip of $\mathbf{v}_1$ in the direction of $\mathbf{v}_2$ the length of $\mathbf{v}_2$. This puts one at the tip of $\mathbf{v}_2$. Often this is referred to as putting the tail of the second vector at the head of the first vector.

This sum is fairly hard to compute using angles and lengths, but it is easy to calculate using rectangular coordinates. Consider two vectors $\mathbf{x} = (x_1,y_1)$ and $\mathbf{y} = (x_2,y_2)$. By following the first vector we move x_1 in the x direction and y_1 in the y direction. The second vector then takes us an additional x_2 in the x direction and an additional y_2 in the y direction. The total movement in the x direction is x_1+x_2 and the total movement in the y direction is y_1+y_2. This gives the vector addition formula

$$(x_1,y_1)+(x_2,y_2) = (x_1+x_2, y_1+y_2). \tag{1.7}$$

This is described as adding coordinate by coordinate.

Example 7. The sum of the vectors $(2,3)$ and $(-3,-1)$ is

$$(2,3)+(-3,-1) = (2-3, 3-1) = (-1,2). \tag{1.8}$$

See Fig. 1.10 on page 8

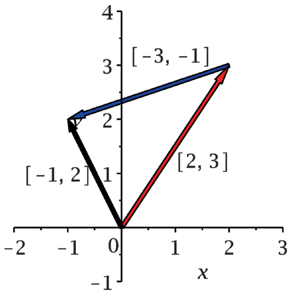


Fig. 1.10 The sum of two vectors

Adding vectors in three dimensions is also done coordinate by coordinate,

$$(x_1,y_1,z_1)+(x_2,y_2,z_2)=(x_1+x_2,y_1+y_2,z_1+z_2). \tag{1.9}$$

Example 8. The sum of $(-2,3,4)$ and $(4,1,-5)$ is

$$(-2,3,4)+(4,1,-5)=(-2+4,3+1,4-5)=(2,4,-1). \tag{1.10}$$

As with real numbers, there is a zero element, the *zero vector* $\mathbf{0}$. It is the vector with all coordinates 0, $(0,0)$ or $(0,0,0)$. With the definition of vector addition one has $\mathbf{u}+\mathbf{0}=\mathbf{u}=\mathbf{0}+\mathbf{u}$.

If one starts at the origin and goes 5 units in the direction of $(3,4)$ twice, where do you finish? Since the length of the vector $(3,4)$ is $\sqrt{3^2+4^2}=5$, the movement is given by $(3,4)+(3,4)=(6,8)$. Note that this is the equivalent of multiplying both components of the vector by 2. This gives us the definition of *scalar multiplication* of vectors,

$$a\,(x,y)=(ax,ay) \tag{1.11}$$

and

$$a\,(x,y,z)=(ax,ay,az). \tag{1.12}$$

Example 9. Find 3 times the vector $\mathbf{x}=(3,1,-2)$.

$$3\mathbf{x}=3\,(3,1,-2)=(3\cdot 3,3\cdot 1,3\cdot(-2))=(9,3,-6). \tag{1.13}$$

This also makes sense for negative numbers since the vector in the opposite direction of (x,y) is $(-x,-y)=-1\,(x,y)$.

Multiplication of vectors by scalars has two important properties that are the same as multiplication properties of numbers by numbers. The first is that this multiplication *distributes over addition*,

$$\alpha\,(\mathbf{v}+\mathbf{w})=\alpha\,\mathbf{v}+\alpha\,\mathbf{w} \qquad \text{and} \qquad (\alpha+\beta)\,\mathbf{v}=\alpha\,\mathbf{v}+\beta\,\mathbf{v}\ .$$

The order of multiplication can also be changed without changing the result,

$$(\alpha\,\beta)\,\mathbf{v}=\alpha\,(\beta\,\mathbf{v})=\beta\,(\alpha\,\mathbf{v})\ .$$

In physics and other fields people use this definition of scalar multiplication to relate the coordinate representation of vectors to another representation. One uses the three vectors of length one in the same directions as the positive x, y, and z axes. These unit vectors are denoted by $\hat{\imath}=(1,0,0)$, $\hat{\jmath}=(0,1,0)$, and $\hat{k}=(0,0,1)$. Sometimes they are also written as $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

We can always write any vector as a sum of these vectors. The translation is easy,

$$\begin{aligned}\mathbf{u}&=(u_1,u_2,u_3)\\&=u_1\hat{\imath}+u_2\hat{\jmath}+u_3\hat{k}.\end{aligned}$$

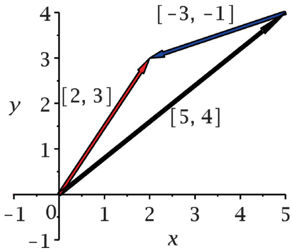


Fig. 1.11 The difference of two vectors

One can also write any vector $\mathbf{u}$ as

$$\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3.$$

For example, $(1,1,1) = \hat{\imath} + \hat{\jmath} + \hat{k}$.

With multiplication of vectors by scalars one can define the difference of two vectors as

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w}.$$

For example $(3,2) - (4,1) = (3,2) + (-4,-1) = (3-4, 2-1) = (-1,1)$.

This definition has other uses besides simply subtracting vectors. The idea is to find the vector that when one starts at a point $\bar{A}$ and follows the vector $\mathbf{v}$ one ends at point $\bar{B}$. Treating the points as vectors for computational purposes, one has

$$\bar{A} + \mathbf{v} = \bar{B}.$$

Rewriting gives

$$\mathbf{v} = \bar{B} - \bar{A}.$$

Example 10. The vector from $(2,-3,1)$ to $(6,-5,-3)$ is given by

$$\begin{aligned}(6,-5,-3) - (2,-3,1) &= (6-2, -5+3, -3-1)\\ &= (4,-2,-4).\end{aligned}$$

Geometrically we can view the difference of $\mathbf{x}_1$ and $\mathbf{x}_2$ by putting the head of $\mathbf{x}_2$ at the head of $\mathbf{x}_1$. The difference $\mathbf{x}_1 - \mathbf{x}_2$ is the vector from the tail of $\mathbf{x}_1$ to the tail of $\mathbf{x}_2$. This is illustrated in Fig. 1.11 on page 9 for the vectors $(2,3)$ and $(-3,-1)$.

For a number of reasons one may want to have a vector of length one in the same direction as a vector $\mathbf{x}$. A vector of length one is called a *unit vector*. Note that, for vectors with two or three coordinates,

$$\begin{aligned}\|\alpha\mathbf{x}\| &= \|(\alpha x, \alpha y, \alpha z)\|\\ &= \sqrt{(\alpha x)^2 + (\alpha y)^2 + (\alpha z)^2}\\ &= |\alpha|\sqrt{x^2+y^2+z^2}\\ &= |\alpha|\,\|\mathbf{x}\|.\end{aligned}$$

This means that, for a nonzero vector $\mathbf{x}$, a unit vector in the same direction as $\mathbf{x}$ is

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x}.$$

Example 11. The unit vector in the same direction as $(3,4)$ is

$$\mathbf{u} = \frac{1}{\sqrt{3^2+4^2}}\,(3,4) = \frac{1}{5}\,(3,4) = \left(\frac{3}{5},\frac{4}{5}\right).$$

One of the uses of vectors in physical situations is the description of forces. Since a force has direction and magnitude, they match exactly with vectors. If we want a force in the plane with the same direction as $\mathbf{v} = (2,-3)$ and magnitude 10 N, we simply take the unit vector in the direction of $\mathbf{v}$ and multiply it by 10,

$$\begin{aligned}\mathbf{F} &= 10\,\frac{\mathbf{v}}{\|\mathbf{v}\|}, N\\ &= \frac{10}{\sqrt{13}}\,(2,-3), N.\end{aligned}$$

One of Newton's laws says that the sum of the forces acting on a stationary object must be $\mathbf{0}$. For example, if you are pushing against a railroad car on a east-west track in a westerly direction with a force of 20 N and the car does not move, the frictional force in an easterly direction must also have magnitude 20 N.

Example 12. Consider a 4 kg mass that is suspended from a ceiling by 2 and 3 m ropes whose ends are 3 m apart. Find the force exerted on the mass by each of the two ropes (Fig. 1.12 on page 10).

Near the surface of the earth the force from the mass is $(0, 4\cdot(-9.8))$ N. If the force from the 3 ft rope is $\mathbf{F}_1$ and the force from the 2 ft rope is $\mathbf{F}_2$ one has

$$\mathbf{F}_1 + \mathbf{F}_2 = (0, 39.2)\,\mathrm{N}.$$

In the Fig. 1.13a on page 11 $\theta = \arccos(1/3)$ and $\phi = 2\ \arcsin(1/3)$. (This can be derived using the fact that this is an isosceles triangle.) This tells us that the lengths of a, b, and c are $7/3$, $2/3$, and $4\sqrt{2}/3$, see Fig. 1.13b on page 11.

If f_1 and f_2 are the magnitudes of $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively, this means that

$$\begin{aligned} f_1\left(\frac{-7}{9}\right) + f_2\,\frac{1}{3} &= 0\\ \text{and}\quad f_1\,\frac{4\sqrt{2}}{9} + f_2\,\frac{2\sqrt{2}}{3} &= 39.2.\end{aligned}$$

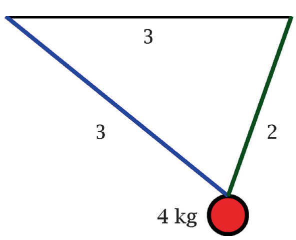
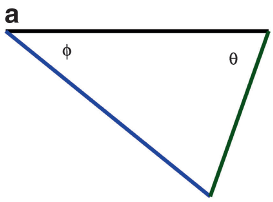
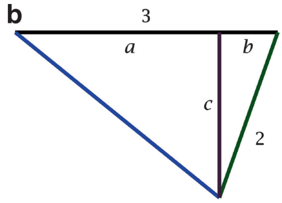


Fig. 1.12 A mass suspended using two ropes



The angles for a mass suspended using two ropes



The distances for a mass suspended using two ropes

Fig. 1.13

Solving these equations for f_1 and f_2 gives $f_1 \approx 13.859$ and $f_2 \approx 32.338$. The vectors are then approximately

$$\begin{aligned} \mathbf{F}_1 &\approx (-10.779, 8.711) \quad \text{and} \\ \mathbf{F}_2 &\approx (10.779, 30.489). \end{aligned}$$

1.2.1 Properties of Addition and Scalar Multiplication

In order to calculate with vectors using vector addition and scalar multiplication some properties of the operations are needed. These properties mirror the properties of addition and multiplication of real numbers. These similarities follow from the fact that the operations on vectors are done coordinate by coordinate.

Theorem 1 (Vector computation rules). *Let* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$ *and let* α *and* β *be real numbers. Then the following are always true.*

(i) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
(ii) $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{v}+\mathbf{u})+\mathbf{w}$
(iii) $\alpha(\mathbf{u}+\mathbf{v})=\alpha\mathbf{u}+\alpha\mathbf{v}$
(iv) $(\alpha+\beta)\mathbf{u}=\alpha\mathbf{u}+\beta\mathbf{u}$
(v) *If* $-\mathbf{u}$ *is the vector such that* $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$ *then* $(-1)\,\mathbf{u}=-\mathbf{u}$.
(vi) $\|\alpha\mathbf{u}\|=|\alpha|\,\|\mathbf{u}\|$

The proofs of most of the properties are similar. Therefore only the first statement is proven.

Proof. Consider the sum $\mathbf{u}+\mathbf{v}$ in $\mathbb{R}^2$. Here

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= (u_1,u_2)+(v_1,v_2) \\ &= (u_1+v_1, u_2+v_2) \\ &= (v_1+u_1, v_2+u_2) \\ &= \mathbf{v}+\mathbf{u}. \end{aligned}$$

An example shows how these rules are used.

Example 13. To find $3(1,-2,3)-7(1,-2,3)$ one can calculate as follows.

$$\begin{aligned} 3(1,-2,3)-7(1,-2,3) &= (3-7)\,(1,-2,3) \\ &= (-4)\,(1,-2,3) \\ &= (-4,8,-12) \end{aligned}$$

Exercises

1. Translate the vector in polar coordinates to rectangular coordinates.

(a) $\mathbf{x}_P = (3, \pi/6)$
(b) $\mathbf{x}_P = (2, 3\pi/2)$
(c) $\mathbf{v}_P = (5, 5\pi/4)$
(d) $\mathbf{v}_P = (-4, 5\pi/6)$
(e) $\mathbf{w}_P = (2, -\pi/4)$

2. Translate the vector in rectangular coordinates to polar coordinates.
To find the r for polar coordinates note that the distance from (x, y) to the origin is $\sqrt{x^2+y^2}$. From this one has $r = \sqrt{x^2+y^2}$. The angle θ is obtained using the facts that $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Taking the ratio of these gives

$$\tan(\theta) = \frac{y}{x}$$

and hence

$$\theta = \arctan\left(\frac{y}{x}\right).$$

(a) $\mathbf{x}_R = (1, -1)$
(b) $\mathbf{x}_R = (-3, 6)$
(c) $\mathbf{v}_R = (6, 12)$
(d) $\mathbf{v}_R = (4, -5)$
(e) $\mathbf{w}_R = (0, -4)$

3. Find the vector from $\bar{A}$ to $\bar{B}$.

(a) $\bar{A} = (2,3)$, $\bar{B} = (0,0)$
(b) $\bar{A} = (-3,-5)$, $\bar{B} = (2,1)$
(c) $\bar{A} = (-5,1)$, $\bar{B} = (0,3)$
(d) $\bar{A} = (3,-6)$, $\bar{B} = (3,-5)$
(e) $\bar{A} = (-2,-1)$, $\bar{B} = (-2,-1)$
(f) $\bar{A} = (1,2,3)$, $\bar{B} = (0,0,0)$
(g) $\bar{A} = (-3,-5,3)$, $\bar{B} = (1,0,2)$
(h) $\bar{A} = (-5,-2,1)$, $\bar{B} = (3,-2,1)$
(i) $\bar{A} = (3,-6,3)$, $\bar{B} = (5,-1,4)$
(j) $\bar{A} = (-2,-1,-3)$, $\bar{B} = (2,1,3)$

4. Find the sum of the vectors.

(a) $\mathbf{w} = (1,2)$, $\mathbf{v} = (0,0)$
(b) $\mathbf{w} = (4,-5)$, $\mathbf{v} = (3,2)$
(c) $\mathbf{w} = (-5,3)$, $\mathbf{v} = (2,-1)$
(d) $\mathbf{w} = (4,-2)$, $\mathbf{v} = (1,7)$
(e) $\mathbf{w} = (-2,-1)$, $\mathbf{v} = (-2,-1)$
(f) $\mathbf{w} = (3,5,-3)$, $\mathbf{v} = (0,0,0)$
(g) $\mathbf{w} = (5,2,8)$, $\mathbf{v} = (1,2,5)$
(h) $\mathbf{w} = (3,6,-1)$, $\mathbf{v} = (-1,7,-15)$
(i) $\mathbf{w} = (4,1,-4)$, $\mathbf{v} = (5,-1,4)$
(j) $\mathbf{w} = (-6,7,-2)$, $\mathbf{v} = (6,-7,2)$

5. Find the difference of the vectors, $\mathbf{w} - \mathbf{v}$.

(a) $\mathbf{w} = (1,2)$, $\mathbf{v} = (0,0)$
(b) $\mathbf{w} = (4,-5)$, $\mathbf{v} = (3,2)$
(c) $\mathbf{w} = (-5,3)$, $\mathbf{v} = (2,-1)$
(d) $\mathbf{w} = (4,-2)$, $\mathbf{v} = (1,7)$
(e) $\mathbf{w} = (-2,-1)$, $\mathbf{v} = (-2,-1)$
(f) $\mathbf{w} = (3,5,-3)$, $\mathbf{v} = (0,0,0)$
(g) $\mathbf{w} = (5,2,8)$, $\mathbf{v} = (1,2,5)$
(h) $\mathbf{w} = (3,6,-1)$, $\mathbf{v} = (-1,7,-15)$
(i) $\mathbf{w} = (4,1,-4)$, $\mathbf{v} = (5,-1,4)$
(j) $\mathbf{w} = (-6,7,-2)$, $\mathbf{v} = (6,-7,2)$

6. Multiply the vector $\mathbf{w}$ by the scalar s.

(a) $\mathbf{w} = (1,2)$, $s = 2$
(b) $\mathbf{w} = (4,-5)$, $s = 4$
(c) $\mathbf{w} = (-5,3)$, $s = -1$
(d) $\mathbf{w} = (4,-2)$, $s = 6$
(e) $\mathbf{w} = (-2,-1)$, $s = -3$
(f) $\mathbf{w} = (3,5,-3)$, $s = 1$
(g) $\mathbf{w} = (5,2,8)$, $s = -2$
(h) $\mathbf{w} = (3,6,-1)$, $s = 4$
(i) $\mathbf{w} = (4,1,-4)$, $s = -10$
(j) $\mathbf{w} = (-6,7,-2)$, $s = 2$

7. Use the vectors $\mathbf{r} = (2,3)$, $\mathbf{v} = (-1,2)$ and $\mathbf{w} = (5,1)$ to compute the following.

(a) $\mathbf{w}+\mathbf{v}+\mathbf{r}$
(b) $\mathbf{w}-\mathbf{v}-\mathbf{r}$
(c) $2\mathbf{w}+\mathbf{v}+3\mathbf{r}$
(d) $\mathbf{r}-2\mathbf{v}+2\mathbf{r}$
(e) $-2\mathbf{w}-\mathbf{v}+5\mathbf{r}$

8. As was noted in this section, adding vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ can be geometrically represented by putting the tail of $\mathbf{v}_2$ at the tip of the $\mathbf{v}_1$ and taking the sum of the vectors to be the vector with tail at the tail of $\mathbf{v}_1$ and tip at the tip of $\mathbf{v}_2$. (See Fig. 1.9 on page 7.) How can the difference of two vectors be represented by a similar figure?

9. Find the lengths, (norms or magnitudes), of the following vectors.

(a) $\mathbf{w} = (3,0)$
(b) $\mathbf{w} = (0,4)$
(c) $\mathbf{x} = (1,-1)$
(d) $\mathbf{x} = (-7,6)$
(e) $\mathbf{v} = (4,-3)$
(f) $\mathbf{v} = (17,12)$
(g) $\mathbf{w} = (3,0,0)$
(h) $\mathbf{w} = (0,4,0)$
(i) $\mathbf{x} = (0,0,-1)$
(j) $\mathbf{x} = (3,4,5)$
(k) $\mathbf{v} = (-1,4,-3)$
(l) $\mathbf{v} = (17,6,12)$

10. Find the unit vector in the same direction as each of the following vectors.

(a) $\mathbf{w} = (3,0)$
(b) $\mathbf{w} = (0,4)$
(c) $\mathbf{x} = (1,-1)$
(d) $\mathbf{x} = (-7,6)$
(e) $\mathbf{v} = (4,-3)$
(f) $\mathbf{v} = (17,12)$
(g) $\mathbf{w} = (3,0,0)$
(h) $\mathbf{w} = (0,4,0)$
(i) $\mathbf{x} = (0,0,-1)$
(j) $\mathbf{x} = (3,4,5)$
(k) $\mathbf{v} = (-1,4,-3)$
(l) $\mathbf{v} = (17,6,12)$

11. In physics a stationary object will stay stationary if the sum of the forces on the object is the zero vector. For each of the following pairs of forces find a third force that will keep an object stationary.

(a) $\mathbf{F}_1 = (-1,1)$, $\mathbf{F}_2 = (-1,-1)$
(b) $\mathbf{F}_1 = (-2,-1)$, $\mathbf{F}_2 = (-1,3)$
(c) $\mathbf{F}_1 = (4,2)$, $\mathbf{F}_2 = (3,7)$
(d) $\mathbf{F}_1 = (-1,2)$, $\mathbf{F}_2 = (-2,-1)$
(e) $\mathbf{F}_1 = (1,1,1)$, $\mathbf{F}_2 = (-2,2,-2)$
(f) $\mathbf{F}_1 = (4,1,-3)$, $\mathbf{F}_2 = (6,-5,6)$
(g) $\mathbf{F}_1 = (0,1,-1)$, $\mathbf{F}_2 = (-1,1,0)$
(h) $\mathbf{F}_1 = (2,-2,1)$, $\mathbf{F}_2 = (3,-3,6)$

12. The following vectors are given in polar coordinates. Find the unit vector in the same direction as each of the following vectors.

(a) $\mathbf{w}_P = (3,\pi)$
(b) $\mathbf{w}_P = (1/4,\pi/6)$
(c) $\mathbf{x}_P = (1,7\pi/6)$
(d) $\mathbf{x}_P = (-5,3\pi/4)$

13. Any two masses have a mutual attraction due to gravity. Given a mass M_1 at $\mathbf{a}$ and a mass M_2 at $\mathbf{b}$ in two or three dimensions, the force that mass M_1 feels from the pull of mass M_2 is

$$\mathbf{F}_g = \frac{GM_1M_2}{\|\mathbf{r}\|^3}\mathbf{r}.$$

Here $\mathbf{r} = \mathbf{a}-\mathbf{b}$, the displacement vector from the center of mass M_2 to the center of mass M_1. The constant G is the gravitational constant $G = 6.67384\times 10^{-11}\,\mathrm{m}^3/\left(\mathrm{kg\,s}^3\right)$. It is assumed this holds throughout the universe. The masses must be measured in kilograms, distance in meters, and time in seconds.

Assume that the center of the earth is at the origin of a coordinate system. If the mass of the earth is $5.9722\times 10^{24}\,\mathrm{kg}$, what is the force that the earth exerts on a mass of the given number of kilograms at the given position in meters. (The radius of the earth is approximately $6,371\,\mathrm{km}$.)

(a) $M = 10^5\,\mathrm{kg}$ and $\mathbf{r} = (10{,}000, 500, 1{,}000)\,\mathrm{km}$
(b) $M = 1{,}000\,\mathrm{kg}$ and $\mathbf{r} = (8{,}000, 1{,}000, 1{,}000)\,\mathrm{km}$

(c) $M = 7.3477 \times 10^{22}$ kg and $\mathbf{r} = (0, 0, 3.844 \times 10^5)$ km. These are approximate numbers for the moon.

(d) $M = 1.98892 \times 10^{30}$ kg and $\mathbf{r} = (1.047 \times 10^8, 1.047 \times 10^8, 0)$ km. These are approximate numbers for the sun.

14. Two objects are placed with their centers at $(-2{,}000, 0)$ and $(0, 4{,}000)$ km in the xy-plane. The objects have masses of 10^5 and 3×10^5 kg respectively. You are to place a 3,000 kg mass on the plane so that the sum of the gravitational forces from the three masses acting on a 100 kg mass at the origin is $\mathbf{0}$.

(a) On what line through the origin should you place the mass?

(b) Where should you place the mass?

1.3 The Dot Product

One of the basic problems in elementary physics classes is finding the force acting on a mass sitting on a frictionless incline (Fig. 1.14 on page 14). This is done with some elementary geometry in physics classes to get the equation

$$a = g\cos(\theta).$$

Here a is the magnitude of the acceleration down the incline and θ is the angle between the vertical and the direction down the incline.

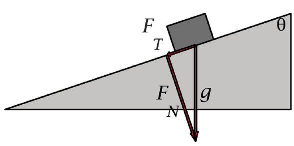


Fig. 1.14 A block on an incline

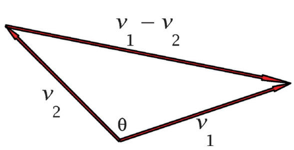


Fig. 1.15 The difference of two vectors

Another way of calculating this is through the dot product, inner product, of two vectors.

Definition 1 (Dot Product). The dot (or inner) product of two vectors, $\mathbf{v}$ and $\mathbf{w}$, in $\mathbb{R}^2$ or $\mathbb{R}^3$ is defined as

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\,\|\mathbf{w}\|\cos(\theta),$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$. If either of the vectors is $\mathbf{0}$, we take the dot product to be 0.

If the vectors are one vectors, real numbers, the dot product is simply multiplication of real numbers. In this setting θ is either 0 or π.

Example 14. The angle between the vectors $\mathbf{v} = (3,0)$ and $\mathbf{w} = (2,2)$ is $\pi/4$. This means that

$$\begin{aligned}\mathbf{v}\cdot\mathbf{w} &= \|\mathbf{v}\|\,\|\mathbf{w}\|\cos(\pi/4)\\ &= 3\cdot 2\sqrt{2}\cdot\frac{\sqrt{2}}{2}\\ &= 6.\end{aligned}$$

The problem with this definition is that we must find the angle between two vectors. In most cases this means taking the arctangent of some quantity. The following theorem gives us a simpler way of calculating the dot product for vectors in rectangular coordinates.

Theorem 2 (Computation of the dot product). *The dot product of vectors is given by*

$$(x_1,y_1)\cdot(x_2,y_2) = x_1x_2 + y_1y_2$$

in $\mathbb{R}^2$ *or*

$$(x_1,y_1,z_1)\cdot(x_2,y_2,z_2) = x_1x_2 + y_1y_2 + z_1z_2$$

in $\mathbb{R}^3$.

Proof. Consider the picture, Fig. 1.15 on page 14, for two vectors in $\mathbb{R}^2$.

By the Law of Cosines,

$$\|\mathbf{v}_1 - \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2\|\mathbf{v}_1\|\,\|\mathbf{v}_2\|\cos(\theta).$$

Let $\mathbf{v}_1 = (x_1,y_1)$ and $\mathbf{v}_2 = (x_2,y_2)$. Solving the equation above for $\|\mathbf{v}_1\|\,\|\mathbf{v}_2\|\cos(\theta) = \mathbf{v}_1\cdot\mathbf{v}_2$, we get

$$\begin{aligned}\mathbf{v}_1\cdot\mathbf{v}_2 = \frac{-1}{2}\Bigg(&\left(\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}\right)^2 - \left(\sqrt{(x_1)^2+(y_1)^2}\right)^2\\ &- \left(\sqrt{(x_2)^2+(y_2)^2}^{\,2}\right)\Bigg)\\ = \frac{-1}{2}&(-2x_1x_2 - 2y_1y_2)\\ = x_1x_2& + y_1y_2.\end{aligned}$$

A similar argument also works for vectors in $\mathbb{R}^3$ and shows that, in $\mathbb{R}^3$

$$\mathbf{v}_1\cdot\mathbf{v}_2 = x_1x_2 + y_1y_2 + z_1z_2.$$

With this theorem it is easy to calculate the dot product of two vectors.

Example 15. The dot product of $(1,-2,3)$ and $(2,3,0)$ is

$$(1,-2,3)\cdot(2,3,0) = 1\cdot 2 + (-2)\cdot 3 + 3\cdot 0 = -4.$$

We can also use this to calculate the angle θ between two vectors. Using the formula

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\;\|\mathbf{v}\|\;\cos(\theta),$$

we have

$$\cos(\theta) = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\;\|\mathbf{v}\|}$$

or

$$\theta = \arccos\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\;\|\mathbf{v}\|}\right). \tag{1.14}$$

For those who are not familiar with the $\arccos(x)$, see Sect. 4.5 for the definition and a graph of the function.

Example 16. Consider the vectors $\mathbf{p} = (2,3,-1)$ and $\mathbf{q} = (0,2,-1)$. If θ is the angle between $\mathbf{p}$ and $\mathbf{q}$,

$$\begin{aligned}\cos(\theta) &= \frac{(2,3,-1)\cdot(0,2,-1)}{\|(2,3,-1)\|\;\|(0,2,-1)\|}\\ &= \frac{2\cdot 0+3\cdot 2+(-1)\cdot(-1)}{\sqrt{14}\,\sqrt{5}}\\ &= \frac{7}{\sqrt{70}}.\end{aligned}$$

In addition

$$\begin{aligned}\theta &= \arccos\left(\frac{7}{\sqrt{70}}\right)\\ &\approx 0.5796\\ &\approx 33.21^\circ.\end{aligned}$$

Remark 2. If the dot product of two nonzero vectors is 0, then the cosine of the angle between the vectors is 0. This implies the angle between the vectors is $\pi/2$. In this case the vectors are called *perpendicular* or *orthogonal*.

Example 17. Consider the vectors $\mathbf{u} = (2,-1,3)$, $\mathbf{v} = (1,5,1)$, and $\mathbf{w} = (1,5,-1)$. Since

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= (2,-1,3)\cdot(1,5,1)\\ &= 2-5+3\\ &= 0,\end{aligned}$$

the vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

If we consider **u** and **w**, we have

$$\begin{aligned}\mathbf{u}\cdot\mathbf{w} &= (2,-1,3)\cdot(1,5,-1)\\ &= 2-5-3\\ &= -6.\end{aligned}$$

The vectors **u** and **w** are not orthogonal.

In order to calculate the force acting on a block on an incline we need some additional facts about the dot product. These are contained in the following result.

Theorem 3 (Dot product and norm). *Let* $\mathbf{v}$ *be a vector in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$. *Then*

(i) $\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}}$
and
(ii) $\mathbf{v}\cdot\mathbf{v} = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$.

Proof. This is done only for $\mathbb{R}^3$ since $\mathbb{R}^2$ is similar and only (i) is considered. The dot product of $\mathbf{v} = (x,y,z)$ with itself is

$$\begin{aligned}\mathbf{v}\cdot\mathbf{v} &= (x,y,z)\cdot(x,y,z)\\ &= x^2+y^2+z^2.\end{aligned}$$

This means that

$$\begin{aligned}\sqrt{\mathbf{v}\cdot\mathbf{v}} &= \sqrt{(x,y,z)\cdot(x,y,z)}\\ &= \sqrt{x^2+y^2+z^2}\\ &= \|(x,y,z)\|.\end{aligned}$$

We can now return to the incline problem. Assume that the incline goes from $(0,0)$ to $(3,1)$. The vector pointing down the incline is $(0,0)-(3,1) = (-3,-1)$. This means that the magnitude of the acceleration down the incline is

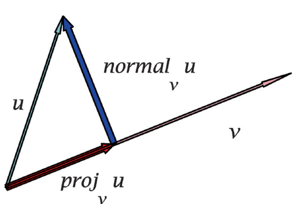


Fig. 1.16 The projection and normal component of a vector

$$\begin{aligned}\|(0,-g)\|\cos(\theta) &= \frac{\|(-3,-1)\|\,\|(0,-g)\|\cos(\theta)}{\|(-3,-1)\|}\\ &= \frac{(-3,-1)\cdot(0,-g)}{\sqrt{10}}\\ &= \frac{g}{\sqrt{10}}.\end{aligned}$$

The actual acceleration vector along the incline is the magnitude of the acceleration times a unit vector in the direction of $(-3,-1)$,

$$\begin{aligned}\mathbf{F}_T &= \frac{\|(-3,-1)\|\,\|(0,-g)\|\cos(\theta)}{\|(-3,-1)\|}\,\frac{(-3,-1)}{\|(-3,-1)\|}\\ &= \frac{(-3,-1)\cdot(0,-g)}{(-3,-1)\cdot(-3,-1)}(-3,-1)\\ &= \frac{g}{10}(-3,-1).\end{aligned}$$

These two quantities are used in a more general setting and are called the *component* of $\mathbf{u}$ in the direction of $\mathbf{v}$,

$$comp_{\mathbf{v}}\,\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{v}\|},$$

and the *projection* of $\mathbf{u}$ onto $\mathbf{v}$, or the component vector of $\mathbf{u}$ in the direction of $\mathbf{v}$,

$$\operatorname{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\,\mathbf{v}. \tag{1.15}$$

Another vector quantity we can calculate is the *normal part* of $\mathbf{u}$ to $\mathbf{v}$. It is given by

$$\operatorname{normal}_{\mathbf{v}}\mathbf{u} = \mathbf{u} - \operatorname{proj}_{\mathbf{v}}\mathbf{u}.$$

It is easy to check that $\operatorname{proj}_{\mathbf{u}}\mathbf{v}$ is *orthogonal* to $\operatorname{normal}_{\mathbf{v}}\mathbf{u}$. In other words, the angle between the vectors is $\pi/2 = 90^\circ$ or their dot product is 0. We can view this as writing $\mathbf{u}$ as a sum of two vectors that are at right angles to each other with the requirement that one is in the direction of $\mathbf{v}$, see Fig. 1.16 on page 17.

Remark 3. In physics the term component is sometimes used both for the vectors $\operatorname{proj}_{\mathbf{v}}\mathbf{u}$ and $\operatorname{normal}_{\mathbf{v}}\mathbf{u}$ and for their lengths. In that case a bold letter $\mathbf{a}$ or a letter with an arrow above $\mathbf{a}$ is used for the vector and the plain letter a is used for the length. It is also common in physics to use a hat, $\hat{a}$ to designate a unit vector in the same direction as $\mathbf{a}$.

Example 18. An example of this is the force a sail provides for a cart on a straight railroad track. Assume the track is going in the direction $(-2,3)$. Also assume the wind is blowing in the same direction as the vector $(3,1)$ and produces a force of 200 N acting on the cart through the sail. What is the force acting on the cart parallel to the track and what is the force produced by the sail that pushes the cart across the tracks?

Call the direction of the track $\mathbf{u}$ and the force from the wind against the sail $\mathbf{F}$. Since $\mathbf{F}$ is in the direction $(3,1)$ with magnitude 200 N, we have

$$\mathbf{F} = 200\,\frac{(3,1)}{\|(3,1)\|}\,\mathrm{N} = 20\sqrt{10}(3,1)\,\mathrm{N}.$$

The force parallel to the track is

$$\begin{aligned}\operatorname{proj}_{\mathbf{u}}\mathbf{F} &= \frac{20\sqrt{10}(3,1)\cdot(-2,3)}{(-2,3)\cdot(-2,3)}(-2,3)\,\mathrm{N}\\ &= \frac{-60\sqrt{10}}{13}(-2,3)\,\mathrm{N}.\end{aligned}$$

The force perpendicular to the track is

$$\begin{aligned}\operatorname{normal}_{\mathbf{u}}\mathbf{F} &= \mathbf{F} - \operatorname{proj}_{\mathbf{u}}\mathbf{F}\\ &= 20\sqrt{10}(3,1) - \frac{-60\sqrt{10}}{13}(-2,3)\,\mathrm{N}\\ &= \frac{20\sqrt{10}}{13}(33,22)\,\mathrm{N}.\end{aligned}$$

There are some properties of the dot product that you should know. These are commonly used and you should be able to use them.

Theorem 4 (Dot product computation rules). *Let* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$ *and let* α *be a real number. Then*

(i) $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$,
(ii) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$,
(iii) $(\alpha\mathbf{u})\cdot\mathbf{v} = \alpha(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(\alpha\mathbf{v})$
and
(iv) $\mathbf{v}\cdot\mathbf{0} = 0$.

Proof. Since the proofs of all of the equations are similar, this is restricted to (i). Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$. Then

$$\begin{aligned}\mathbf{v}\cdot\mathbf{w}\,&(v_1,v_2,v_3)\cdot(w_1,w_2,w_3)\\ &= v_1w_1+v_2w_2+v_3w_3\\ &= w_1v_1+w_2v_2+w_3v_3\\ &= \mathbf{w}\cdot\mathbf{w}.\end{aligned}$$

In elementary physics work is usually defined as force times distance. Here we use the standard mks units: force is in Newtons (N), distance is in meters (m), and work is in Joules (J). For example, if you move down a track with a constant force of 10 N for 100 m you use 1,000 J. This is not as simple if the force being applied is not in the direction of the motion.

Consider a car on a straight east-west railroad track. If the force applied to the car by the wind is 200 N in a south-easterly direction, what is the work done moving the car 200 m along the track? If $\mathbf{v} = (1,0)$ is the direction of the track and $\mathbf{F} = 100\sqrt{2}(1,-1)$ N, then the force along the track is

$$comp_{\mathbf{v}}\mathbf{F} = \frac{\mathbf{v}\cdot\mathbf{F}}{||\mathbf{v}||}N\ .$$

Since the cart moves 200 m along the track, the displacement of the car can be taken as $\mathbf{w} = (200,0) = 200\mathbf{v}$ m. This means the work is

$$\begin{aligned}W &= 200\frac{\mathbf{v}\cdot\mathbf{F}}{||\mathbf{v}||}\\ &= (200\mathbf{v})\cdot\mathbf{F}\\ &= \mathbf{w}\cdot\mathbf{F}\\ &= 20{,}000\sqrt{2}J.\end{aligned}$$

This idea leads to the formula that the work done moving a mass along a straight line through a displacement of $\mathbf{w}$ under the influence of a constant force $\mathbf{F}$ is given by

$$W = \mathbf{w}\cdot\mathbf{F}.$$

Example 19. Find the work done moving a mass in a straight line from $(-1,2,-1)$ m to $(3,4,7)$ m with a constant force of $\mathbf{F} = (1,-1,2)$ N acting on the mass.

The displacement vector is $\mathbf{v} = (3,4,7)\,\mathrm{m} - (-1,2,-1)\,\mathrm{m} = (4,2,8)\,\mathrm{m}$. Calculating gives the work as

$$W = \mathbf{F}\cdot\mathbf{v} = (1,-1,2)\,\mathrm{N}\cdot(4,2,8)\,\mathrm{m} = 18J.$$

Exercises

1. Find the dot products of the following pairs of vectors.

(a) $\mathbf{x} = (3,2)$, $\mathbf{y} = (4,5)$
(b) $\mathbf{x} = (-2,1)$, $\mathbf{y} = (-2,1)$
(c) $\mathbf{x} = (4,-5)$, $\mathbf{y} = (-3,1)$
(d) $\mathbf{x} = (1,1,1)$, $\mathbf{y} = (1,2,3)$
(e) $\mathbf{x} = (0,1,-2)$, $\mathbf{y} = (1,0,0)$
(f) $\mathbf{x} = (6,7,-3)$, $\mathbf{y} = (-2,4,-1)$

2. Find the angles between the pair of vectors.

(a) $\mathbf{x} = (3,2)$, $\mathbf{y} = (4,5)$
(b) $\mathbf{x} = (-2,1)$, $\mathbf{y} = (-2,1)$
(c) $\mathbf{x} = (4,-5)$, $\mathbf{y} = (-3,1)$
(d) $\mathbf{x} = (1,1,1)$, $\mathbf{y} = (1,2,3)$
(e) $\mathbf{x} = (0,1,-2)$, $\mathbf{y} = (1,0,0)$
(f) $\mathbf{x} = (6,7,-3)$, $\mathbf{y} = (-2,4,-1)$

3. Find projection of the first vector onto the second.

(a) $\mathbf{x} = (3,2)$, $\mathbf{y} = (4,5)$
(b) $\mathbf{x} = (-2,1)$, $\mathbf{y} = (-2,1)$
(c) $\mathbf{x} = (4,-5)$, $\mathbf{y} = (-3,1)$
(d) $\mathbf{x} = (1,1,1)$, $\mathbf{y} = (1,2,3)$
(e) $\mathbf{x} = (0,1,-2)$, $\mathbf{y} = (1,0,0)$
(f) $\mathbf{x} = (6,7,-3)$, $\mathbf{y} = (-2,4,-1)$

4. Find the part of the first vector normal to the second vector.

(a) $\mathbf{x} = (3,2)$, $\mathbf{y} = (4,5)$
(b) $\mathbf{x} = (-2,1)$, $\mathbf{y} = (-2,1)$
(c) $\mathbf{x} = (4,-5)$, $\mathbf{y} = (-3,1)$
(d) $\mathbf{x} = (1,1,1)$, $\mathbf{y} = (1,2,3)$
(e) $\mathbf{x} = (0,1,-2)$, $\mathbf{y} = (1,0,0)$
(f) $\mathbf{x} = (6,7,-3)$, $\mathbf{y} = (-2,4,-1)$

5. Find the work done in moving from point $\bar{A}$ to point $\bar{B}$ with a constant force vector $\mathbf{F}$.

(a) $\bar{A} = (2,1)$, $\bar{B} = (3,-2)$, $\mathbf{F} = (1,-1)$
(b) $\bar{A} = (-3,0)$, $\bar{B} = (0,2)$, $\mathbf{F} = (3,6)$
(c) $\bar{A} = (-2,-2)$, $\bar{B} = (5,5)$, $\mathbf{F} = (-4,4)$
(d) $\bar{A} = (-2,3)$, $\bar{B} = (7,-1)$, $\mathbf{F} = (0,-1)$
(e) $\bar{A} = (1,2,1)$, $\bar{B} = (2,3,-2)$, $\mathbf{F} = (1,1,-1)$
(f) $\bar{A} = (-1,-2,2)$, $\bar{B} = (7,-7,3)$, $\mathbf{F} = (2,2,-4)$
(g) $\bar{A} = (2,1,3)$, $\bar{B} = (4,3,1)$, $\mathbf{F} = (1,1,2)$
(h) $\bar{A} = (3,2,-1)$, $\bar{B} = (-1,-1,-1)$, $\mathbf{F} = (4,-4,1)$

6. In the following the vector $\mathbf{T}$ points down an incline. A M kg mass is on the incline with an acceleration due to gravity $\mathbf{g} = (0,-9.8)\,\mathrm{m/s}$. Find the force pulling the mass down the incline.

(a) $\mathbf{T} = (2,-2)$, $M = 2$
(b) $\mathbf{T} = (-4,-1)$, $M = 4$
(c) $\mathbf{T} = (7,-2)$, $M = 1$
(d) $\mathbf{T} = (6,-1)$, $M = 3$
(e) $\mathbf{T} = (-7,-1)$, $M = 5$
(f) $\mathbf{T} = (4,-1)$, $M = 10$

7. Find two units vectors orthogonal to each of the following vectors.

(a) $\mathbf{w} = (1,2)$
(b) $\mathbf{w} = (-4,0)$
(c) $\mathbf{v} = (-1,1)$
(d) $\mathbf{v} = (-6,-5)$

8. Find two unit vectors orthogonal to the following pairs of vectors in $\mathbb{R}^3$.

(a) $\mathbf{w} = \hat{\imath}$, $\mathbf{z} = \hat{k}$
(b) $\mathbf{w} = (1,1,0)$, $\mathbf{v} = (-1,1,0)$
(c) $\mathbf{w} = (1,1,-2)$, $\mathbf{v} = (-1,1,0)$
(d) $\mathbf{w} = (3,2,1)$ $\mathbf{v} = (-1,2,1)$

9. Show that the dot product distributes over vector addition. For vectors **v**, **w**, and **z** one has

$$\mathbf{v} \cdot (\mathbf{w} + \mathbf{z}) = \mathbf{v} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{z}.$$

10. Find formulas for the cosines of the angles of a vector $\mathbf{v} = (v_1, v_2, v_3)$ with each of the coordinate axes.

11. The following vectors are given in polar coordinates. Find the dot product of the two vectors without translating to rectangular coordinates.

(a) $\mathbf{u}_P = (1, 7\pi/6)$, $\mathbf{v}_P = (3, \pi)$
(b) $\mathbf{u}_P = (6, 5\pi/6)$, $\mathbf{v}_P = (4, \pi/4)$
(c) $\mathbf{u}_P = (4, 5\pi/4)$, $\mathbf{v}_P = (3, -\pi/2)$
(d) $\mathbf{u}_P = (2, 5\pi/6)$, $\mathbf{v}_P = (-5, \pi/4)$

12. In each of the following a mass of M kilograms is on an incline without friction that goes from A to B. Assume that the mass is under the influence of gravity at the surface of the earth. Find the force acting on the mass.

(a) $M = 1$, $A = (0,0)$, and $B = (4,3)$.
(b) $M = 5$, $A = (0,3)$, and $B = (1,4)$.
(c) $M = 2$, $A = (1,1)$, and $B = (7,2)$.
(d) $M = 1$, $A = (0,5)$, and $B = (4,5)$.

1.4 Vectors in n Dimensions

Many times one needs to use more than two or three dependent variables to describe a physical situation. An example is describing the motion of three masses in three dimensions. It is common to assume that one of the masses is fixed at $\mathbf{0}$. For each of the other two objects we need both a position and a velocity. This gives a total of $2 \cdot (3+3) = 12$ coordinates needed for this problem. Because of this common need, we will be using vectors with n components.

Definition 2. The space of vectors with n coordinates, $\mathbb{R}^n$, is the set of all ordered n-tuples of real numbers,

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) | x_i \in \mathbb{R}, i = 1, 2, \ldots, n\}.$$

These vectors are used the same way that vectors from $\mathbb{R}^2$ and $\mathbb{R}^3$ are. For example, we define addition of two vectors in $\mathbb{R}^n$ by

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$$

and we define scalar multiplication by

$$\alpha \mathbf{w} = (\alpha w_1, \alpha w_2, \ldots, \alpha w_n).$$

The length, norm, of a vector also uses the same formula as in $\mathbb{R}^3$,

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Example 20. Let $\mathbf{v} = (1, 2, -2, 1, 3, 5)$, $\mathbf{w} = (4, -3, 2, 1, -3, 2)$ and $\alpha = -3$. Then

$$\mathbf{v} + \mathbf{w} = (1+4, 2-3, -2+2, 1+1, 3-3, 5+2) = (5, -1, 0, 2, 0, 7)$$

and

$$\alpha \mathbf{v} = 3(1, 2, -2, 1, 3, 5) = (3, 6, -6, 3, 9, 15).$$

We also have

$$\|\mathbf{v}\| = \sqrt{1^2 + 2^2 + (-2)^2 + 1^2 + 3^2 + 5^3} = \sqrt{44}.$$

As in $\mathbb{R}^2$ and $\mathbb{R}^3$, the zero vector is $\mathbf{0} = (0,0,\ldots,0)$ with $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all vectors $\mathbf{v}$ in $\mathbb{R}^n$.

The same properties of addition and scalar multiplication that we have for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ also hold in $\mathbb{R}^n$.

Theorem 5 (Vector computation rules, $\mathbb{R}^n$). *Let* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$ *and let* α *and* β *be real numbers. Then*

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$,
(ii) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$,
(iii) $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
and
(iv) $(\alpha + \beta)\mathbf{v} = (\alpha\mathbf{v}) + (\beta\mathbf{v})$.

The proofs of these equalities are left to the reader. We do them coordinate by coordinate where the same properties hold for real numbers.

The angle definition of the dot product we used in $\mathbb{R}^2$ and $\mathbb{R}^3$ is not valid in $\mathbb{R}^n$ if $n > 3$ since we do not know how to define angles in the way they were defined for $\mathbb{R}^2$ and $\mathbb{R}^3$. Instead, the dot product is defined as follows. The angle between two vectors is then defined using the formula from the dot product definition in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 3 (Dot product, $\mathbb{R}^n$). Let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$. The dot product of v and w is defined as

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

This makes it easy to compute a dot production.

Example 21. The dot product of $(1,2,-2,4,-1)$ and $(2,1,2,1,5)$ is

$$\begin{aligned}(1,2,-2,4,-1)\cdot(2,1,2,1,5) &= 1\cdot 2 + 2\cdot 1 + (-2)\cdot 2 + 4\cdot 1 + (-1)\cdot 5\\ &= -1.\end{aligned}$$

Using the $\mathbb{R}^2$ and $\mathbb{R}^3$ definition of the dot product,

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\,\|\mathbf{w}\|\cos(\theta),$$

as a model, the angle θ between two vectors in $\mathbb{R}^n$ is defined by

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|}.$$

Since $|\cos(\theta)| \leq 1$, this definition requires the Cauchy-Schwarz Inequality, for any two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|.$$

A quick proof of the Cauchy-Schwarz Inequality is given at the end of this section, Theorem 8 on page 24.

Using the example above we have that the cosine of the angle between $(1,2,-2,4,-1)$ and $(2,1,2,1,5)$ is

$$\cos(\theta) = \frac{-1}{\sqrt{26}\sqrt{35}}.$$

This gives an angle θ, usually taken in $[0,\pi]$, that is approximately 1.604 rad or approximately 91.9°.

As in $\mathbb{R}^2$ and $\mathbb{R}^3$ two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ are said to be orthogonal if $\mathbf{v}\cdot\mathbf{w}=0$. This is the equivalent of the vectors being at right angles.

Going from $\mathbb{R}^3$ to $\mathbb{R}^n$ does not change the properties of the dot product. This means that the same theorems on the properties of the dot product hold.

Theorem 6 (Dot product and norm, $\mathbb{R}^n$). *Let* $\mathbf{v}$ *be a vector in* $\mathbb{R}^n$. *Then*

(i) $\|\mathbf{v}\|=\sqrt{\mathbf{v}\cdot\mathbf{v}}$
and
(ii) $\mathbf{v}\cdot\mathbf{v}=0$ *if and only if* $\mathbf{v}=\mathbf{0}$.

Theorem 7 (Dot product computation rules, $\mathbb{R}^n$). *Let* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$ *and let* α *be a real number. Then*

(i) $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$,
(ii) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$,
(iii) $(\alpha\mathbf{u})\cdot\mathbf{v}=\alpha\,(\mathbf{u}\cdot\mathbf{v})=\mathbf{u}\cdot(\alpha\mathbf{v})$
and
(iv) $\mathbf{v}\cdot\mathbf{0}=0\,.$

Proof. The proofs of all of these equalities are similar. As an example, the fact that $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$ is proven. Here

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v}&=u_1v_1+u_2v_2+\cdots+u_nv_n\\&=v_1u_1+v_2u_2+\cdots+v_nu_n\\&=\mathbf{v}\cdot\mathbf{u}.\end{aligned}$$

As in $\mathbb{R}^2$ and $\mathbb{R}^3$ we define the *projection* of $\mathbf{v}$ onto $\mathbf{w}$ as

$$\operatorname{proj}_{\mathbf{w}}\mathbf{v}=\frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}\,\mathbf{w}.$$

Again, this is the vector $\mathbf{z}$ in the direction of $\mathbf{w}$ such that $\mathbf{v}-\mathbf{z}$ is orthogonal to $\mathbf{w}$. The vector $\mathbf{v}-\mathbf{z}$ is called the *normal part* to the projection of $\mathbf{v}$ onto $\mathbf{w}$,

$$\operatorname{normal}_{\mathbf{w}}\mathbf{v}=\mathbf{v}-\operatorname{proj}_{\mathbf{w}}\mathbf{v}.$$

Example 22. Consider the vectors $\mathbf{v}=(2,6,-3,4,1)$ and $\mathbf{w}=(1,2,1,2,1)$. The component of $\mathbf{v}$ in the direction of $\mathbf{w}$ is

$$\begin{aligned}\operatorname{comp}_{\mathbf{w}}\mathbf{v}&=\frac{(2,6,-3,4,1)\cdot(1,2,1,2,1)}{\|(1,2,1,2,1)\|}\\&=\frac{2+12-3+8+1}{\sqrt{1+4+1+4+1}}\\&=\frac{20}{\sqrt{11}}.\end{aligned}$$

The projection of $\mathbf{v}$ onto $\mathbf{w}$ is

$$\begin{aligned}\operatorname{proj}_{\mathbf{w}}\mathbf{v}&=\frac{(2,6,-3,4,1)\cdot(1,2,1,2,1)}{(1,2,1,2,1)\cdot(1,2,1,2,1)}\,(1,2,1,2,1)\\&=\frac{20}{11}\,(1,2,1,2,1).\end{aligned}$$

Finally, the normal to the projection of $\mathbf{v}$ onto $\mathbf{w}$ is

$$\begin{aligned}\operatorname{normal}_{\mathbf{w}} \mathbf{v} &= \mathbf{v} - \operatorname{proj}_{\mathbf{w}} \mathbf{v} \\ &= (2,6,-3,4,1) - \frac{20}{11}(1,2,1,2,1) \\ &= \left(\frac{2}{11}, \frac{26}{11}, -\frac{53}{11}, \frac{4}{11}, -\frac{9}{11}\right).\end{aligned}$$

What follows is the statement and a proof of the Cauchy-Schwarz Inequality. It is optional material.

Theorem 8 (Cauchy-Schwarz Inequality). *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be any vectors in* $\mathbb{R}^n$. *Then*

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| .$$

Proof. Consider the quadratic polynomial in t,

$$\begin{aligned}(u_1 t + v_1)^2 + (u_2 t + v_2)^2 + \cdots + (u_n t + v_n)^2 \\ &= \left(u_1^2 + u_2^2 + \cdots + u_n^2\right) t^2 + 2\left(u_1 v_1 + u_2 v_2 + \cdots + u_n v_n\right) t \\ &\qquad + \left(v_1^2 + v_2^2 + \cdots + v_n^2\right) \\ &= \|\mathbf{u}\|^2 t^2 + 2|\mathbf{u} \cdot \mathbf{v}| t + \|\mathbf{v}\|^2 .\end{aligned}$$

Since this quadratic polynomial is always nonnegative, the discriminant of the quadratic (the part under the square root in the quadratic formula, Eq. A.2 on page 386) must be nonpositive. This means that

$$0 \geq 4|\mathbf{u} \cdot \mathbf{v}|^2 - 4\|\mathbf{u}\|^2 \, \|\mathbf{v}\|^2 .$$

Rewriting this inequality as

$$\|\mathbf{u}\|^2 \, \|\mathbf{v}\|^2 \geq |\mathbf{u} \cdot \mathbf{v}|^2$$

and taking the square roots of both sides gives the Cauchy-Schwarz Inequality,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| .$$

Exercises

1. Find the dot products of the following pairs of vectors.

(a) $\mathbf{x} = (3,2,5,1)$, $\mathbf{y} = (4,5,1,-4)$
(b) $\mathbf{x} = (-2,1,1,-2)$, $\mathbf{y} = (-2,1,1,-2)$
(c) $\mathbf{x} = (4,-5,2,1,3)$, $\mathbf{y} = (-3,1,-3,4,-2)$
(d) $\mathbf{x} = (1,1,1,4,5)$, $\mathbf{y} = (1,2,3,-2,2)$
(e) $\mathbf{x} = (0,1,-2,0,1,2)$, $\mathbf{y} = (1,0,0,6,0,0)$
(f) $\mathbf{x} = (6,7,-3,2,1,1)$, $\mathbf{y} = (-2,4,-1,1,-2,3)$

2. Find the angles between the pair of vectors.

(a) $\mathbf{x} = (3,2,5,1)$, $\mathbf{y} = (4,5,1,-4)$
(b) $\mathbf{x} = (-2,1,1,-2)$, $\mathbf{y} = (-2,1,1,-2)$
(c) $\mathbf{x} = (4,-5,2,1,3)$, $\mathbf{y} = (-3,1,-3,4,-2)$

(d) $\mathbf{x} = (1,1,1,4,5)$, $\mathbf{y} = (1,2,3,-2,2)$
(e) $\mathbf{x} = (0,1,-2,0,1,2)$, $\mathbf{y} = (1,0,0,6,0,0)$
(f) $\mathbf{x} = (6,7,-3,2,1,1)$, $\mathbf{y} = (-2,4,-1,1,-2,3)$

3. Find projection of the first vector onto the second.

(a) $\mathbf{x} = (3,2,5,1)$, $\mathbf{y} = (4,5,1,-4)$
(b) $\mathbf{x} = (-2,1,1,-2)$, $\mathbf{y} = (-2,1,1,-2)$
(c) $\mathbf{x} = (4,-5,2,1,3)$, $\mathbf{y} = (-3,1,-3,4,-2)$
(d) $\mathbf{x} = (1,1,1,4,5)$, $\mathbf{y} = (1,2,3,-2,2)$
(e) $\mathbf{x} = (0,1,-2,0,1,2)$, $\mathbf{y} = (1,0,0,6,0,0)$
(f) $\mathbf{x} = (6,7,-3,2,1,1)$, $\mathbf{y} = (-2,4,-1,1,-2,3)$

4. Find normal to the projection of the first vector onto the second.

(a) $\mathbf{x} = (3,2,5,1)$, $\mathbf{y} = (4,5,1,-4)$
(b) $\mathbf{x} = (-2,1,1,-2)$, $\mathbf{y} = (-2,1,1,-2)$
(c) $\mathbf{x} = (4,-5,2,1,3)$, $\mathbf{y} = (-3,1,-3,4,-2)$
(d) $\mathbf{x} = (1,1,1,4,5)$, $\mathbf{y} = (1,2,3,-2,2)$
(e) $\mathbf{x} = (0,1,-2,0,1,2)$, $\mathbf{y} = (1,0,0,6,0,0)$
(f) $\mathbf{x} = (6,7,-3,2,1,1)$, $\mathbf{y} = (-2,4,-1,1,-2,3)$

5. If one has two points $\bar{A}$ and $\bar{B}$ in $\mathbb{R}^n$ one defines the distance between them as the length of the vector $\mathbf{v} = \bar{B} - \bar{A}$. Find the distances between the following pairs of points.

(a) $\bar{A} = (2,1,3,5)$, $\bar{B} = (3,-2,0,1)$
(b) $\bar{A} = (-3,0,2,1)$, $\bar{B} = (0,2,4,-2)$
(c) $\bar{A} = (-2,-2,2,3,1)$, $\bar{B} = (5,5,5,5,0)$

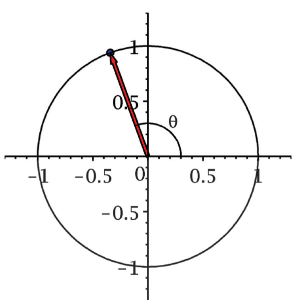


Fig. 1.17 The function $\mathbf{W}(\theta)$

(d) $\bar{A} = (-2,3,-7,2,1)$, $\bar{B} = (7,-1,0,-1,5)$
(e) $\bar{A} = (1,2,1,-1,-2,2)$, $\bar{B} = (2,3,-2,7,-7,3)$
(f) $\bar{A} = (2,1,3,3,2,-1)$, $\bar{B} = (4,3,1,-1,-1,-1)$

6. Show that the dot product in $\mathbb{R}^n$ distributes over vector addition. For vectors $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$ one has

$$\mathbf{v} \cdot (\mathbf{w} + \mathbf{z}) = \mathbf{v} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{z}.$$

7. Find formulas for the cosines of the angles of a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ with each of the coordinate axes in $\mathbb{R}^n$.

1.5 Vector-Valued Functions

Many things are described by functions that have one input and a vector quantity as the output. This section gives some examples of such functions. They include lines and circles. Such functions, $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$, give what are called parametrized curves.

The first example in this section is one of the most important and one of the most common parametrized curves.

Example 23. The circle $x^2 + y^2 = 1$ is the image of the function

$$\mathbf{W}(\theta) = (\cos(\theta), \sin(\theta)) \ .$$

What you should have learned is that each point on the unit circle is defined by $(x, y) = (\cos(\theta), \sin(\theta))$ where θ is the counter clockwise angle from the positive x-axis to the line segment from the origin to (x, y), see Fig. 1.17 on page 25.

If the distance from the origin is changed from 1 to r, the result is a circle of radius r. Since the vector $(\cos(\theta), \sin(\theta))$ is multiplied by r, this circle can be parametrized by $(r\cos(\theta), r\sin(\theta))$. We can also parametrize a circle centered at a point (x_0, y_0) with radius r. The circle has equation $(x - x_0)^2 + (y - y_0)^2 = r^2$. Setting $x - x_0 = r\cos(\theta)$ and $y - y_0 = r\sin(\theta)$ gives a parametrization, $\mathbf{w}(\theta) = (x_0 + r\cos(\theta), y_0 + r\sin(\theta))$.

Example 24. Find a parametrization of the circle of radius 5 centered at $(-2, 3)$. Using the parametrization of the circle above we have

$$\mathbf{x}(t) = (-2 + 5\cos(t), 3 + 5\sin(t)).$$

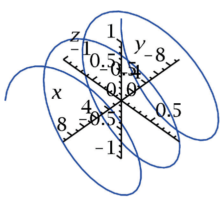


Fig. 1.18 The helix $\mathbf{H}(t) = (t, \cos(t), \sin(t))$

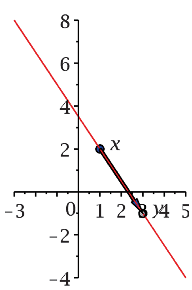


Fig. 1.19 The line containing $\mathbf{x}$ and $\mathbf{y}$

If we go in the opposite direction around the circle we have a parametrization $\mathbf{x}(t) = (-2 + 5\cos(-t), 3 + 5\sin(-t)) = (-2 + 5\cos(t), 3 - 5\sin(t))$ and if we go around the circle twice as fast we have the parametrization $\mathbf{x}(t) = (-2 + 5\cos(2t), 3 + 5\sin(2t))$.

The graph of a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ can be parametrized easily.

Example 25. The graph of a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ can be parametrized as $(x, \mathbf{f}(x))$. This is true since the graph of f is all points of the form $(x, \mathbf{f}(x))$ where x is in the domain of $\mathbf{f}$.

Example 26. The graph of the function $\mathbf{W}(t) = (\cos(t), \sin(t))$ is the helix $\mathbf{H}(t) = (t, \cos(t), \sin(t))$, see Fig. 1.18 on page 26 where the function is plotted on the interval $[-3\pi, 3\pi]$.

A straight line in $\mathbb{R}^n$ through two points $\mathbf{x}$ and $\mathbf{y}$ can easily be parametrized.

Example 27. If $\mathbf{x}$ and $\mathbf{y}$ are two points in $\mathbb{R}^n$ and $\mathbf{v} = \mathbf{y} - \mathbf{x}$, then the line through the two points can be parametrized as $\ell(t) = \mathbf{x} + t\,\mathbf{v}$. This follows from the fact that $\mathbf{v}$ is the direction from $\mathbf{x}$ to $\mathbf{y}$. Since each point on the line is reached from $\mathbf{x}$ by going some distance in the direction from $\mathbf{x}$ to $\mathbf{y}$, or in the opposite direction, our equation must give a parametrization of the line, see Fig. 1.19 on page 26.

Example 28. Find a parametrization of the line containing $(1, 4, -2)$ and $(2, 2, 2)$. A direction along the line is $(1, 4, -2) - (2, 2, 2) = (-1, 2, -4)$. This means that the line can be parametrized as $\ell(t) = (1, 4, -2) + t(-1, 2, -4)$.

t	0	$\frac{\pi}{18}$	$\frac{\pi}{9}$	$\frac{\pi}{6}$	$\frac{2\pi}{9}$	$\frac{5\pi}{18}$	$\frac{\pi}{3}$	$\frac{7\pi}{18}$	$\frac{4\pi}{9}$	$\frac{\pi}{2}$
x	1.	0.850	0.470	0.	−0.384	−0.555	−0.500	−0.298	−0.090	0.
y	0.	0.150	0.171	0.	−0.321	−0.665	−0.865	−0.810	−0.492	0.

Table 1.1 Values of $\mathbf{x}(t) = (\cos(3t)\cos(t), \cos(3t)\sin(t))$

t	$\frac{5\pi}{9}$	$\frac{11\pi}{18}$	$\frac{2\pi}{3}$	$\frac{13\pi}{18}$	$\frac{7\pi}{9}$	$\frac{5\pi}{6}$	$\frac{8\pi}{9}$	$\frac{17\pi}{18}$	π
x	−0.090	−0.298	−0.500	−0.555	−0.384	0.	0.470	0.850	1.
y	0.492	0.810	0.865	0.665	0.321	0.	−0.171	−0.150	0.

Table 1.2 Values of $\mathbf{x}(t) = (\cos(3t)\cos(t), \cos(3t)\sin(t))$

Example 29. The path taken by a mass with constant velocity is a line. If a mass is at a point $\mathbf{x}_0$ when $t = t_0$ and travels with constant velocity $\mathbf{v}$, its position as a function of time is $\mathbf{x}(t) = \mathbf{x}_0 + (t - t_0)\,\mathbf{v}$.

This is a parametrization of a line. Such a parametrization always has the form $\ell(t) = \mathbf{a} + t\,\mathbf{v}$ where $\mathbf{a}$ is a point on the line and $\mathbf{v}$ is a direction along the line.

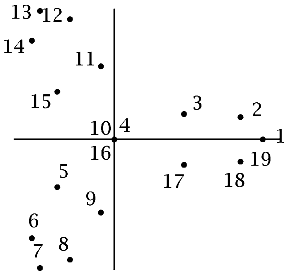


Fig. 1.20 Some points for plotting $\mathbf{x}(t) = (\cos(3t)\cos(t), \cos(3t)\sin(t))$

When plotting vector-valued functions we often do not plot both the inputs and outputs. Often only the outputs are plotted. We make a table of output vectors of the function and then plot them. The plotted points are then connected in an appropriate manner, in the same order as the order of the input points.

Example 30. Plot the curve parametrized by

$$\mathbf{x}(t) = (\cos(3t)\cos(t), \cos(3t)\sin(t)).$$

First make a table of values as in Tables 1.1 on page 27 and 1.2 on page 27.

Then plot the points, see Fig. 1.20 on page 28.

Finally, connect the points in order, see Fig. 1.21 on page 28.

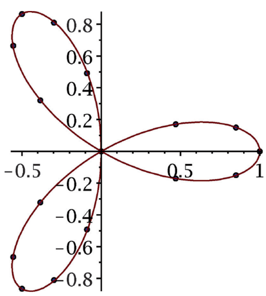


Fig. 1.21 The graph of $\mathbf{x}(t) = (\cos(3t)\cos(t), \cos(3t)\sin(t))$

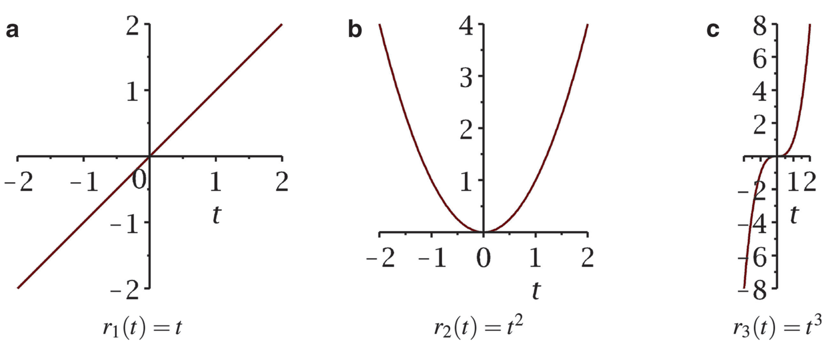


Fig. 1.22

Another way of getting some idea of the shape of a parametrized curve $\mathbf{r}(t)$ is to use plots of the coordinate functions that define $\mathbf{r}(t)$. The coordinate functions for a function $\mathbf{r}(t) : \mathbb{R} \to \mathbb{R}^n$ are the functions $r_i(t)$ such that $\mathbf{r}(t) = (r_1(t), r_2(t), \ldots, r_n(t))$. The information obtained from the plots of the functions $r_i(t)$ and parametrized curves $\mathbf{r}_{i,j} = (r_i(t), r_j(t))$ can help greatly. The following example illustrates this for a parametrized curve in $\mathbb{R}^3$.

Example 31. Consider the vector valued function $\mathbf{r}(t) = \left(t, t^2, t^3\right)$. The graphs of the coordinate functions on the interval $[-2, 2]$ are in Fig. 1.22 on page 28.

Those graphs indicate that at $t = -2$ the curve is at $(-2, 4, -8)$ and moves to $(0, 0, 0)$ at $t = 0$. Over this range x is changing at a constant speed, y is changing at a decreasing speed, and the rate of change of z is decreasing faster than the rate of change of y is decreasing, but the speed also goes to 0. As t goes from 0 to 2 the speeds reverse what they did from -2 to 0 and the curve goes from $(0, 0, 0)$ to $(2, 4, 8)$.

We can also plot the parametrized curves given by $\mathbf{r}_{i,j} = (r_i(t), r_j(t))$. These plots give additional information about the parametrized curve. The graphs of these curves are given in Fig. 1.23 on page 29.

These curves show what the parametrized curve looks like when looking at the graph from each of the positive coordinate axes. Looking from the z-axis it is a parabola, from the y-axis it is a cubic curve, and from the x-axis it is a $y = z^{2/3}$ curve. (Why does the last of these graphs take the form $y = z^{2/3}$?) This means that we can think of the parametrized curve in $\mathbb{R}^3$ as similar to a cubic in z along the curve $y = x^2$ in xy-plane. All of this information matches the correct parametrized curve in Fig. 1.24 on page 29.

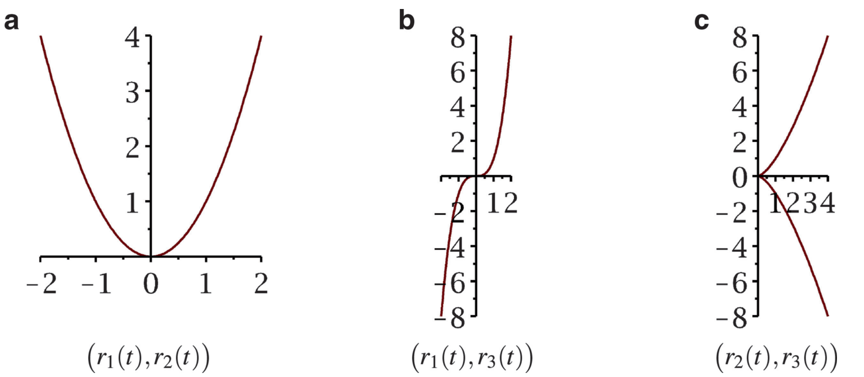


Fig. 1.23

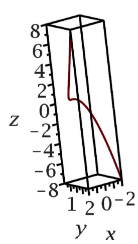


Fig. 1.24 The curve $\mathbf{r}(t) = \left(t, t^2, t^3\right)$

Exercises

1. Match the following vector valued functions with their images.

(a) $\mathbf{f}(t) = (\cos(t), 3\sin(t))$

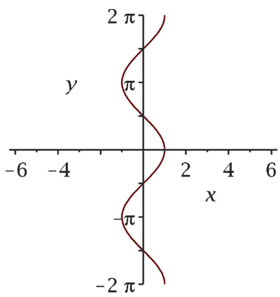
(b) $\mathbf{g}(t) = (t^2, t^3)$

(c) $\mathbf{h}(s) = (\cos(s), s)$

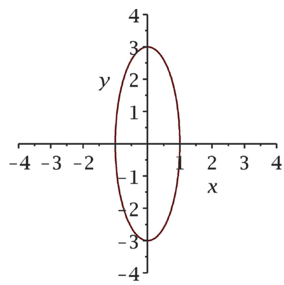
(d) $\ell(s) = (2s - 1, s + 2)$

(e) $\mathbf{k}(z) = (4\cos^2(z), 4\cos(z)\ \sin(z))$

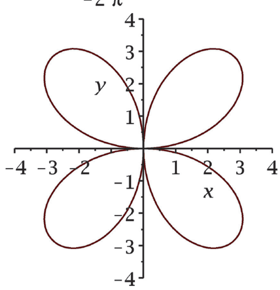
(f) $\mathbf{r}(z) = (4\sin(2z)\cos(z), 4\sin(2z)\sin(z))$



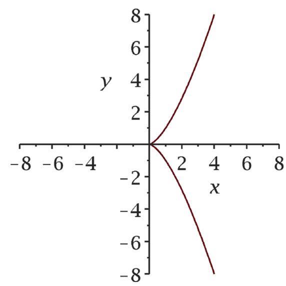
(I)



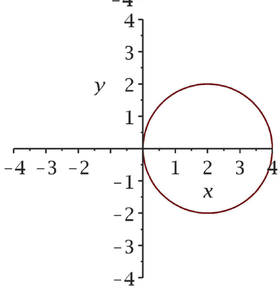
(II)



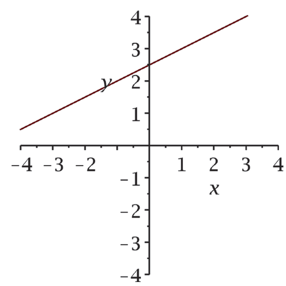
(III)



(V)



(IV)



(VI)

2. Match the functions with their images.

(a) $\mathbf{f}(t) = (t, 2\cos(t), 4\sin(t))$

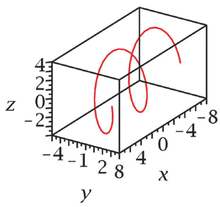
(b) $\mathbf{g}(t) = (t^2, t^2, t^3)$

(c) $\mathbf{h}(s) = (3\sin(s), 3\cos(s), s)$

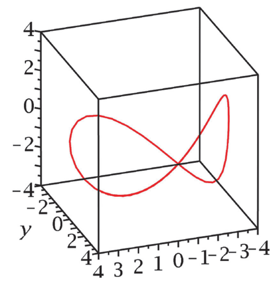
(d) $\ell(s) = (-s, s-1, 2s+1)$

(e) $\mathbf{k}(z) = (4\sin(z), 4\cos^2(z), 4\cos(z)\ \sin(z))$

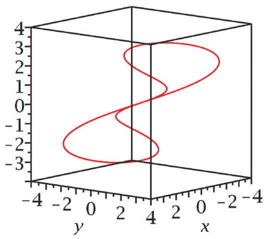
(f) $\mathbf{r}(z) = (4\sin(2z)\cos(z), 4\sin(2z)\sin(z), 3\cos(z))$



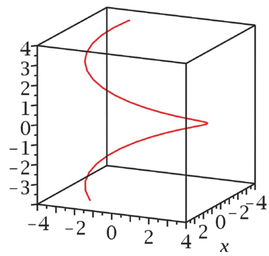
(I)



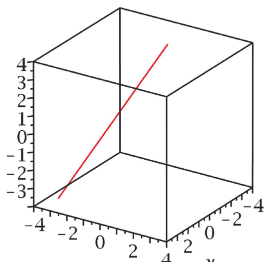
(III)



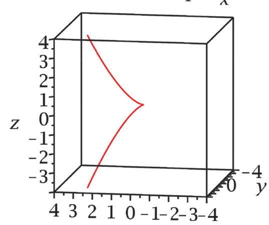
(II)



(IV)



(V)



(VI)

3. Find parametrizations $\mathbf{R}(t)$ of the unit circle based on $\mathbf{W}(t) = (\cos(t), \sin(t))$ that do the following things.

(a) Goes in the opposite direction of $\mathbf{W}(t)$ with $\mathbf{R}(0) = \mathbf{W}(0)$.

(b) Goes around in the same direction as $\mathbf{W}(t)$, takes half the time to complete a cycle, and $\mathbf{R}(0) = \mathbf{W}(0)$.

(c) Goes around in the same direction as $\mathbf{W}(t)$, takes three times the time to complete a cycle, and $\mathbf{R}(0) = \mathbf{W}(0)$.

(d) Goes around in the same direction as $\mathbf{W}(t)$, takes the same time to complete a cycle, and $\mathbf{R}(0) = \mathbf{W}(\pi)$.

(e) Goes around in the opposite direction as $\mathbf{W}(t)$, takes three times the time to complete a cycle, and $\mathbf{R}(0) = \mathbf{W}(\pi/2)$.

(f) Goes around in the same direction as $\mathbf{W}(t)$, $\mathbf{R}(0) = \mathbf{W}(0)$, and $\mathbf{R}(2) = \mathbf{W}(0)$.

4. A mass M starting at a point $\mathbf{r}_0$ with an initial velocity $\mathbf{v}$ that has no forces acting on it will maintain a constant velocity. Its position as a function of time is given by $\mathbf{r}(t) = \mathbf{r}_0 + t\,\mathbf{v}$. The speed of the mass is also constant and is given by $s(t) = \|\mathbf{v}\|$. Assume that the original position of a mass is $(1, -2)$ and its velocity is always $(2, 1)$.

(a) Find the position of the mass M as a function of time.

(b) A mass starts at the same position as M but goes in the opposite direction at the same speed. Find the position of this mass as a function of time.

(c) A mass starts at the same position as M, goes in the opposite direction as M, but goes at speed one. Find the position of this mass as a function of time.

(d) A mass starts at the same position as M, goes in the same direction as M, but goes at speed one. Find the position of this mass as a function of time.

5. A bead is restricted to travel on the path $\ell(t) = (2t-5, 5-3t, 4+t)$. If the bead is at $(-1,-1,6)$ and a force $\mathbf{F} = (-3,5,-2)$ is acting on the bead, what is the force pushing the bead along the path?

6. Find a parametrization of the line through $(2,1)$ and $(-3,5)$. (See Example 29.)

Chapter 2
Limits and Derivatives

The idea of a limit is central to all of calculus. Throughout the rest of your calculus classes it will be behind everything you learn. Most people do not understand this concept the first time they see it, but they can get a feeling for the basic idea with some effort.

After looking at limits, this chapter moves on to the idea of a derivative. This idea is more natural and many more people understand the concept of a derivative. In the context of physics it is often used as the instantaneous rate change of position, velocity, or the instantaneous rate of change of velocity, acceleration. Even though physical motion is where Newton originally defined the concept of a derivative, the derivative has found uses in many other areas, including economics and biology.

2.1 Sequences and Limits

The idea of a limit was approached for thousands of years without ever being reached. The concept, when finally realized, helped to revolutionize mathematics.

One of the problems that leads toward the concept of a limit is what mathematicians call Zeno's paradox. A person is walking at a constant speed across a room. At some time the person is half way across the room. At another time they will again have halved their distance to the opposite wall. At a third time they will, again, have halved their distance from the opposite wall. This process continues forever. Does the person ever reach the opposite wall?

The problem here is that the distances left to the far wall are, assuming the original distance was 1,

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots$$

If we consider only these distances, they are never zero. Does this mean that the person does not reach the wall?

On the other hand, if the person is traveling at one distance unit per minute, what are the times when the person has half the distance left to the far wall, $1/4$ the distance left, $1/8$ the distance left, etc. These times are $1/2$ min at $1/2$ the distance left, $3/4$ min at $1/4$ the distance left, $7/8$ min at $1/8$ the distance left, etc. Does the fact that these times are less than one make a difference?

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_2) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_2

Using the concept of a limit these questions will be answered. First the definition of a sequence is presented. The definition may seem strange, but the examples should make the definition fairly clear.

Definition 4 (Infinite sequence). An *infinite sequence* is a function $\mathbf{a}$ from the natural numbers to a set, usually $\mathbb{R}$ or $\mathbb{R}^n$. The values of this function are often denoted by $\{\mathbf{a}_n\}_{n=1}^{\infty}$ instead of writing $\mathbf{a}(n)$.

The image space of $\mathbf{a}(n)$ does not need to be $\mathbb{R}$ or $\mathbb{R}^n$, but those are the only sets we will use.

A *finite sequence* is a function $\mathbf{a}$ from $1, 2, 3, \ldots, N$ to a set. They are very important, but in calculus we deal mostly with infinite processes, so finite sequences are not used in this section on limits.

Here are some examples of infinite sequences.

Example 32. The natural numbers are a sequence with $a(n) = a_n = n$. This is written as $\{n\}_{n=1}^{\infty}$ or, being less precise, $\{1, 2, 3, \ldots\}$.

Often we write down only the first few terms of a sequence to get a feeling for the pattern. The first few terms do not, however, give a precise definition of any sequence.

Example 33. The function $a(n) = a_n = 1/n^2$ defines a sequence. This is written as $\left\{1/n^2\right\}_{n=1}^{\infty}$ or, being less precise, $\{1, 1/4, 1/9, \ldots\}$.

Example 34. If we take the distances to the wall and the time traveled from Zeno's paradox we get a sequence of vectors $\{(1/2^n, 1 - 1/2^n)\}_{n=1}^{\infty}$.

The question dealt with in this section is what happens to the sequence $\{a_n\}$ as n heads toward infinity. This is one of the questions raised by Zeno's paradox. We can look at a simpler version of the question, does a_n approach a given value, number, point or vector, as n gets large. The following definition codifies this idea.

Definition 5 (Infinite sequence). A sequence $\{\mathbf{a}_n\}_{n=1}^{\infty}$ of vectors in $\mathbb{R}^k$ *converges* to $\mathbf{L}$ if, given any distance $r > 0$, there is an N, such that when $n \geq N$, $\mathbf{a}_n$ is within the distance r of $\mathbf{L}$. If this condition is not satisfied, the sequence *diverges*.

If a sequence $\{\mathbf{a}_n\}_{n=1}^{\infty}$ converges to $\mathbf{L}$ we say that $\mathbf{L}$ is the *limit* of the sequence. We use the notation

$$\lim_{n \to \infty} \mathbf{a}_n = \mathbf{L}$$

to denote that $\mathbf{L}$ is the limit of $\{\mathbf{a}_n\}_{n=1}^{\infty}$.

Example 35. Consider the sequence $\left\{1/n^2\right\}$. This sequence converges to 0. For example, if we want to be within $r = 0.000001$ of 0, we ask which n's will make $|1/n^2 - 0| < 10^{-6}$? Rewriting this inequality gives

$$n^2 > 10^6.$$

See Fig. 2.1 on page 34 where a log scale is used on the n-axis.

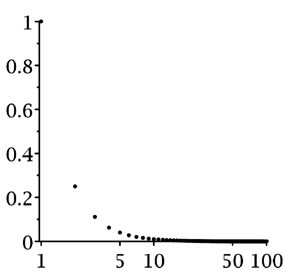
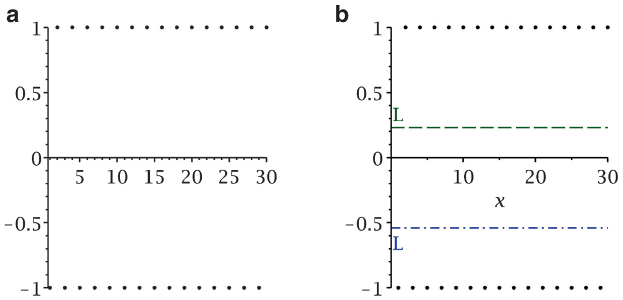


Fig. 2.1 The points $\left(n, 1/n^2\right)$



The points $(n,(-1)^n)$

The points $(n,(-1)^n)$ with potential limits

Fig. 2.2

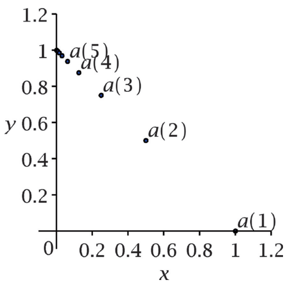


Fig. 2.3 The sequence $\{(1/2^n, 1-1/2^n)\}$

Solving for n leads to $n > 1{,}000$. Thus an N greater than 1,000 is what we need in the definition of a limit for a distance of 10^{-6}.

The next example shows that limits may not exist for some sequences.

Example 36. Consider the sequence $a_n = (-1)^n$. See Fig. 2.2a on page 35. If L is the limit of this sequence, then either $L \geq 0$ or $L \leq 0$. Since r is arbitrary, we can fix the distance to consider at $r = 1/2$. If $L \geq 0$ then, for odd n the inequality $|a_n - L| = |-1-L| \geq 1 > 1/2$ holds. Similarly, if $L < 0$, for even n we have $|a_n - L| \geq 1 > 1/2$. See Fig. 2.2a on page 35.

This means that the sequence never stays near a given L, and the limit does not exist.

Now that the idea of a limit for sequence has been defined, the sequence in Zeno's paradox can be examined more closely. Recall that the sequence is $\{(1/2^n, 1-1/2^n)\}$. See Fig. 2.3 on page 35.

The distance of the nth term, $\mathbf{a}(n)$, in the sequence from $(0,1)$ is

$$\sqrt{1/((2^n)^2) + 1/((2^n)^2)} = \sqrt{2}/2^n\ .$$

Since $2^n \geq 2n$ for $n \geq 1$, we have $\sqrt{2}/2^n < 1/n$. This implies that for any fixed r, if $N > 1/r$ and $n > N$, the distance from $(1/2^n, 1-1/2^n)$ to $(0,1)$ is less than r. Thus, the sequence converges to $(0,1)$. Does this mean that at time 1 the person has reached the opposite wall? What does this say about Zeno's paradox?

To effectively use this definition of the limit for a sequence, several basic rules are needed. They are fairly simple, so they are grouped together.

Theorem 9 (Computations with sequences). *Let* $\{\mathbf{a}_n\}$, $\{\mathbf{b}_n\}$, *and* $\{c_n\}$ *be sequences that converge to* $\mathbf{K}$, $\mathbf{L}$, *and* M *respectively. Also, let* s *be a number and assume that* $M \neq 0$. *Then the following hold:*

(i) $\lim_{n\to\infty} (\mathbf{a}_n + \mathbf{b}_n) = \mathbf{K} + \mathbf{L}$
(ii) $\lim_{n\to\infty} (s\mathbf{a}_n) = s\mathbf{K}$
(iii) $\lim_{n\to\infty} (\mathbf{a}_n \cdot \mathbf{b}_n) = \mathbf{K} \cdot \mathbf{L}$
(iv) $\lim_{n\to\infty} \dfrac{\mathbf{a}_n}{c_n} = \dfrac{\mathbf{K}}{M}$

Proof. Only the first equality is proven in the scalar case. The proofs of the other results are similar and are left to the reader.

Fix an $r > 0$. Choose N and P such that if $n > N$ and $p > P$ then $|a_n - L| < r/2$ and $|b_p - M| < r/2$. Let R be the larger of N and P. If $n > R$, then

$$\begin{aligned} |(a_n + b_n) - (L + M)| &= |(a_n - L) + (b_n - M)| \\ &\leq |(a_n - L)| + |(b_n - M)| \\ &< r/2 + r/2 \\ &< r. \end{aligned}$$

By our definition, $\{a_n + b_n\}$ converges to $L + M$.

We can use these rules to extend the sequences we can consider.

Example 37. Consider the sequence $\left\{\left(n^2+n\right)/n^2\right\}$. We have $\left(n^2+n\right)/n^2 = 1+n/n^2 = 1+1/n$. Since the sequence $\{1\}_{n=1}^{\infty}$ has limit 1 and $\{1/n\}_{n=1}^{\infty}$ converges to 0, the original sequence converges to $1+0=1$.

There are more complicated ways of doing this.

Example 38. Let

$$a_n = \frac{n+6n^3}{n^3+n^2-10n-5}.$$

Assuming that a_n exists, as it does for all large n, we can divide the numerator and denominator by n to the largest power of n in the denominator. This is n^3 for this example. Using this gives

$$a_n = \frac{\frac{1}{n^2}+6}{1+\frac{1}{n}-\frac{10}{n^2}-\frac{5}{n^3}}.$$

Since $1/n$, $1/n^2$, and $1/n^3$ all go to zero, we have

$$\lim_{n\to\infty} \frac{1}{n^2} + 6 = 6 \qquad \text{and} \qquad \lim_{n\to\infty} 1 + \frac{1}{n} - \frac{10}{n^2} - \frac{5}{n^3} = 1.$$

This means that $\lim_{n\to\infty} a_n = 6$.

The next result is used to show a sequence converges by comparing it to a convergent sequence. The proof is not difficult, but it is a little technical. The proof is omitted.

Theorem 10. *Let $\{\mathbf{a}_n\}_{n=1}^{\infty}$ be a sequence of vectors converging to $\mathbf{L}$ and let $\{\mathbf{b}_n\}_{n=1}^{\infty}$ be a sequence such that, for n after some N,*

$$\|\mathbf{b}_n - \mathbf{L}\| \leq \|\mathbf{a}_n - \mathbf{L}\|.$$

Then

$$\lim_{n\to\infty} \mathbf{b}_n = \mathbf{L}.$$

A couple examples will illustrate the uses of this result.

Example 39. It has been demonstrated that the sequence $\left\{1/n^2\right\}_{n=1}^{\infty}$ converges to 0. Since $n^2+2n-1 > n^2$ if $n > 1$,

$$0 < \frac{1}{n^2+2n-1} < \frac{1}{n^2}$$

if $n > 1$.

This means that

$$\left|\frac{1}{n^2+2n-1} - 0\right| < \left|\frac{1}{n^2} - 0\right|$$

when $n > 1$ and the sequence $\left\{1/\left(n^2+2n-1\right)\right\}_{n=1}^{\infty}$ converges to 0.

The result can also be used for vector valued sequences.

Example 40. Consider the sequence $\mathbf{b}_n = (1/n, \sin(n)/n)$. The distance from $\mathbf{b}_n$ to $\mathbf{0}$ is

$$\begin{aligned}
\left\|\left(\frac{1}{n}, \frac{\sin(n)}{n}\right) - (0,0)\right\| &= \left\|\left(\frac{1}{n}, \frac{\sin(n)}{n}\right)\right\| \\
&= \sqrt{\left(\frac{1}{n}\right)^2 + \left(\frac{\sin(n)}{n}\right)^2} \\
&= \frac{\sqrt{1+\sin^2(n)}}{n} \\
&\leq \frac{\sqrt{2}}{n}.
\end{aligned}$$

Since $\left\{\sqrt{2}/n\right\}_{n=1}^{\infty}$ converges to 0, the sequence of distances from $\mathbf{b}_n$ to $\mathbf{0}$ converges to 0. This means that $\mathbf{b}_n$ converges to $\mathbf{0}$.

We can state a slightly more general result that allows us to show convergence by "squeezing" a sequence between two convergent series. For scalar sequences, Theorem 10 on page 37 can be viewed a squeezing the sequence of b_n's between the constant sequence with terms L and a sequence with terms a_n that converges to L.

Theorem 11 (Squeeze). *Let $\{a_n\}_{n=0}^{\infty}$ and $\{b_m\}_{m=0}^{\infty}$ be two sequences that converge to some L. If $\{c_k\}_{k=0}^{\infty}$ is a sequence such that c_n is between a_n and b_n for all integers n greater than some N, then $\{c_k\}_{k=0}^{\infty}$ converges to L.*

Proof. Fix an $r > 0$. For some N, we have both $|a_n - L| < r$ and $|b_n - L| < r$ when $n > N$. Since c_n is between a_n and b_n, we have $|c_n - L| < r$ if $n > N$ and the sequence $\{c_k\}_{k=0}^{\infty}$ converges to L.

Example 41. Consider the sequences with terms $a_n = 1 - 1/\sqrt{n}$ and $b_n = 1 + 1/\sqrt{n}$. Both of these series converge to 1. Since $\sin(n)/\sqrt{n} \in [-1/\sqrt{n}, 1/\sqrt{n}]$ for all n, Theorem 11 tells us that the sequence $c_n = 1 + \sin(n)/\sqrt{n}$ converges to 1. This is illustrated in Fig. 2.4 on page 38.

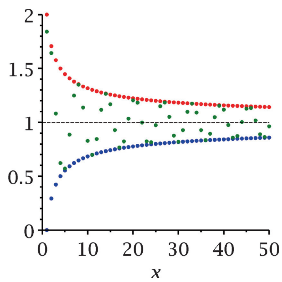


Fig. 2.4 Squeezing a sequence to a limit

A useful fact for showing that the limit of a sequence does not exist is the following result. The idea is that if the sequence always "moves" at least a minimum distance after every N, then the distance of the sequence from a fixed $\mathbf{L}$ cannot go to 0. The proof is omitted, but an example of its use is supplied.

Theorem 12. *Let $\{\mathbf{a}_n\}_{n=1}^{\infty}$ be a sequence of vectors. The sequence does not converge if and only if there is an $r > 0$ such that for every N, there are $n, m > N$ with*

$$|\mathbf{a}_n - \mathbf{a}_m| > r.$$

Example 42. Consider the sequence $a_n = \sqrt[3]{n}$. If we can show that after any $N \in \mathbb{N}$ there are k and m such that $|a_k - a_m| = 1$, this will mean that $\{a_n\}$ does not converge. Fix an integer N and let i be an integer with $i > \sqrt[3]{N}$. Then $k = i^3 > N$ and $m = (i+1)^3 > N$. We also have

$$\begin{aligned} |a_k - a_m| &= \left|a_{i^3} - a_{(i+1)^3}\right| \\ &= \left|\sqrt[3]{i^3} - \sqrt[3]{(i+1)^3}\right| \\ &= |i - (i+1)| \\ &= 1. \end{aligned}$$

This means that the sequence does not converge.

A final, but very useful result for this section concerns calculating limits of vector valued sequences. As was done with the example concerning Zeno's paradox, we can calculate the distance to the limit value using the vector norm. It is much easier to calculate the limits for each component separately. This result justifies that technique. The proof is omitted.

Theorem 13 (Sequence convergence by components). *Let $\{\mathbf{a}_n\}$ be a sequence with $\mathbf{a}_n$ having m components. Then $\lim_{n\to\infty} \mathbf{a}_n = \mathbf{L}$ if and only if each of the components of $\mathbf{a}_n$ converges to the corresponding component of $\mathbf{L}$. Writing this out gives $\lim_{n\to\infty} (\mathbf{a}_n)_i = \mathbf{L}_i$.*

Example 43. Consider the sequence with $\mathbf{a}_n = \big(1/n,\ n^2 - 1/n^2,\ n^3 - 64/\big(n^4 - 1{,}000\big)\big)$. We can calculate that

$$\lim_{n\to\infty} \frac{1}{n} = 0,$$
$$\lim_{n\to\infty} \frac{n^2-1}{n^2} = 1, \text{ and}$$
$$\lim_{n\to\infty} \frac{n^3-64}{n^4-1{,}000} = 0,$$

This means that

$$\lim_{n\to\infty} \mathbf{a}_n = (0,1,0).$$

Exercises

1. Find the limits of the following sequences as n goes to infinity.

(a) $a_n = \frac{2}{n}$

(b) $a_n = \frac{2}{n^2}$

(c) $b_n = \frac{n^{3/2}}{n}$

(d) $b_n = \frac{n^{3/2}}{n^2}$

(e) $c_n = \frac{2}{\sqrt{n}}$

(f) $c_n = \frac{(-1)^n}{n^3}$

2. Find the limits of the following sequences as n goes to infinity.

(a) $a_n = \frac{2}{n+5}$

(b) $a_n = \frac{n}{n-5}$

(c) $a_n = \frac{n-10}{n+6}$

(d) $a_n = \frac{n^2+n-1}{n^3}$

(e) $a_n = \frac{n^2+n-1}{10-n+6n^2}$

(f) $a_n = \frac{n^2+n-1}{6-n}$

3. Find the limits of the following sequences as n goes to infinity.

(a) $a_n = \frac{2}{n+5} + \frac{1}{n^2}$

(b) $a_n = \frac{n}{n-5} + \frac{3n}{2n+2}$

(c) $a_n = \left(\frac{n-10}{n+6}\right)\left(\frac{4}{n}\right)$

(d) $a_n = \left(\frac{n^2+n-1}{2n^2}\right)\left(\frac{6n+1}{n-5}\right)$

(e) $a_n = \frac{n^2+n-1}{10-n+6n^2} + \frac{n^2}{100n-1}$

(f) $a_n = \frac{n^2+n-1}{6-n}\,\frac{10n}{3n^2+4}$

4. Explain why the following sequences do not converge as n goes to infinity.

(a) $a_n = \frac{n}{1{,}000}$

(b) $a_n = \frac{n^2}{2}$

(c) $b_n = \frac{n^{3/2}}{n}$

(d) $b_n = \frac{\sqrt{n}}{\sqrt[3]{n}}$

(e) $c_n = \frac{n}{\sqrt{n}}$

(f) $c_n = \frac{(-1)^n n^2}{6n-1}$

5. Find the limits of the following sequences as n goes to infinity.

(a) $\mathbf{a}_n = \left(1, \frac{2}{n}\right)$

(b) $\mathbf{a}_n = \left(\frac{2n+1}{n}, \frac{2n+1}{n^2}\right)$

(c) $\mathbf{b}_n = \left(\frac{n+1}{\sqrt{n}}, \frac{2\sqrt{n}}{n}\right)$

(d) $\mathbf{b}_n = \left(\frac{n}{n^{3/2}}, 3, \frac{5n}{10+2n}\right)$

(e) $\mathbf{c}_n = \left(\frac{n^5}{4+n^6}, 1-\frac{2}{n}, \frac{n^{1/2}}{n+2}\right)$

(f) $\mathbf{a}_n = \left(\frac{n^{3/2}}{2n+2n^2}, \frac{\sqrt{n}-1}{4n+6}\right)$

6. Use the fact that $\left|\sin\left(\frac{1}{n}\right)\right| < 1$ and the squeeze theorem, Theorem 11, to show that the sequence defined by

$$a_n = \frac{1}{n}\sin\left(\frac{1}{n}\right)$$

converges.

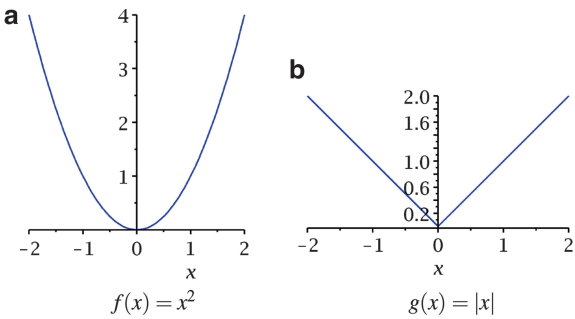


Fig. 2.5

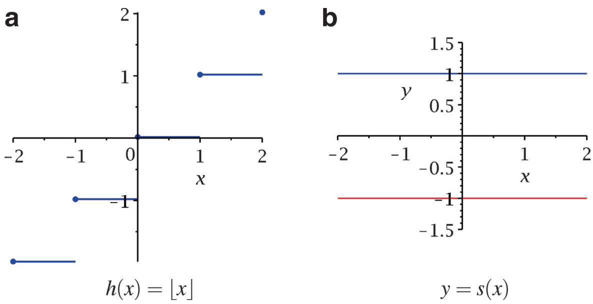


Fig. 2.6

2.2 Limits of Functions and Continuity

In these notes most of the material concerns functions from the real numbers to the real numbers or functions from the real numbers to vectors of real numbers. These functions can be nice, like $f(x) = x^2$, see Fig. 2.5a on page 40. Some are not quite as nice, $g(x) = |x|$, Fig. 2.5b on page 40,

Some have jumps. The *floor* or *greatest integer* function, $h(x) = \lfloor x \rfloor$, is the function that assigns to x the largest integer less than or equal to x, see Fig. 2.6a on page 40.

There are even rather ugly functions, see Fig. 2.6b on page 40,

$$s(x) = \begin{cases} 1 & x \text{ is rational} \\ -1 & x \text{ is irrational}\,. \end{cases}$$

It is important to note that the lines at $s = -1$ and $s = 1$ in Fig. 2.6b are not solid lines. Each has an infinite number of holes in it. These holes are offset so that each line $x = c$ hits one and only one of $s = -1$ and $s = 1$.

In this section the ideas of limits of functions and continuity of functions are used to help sort functions into reasonable and unreasonable classes of functions for calculus.

When dealing with sequences the limit was defined for a_n approaching L as $n \to \infty$. If we have a function $\mathbf{f}(x)$ defined around some point a, the question is what happens to $\mathbf{f}(x)$ as x approaches a? If we have a sequence $a_n \to a$, we can look at what happens to the sequence of function values $\mathbf{b}_n = \mathbf{f}(a_n)$. Looking at the functions above we can see that in the first two functions, no matter what sequence $\{a_n\}$ converging to $a = 0$ that we choose, the sequence $f(a_n)$ converges to 0. For the third function, what happens depends on the sequence $\{a_n\}$ chosen that approaches 0.

Example 44. Consider $f(x) = x^2$ and any sequence $a_n \to 0$. If $0 < |a_n| < 1$, then $0 < (a_n)^2 < |a_n|$. Thus $f(a_n) \to 0$.

Example 45. Consider $g(x) = |x|$ and any sequence $a_n \to 0$. Since $\big||a_n| - 0\big| = |a_n|$, the sequence $g(a_n) \to 0$.

Example 46. Consider $h(x) = \lfloor x \rfloor$ and let $a_n = (-1)^n/n$. Here $\lim_{n\to\infty} a_n = 0$, $h(a_n) = -1$ if n is odd and $h(a_n) = 0$ if n is even. This means that $\lim_{n\to\infty} h(a_n)$ does not exist.

Example 47. Consider the sequences $a_n = 1/n$, $b_n = \pi/n$, and $c_n = 1/\sqrt{n}$ that all converge to 0. All of the a_n's are rational. This means that $s(a_n) = 1$ for all n and $s(a_n) \to 1$. Similarly, all of the b_n's are irrational and $s(b_n) = -1$ for all n. This means that $s(b_n) \to -1$.

Consider the c_n's. If n is a perfect square, $n = m^2$ for some integer m, then c_n is rational. Otherwise, c_n is irrational. This means that for almost all n, $c_n = -1$. But, as n goes toward infinity, n will occasionally be a perfect square and c_n will be 1. This means that $s(c_n)$ does not converge since no matter how large N is, there are $n, m > N$ such that $g(c_n) = 1$ and $g(c_m) = -1$.

These four examples show what we want is for the function values to approach a single value. The problems with $s(x)$ mean that we need to consider all sequences, not just one, or a few, sequences. Because of this, the following is the definition of a limit used in these notes.

Definition 6 (Function limit). Let $\mathbf{f}$ be a function on an interval around a. *The interval may exclude a.* We say that the limit as x approaches a of $\mathbf{f}(x)$ equals $\mathbf{L}$ if for **all** sequences $a_n \to a$, with $a_n \neq a$ for all n, we have

$$\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L}.$$

A common notation for the limit is

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

The four examples before this definition show that $\lim_{x\to 0} f(x)$ and $\lim_{x\to 0} g(x)$ exist whereas $\lim_{x\to 0} h(x)$ and $\lim_{x\to 0} s(x)$ do not exist. Showing that $f(x)$ and $g(x)$ have limits at $x = 0$ was fairly difficult. Demonstrating that $h(x)$ and $s(x)$ do not have limits was much easier since all that was required was finding *a* sequence $a_n \to 0$ where $h(a_n)$ or $s(a_n)$ does not converge.

Remark 4. An intuitive way of interpreting this definition is that the limit of $\mathbf{f}$ as x approaches a is $\mathbf{L}$, if, no matter how we approach a, the function values always go to $\mathbf{L}$.

For the function $h(x)$ from above, if we approach 0 from the negative side, the function values go to -1. If we approach 0 from the positive side, the values of $h(x)$ go to 0. Since the values of $h(x)$ can approach -1 or 0, the limit does not exist.

The condition for a limit of a function to not exist is stated as the following theorem. The proof is omitted.

Theorem 14. *Limit does not exist by sequence Let* $\mathbf{f}(x)$ *be a function from* $\mathbb{R}$ *to* $\mathbb{R}^n$*. The limit* $\lim_{x\to a}\mathbf{f}(x)$ *does not exist if and only if there is a sequence* $x_m \to a$ *such that the sequence* $\mathbf{f}(x_m)$ *does not converge.*

As an example to illustrate this consider the floor function.

Example 48. Take $h(x) = \lfloor x \rfloor$ and let $a_n = (-1)^n/n$ for $n \geq 2$. Then $\lim_{n\to\infty} a_n = 0$. If n is odd $a_n = -1/n$ is between -1 and 0. This means that $-1 \leq a_n < 0$ and $h(a_n) = -1$.

Similarly, if n is even $0 < a_n < 1$. This implies that $h(a_n) = 0$. Combining the last two conclusions we have

$$h(a_n) = \begin{cases} -1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

This sequence cannot converge since $|h(a_{n+1}) - h(a_n)| = 1$ for all n. This means that $\lim_{x\to 0} h(x)$ does not exist.

Working directly from the definition to find a limit is not easy. However, with a few rules, many limits are easy to calculate. The rules are very similar to the rules for limits of sequences.

Theorem 15 (Function limit rules). *Let* $\mathbf{f}(x)$ *and* $\mathbf{g}(x)$ *be functions from* $\mathbb{R}$ *to* $\mathbb{R}^m$ *with limits* $\mathbf{K}$ *and* $\mathbf{L}$ *at* $x = a$*, and let* $h(x)$ *be a function from* $\mathbb{R}$ *to* $\mathbb{R}$ *with limit* M *at* a*. Then the following hold:*

(i) $\lim_{x\to a}(\mathbf{f}+\mathbf{g})(x) = \mathbf{K}+\mathbf{L}$

(ii) $\lim_{x\to a} h(x)\mathbf{f}(x) = M\mathbf{K}$

(iii) $\lim_{x\to a}(\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{K}\cdot\mathbf{L}$

(iv) *If* $M \neq 0$*, then* $\lim_{x\to a}\dfrac{\mathbf{f}(x)}{h(x)} = \mathbf{K}/M$

Proof. All of the proofs are similar, so only (i) is considered. Let a_n be any sequence converging to a. Then the sequences $\mathbf{f}(a_n)$ and $\mathbf{g}(a_n)$ converge to $\mathbf{K}$ and $\mathbf{L}$ respectively. Fix an $r > 0$. For some N_1 and N_2, if $n_1 > N_1$ and $n_2 > N_2$, we have $\|\mathbf{f}(a_{n_1}) - \mathbf{K}\| < r/2$ and $\|\mathbf{g}(a_{n_2}) - \mathbf{L}\| < r/2$.

Assume that $N = \max\{N_1, N_2\}$ and let $n > N$. Then, if $n > N$,

$$\begin{aligned} \|(\mathbf{f}+\mathbf{g})(a_n) - (\mathbf{K}+\mathbf{L})\| &= \|(\mathbf{f}(a_n) - \mathbf{K}) + (\mathbf{g}(a_n) - \mathbf{L})\| \\ &\leq \|\mathbf{f}(a_n) - \mathbf{K}\| + \|\mathbf{g}(a_n) - \mathbf{L}\| \\ &< r/2 + r/2 = r. \end{aligned}$$

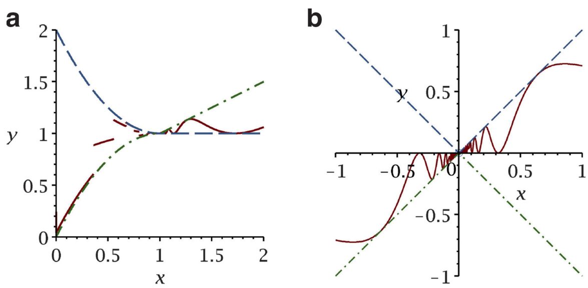
By the definition of convergence of sequences, $(\mathbf{f}+\mathbf{g})(a_n)$ converges to $\mathbf{K}+\mathbf{L}$. Since the sequence was arbitrary, this holds for all sequences, and the limit is $\mathbf{K}+\mathbf{L}$.

It should be fairly clear that a constant function $f(x) = c$ has limit c at any point a. Somewhat harder, but not too hard, is the fact that $f(x) = x$ has limit a at any point a. We can use this in a simple application of Theorem 15.

Example 49. Using the just stated facts about constant functions and $f(x) = x$, we can show that $\lim_{x\to a} cx^2 = ca^2$. First, since $x^2 = x \cdot x$, the product of $f(x)$ with itself, we have

$$\lim_{x\to a} cx^2 = (\lim_{x\to a} c)(\lim_{x\to a} x)(\lim_{x\to a} x) = c \cdot a \cdot a = ca^2.$$

We also have an analog of Theorem 11 on page 37. Assume we have functions $f(x)$, $g(x)$, and $h(x)$ defined around some point $x = a$ such that $h(x)$ is between $f(x)$ and $g(x)$ near a and such that $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = L$. For any sequence $x_n \to a$, after some N, the values of the sequence $h(x_n)$ are between $f(x_n)$ and $g(x_n)$. Since $\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} g(x_n) = L$, Theorem 11 applies and the sequence of values $h(x_n)$ converges to L. This implies that $\lim_{x\to a} h(x_n) = L$. An illustration of this idea is in Fig. 2.7a on page 43. We state this as the following theorem.



Idea of the squeeze theorem   Squeezing $x \sin^2(1/x)$

Fig. 2.7

Theorem 16 (Squeeze). *Assume the functions $f(x)$, $g(x)$, and $h(x)$ are defined for all points with $|x-a| \in (0,\delta)$ for a fixed $\delta > 0$. Also assume that $h(x)$ is between $f(x)$ and $g(x)$ for all x with $|x-a| \in (0,\delta)$ and assume that $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = L$. Then $\lim_{x\to a} h(x) = L$.*

A simple example illustrates how this works.

Example 50. Let $h(x) = x\sin^2(1/x)$, $f(x) = |x|$, and $g(x) = -|x|$. Here, since $|\sin^2(y)| \in [-1,1]$ for all $y \in \mathbb{R}$, we have $h(x)$ is between $f(x)$ and $g(x)$ for all $x \neq 0$. See Fig. 2.7b on page 43 for a graph of the functions. This means that, since $\lim_{x\to 0} |x| = \lim_{x\to 0} -|x| = 0$, $\lim_{x\to 0} h(x) = 0$.

It is also useful to evaluate limits of compositions of functions. The next result shows how this is done.

Theorem 17 (Limit of Composition). *Assume $f(x)$ has a limit L as $x \to a$ and assume $g(y)$ is defined on an interval around L and has a limit $M = g(L)$ as $y \to L$, then $g \circ f$ has a limit as $x \to a$ and*

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

Proof. Assume that x_n is any sequence such that $x_n \to a$ and no $x_m = a$. Then the sequence $y_n = f(x_n)$ is defined after some N and y_n converges to L. Thus, the sequence $g(y_n) = g(f(x_n))$ converges to M for any sequence $x_n \to a$ with $x_n \neq a$ for all m. This means that $\lim_{x\to a}(g \circ f)(x) = M$.

A couple examples will help understand this result.

Example 51. Let $g(y) = y^2 + 4$ and let $f(x) = \left(x^2 - 1\right)/(x+1)$. Here $\lim_{x\to -1} f(x) = -2$ and $\lim_{y\to -2} g(y) = 8$. Using Theorem 17, we have

$$\lim_{x\to -1}\left(\left(\frac{x^2-1}{x+1}\right)^2 + 4\right) = 8\ .$$

Example 52. Here the limit exists without all of the criteria being met. The assumptions of Theorem 17 are not met at $x = 0$ if $g(y) = \cos(y)$ and $f(x) = 4\pi\left(\lfloor x \rfloor - 1/2\right)$ since $\lim_{x\to 0} f(x)$ does not exist. However, since $f(x) = 2\pi$ if $x \in [0,1)$ and $f(x) = -2\pi$ is $x \in [-1,0)$, $\cos(f(x)) = 1$ on $[-1,1)$ and

$$\lim_{x\to 0}(g\circ f)(x) = 1\ .$$

Example 53. Here the limit does not exist when the criteria of the theorem are not met. Consider the function $g(y) = \left(y^2 - 1\right)/(y+1)$ with $\lim_{y\to -1} g(y) = -2$ where $g(-1)$ does not exist and consider the function $f(x) = -1$ that has $\lim_{x\to -a} f(x) = -1$ for any real number a. Since $g(f(x))$ does not exist anywhere, the composition $g\circ f$ cannot have a limit as $x \to a$ for any a.

As you will see in a couple sections, avoiding the point a in the definition of $\lim_{x\to a}\mathbf{f}(x) = \mathbf{L}$ is necessary. However, the nice functions $f(x) = x^2$ and $g(x) = |x|$ satisfy the property that $\lim_{x\to 0} f(x) = 0 = f(0)$ and $\lim_{x\to 0} g(x) = 0 = g(0)$. This nice property is captured in the definition of continuity.

Definition 7 (Continuous function). A function $\mathbf{f}(x)$ is *continuous* at a point a if $\lim_{x\to a}\mathbf{f}(x) = \mathbf{f}(a)$. If the function is continuous at every point of an interval, $\mathbf{f}$ is said to be continuous on that interval.

A function that is not continuous at a point a is said to be *discontinuous* at a.

Example 54. Again consider the floor function, see Fig. 2.6a on page 40. At every integer n, $\lim_{x\to n}\lfloor x \rfloor$ does not exist. This means that the floor is discontinuous at every integer. On the other hand, since $\lfloor x \rfloor$ is constant between consecutive integers, the floor function is continuous at every point that is not an integer.

The idea that a limit of a sequence $\mathbf{b}_n$ is taken coordinate by coordinate is used to prove the following theorem. This theorem is used for evaluating limits of vector valued functions and for deciding on the continuity of vector-valued functions.

Theorem 18 (Continuous vector valued functions).
Let $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_m(x))$ *be defined on an interval around a. The limit* $\lim_{t\to a}\mathbf{f}(t) = \mathbf{L}$ *if and only if* $\lim_{t\to a} f_i(t) = L_i$ *for* $i = 1,2,\ldots,m$. *The function* $\mathbf{f}(x)$ *is continuous at a if and only if* $f_i(x)$ *is continuous at a for* $i = 1,2,\ldots,m$.

Example 55. Let $\mathbf{f}(t) = (4t^2, 1, t)$. Then

$$\begin{aligned}\lim_{t\to 2}\mathbf{f}(t) &= \left(\lim_{t\to 2} 4t^2,\ \lim_{t\to 2} 1,\ \lim_{t\to 2} t\right)\\ &= (16,1,2).\end{aligned}$$

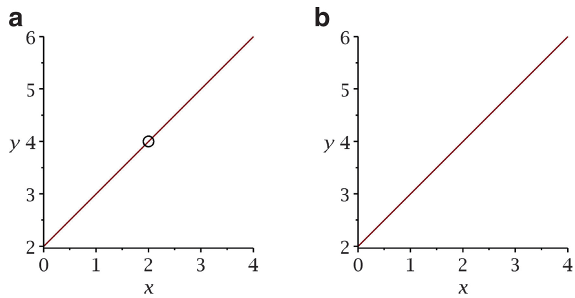
Many elementary functions are continuous at every point in their domains. It is not demonstrated now that these are continuous, but any polynomial is continuous on $\mathbb{R}$, $\sin(x)$ and $\cos(x)$ are continuous on $\mathbb{R}$, e^x is continuous on $\mathbb{R}$, and $\ln(x)$ is continuous on $(0,\infty)$.

Using the theorem on limits above, we can prove the following result. The proof is left to the ambitious reader.

Theorem 19 (Continuous vector valued functions, rules). *Let* $\mathbf{f}(x)$, $\mathbf{g}(x)$, $r(x)$ *and* $h(x)$ *be functions that are continuous at a. Assume that* $h(a) \neq 0$. *Then the following hold:*

(i) $\mathbf{f}+\mathbf{g}$ *is continuous at a,*
(ii) $r\mathbf{f}$ *is continuous at a,*
(iii) $\mathbf{f}\cdot\mathbf{g}$ *is continuous at a, and*
(iv) $\dfrac{\mathbf{f}}{h}$ *is continuous at a.*

This theorem can be used to demonstrate that many elementary functions are continuous on their domains. For example, the polynomials are continuous on $\mathbb{R}$, any rational function is continuous on its domain, and $\tan(x)$, $\sec(x)$, $\cot(x)$, and $\csc(x)$ are continuous on their domains.



The discontinuous function $y = (x^2-4)/(x-2)$

The continuous function $y = x+2$

Fig. 2.8

Example 56. Consider $\tan(x) = \dfrac{\sin(x)}{\cos(x)}$. Since both $\sin(x)$ and $\cos(x)$ are continuous everywhere, $\tan(x)$ is continuous wherever $\cos(x) \neq 0$. This is exactly where $\tan(x)$ is defined, $\{x|x \neq n\pi/2 \text{ where n is an odd integer.}\}$.

Example 57. Consider the functions $\mathbf{f}(x) = (\sin(x), x^2, 4)$ and $h(x) = x^2+1$. Using the theorem above we have that the function $\mathbf{f}(x)/h(x)$ is continuous at all x.

There are also many functions that have points where they are discontinuous.

Example 58. The function

$$s(x) = \begin{cases} 1 & x \text{ is rational} \\ -1 & x \text{ is irrational}\ . \end{cases}$$

is discontinuous everywhere. We will not prove it here, but every open interval in the real line contains both rational and irrational numbers. This means that $s(x)$ takes on the values -1 and 1 in every interval and the limit $\lim_{x\to a} s(x)$ does not exist for any a.

A more common case in elementary calculus is the case where a function is not continuous since it does not exist at a point. Sometimes we can make a function continuous at a point where it is discontinuous by changing the value of the function or adding a value for the function at the point. Such a discontinuity is called a *removable discontinuity*.

Example 59. Consider the function $f(x) = \left(x^2-4\right)/(x-2)$. This function is discontinuous at $x=2$ since $f(x)$ is not defined at $x=2$. However, if $x \neq 2$

$$\begin{aligned} f(x) &= \frac{(x-2)(x+2)}{x-2} \\ &= x+2. \end{aligned}$$

By adding the point $(2,4)$ to the graph of $f(x)$ we get a new function $g(x) = x+2$ that is continuous at $x=2$. This means the discontinuity is a removable discontinuity. See Fig. 2.8 on page 45

A discontinuity that is not removable is called an *essential discontinuity*. Two simple examples of essential discontinuities are the floor function at every integer and the function $f(x) = 1/x$ at $x=0$.

The following result allows us to show that many more functions are continuous.

Theorem 20 (Continuity of compositions). *Let $f(x)$ be defined on an interval around a and assume that $f(x)$ is continuous at $x=a$. Also let $\mathbf{g}(y)$ be defined on an interval around $f(a)$ and be continuous at $y=f(a)$. Then the composition $(\mathbf{g} \circ f)(x)$ is continuous at $x=a$.*

Proof. Let a_n be a sequence of points converging to a. After some N_1, $b_n = f(a_n)$ is defined and $b_n \to f(a)$ by the definition of continuity. Since $b_n \to f(a)$, $\mathbf{c}_n = \mathbf{g}(b_n) = (\mathbf{g} \circ f)(a_n)$ is defined after some N and c_n converges to $\mathbf{g}(f(a))$. Therefore, $\mathbf{g} \circ f$ is continuous at a.

This is very useful when combining functions. It will be used without stating the result throughout this chapter.

Example 60. The function $f(x) = \sin(2x+10)$ is continuous everywhere. This follows from Theorem 20 since both $g(y) = \sin(y)$ and $h(x) = 2x+10$ are continuous everywhere.

There are several important applications of continuity that will be considered later. Here is one consequence of continuity that will be used when derivatives are discussed.

Theorem 21. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous at a and assume that $f(a) > 0$ (or $f(a) < 0$). Then there is an interval (b,c) containing a such $f(x) > 0$ (or $f(x) < 0$) on (b,c).*

Proof. Assume that, to the contrary, that $f(x)$ is continuous at a, $f(a) > 0$, and there is a sequence $a_n \to a$ such that $f(a_n) \leq 0$. Then $|f(a_n) - f(a)| \geq f(a) > 0$. This means that $\lim_{x \to a} f(x)$ either does not exist or is not equal to $f(a)$. Since $f(x)$ is continuous at a, neither of these conclusions is true and $f(x) > 0$ on some open interval containing a.

Example 61. Let $f(x) = \sin(x)$. Since this function is continuous and $f(\pi/2) > 0$, $\sin(x) > 0$ on an interval containing $\pi/2$. We can use the interval $(0, \pi)$.

Example 62. Consider the function

$$f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases},$$

See Fig. 2.9 on page 47

This function has $f(0) > 0$ but is not continuous at 0. It is clear that there are points x close to 0 where $f(x) = -1 < 0$.

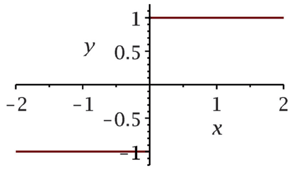


Fig. 2.9 $y = f(x)$ from Example 62

This section concludes with several examples concerning numerical computation of limits and concerning discontinuities and graphs. It is common to use a table of values to estimate the limit of a function $f(x)$ as $x \to a$. This is similar to saying the limit of a sequence exists by considering only the first few terms of a sequence of a_n's. The next two examples show that this may give the impression that there is a limit, even when no limit exists.

First consider a sequence with no limit.

Example 63. Consider the sequence $\{a_n\}_{n=0}^{\infty}$ with $a_n = 1$ if $n \le 10{,}000$ and with $a_n = (-1)^n$ if $n > 10{,}000$. If we look only at the first 1,000 elements, the limit appears to be 1. However, the limit does not exist since the a_n's alternate between -1 and 1 after $n = 10{,}000$.

For functions the situation is more complex. Some sequences of function values at points converging to the a of interest may converge when others diverge or converge to different values.

Example 64. Consider the function $f(x) = \sin(\pi/x)$, Fig. 2.10 on page 47.

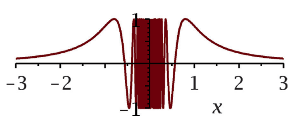


Fig. 2.10 $f(x) = \sin(\pi/x)$

If we choose any sequence of the form $a_n = 1/m(n)$ where $m(n)$ is an increasing sequence of integers that goes to infinity, for example $a_n = 10^{-n}$, then $f(a_n) = 0$ for all n. From the graph, this does not capture the behavior of the function. In fact, if we take $a_n = 1/(2n + 1/2)$, we $f(a_n) = 1$ for all n. This is a different possible limit.

Since the sample sequences do not converge to the same value, the function does not have a limit as x goes to 0.

The previous example shows that using only one sequence to show that a limit exists does not work. If we know that the of a limit function exists, we can use any sequence to estimate the limit.

Theorem 22. *If* $\lim_{x\to a} \mathbf{f}(x) = \mathbf{L}$, *then for any sequence* $a_n \to a$, *we have*

$$\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L}.$$

Here is how we can use this theorem.

Example 65. Let $f(x) = \cos(\pi x)$ and consider the sequence $a_n = (4^{n-1} - 1)/4^n$. We can show that $|a_n - 1/4| = 1/4^n$ and that $a_n \to 1/4$. Since $f(x)$ is continuous everywhere, $f(a_n) \to f(1/4) = \cos(\pi/4) = \sqrt{2}/2$. Here The values for the first 8 terms of the sequence are in Table 2.1 on page 48. From the table we can guess that the limit is approximately 0.707. Since this is $\sqrt{2}/2$ to three decimal places, it is correct.

x	0	$\frac{3}{16}$	$\frac{15}{64}$	$\frac{63}{256}$	$\frac{255}{1{,}024}$	$\frac{1{,}023}{4{,}096}$	$\frac{4{,}095}{16{,}384}$	$\frac{16{,}383}{65{,}536}$
$f(x)$	1	0.831470	0.740951	0.715731	0.709273	0.707649	0.707242	0.707141

Table 2.1 Values of $\cos(\pi x)$ as x approaches $1/4$

t	$1/2$	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}
$\sin(t)/t$	0.958851	0.989616	0.997398	0.999349	0.999837	0.999959	0.999990	0.999997
$(\cos(t)-1)/t^2$	−0.4896698	−0.497401	−0.499349	−0.499837	−0.499959	−0.5000	−0.499990	−0.500

Table 2.2 Values of $\left(\sin(t)/t, (\cos(t)-1)/t^2\right)$ as t approaches 0

Example 66. Consider the case of the function $\mathbf{r}(t) = \left(\sin(t)/t, (\cos(t)-1)/t^2\right)$. It is known that both $\lim_{t\to 0} \sin(t)/t$ and $\lim_{t\to 0} (\cos(t)-1)/(t^2)$ exist. (These limits are considered in Sect. 7.4.) To get an estimate for the values of the limits we can create Table 2.2 on 48.

From this table we can conclude that $\lim_{t\to 0} \sin(t)/t$ is probably 1 and that $\lim_{t\to 0} (\cos(t)-1)/t^2$ is probably $-1/2$. This means that $\lim_{t\to 0} \mathbf{r}(t)$ is probably $(1, -1/2)$.

Frequently we can see discontinuities and possible discontinuities from the graph of a function. The three basic things that can cause a function $f(x)$ to be discontinuous at a are that the function may not be defined at a, the limit of $f(x)$ does not exist at a, and the limit exists but does not equal $f(a)$.

It is important to remember that graphs are only partial representations of a function. As such, not all the features of a function may be visible. Therefore we must do more than look at a graph to guarantee that a function is continuous at a point.

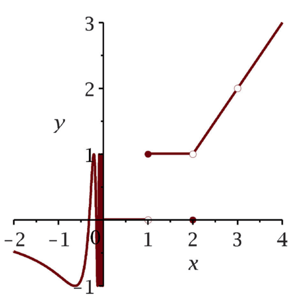


Fig. 2.11 The graph of a function with different types of discontinuities

Example 67. Figure 2.11 on page 48 is the graph of a function with several discontinuities.

The apparent discontinuities are at $x = 0, 1, 2,$ and 3. *If we assume that the graph represents all of the behavior of $f(x)$ we can make some conclusions.* At both $x = 0$ and $x = 1$ the limit of $f(x)$ does not exist. This means that $f(x)$ is not continuous at either point.

At both $x = 2$ and $x = 3$ the limit of $f(x)$ exists The value $f(2) = 1$ does not equal $\lim_{x\to 2} f(x) = 0$. Since $f(x)$ is not defined at $x = 3$, it is not continuous there. The discontinuities at $x = 2$ and $x = 3$ are removable discontinuities.

Exercises

1. Find the following limits of functions.

(a) $\lim_{x\to 3} 4$

(b) $\lim_{x\to -2} \frac{\pi}{4}$

(c) $\lim_{x\to 7} \frac{x}{2}$

(d) $\lim_{z\to -1} 2z$

(e) $\lim_{z\to \frac{3}{2}} 3z-2$

(f) $\lim_{z\to 0} \frac{3-2z}{5}$

(g) $\lim_{w\to 0} 10w-6$

(h) $\lim_{w\to 1} (3w+2)^2$

(i) $\lim_{w\to -1} w^2-3w+2$

(j) $\lim_{y\to 0} y^3+7y-2$

(k) $\lim_{y\to -2} y^4-y^6$

(l) $\lim_{y\to 0} \frac{y^2-2}{3-y}$

(m) $\lim_{t\to \pi} \frac{\cos(t)}{t^2+2}$

(n) $\lim_{w\to 1} (w^1-3, w+1, w^2)$

(o) $\lim_{y\to 0} (y^2-6y^3, 3y-5, 4y^2-1)$

(p) $\lim_{y\to 1} (y^4-y^6+1, -6, 5y+10)$

(q) $\lim_{y\to 2} \left(\frac{y^2-2}{3-y}, \frac{6y^2-1}{y^3+2}\right)$

2. Find the limits of the following functions.

(a) $\lim_{x\to 1} \ln(3x-2)$

(b) $\lim_{t\to 2} \cos(t^2+t)$

(c) $\lim_{y\to \pi} \tan\left(\frac{y}{2}-\frac{\pi}{4}\right)$

(d) $\lim_{x\to -1} e^{2x+2}$

(e) $\lim_{w\to 0} \sec\left(w^2+\frac{\pi}{6}\right)$

(f) $\lim_{t\to 6} \left(t^2-7t+1\right)^{110}$

(g) $\lim_{x\to -1} \left(e^{2x+2}, \cos(\pi x), \ln(x+2)\right)$

(h) $\lim_{w\to 0} \left(e^{w+1}, \sec\left(w^2+\frac{\pi}{6}\right), w^3+3\right)$

(i) $\lim_{t\to 6} \left(\frac{t+2}{t-3}, \left(t^2-7t+1\right)^{110}\right)$

3. Show that the following limits do not exist.

(a) $\lim_{x\to 0} \cos\left(\frac{1}{x}\right)$

(b) $\lim_{x\to 0} \frac{1}{x}$

(c) $\lim_{x\to \frac{\pi}{2}} \tan(x)$

(d) $\lim_{x\to -1} \frac{1}{1-e^{-(2x+2)}}$

(e) $\lim_{x\to 0} \cot(x)$

(f) $\lim_{x\to 0} \left(\frac{1}{x^2}\right)$

(g) $\lim_{x\to 0} \left(x^2+1, \cot(x)\right)$

(h) $\lim_{x\to 0} \left(\frac{1}{x^2}, \cos(x^2)\right)$

4. Estimate the following limits numerically.

(a) $\lim\limits_{x\to 2} \dfrac{x^2-4}{x^2+x-6}$

(b) $\lim\limits_{x\to 0} \dfrac{\cos(x)}{x^2}$

(c) $\lim\limits_{x\to \pi} \dfrac{\sin(x)}{x-\pi}$

(d) $\lim\limits_{x\to 3} \dfrac{x^3-27}{x^2-9}$

(e) $\lim\limits_{x\to \pi} \left(\dfrac{\sin(x)}{x-\pi}, \sqrt{x}\right)$

(f) $\lim\limits_{x\to 3} \left(\sin\left(\dfrac{\pi x}{18}\right), \dfrac{x^3-27}{x^2-9}\right)$

5. Where are the following functions continuous?

(a) $f(x) = x^2+3x-6$

(b) $g(x) = \cot(x)$

(c) $h(x) = \dfrac{x+2}{x^2-1}$

(d) $f(x) = \dfrac{x}{\cos(x)}$

(e) $g(x) = \left\lfloor \dfrac{x}{2} \right\rfloor$

(f) $\mathbf{h}(w) = \left(w^5-w^2, \sin(w)+\cos(2w)\right)$

(g) $\mathbf{g}(w) = \left(\left\lfloor \dfrac{w}{2} \right\rfloor, \tan(w)\right)$

(h) $\mathbf{h}(r) = \left(\dfrac{1}{r^2-4}, |r|\right)$

6. Are the discontinuities of the following functions at the given points removable discontinuities?

(a) $f(x) = \lfloor x \rfloor,\ x = 2$

(b) $g(h) = \dfrac{(3+h)^2-9}{h},\ h = 0$

(c) $h(x) = \dfrac{1}{x},\ x = 0$

(d) $f(w) = \dfrac{w}{\cos(w)},\ w = \pi/2$

(e) $g(w) = \dfrac{(w-1)^3+1}{w},\ w = 0$

(f) $h(x) = \dfrac{x+1}{x^2-1},\ x = 1$

(g) $h(x) = \dfrac{x^2-5x+6}{x^2-4x+4},\ x = 2$

(h) $h(x) = \dfrac{x^2+3x-4}{x^3-3x^2+3x-1},\ x = 1$

7. Where are the following functions continuous?

(a) $f(x) = \begin{cases} x^2 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$

(b) $g(t) = \begin{cases} 4t-2 & \text{if } t \leq 1 \\ t+2 & \text{if } t > 1 \end{cases}$

(c) $h(z) = \begin{cases} 4-z^2 & \text{if } |z| < 2 \\ z^2-4 & \text{if } |z| \geq 2 \end{cases}$

(d) $s(w) = \begin{cases} w^3-4w & \text{if } w < 2 \\ w-1 & \text{if } w \geq 2 \end{cases}$

8. Use the squeeze theorem, Theorem 16 on page 43, to show that if

$$g(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ -x & \text{if } x \text{ is irrational}, \end{cases}$$

then $g(x)$ is continuous at $x = 0$. How does this relate to the idea of a function being continuous at $x = a$ "if you can draw the graph of the function without taking your pencil off the paper?"

2.3 Rates of Change and Derivatives

Differential calculus can be considered to be a study of how things change. As the first part of this study of change in this book, the idea of instantaneous rate of change is considered. The preliminary example here is that of a car driving along a straight east-west road. At a time, say t_0, how fast is the car going?

Assume that the position of the car is know as a function of time, say $r(t)$. Here r is the signed distance from a fixed point on the road, positive for east and negative for west. Consider the distance traveled from time t_0 to time t_1, $r(t_1) - r(t_0)$. If the car is traveling at a constant velocity, that velocity is

$$v = \frac{r(t_1) - r(t_0)}{t_1 - t_0}.$$

The *velocity* is the speed of the car with a sign to determine direction. In this case, if the car is moving east in the positive direction, the velocity is positive, and if the car is moving west in the negative direction, the velocity is negative.

If the car is not traveling at a constant velocity, we say that the *average velocity* of the car from t_0 to t_1 is

$$v_{ave} = \frac{r(t_1) - r(t_0)}{t_1 - t_0}.$$

This is the constant velocity the car would need to travel from $r(t_0)$ to $r(t_1)$ in the same length of time, $t_1 - t_0$. The graph in Fig. 2.12 on page 51 illustrates this idea.

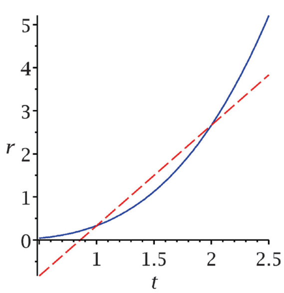


Fig. 2.12 A secant line representing average velocity

The solid curve represents the position $r(t)$ and the dashed line represents position with a constant velocity that has the same distance values as $r(t)$ at $t = 1$ and $t = 2$. The line including the points $(t_0, r(t_0))$ and $(t_1, r(t_1))$ is called the *secant line* to the graph of $r(t)$ through $(t_0, r(t_0))$ and $(t_1, r(t_1))$.

The ratio used to find an average rate of change for any function

$$\frac{\mathbf{f}(y)-\mathbf{f}(x)}{y-x}$$

is called a *difference quotient*. It is often written in the form

$$\frac{\mathbf{f}(x+h)-\mathbf{f}(x)}{h} \tag{2.1}$$

by replacing y with $x+h$. This quotient is fundamental for differential calculus.

Example 68. Assume the position of a car is given by $r(t) = 50t + 3\cos(t/5)$. The average velocity of the car from $t = 0$ to $t = 2$ is given by

$$\begin{aligned} v_{\text{avg}} &= \frac{r(2)-r(0)}{2-0} \\ &= \frac{50\cdot 2 + 3\cos(2/5) - (50\cdot 0 + 3\cos(0))}{2} \\ &= 50 + \frac{3\cos(2)-3}{2} \\ &\approx 47.88. \end{aligned}$$

If this is a real car, we assume that the average velocity over the interval from t_0 to t_1 gets closer to the instantaneous velocity *at* t_0 as t_1 gets closer to t_0. Assuming that this is true, we define the *instantaneous velocity* of the car at t_0 as

$$v(t_0) = \lim_{t_1\to t_0} \frac{r(t_1)-r(t_0)}{t_1-t_0}.$$

The same definition is used for vector valued position functions to define velocity and to define acceleration for a vector valued velocity function. *Acceleration* is the rate of change, derivative, of velocity.

Example 69. Let the position function of a mass be given by $r(t) = -16t^2 + 20t + 6$. The velocity at $t = 1$ is given by

$$\begin{aligned} v(1) &= \lim_{t\to 1} \frac{r(t)-r(1)}{t-1} \\ &= \lim_{t\to 1} \frac{-16t^2+20t+6-(-16+20+6)}{t-1} \\ &= \lim_{t\to 1} \frac{-16(t^2-1)+20(t-1)}{t-1} \\ &= \lim_{t\to 1} -16(t+1)+20 \\ &= -16\,(1+1)+20 \\ &= -12. \end{aligned}$$

As you will see in the next section, the velocity of the mass is $v(t) = -32t+20$. This allows us to calculate the acceleration of the mass at $t = 1$, $a(1)$:

$$\begin{aligned} a(1) &= \lim_{t\to 1} \frac{v(t)-v(1)}{t-1} \\ &= \lim_{t\to 1} \frac{-32t+20-(-32+20)}{t-1} \\ &= \lim_{t\to 1} \frac{-32(t-1)}{t-1} \\ &= -32. \end{aligned}$$

We can do the same type of example with vector valued motion.

Example 70. Assuming there is no friction, the position of a projectile fired from $(0,0)$ at an angle θ above horizontal with an initial speed of $v\,\mathrm{m/sec}$ in the positive x direction can be written as

$$\mathbf{r}(t) = \big(v\,t\,\cos(\theta), -9.8t^t + v\,t\,\sin(\theta)\big)\ .$$

The average velocity over the interval from $t = 0$ to $0+h$ is

$$\begin{aligned}
\mathbf{v}_{ave} &= \frac{\big(v\,h\cos(\theta), -9.8h^2 + v\,h\sin(\theta)\big) - (0,0)}{h} \\
&= \frac{v\,h\big(\cos(\theta),\sin(\theta)\big) - 9.8\,h^2\,(0,-1)}{h} \\
&= v\,\big(\cos(\theta),\sin(\theta)\big) - 9.8\,h\,(0,-1)\ .
\end{aligned}$$

Taking the limit as $h \to 0$ of $\mathbf{v}_{ave}$ we get

$$\mathbf{v}(0) = v\,\big(\cos(\theta),\sin(\theta)\big)\ .$$

This gives us that the velocity is a vector of length v, the speed, in the direction of travel, $\big(\cos(\theta),\sin(\theta)\big)$.

Following this idea for general functions we define the derivative of a vector (scalar) valued function.

Definition 8 (Derivative). Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ be a function that is defined on an interval $(a-r, a+r)$ with $r > 0$. If the limit exists, the *derivative* of f at a is defined as

$$\frac{d\,\mathbf{f}(x)}{dx}(a) = \lim_{x\to a} \frac{\mathbf{f}(x) - \mathbf{f}(a)}{x-a}.$$

The derivative of $\mathbf{f}$ at a is also denoted by

$$D\mathbf{f}(a), \quad \mathbf{f}'(a), \quad \text{and} \quad \frac{d\,\mathbf{f}(a)}{dx}.$$

For a few functions this is easy to calculate.

Example 71. Consider any function of the form $\mathbf{f}(x) = \mathbf{m}x + \mathbf{b}$. The derivative at any point a is given by

$$\begin{aligned}
\mathbf{f}'(a) &= \lim_{x\to a} \frac{(\mathbf{m}x+\mathbf{b}) - (\mathbf{m}a+\mathbf{b})}{x-a} \\
&= \lim_{x\to a} \frac{(\mathbf{m}\,(x-a))}{x-a} \\
&= \mathbf{m}.
\end{aligned}$$

This is the slope of the line $y = mx+b$ if m and b are scalars and it is the velocity vector that includes speed and direction of travel if $\mathbf{f}(x)$ is a position function.

This gives a common interpretation of the derivative of a scalar valued function f at a. It is the slope of the tangent line to the graph of f at $(a, f(a))$. This can also be seen by looking at the "limit" of secant lines.

Example 72. Consider the function $f(x) = x^3 - x^2/2 - x/2$ at the point $x_0 = 1$. We can show that $f'(1) = 3/2$. We can look at the slopes of the secant lines for the f between the points $(1, f(1))$ and $(1 + (1/2)^n, f(1 + (1/2)^n))$. The picture indicates that the lines "approach" the tangent line to the graph of f at $(1,0)$. Figure 2.13 on page 54 is a graph that includes that tangent line and the first five secant lines.
A table of slopes, Table 2.3 on page 54, also indicates that the slopes of the secant lines approach $3/2$.

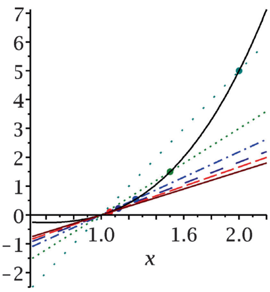


Fig. 2.13 Secant lines with slopes decreasing to the slope of a tangent line

$x+h$	2.	1.50000	1.25000	1.12500	1.06250	1.03125	1.01562	1.00781	1.00391	1.00195
Slope	5.	3.	2.18750	1.82812	1.66016	1.57910	1.53931	1.51959	1.50978	1.50489

Table 2.3 Slopes of secant lines for $f(x) = x^3 - x^2/2 - x/2$

The way in which the slope of the *tangent line* was introduced above gives an easy definition in the case of functions from $\mathbb{R}$ to $\mathbb{R}$.

Definition 9 (Tangent line). Assume the function $f : \mathbb{R} \to \mathbb{R}$ has a derivative at a. The *tangent line* to the graph of $f(x)$ at $(a, f(a))$ is the line through $(a, f(a))$ with slope $f'(a)$.

Assume the function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ has a derivative at a. The *tangent line* to the curve parametrized by $\mathbf{f}(x)$ at $\mathbf{f}(a)$ is the line through $\mathbf{f}(a)$ with direction $\mathbf{f}'(a)$.

Example 73. Consider the function $r(t) = -16t^2 + 20t + 6$ from Example 69 on page 52. The derivative of $r(t)$ at $t = 1$ is $r'(1) = -12$. The point where the tangent line meets the graph is $(1, r(1)) = (1, 10)$. The the equation of the tangent line is $y - 10 = -12(t-1)$ or $y = -12t + 22$, see Fig. 2.14 on page 55.

The derivative can also be interpreted as the rate of change of a quantity with respect to a variable. Examples of this are the rate of change of the surface of a sphere with respect to change in the volume and the rate of change of the profit for a company with respect to the price the company charges for a product.

An important result that is often used without statement is the relationship between derivatives and continuity. If the derivative of f exists at a, then the numerator in the limit of the definition of the derivative must go to 0 as $x \to a$. (Why is this true?). This means that

$$0 = \lim_{x \to a} \big(f(x) - f(a)\big) = \lim_{x \to a} \big(f(x)\big) - f(a).$$

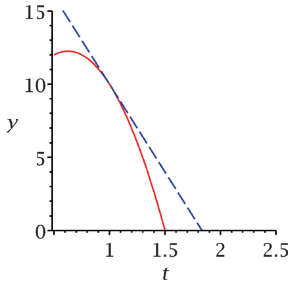


Fig. 2.14 A tangent line to $r(t) = -16t^2 + 20t + 6$

We then conclude that

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

and f is continuous at a. Restated, this says:

Theorem 23 (Differentiability implies continuity). *If $f(x)$ is differentiable at a, then $f(x)$ is continuous at a.*

Example 74. Consider the function $f(x) = mx + b$. Since it is differentiable everywhere, it is continuous everywhere.

There are cases when $f(x)$ is continuous at a point, but not differentiable at that point.

Example 75. Let $f(x) = |x|$. Then

$$\lim_{x \to 0} |x| = 0.$$

On the other hand, if $x < 0$, then

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

and if $x > 0$, then

$$\frac{|x|}{x} = 1.$$

Combined, these equations mean that $f'(0)$ does not exist.

Working with derivatives of vector-valued functions is very similar to working with limits of vector-valued functions. It is done coordinate by coordinate.

Theorem 24 (Differentiability of vector valued functions).
Let $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_n(x))$ be defined on an interval around a. The derivative of $\mathbf{f}(x)$ at a exists if and only if $f_i'(a)$ exist for $i = 1, 2, \ldots, n$. In that case we have

$$\frac{d\,\mathbf{f}(x)}{dx}(a) = (f_1'(a), f_2'(a), \ldots, f_n'(a)).$$

Example 76. The derivative of $\mathbf{f}(t) = (6t + 2, 2 - t, 3)$ is $\mathbf{f}'(t) = (6, -1, 0)$ since

$$\frac{d\,(6t+2)}{dt} = 6, \qquad \frac{d\,(2-t)}{dt} = -1, \qquad \text{and} \qquad \frac{d\,(3)}{dt} = 0.$$

This example is important in physical situations. Interpreting $\mathbf{f}'(t)$ as velocity and $\mathbf{f}(t)$ as position, we can say that an object moving with constant velocity, $(6,-1,0)$ starting at $\mathbf{f}(0)=(2,2,3)$ moves along the straight line $\mathbf{f}(t)=(6t+2,2-t,3)$. This characterizes the motion of a mass in the absence of any forces.

We can visualize this in a manner similar to that done for functions of one variable. Here the secant vectors,

$$\mathbf{v}=\frac{\mathbf{f}(x)-\mathbf{f}(t)}{x-t},$$

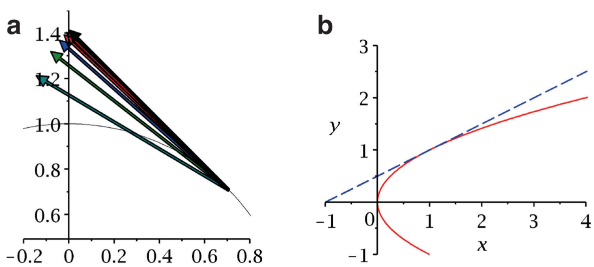
are used instead of slopes. (Why are slopes not appropriate for vector-valued functions with more than one output?) For vector-valued functions the idea is that the secant vectors converge to a tangent vector.

Example 77. Figure 2.15 on page 56 is the graph of $\mathbf{W}(t)=(\cos(t),\sin(t))$ around $\mathbf{W}(\pi/4)\approx(0.707,0.707)$. Recall that the tangent to a circle at a point is perpendicular to the radius of the circle. In this case the tangent vector in the picture is $\mathbf{v}=(-\sqrt{2}/2,\sqrt{2}/2)$, a unit vector that is showing counter clockwise motion around the circle.

The secant vectors are taken between the point $\mathbf{W}(\pi/4)$ and the points $\mathbf{W}(\pi/4+(1/2)^n)$ for $n=1,2,\ldots,5$ with length

$$\left\|\frac{\mathbf{W}(\pi/4+(1/2)^n)-\mathbf{W}(\pi/4)}{(1/2)^n}\right\| .$$

The graph in Fig. 2.15a on page 56 shows how the secant vectors approach this tangent vector.



Secant vectors going to a tangent vector

A tangent line to $\mathbf{r}(t)=(t^2,t)$

Fig. 2.15

Example 78. Consider the function $\mathbf{r}(t)=(t^2,t)$ that has derivative $\mathbf{r}'=(2t,1)$. The value of $\mathbf{r}(t)$ at $t=1$ is $\mathbf{r}(1)=(1,1)$ and derivative of $\mathbf{r}(t)$ at $t=1$ is $\mathbf{r}'(1)=(2,1)$. This means that a parametrization of the tangent line to the curve parametrized by $\mathbf{r}(t)$ at $t=1$ is $\ell(s)=(1,1)+s(2,1)$, see Fig. 2.15b on page 56.

Example 79. Consider the position curve of the projectile and velocity at $t=0$ in Example 70.

$$\mathbf{r}(t)=\left(vt\cos(\theta),-9.8t^t+vt\sin(\theta)\right)$$
$$\mathbf{v}(0)=v\left(\cos(\theta),\sin(\theta)\right) .$$

In this case the velocity vector at $t=0$ is tangent to the curve of motion of the projectile. Its direction is the direction of travel of the projectile at $t=0$ and its length is equal to the speed of the projectile, see Fig. 2.16 on page 57.

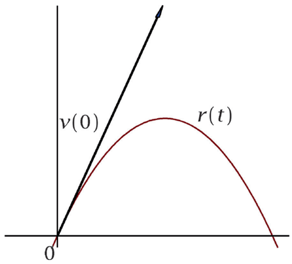


Fig. 2.16 The initial velocity vector and path of motion for Example 78

Exercises

1. What is the average rate of change of the function from $x = a$ to $x = b$ as given below?

(a) $f(x) = x^2 + 2$, $a = 1$, $b = 4$
(b) $f(x) = \sin(\pi x)$, $a = 0$, $b = \frac{1}{4}$
(c) $f(x) = e^{x-1}$, $a = 1$, $b = 2$
(d) $f(x) = x^3 + \frac{1}{x}$, $a = -3$, $b = -1$
(e) $g(x) = \frac{x^2+1}{4+x^3}$, $a = -1$, $b = 5$
(f) $g(x) = x^2 + 2x$, $a = -2$, $b = 0$
(g) $g(x) = \ln(x+1)$, $a = 0$, $b = 3$
(h) $g(x) = \cos(\pi(x+1)/3)$, $a = -2$, $b = -1$

2. Use the definition of the derivative to find the derivatives of the following functions at the given point.

(a) $f(x) = 0$, $a = 3$
(b) $f(x) = -3$, $a = 1$
(c) $g(x) = x + 1$, $a = 0$
(d) $g(x) = 2x - 1$, $a = -1$
(e) $h(z) = 3z + 2$, $a = 2$
(f) $h(z) = -4z - 4$, $a = 6$
(g) $h(z) = z^2 + 2$, $a = 1$
(h) $h(z) = -z^2 - 4$, $a = 0$
(i) $h(z) = z^2 - 2z + 1$, $a = 0$
(j) $h(z) = 2z^2 + z - 3$, $a = -1$

3. Find the average velocity of a mass whose position is given by $\mathbf{r}(t)$ from a to b.

(a) $\mathbf{r}(t) = (t, 3t - 1)$, $a = 0$, $b = 3$
(b) $\mathbf{r}(t) = (e^t, \ln(3t - 1))$, $a = 1$, $b = 2$
(c) $\mathbf{r}(t) = (t^2, \cos(\pi t/4))$, $a = 1$, $b = 7$
(d) $\mathbf{r}(t) = (3t, t^2, 1 - 2t)$, $a = -1$, $b = 1$
(e) $\mathbf{r}(t) = (e^{3t}, \ln(t^2 + 1), 1 - 2t^3)$, $a = 3$, $b = 5$
(f) $\mathbf{r}(t) = (\sin(\pi t/6), \cos(\pi t/6), t - 1)$, $a = 2$, $b = 4$

4. Use the definition of the derivative to find the derivatives of the following functions at the given point.

(a) $\mathbf{f}(t) = (0, 1, 3)$, $a = 2$
(b) $\mathbf{f}(t) = (-3, t, t)$, $a = 1$
(c) $\mathbf{g}(t) = (2t, t^2)$, $a = -2$
(d) $\mathbf{g}(t) = (t^2 + 1, \frac{1}{2}t)$, $a = -1$
(e) $\mathbf{h}(s) = (3s + 2, -3s - 2)$, $a = 2$
(f) $\mathbf{h}(s) = (-s^2, s^2 + 2)$, $a = 6$

5. Explain why the following functions do not have derivatives at the designated points.

(a) $f(x) = \frac{1}{x}$, $x = 0$
(b) $f(x) = |x-1|$, $x = 1$
(c) $g(x) = x - |x|$, $x = 0$
(d) $g(x) = \frac{\sin(x)}{x}$, $x = 0$
(e) $h(x) = x^{2/3}$, $x = 0$
(f) $h(x) = x^{1/3}$, $x = 0$
(g) $\mathbf{r}(t) = \left(3t^2, |t+2|, \cos(t)\right)$, $t = -2$
(h) $\mathbf{r}(t) = (\tan(t), \sec(t))$, $t = \frac{\pi}{2}$

6. The following functions have derivatives at the given points. Estimate the derivative to two decimal places using a numerical technique.

(a) $f(x) = x^{1/3}$, $x = 4$
(b) $f(x) = \cos(x)$, $x = \frac{\pi}{3}$
(c) $g(x) = \ln(x)$, $x = 1$
(d) $r(t) = \tan(x^3 - x)$, $x = 1$
(e) $g(z) = e^z$, $z = 0$
(f) $h(z) = z^{-1/3}$, $z = 4$
(g) $h(z) = \sqrt{z+1}$, $x = 6$
(h) $\mathbf{r}(t) = (t^{1/3}, t^6)$, $t = 1$
(i) $\mathbf{r}(t) = (\sin(\pi t/6), t^{-5}, t^2 + 1)$, $t = 2$
(j) $\mathbf{s}(t) = \left(\exp\left(\frac{t}{3}\right), \cos\left(t^{-2}\right), \ln(t^2+1)\right)$, $t = 2$

7. Explain why the following functions do not have a derivative at the given point.

(a) $f(x) = \lfloor x \rfloor$, $x = 4$
(b) $f(x) = \sqrt{x}$, $x = 0$
(c) $g(x) = \frac{1}{x^2}$, $x = 0$
(d) $g(z) = \frac{z^2-4}{z-2}$, $z = 2$
(e) $h(z) = z^{-1/3}$, $z = 0$
(f) $h(z) = \sqrt{z^2 - 2z + 1}$, $x = 1$
(g) $\mathbf{r}(t) = \left(3t^2 + 2, \cos(t), \lfloor \frac{t}{2} \rfloor\right)$, $z = 0$
(h) $\mathbf{h}(s) = \left(\frac{1}{s+1}, \frac{s+1}{s^2+3s+2}\right)$, $s = -1$

8. Use the definition of the derivative and the squeeze theorem, Theorem 16 on page 43, to show that the function

$$h(w) = \begin{cases} w^2 & \text{if } w \text{ is rational} \\ -w^2 & \text{if } w \text{ is irrational} \end{cases}$$

has a derivative at $w = 0$.

9. Why does the function

$$h(w) = \begin{cases} w^2 & \text{if } w \text{ is rational} \\ -w^2 & \text{if } w \text{ is irrational} \end{cases}$$

not have a derivative at any point besides $w = 0$.

10. Assume that $g(y)$ has a derivative at $y = 3.5$ with $g'(-3.5) = -3$. Using the definition of the derivative, explain why the function $f(y) = g(y) - 10$ has a derivative at $y = 3.5$ with $f'(3.5) = -3$.

11. Assume that $g(y)$ has a derivative at $y = 3.5$ with $g'(-3.5) = -3$. Use geometry to explain why the function $f(y) = g(y) - 10$ has a derivative at $y = 3.5$ with $f'(3.5) = -3$.

2.4 Derivatives of a Few Common Functions

As you should be able to tell from the last section, only having the derivatives of polynomials is very limiting. Even though not all of the derivatives of the functions in this section will be derived here, having them available for the rest of this chapter allows us to consider many more problems and applications.

The first functions we consider are $\sin(\theta)$ and $\cos(\theta)$. Recall the definition of $\sin(\theta)$ and $\cos(\theta)$ in terms of radians. The $\sin(\theta)$ and $\cos(\theta)$ are the x and y coordinates of the point on the unit circle centered at $(0,0)$ obtained by going θ units counter clockwise around the circle from the point $(1,0)$. This means that the function $\mathbf{W}(\theta) = (\cos(\theta), \sin(\theta))$ has us traveling counter clockwise around the circle at unit speed.

The above means that the derivative of **W** must be tangent to the unit circle and have length 1. The speed is 1 since the rotation goes at one radian per time unit. The only unit vectors tangent to the unit circle at $\mathbf{W}(\theta)$ are $(-\sin(\theta), \cos(\theta))$ and $(\sin(\theta), -\cos(\theta))$. From Fig. 2.17 on page 59, the choice should be
$(-\sin(\theta), \cos(\theta))$. It matches direction of travel around the unit circle.

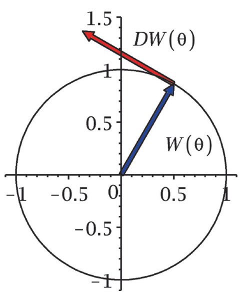


Fig. 2.17 $\mathbf{W}(\theta)$ and $D\mathbf{W}(\theta) = \mathbf{W}'(\theta)$

In particular, this means that

$$\frac{d}{d\theta}\cos(\theta) = -\sin(\theta)$$

and

$$\frac{d}{d\theta}\sin(\theta) = \cos(\theta).$$

It is easy to use these formulas.

Example 80. Find the derivative of $f(\theta) = 4\cos(\theta)$.

$$\begin{aligned}\frac{d}{d\theta}f(\theta) &= \lim_{h\to 0}\frac{4\cos(\theta+h) - 4\cos(\theta)}{h}\\ &= 4\lim_{h\to 0}\frac{\cos(\theta+h) - \cos(\theta)}{h}\\ &= -4\sin(\theta).\end{aligned}$$

Another two functions considered in this section are the natural exponential and natural logarithm functions. You should be familiar with the graphs of $1/x$ and $\ln(x)$, Fig. 2.18 on page 60.

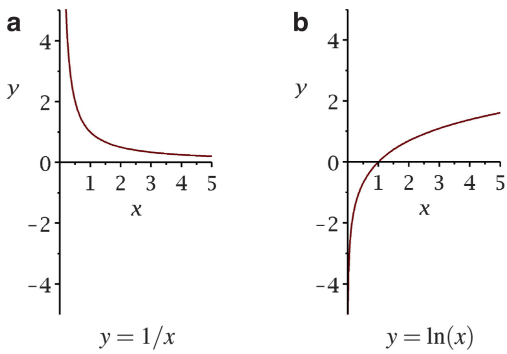


Fig. 2.18

The graph of $\ln(x)$ indicates that as x goes to 0 from the right that the derivative of $\ln(x)$ goes to infinity. As x goes to infinity the derivative of $\ln(x)$ goes to 0. Pictorially, $1/x$ looks like the derivative of $\ln(x)$. In fact, the derivative of $\ln(x)$ is $1/x$.

The final function in this section is the natural exponential, $\exp(x) = e^x$. Plotting the slopes of the tangent lines gives an indication of the function. Figure 2.19 on page 60 contains the plots of e^x and approximations of its derivative at intervals of 0.2. The approximations were done with the difference quotient

$$\frac{e^{x+0.001} - e^x}{0.001}.$$

Given the similarity in the graphs, we can guess that the derivative of e^x is e^x. In fact, that is true.

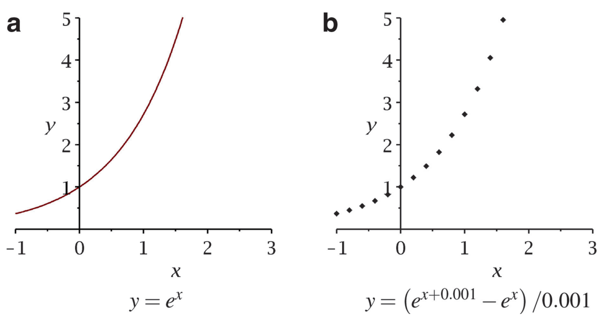


Fig. 2.19

Example 81. The derivative of $g(x) = \ln(x) + e^x$ is given by

$$\begin{aligned} g'(a) &= \lim_{x \to a} \frac{\ln(x) + e^x - (\ln(a) + e^a)}{x - a} \\ &= \lim_{x \to a} \frac{\ln(x) - \ln(a)}{x - a} + \lim_{x \to a} \frac{e^x - e^a}{x - a} \\ &= \frac{1}{a} + e^a. \end{aligned}$$

The natural exponential function has many uses in applications.

Example 82. A common assumption for the growth of a simple population is that the rate of change of the population, $P(t)$, is proportional to the population. If the proportionality constant is one, the equation describing this is

$$\frac{d\,P(t)}{dt} = P(t).$$

Since

$$\begin{aligned} \frac{d}{dt} Ce^t &= \lim_{x \to t} \frac{Ce^x - Ce^t}{x - t} \\ &= C \lim_{x \to t} \frac{e^x - e^t}{x - t} \\ &= C \frac{d}{dt} e^t \\ &= Ce^t, \end{aligned}$$

every function of the form $P(t) = Ce^t$ is a solution for this mathematical model of population growth.

Another class of functions that is frequently used is the class of polynomials. In the next chapter you will see how the rule for finding the derivative of a polynomial is derived. For now, the rule is simply stated.

Theorem 25 (Derivatives of polynomials). *Let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be any polynomial of degree n. The derivative of $p(x)$ is given by*

$$\frac{d\,p(x)}{dx} = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + m a_m x^{m-1} + \cdots + n a_n x^{n-1}.$$

This is easy to use.

Example 83. The derivative of $p(x) = 3 - 2x + 4x^2 + 6x^3$ is

$$\begin{aligned} p'(x) &= -2 + 2 \cdot 4x^{2-1} + 3 \cdot 6x^{3-1} \\ &= -2 + 8x + 18x^2. \end{aligned}$$

Example 84. The derivative of $q(x) = x^{101}$ is

$$\frac{d\,q(x)}{dx} = 101x^{101-1} = 101x^{100}.$$

Example 85. If there was no air resistance and no wind, the position of a projectile near the surface of the earth would be

$$\mathbf{s}(t) = \left(v_x t + x_0, -4.9t^2 + v_y t + y_0\right).$$

Here (v_y, v_x) is the initial velocity at $t = 0$ and (x_0, y_0) is the initial position. Since the only force acting on the projectile is gravity, the horizontal velocity should be constant.

The velocity is

$$\begin{aligned} v(t) &= \frac{d\,s(t)}{dt} \\ &= \left(\frac{d}{dt}\,(v_x t + x_0)\right), \frac{d}{dt}\,\left(-4.9t^2 + v_y t + y_0\right) \\ &= (v_x, -9.8t + v_y). \end{aligned}$$

The result matches what is taught in elementary physics classes.

The section ends with a function whose derivative can be found using a little skill and the definition of the derivative. It will be used for problems involving various differentiation rules. There are mathematical models where it is used.

Example 86. The derivative of $\sqrt{x}$ is given by

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}.$$

The key to the derivation of this rule is that, if the square roots are defined, $(\sqrt{a} - \sqrt{b})\,(\sqrt{a} + \sqrt{b}) = a - b$.

The following is the derivation of this formula. Assuming that x and $x + h$ are both positive,

$$\begin{aligned} \frac{d}{dx}\sqrt{x} &= \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \\ &= \lim_{h\to 0} \frac{(\sqrt{x+h} - \sqrt{x})\,(\sqrt{x+h} + \sqrt{x})}{h\,(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h\to 0} \frac{x + h - x}{h\,(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h\to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} \\ &= \frac{1}{2\sqrt{x}}. \end{aligned}$$

Exercises

1. Find the derivatives of the following functions at the given point a.

(a) $f(x) = x^2 + 2$, $a = 1$

(b) $f(x) = x^3 - 3x^2 + 6x - 2$, $a = 0$

(c) $f(x) = 10x + 6 - 3x^2 - x^5$, $a = 2$

(d) $f(x) = 101x^{201} + 55x^{50} - 1{,}000x$, $a = -1$

(e) $f(x) = \cos(x)$, $a = \frac{\pi}{4}$

(f) $f(x) = e^x$, $a = 1$

(g) $f(x) = \sin(x)$, $a = \frac{\pi}{2}$

(h) $f(x) = e^x$, $a = -1$

(i) $\mathbf{r}(t) = \left(t^2 - 2t + 3, \cos(t), e^t\right)$, $a = -2$

(j) $\mathbf{r}(t) = \left(t^2 - 2t + 3, \cos(t), e^t\right)$, $a = -2$

(k) $\mathbf{r}(t) = \left(t^5 - 5t^3, \ln(t), \sin(t)\right)$, $a = 2$

(l) $\mathbf{r}(t) = \left(t, t^2, t^3, t^4\right)$, $a = 1$

(m) $\mathbf{r}(t) = (e^t, \ln(t), \cos(t))$, $a = 4$

2. Use the same type of reasoning as was used to find the derivatives of $\sin(\theta)$ and $\cos(\theta)$ to find the derivatives of the following functions.

(a) $\sin(-\theta)$

(b) $\cos(2\theta)$

(c) $\cos\left(\frac{\theta}{3}\right)$

(d) $\sin(\theta + \frac{\pi}{2})$

3. Each of the following is the velocity of an object. Find its acceleration.

(a) $\mathbf{v}(t) = (5t - 6, -16t + 32)$

(b) $\mathbf{v}(t) = (-\cos(t), \sin(t), 2)$

(c) $\mathbf{v}(t) = \left(e^t, \ln(t), 4t^2 - 6t\right)$

(d) $\mathbf{v}(t) = \left(5t^2 - 26t + 10, -16t^2 + 150\right)$

2.5 Derivatives, Graphs, and Approximations

Recall that the derivative of $\mathbf{f}(x)$ at a, if it exists, is defined as

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

Letting $h = x - a$, this can be rewritten as

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

In many cases this form of the definition of the derivative is easier to use.

Since the limit is the derivative we have that if h is close to 0, then

$$\mathbf{f}'(a) \approx \frac{\mathbf{f}(a+h) - \mathbf{f}(a)}{h}.$$

Multiplying both sides by h and rearranging gives

$$\mathbf{f}(a+h) \approx \mathbf{f}(a) + \mathbf{f}'(a)\,h. \tag{2.2}$$

Given a function value and a derivative at a, this enables us to approximate values of the function near a. This idea is used throughout calculus.

Example 87. Given that $f(1) = 4$ and $f'(1) = -0.5$, approximate $f(1.2)$. Here we have $h = 1.2 - 1 = 0.2$. Thus

$$\begin{aligned} f(1.2) &\approx f(1) + f'(1)\,h \\ &\approx 4 + (-0.5)\,(0.2) \\ &\approx 3.9 \end{aligned}$$

This can be a good approximation for points close to a. As the following example shows, the approximation may not be good for points away from a.

Example 88. Consider the function $f(x) = \cos(x) - 4x^2$ around $a = 0$, see Fig. 2.20 on page 64. Since $f'(x) = -\sin(x) - 8x$, we have $f'(0) = 0$. The linear approximation is $f(0+h) \approx \ell(h) = 1 + 0h = 1$. This means that the error at $h = -0.01$ is $f(-0.01) - \ell(-0.01) \approx -0.00045$ and at $h = 0.2$ the error is $f(0.2) - \ell(0.2) \approx -0.17993$. The error is growing very fast in comparison with the size of h.

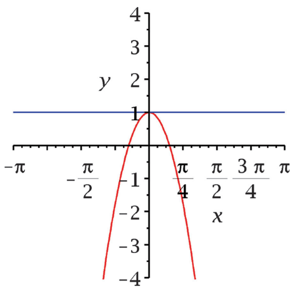


Fig. 2.20 Tangent line to $f(x) = \cos(x) - 4x^2$ at $x = 0$

This idea of approximation is also used in differential notation. This notation is commonly used in the sciences. If we take Eq. (2.2) and rewrite it we get

$$\mathbf{f}(a+h) - \mathbf{f}(a) \approx \mathbf{f}'(a)\,h.$$

Setting $\Delta\mathbf{f} = \mathbf{f}(a+h) - \mathbf{f}(a)$ and $\Delta x = h$, this approximation becomes

$$\Delta\mathbf{f} \approx \mathbf{f}'(a)\,\Delta x. \tag{2.3}$$

The Δ's are small changes in the values of $\mathbf{f}$ and x.

If we take the limit in the sense of making the distances $\Delta\mathbf{f}$ and Δx infinitely small, infinitesimals, we get the equation

$$d\mathbf{f} = \mathbf{f}'(x)\,dx. \tag{2.4}$$

We can use the equation in the same way that we use Eq. (2.3).

Example 89. Assume that the position of a mass is $\mathbf{s}(t) = \left(t^2 - t, 3t+5, \cos(t)\right)$ when written as a function of time. Approximate the position of the mass at $t = -0.1$.

The derivative of $\mathbf{s}(t)$ is

$$\mathbf{s}'(t) = (2t-1, 3, -\sin(t)),$$

$s(0) = (0,5,1)$ and $s'(0) = (-1,3,0)$. Using differential notation we have $dt = -0.1 - 0 = -0.1$ and

$$\begin{aligned} d\mathbf{s} &= \mathbf{s}'(0)\,dt \\ &\approx (-1,3,0)\,(-0.1) \\ &\approx (0.1,-0.3,0). \end{aligned}$$

Since $d\mathbf{s} = \mathbf{s}(-0.1) - \mathbf{s}(0)$, we have

$$\begin{aligned} \mathbf{s}(-0.1) &\approx \mathbf{s}(0) + d\mathbf{s} \\ &\approx (0,5,1) + (0.1,-0.3,0) \\ &\approx (0.1,4.7,1). \end{aligned}$$

We can also plot both sides of Eq. (2.2) to get a geometric view of the derivative.

Example 90. Consider the function $f(x) = e^x$. Then $f'(x) = e^x$ and $f(0) = f'(0) = 1$. This gives us

$$e^x \approx 1 + 1 \cdot x.$$

The plot shows that the line $y = 1 + x$ is a good approximation to $y = e^x$ around the point $(0,1)$, see Fig. 2.21 on page 65.

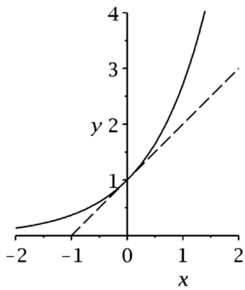


Fig. 2.21 Tangent line approximation to $y = e^x$ at $x = 0$

The same basic picture works for vector valued functions.

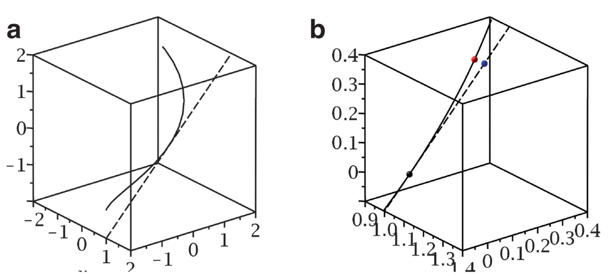
Example 91. Consider the function $\mathbf{f}(t) = (\cos(t), \sin(t), 2t)$. The derivative of this function is $\mathbf{f}'(t) = (-\sin(t), \cos(t), 2)$. At $t = 0$ we get that

$$\begin{aligned}\mathbf{f}(t) &\approx \mathbf{f}(0) + \mathbf{f}'(0)t \\ &\approx (1,0,0) + t\,(0,1,2) \\ &\approx (1,t,2t).\end{aligned}$$

Plotting the left and right sides of the last equation gives Fig. 2.22a on page 65. When we try to approximate $\mathbf{f}(0.2)$ we get

$$\begin{aligned}\mathbf{f}(0.2) &\approx \mathbf{f}(0) + \mathbf{f}'(0)\,(0.2) \\ &(1,0.2,0.4).\end{aligned}$$

This is plotted in Fig. 2.22b on page 65.



The tangent line to $\mathbf{f}(t) = (\cos(t), \sin(t), 2t)$

The tangent line approximation to $\mathbf{f}(t) = (\cos(t), \sin(t), 2t)$

Fig. 2.22

Exercises

1. In each part of this problem a function value and derivative of a function are given at a point a. Find an approximation for the function value at the given b.

(a) $f(1)=2$, $f'(1)=3$, $b=1.3$
(b) $f(0)=0$, $f'(0)=-2$, $b=-0.5$
(c) $f(-1)=2$, $f'(-1)=1/4$, $b=-1.3$
(d) $f(3)=2$, $f'(3)=-2$, $b=2.9$
(e) $\mathbf{r}(1)=(1,-2)$, $\mathbf{r}'(1)=(-1,1)$, $b=1.2$
(f) $\mathbf{r}(0)=(-1,0)$, $\mathbf{r}'(0)=(2,-1)$, $b=-0.2$
(g) $\mathbf{r}(-10)=(3,2,-1)$, $\mathbf{r}'(-10)=(4,-2,1)$, $b=-9.7$
(h) $\mathbf{r}(3)=(0,-1,1)$, $\mathbf{r}'(3)=(-1,2,-2)$, $b=2.7$

2. Given are a function value at a, a derivative value at a, and a b. Use this information to approximate $f(b)$.

(a) $f(1)=1$, $f'(1)=-2$, $b=1.5$
(b) $f(2)=-1$, $f'(2)=-1$, $b=2.5$
(c) $f(3)=1.3$, $f'(3)=0.25$, $b=1.35$
(d) $\mathbf{r}(1)=(1,0)$, $\mathbf{r}'(1)=(3,1/3)$, $b=0.95$
(e) $\mathbf{f}(0)=(2,-1)$, $\mathbf{f}'(0)=(0.3,-0.3)$, $b=0.33$
(f) $\mathbf{r}(5)=(-1,2,1)$, $\mathbf{r}'(5)=(-0.2,0.2,0)$, $b=5.15$

3. Given are a function, an a, and a b. Use this information to approximate $f(b)$ using the value and the derivative of the function at a.

(a) $f(x)=3x+3$, $a=2$, $b=2.5$
(b) $f(x)=\sin(x)+\cos(x)$, $a=\pi/4$, $b=0.9$
(c) $g(w)=w^3-4w^2+4$, $a=3$, $b=2.9$
(d) $\mathbf{r}(t)=(t^2,t^3-3t)$, $a=-2$, $b=-1.95$
(e) $\mathbf{f}(t)=(\sin(t),\cos(t),t)$, $a=3\pi/2$, $b=4.8$
(f) $\mathbf{r}(s)=(s^2,2-s^3,s^2-4s+2)$, $a=-1$, $b=-1.2$

4. Use differential notation to write df for the following functions.

(a) $f(x)=3x+3$
(b) $f(x)=\sin(x)$
(c) $f(x)=e^x+2x$
(d) $f(x)=\cos(x)+\ln(x)$
(e) $\mathbf{f}(x)=(x^3-3x+4,\ln(x)+3)$
(f) $\mathbf{f}(w)=\left(6w^2,5w^{10},6e^w\right)$

5. The surface area of a sphere with radius r is $A(r)=4\pi r^2$. Use a linear approximation for $A(r)$ at $r=4$ to approximate the area of a sphere with radius $r=4.08$. What is the error in this approximation?

6. The volume of a sphere with radius r is $V(r)=\frac{4}{3}\pi r^3$. Use a linear approximation for $V(r)$ at $r=2$ to approximate the volume of a sphere with radius $r=1.94$. What is the error in this approximation?

7. The surface are of a right circular cylinder with height 5 cm and radius r is $A(r) = 2\pi r^2 + 10\pi r\,\mathrm{cm}$. Use a linear approximation to $A(r)$ at $r = 3\,\mathrm{cm}$ to approximate the volume of a cylinder with radius $2.95\,\mathrm{cm}$.

8. The volume of a right circular cone with height sm and base diameter sm is $V(s) = \frac{1}{12}\pi s^3\,\mathrm{m}^3$. Using a linear approximation to $V(s)$ at $s = 5\,\mathrm{m}$, approximate the volume of a cone with $s = 5.11\,\mathrm{m}$.

9. The position of a projectile is given by $\mathbf{r}(t) = (10t, -4.9t^2 + 60t)\,\mathrm{m}$ where the first coordinate is the distance down range from the firing point and the second coordinate is the height above the ground of the projectile. (Distances are in meters and time is measured in seconds.) Using a linear approximation to the position function at $t = 8\,\mathrm{s}$, estimate when and where the projectile lands.

Chapter 3
More on Limits

3.1 One-Sided Limits

There are several other situations where the idea of a limit is needed. This section states the definitions for some of these situations and gives a theorem that is useful when trying to find limits.

The first definition is that of a *one-sided limit.* A simple example of this is what happens to $f(x) = \sqrt{x}$ as $x \to 0$, see Fig. 3.1 on page 69.
Since the domain of $f(x)$ is $[0,\infty)$, we cannot use any sequences with $a_n < 0$ for some n. This leads to the following definition.

Definition 10. Let $\mathbf{f}(x)$ be defined on an interval (a,b), then $\mathbf{L}$ is the limit of $\mathbf{f}(x)$ as x approaches a from the right,

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

if for all sequences $a_n \to a$ with $a_n > a$ we have

$$\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L},$$

The left limit is defined in a similar way, a vector $\mathbf{L}$ is the limit of $\mathbf{f}(x)$ as x approaches a from the left,

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

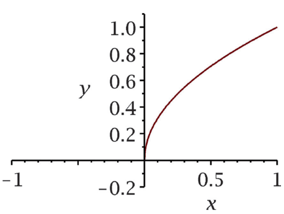


Fig. 3.1 $y = \sqrt{x}$

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_3) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_3

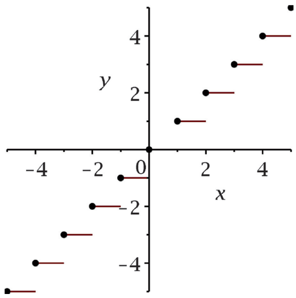


Fig. 3.2 $y = \lfloor x \rfloor$

if for all sequences $a_n \to a$ with $a_n < a$ we have

$$\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L},$$

Example 92. Consider the floor function $f(x) = \lfloor x \rfloor$. For a given x this is the largest integer less than or equal to x. This means that $\lfloor n \rfloor = n$ for any integer. If $x \in (n, n+1)$, then $\lfloor x \rfloor = n$ for any integer n, Fig. 3.2 on page 70.

If we take values in $(1,2)$, $\lfloor x \rfloor = 1$. For any sequence $a_n \to 1$ with $a_n > 1$ we have $\lim_{n\to\infty} \lfloor a_n \rfloor = 1$. This means that $\lim_{x\to 1^+} \lfloor x \rfloor = 1$.

Similarly, we can show that $\lim_{x\to 1^-} \lfloor x \rfloor = 0$. This means that both of the one-sided limits for $\lfloor x \rfloor$ exist at $x = 1$, but the limit as $x \to 1$ of $\lfloor x \rfloor$ does not exist. (Why?)

Example 93. The vector valued function

$$\mathbf{r}(t) = \begin{cases} \left(\cos(t) - 1, t+1, t^2\right) & \text{if} \quad t < 0 \\ (\sin(t), -t+1, 0) & \text{if} \quad t > 0 \end{cases}$$

has right and left limits at $t = 0$ that are the same.

Since all of the coordinate functions in the definition of $\mathbf{r}(t)$ are continuous everywhere, all of their one sided limits at $t = 0$ exist. (This is because sequences converging to $t = 0$ from one side are still sequences converging to $t = 0$.) This means that

$$\lim_{t\to 0^-} \mathbf{r}(t) = (0,1,0)$$

and

$$\lim_{t\to 0^+} \mathbf{r}(t) = (0,1,0).$$

This is illustrated in Fig. 3.3 on page 70.

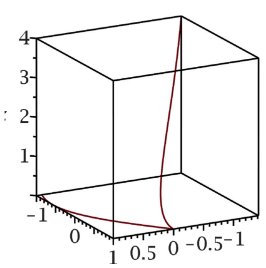


Fig. 3.3 Image of $\mathbf{r}(t)$ for Example 93

Note that since sequences converging to a from one side are still sequences converging to a, if $\lim_{x\to a}\mathbf{f}(x)=\mathbf{L}$ then both of the limits $\lim_{x\to a^-}\mathbf{f}(x)$ and $\lim_{x\to a^+}\mathbf{f}(x)$ exist and equal $\mathbf{L}$,

Example 94. Recall the definition $\mathbf{W}(\theta)=\big(\cos(\theta),\sin(\theta)\big)$. Since both $\sin(\theta)$ and $\cos(\theta)$ are continuous everywhere, $\mathbf{W}(\theta)$ is continuous everywhere. In particular, $\lim_{\theta\to\frac{\pi}{4}}\mathbf{W}(\theta)=\frac{\sqrt{2}}{2}(1,1)$. This means that $\lim_{\theta\to\frac{\pi}{4}^-}\mathbf{W}(\theta)=\frac{\sqrt{2}}{2}(1,1)$ and $\lim_{\theta\to\frac{\pi}{4}^+}\mathbf{W}(\theta)=\frac{\sqrt{2}}{2}(1,1)$.

The above example illustrates part of the following theorem. It relates one sided limits with the limit. This result is used both to find limits and to show that limits do not exist.

Theorem 26. *Let $\mathbf{f}(x):\mathbb{R}\to\mathbb{R}^n$ be a function defined on a set $(a-r,a+r)-\{a\}$. The limit $\lim_{x\to a}\mathbf{f}(x)$ exists and equals $\mathbf{L}$ if and only if the limits $\lim_{x\to a^-}\mathbf{f}(x)$ and $\lim_{x\to a^+}\mathbf{f}(x)$ both exist and both equal $\mathbf{L}$.*

Proof. The ($\Rightarrow$) part is simple. Since $\lim_{x\to a}\mathbf{f}(x)=\mathbf{L}$, all sequences $a_n\to a$ with either $a_n>a$ or $a_n<a$ satisfy $\lim_{n\to\infty}\mathbf{f}(a_n)=\mathbf{L}$.

The ($\Leftarrow$) part is a little harder. There are three cases to consider; when, after some N, all a_n are greater that a; when, after some N, all a_n are less that a; and when after all N there are $a_n<a$ and $a_n>a$. The first two cases are easier and left to the reader.

Assume that the limits $\lim_{x\to a^-}\mathbf{f}(x)$ and $\lim_{x\to a^+}\mathbf{f}(x)$ both exist for all such sequences and all limits of sequences equal $\mathbf{L}$. Let $a_n\to a$, $a_n\neq a$ be an arbitrary sequence converging to a that can be divided into two sequences $\{b_m\}$ and $\{c_k\}$ where a_n is put into $\{b_m\}$ if $a_n>a$ and a_n is put into $\{c_m\}$ if $a_n>a$.

Since both one-sided limits are $\mathbf{L}$, for any $r>0$ there are M and K such that if $m>M$ and $k>K$ we have

$$\|\mathbf{f}(b_m)-\mathbf{L}\|<r \quad\text{and}\quad \|\mathbf{f}(c_k)-\mathbf{L}\|<r.$$

If we take an N such that for $n>N$, a_n was put into one of the sequences $\{b_m\}$ and $\{c_k\}$ as b_m with $m>M$ or c_k with $k>K$, we have

$$\|\mathbf{f}(a_n)-\mathbf{L}\|<r$$

for all $n>N$. This means that $\lim_{n\to\infty}\mathbf{f}(a_n)=\mathbf{L}$ for any sequence $a_n\to a$.

The following is an example of how we can use this result.

Example 95. Let

$$g(t)=\begin{cases}\sin(t) & \text{if } t\le 0\\ \cos(t)-1 & \text{if } t>0\end{cases},$$

see Fig. 3.4 on page 71

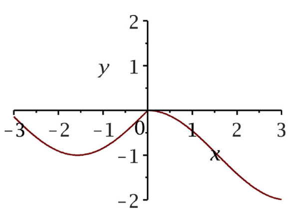


Fig. 3.4 $y=g(t)$ for Example 95

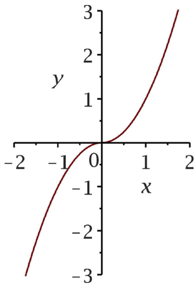


Fig. 3.5 $y = f(x)$ for Example 96

Since $\sin(t)$ and $\cos(t) - 1$ are differentiable at $t = 0$, they are both continuous at 0. This means that

$$\begin{aligned}\lim_{t\to 0^-} g(t) &= \lim_{t\to 0^-} \sin(t)\\ &= \lim_{t\to 0} \sin(t)\\ &= 0\end{aligned}$$

and

$$\begin{aligned}\lim_{t\to 0^+} g(t) &= \lim_{t\to 0^+} \cos(t) - 1\\ &= \lim_{t\to 0} \cos(t) - 1\\ &= 0.\end{aligned}$$

Therefore $\lim_{t\to 0} g(t) = 0$.

This theorem is also used for finding derivatives.

Example 96. Let

$$f(x) = \begin{cases} -x^2 & \text{if } x \le 0 \\ x^2 & \text{if } x > 0 \end{cases},$$

see Fig. 3.5 on page 72.

This function is continuous at $x = 0$ with $f(0) = 0$. (This is left for the reader to prove.)

To show that f is differentiable at $x = 0$ we evaluate the limits $\lim_{x\to 0^-} (f(x) - f(0))/(x - 0)$ and $\lim_{x\to 0^+} (f(x) - f(0))/x - 0$. The limits are

$$\begin{aligned}\lim_{x\to 0^-} \frac{f(x) - f(0)}{x - 0} &= \lim_{x\to 0^-} \frac{-x^2 - 0}{x - 0}\\ &= \lim_{x\to 0^-} -x\\ &= 0,\end{aligned}$$

and

$$\begin{aligned}\lim_{x\to 0^+}\frac{f(x)-f(0)}{x-0} &= \lim_{x\to 0^-}\frac{x^2-0}{x-0}\\ &= \lim_{x\to 0^-} x\\ &= 0.\end{aligned}$$

This means that $f'(0)=0$.

Among the major uses of one sided limits is finding the largest and smallest values of functions on an interval $[a,b]$ that includes its endpoints, see the section on extrema on page 137, and the statement of the Intermediate Value Theorem, Theorem 43 on page 121. These results depend on a function being continuous on the interval $[a,b]$. The idea is given in the following definition.

Definition 11. Let $[a,b]$ be an interval that includes its endpoints and let f be a function defined on $[a,b]$. The function f is continuous at a if

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

and the function f is continuous at b if

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

The function f is continuous on $[a,b]$ if f is continuous on (a,b) and if f is continuous at both a and b.

To simplify our language about intervals we will use two terms. An interval $\mathscr{I}\subset\mathbb{R}$ is *bounded* if both endpoints are finite. An interval is *closed* if it contains any finite endpoints. For example, the interval $[-10,101]$ is bounded and $(-\infty,0]$ is not bounded. The intervals $[-1,101]$ and $(-\infty,10]$ are closed whereas the intervals $(-10,101]$ and $(0,\infty)$ are not closed. This means the interval $[-10,101]$ is closed and bounded.

Here are two examples to illustrate this idea of a function being continuous on a closed bounded interval.

Example 97. Let $f(x)=x^2+\sin(x^{64})$. Since f is continuous at all $x\in\mathbb{R}$, for all real numbers a and b we have

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

This means f is continuous on all closed bounded intervals $[a,b]$.

Example 98. Let $f(x)=\lfloor x\rfloor$. This is not continuous on $[0,1]$ since

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

The rules for calculating one-sided limits are the same as the rules for calculating limits. The statement is the same except for the types of limits considered.

Theorem 27. *Let $\mathbf{f}(x)$ and $\mathbf{g}(x)$ be functions from $\mathbb{R}$ to $\mathbb{R}^m$ with limits $\mathbf{K}$ and $\mathbf{L}$ from the right (left) at $x=a$, and let $h(x)$ be a function from $\mathbb{R}$ to $\mathbb{R}$ with right sided (left sided) limit M at a. Then the following hold:*

(i) $$\lim_{x\to a^+}(\mathbf{f}+\mathbf{g})(x) = \mathbf{K}+\mathbf{L} \quad \left(\lim_{x\to a^-}(\mathbf{f}+\mathbf{g})(x) = \mathbf{K}+\mathbf{L}\right)$$

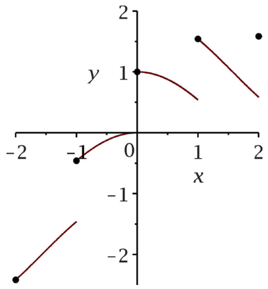


Fig. 3.6 $f(x) = \lfloor x \rfloor + \cos(x)$

(ii) $\lim_{x\to a^+} h(x)\mathbf{f}(x) = M\mathbf{K} \quad \left(\lim_{x\to a^-} h(x)\mathbf{f}(x) = M\mathbf{K}\right)$

(iii) $\lim_{x\to a^+} (\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{K}\cdot\mathbf{L} \quad \left(\lim_{x\to a^-} (\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{K}\cdot\mathbf{L}\right)$

(iv) *If* $M \neq 0$, *then* $\lim_{x\to a^+} \frac{\mathbf{f}(x)}{h(x)} = \mathbf{K}/M \quad \left(\lim_{x\to a^-} \frac{\mathbf{f}(x)}{h(x)} = \mathbf{K}/M\right)$

Example 99. Consider the limits $\lim_{x\to 0^+} f(x)$ and $\lim_{x\to 0^-} f(x)$ where

$$f(x) = \lfloor x \rfloor + \cos(x).$$

Recall that $\lim_{x\to 0}\cos(x) = 1$. Since $\lfloor x \rfloor = 0$ if $x \in (0,1)$, $\lim_{x\to 0^+}\lfloor x \rfloor = 0$ and since $\lfloor x \rfloor = -1$ if $x \in (-1,0)$, $\lim_{x\to 0^-}\lfloor x \rfloor = -1$. These facts imply that

$$\begin{aligned}\lim_{x\to 0^+} f(x) &= \lim_{x\to 0^+}\lfloor x \rfloor + \lim_{x\to 0^+}\cos(x)\\ &= 0+1\\ &= 1\end{aligned}$$

and that

$$\begin{aligned}\lim_{x\to 0^-} f(x) &= \lim_{x\to 0^-}\lfloor x \rfloor + \lim_{x\to 0^-}\cos(x)\\ &= -1+1\\ &= 0.\end{aligned}$$

See Fig. 3.6 on page 74.

Exercises

1. Find the following one-sided limits.

(a) $\lim_{x\to 0^+} \sqrt{x} - 4$

(b) $\lim_{x\to 3^-} \sqrt{9-x^2}$

(c) $\lim_{x\to 2^+} \lfloor x \rfloor$

(d) $\lim_{x\to 2^-} \lfloor x \rfloor$

(e) $\lim_{x\to 2^+} \lfloor -2x \rfloor$

(f) $\lim_{x\to 2^-} \lfloor -2x \rfloor$

2. Find the left and right sided limits of the following functions at the given point. Are the functions continuous at the point?

(a) $f(x) = x^2 + 2,\, x = 1$
(b) $f(x) = \sin(\pi x),\, x = 0$
(c) $f(x) = e^{\lfloor x \rfloor},\, x = 1$
(d) $f(x) = \frac{|x|}{x},\, x = 0$
(e) $g(x) = \frac{x^2+1}{4+x^3},\, x = -1$
(f) $g(x) = \ln(x),\, x = e^3$

3. Find the left and right sided limits of the following functions at the given point. Are the functions continuous at the point?

(a) $\mathbf{r}(t) = \left(0, t, t^2 + 1\right),\, t = 3$
(b) $\mathbf{f}(t) = \left(\sin(t), \cos(t), \tan(t)\right),\, t = 1/2$
(c) $\mathbf{r}(t) = \left(t+1, t^3+t^2+t+1, \lfloor t+1 \rfloor\right),\, t = -3$
(d) $\mathbf{f}(x) = \left(\sin(1/t), t+1\right),\, t = 0$

4. Why do both one sided limits for a function $\mathbf{f}(x)$ exist at a point a where $f(x)$ is continuous?

5. A projectile has position $\mathbf{r}(t) = (10t, -4.9t^2 + 60t)$ m. Here the first coordinate is the position down range and the second coordinate is height above the ground. (Time is measured in seconds.) Explain why finding the velocity when the projectile lands involves a one-sided limit.

6. Are the given functions continuous on the given intervals?

(a) $f(x) = \cos(3x) - \ln(x)$ on $[1,2]$
(b) $f(x) = \cos(3x) - \ln(x)$ on $[0,1]$
(c) $g(z) = \frac{1}{z}$ on $[-1,1]$
(d) $g(z) = \frac{1}{z}$ on $[-3,-1]$
(e) $h(w) = \frac{w^2-1}{w+1}$ on $[-1,0]$
(f) $h(w) = \frac{w^2-1}{w+1}$ on $[0,1]$
(g) $r(t) = \frac{t^2-1}{t+1}$ on $(0,1)$
(h) $s(t) = \tan\left(\frac{\pi t}{2}\right)$ on $(-1,1)$

7. Explain how one can extend Definition 11 on page 73 to vector valued functions.

8. Show that if a function $f(x)$ is even and differentiable at $x = 0$, then $f'(0) = 0$.

9. If a function $f(x)$ is odd and differentiable at $x = 0$, can $f'(0) = 0$?

3.2 Limits Involving Infinity

There are two simple interesting situations involving limits of functions where infinity plays a role. One is the case when x goes toward ∞ or $-\infty$. The other situation is when the values of $f(x)$ go toward ∞ or $-\infty$. In fact, the definition of the limit of a sequence $\{a_n\}_{n=1}^{\infty}$ is already in the form we need.

Recall that $\lim_{n\to\infty} \mathbf{a}_n = \mathbf{L}$ if for all $r > 0$ there is an N such that if $n > N$ then $\|\mathbf{a}_n - \mathbf{L}\| < r$. If we take $\mathbf{a}_n = \mathbf{f}(n)$ for some function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$, this definition can be viewed as taking one sequence of $\mathbf{f}(x)$ values as x gets infinitely large. In a manner corresponding to the definition of the limit of a function, we only need to include the convergence of all sequences. First the definition of a sequence going to infinity is needed.

Definition 12 (Sequence limit is infinity). A sequence a_n goes to infinity, $a_n \to \infty$, if for all $M > 0$ there is an N such that if $n > N$, then $a_n > M$.

A sequence b_n goes to negative infinity, $b_n \to -\infty$, if for all $M > 0$ there is an N such that if $n > N$, then $b_n < -M$.

This concept can be illustrated using a couple simple examples.

Example 100. Let $a_n = n$. Then, for any $M > 0$, there is an integer $N > M$. If $n > N > M$, we have $a_n > M$. This means that $a_n \to \infty$.

On the other hand, let $b_n = -n$. Then, for any $M > 0$, there is an integer $N > M$. If $n > N > M$, we have $b_n < -M$. This means that $b_n \to -\infty$.

Example 101. Let $a_n = \sqrt{n}$. Then, for any $M > 0$, there is an integer $N > M^2$. If $n > N$, we have $a_n = \sqrt{n} \geq \sqrt{N} \geq \sqrt{M^2} = M$. This means that $a_n \to \infty$.

On the other hand, let $b_n = -\sqrt{n}$. Then, for any $M > 0$, there is an integer $N > M^2$. If $n > N$, we have $b_n = -\sqrt{n} \leq -\sqrt{N} \leq -\sqrt{M^2} = -M$. This means that $b_n \to -\infty$.

With this, the definition of a function $\mathbf{f}(x)$ converging as $x \to \infty$ can be stated.

Definition 13 (Function limit at infinity). The limit of a function $\mathbf{f}(x) : \mathbb{R} \to \mathbb{R}^n$ as $x \to \infty$ is $\mathbf{L}$,

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

if for all sequences $a_n \to \infty$, $\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L}$.

The limit of a function $\mathbf{f}(x) : \mathbb{R} \to \mathbb{R}^n$ as $x \to -\infty$ is $\mathbf{L}$,

$$\lim_{x\to-\infty} \mathbf{f}(x) = \mathbf{L},$$

if for all sequences $a_n \to -\infty$ we have $\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L}$.

This idea can be illustrated with a couple simple examples.

Example 102. Consider the function $f(x) = 1/x$. The claim is that

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

and

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

Let a_n be any sequence going to infinity. For a given $r > 0$ there is an $M > 0$ such that $r > 1/M$. There is an N such that if $n > N$ then $a_n > M$ and $1/a_n - 0 < 1/M < r$. This means that the arbitrary sequence has $f(a_n) \to 0$ and the function converges to 0 as $x \to \infty$, see Fig. 3.7 on page 76. In this figure the x-axis has been compressed in a way that approximates the arctangent function, (x, y) is plotted as $(c \arctan(x) + b, y)$.

Let a_n be any sequence going to negative infinity. For a given $r > 0$ there is an $M > 0$ such that $r > 1/M$. There is an N such that if $n > N$ then $|a_n| > M$ and $|1/a_n - 0| = |1/a_n| < 1/M < r$. This means that the arbitrary sequence has $f(a_n) \to 0$ and the function converges to 0 as $x \to -\infty$.

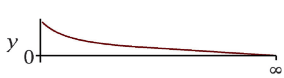


Fig. 3.7 $f(x) = 1/x$

The next example shows one way that a limit at infinity may fail to exist.

Example 103. Consider the function $f(x) = \sin(x)$ and the limit as $x \to \infty$ of $f(x)$. Define the sequences $a_n = 2n\pi + \pi/2$ and $b_n = 2n\pi + 3\pi/2$. For all n we have $f(a_n) = 1$ and $f(b_n) = -1$. Since the sequences of f values converge to different L's, the limit does not exist, see Fig. 3.8 on page 77. As in Fig. 3.7, in this figure the x-axis has been compressed in a way that approximates the arctangent function, (x,y) is plotted like $(c\arctan(x) + b, y)$.

A similar argument also shows that the limit of $f(x)$ does not exist as $x \to -\infty$.

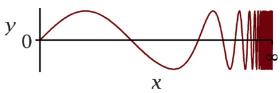


Fig. 3.8 $f(x) = sin(x)$

The rules for taking limits of functions as x goes to $\pm\infty$ are the same as the rules for taking limits at ordinary points. The rules are stated without proofs. The proofs follow lines similar to the proofs from standard limits.

Theorem 28 (Rules for limits at infinity). *Let* $\mathbf{f}(x)$ *and* $\mathbf{g}(x)$ *be functions from* $\mathbb{R}$ *to* $\mathbb{R}^m$ *with limits* $\mathbf{K}$ *and* $\mathbf{L}$ *as* x *goes to* ∞ $(-\infty)$*, and let* $h(x)$ *be a function from* $\mathbb{R}$ *to* $\mathbb{R}$ *with limit* M *as* x *goes to* ∞ $(-\infty)$*. Then the following hold:*

(i) $\lim_{x\to\infty}(\mathbf{f}+\mathbf{g})(x) = \mathbf{K}+\mathbf{L} \quad \left(\lim_{x\to-\infty}(\mathbf{f}+\mathbf{g})(x) = \mathbf{K}+\mathbf{L}\right)$

(ii) $\lim_{x\to\infty} h(x)\mathbf{f}(x) = M\mathbf{K} \quad \left(\lim_{x\to-\infty} h(x)\mathbf{f}(x) = M\mathbf{K}\right)$

(iii) $\lim_{x\to\infty}(\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{K}\cdot\mathbf{L} \quad \left(\lim_{x\to-\infty}(\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{K}\cdot\mathbf{L}\right)$

(iv) *If* $M \neq 0$*, then* $\lim_{x\to\infty}\frac{\mathbf{f}(x)}{h(x)} = \mathbf{K}/M \quad \left(\lim_{x\to-\infty}\frac{\mathbf{f}(x)}{h(x)} = \mathbf{K}/M\right)$

This result can be used in the case of rational functions of x as x goes to $\pm\infty$. Using a technique similar to that used for limits of sequences that are rational functions of n, we can find limits of rational functions $f(x)$ as x goes to ∞.

Example 104. Consider the rational function $h(z) = (z^2 + z - z^3)/(4z^3 + 5z + 10)$. The limit as z goes to $-\infty$ of $h(z)$ can be computed as follows.

$$\begin{aligned}
\lim_{z\to-\infty} h(z) &= \lim_{z\to-\infty}\frac{z^2+z-z^3}{4z^3+5z+10} \\
&= \lim_{z\to-\infty}\frac{\frac{1}{z}+\frac{1}{z^2}-1}{4+\frac{5}{z^2}+\frac{10}{z^3}} \quad \text{(Dividing the numerator and denominator by } z^3\text{.)} \\
&= \frac{0+0-1}{4+0+0} \quad \text{(Using Theorem 28.)} \\
&= -\frac{1}{4}.
\end{aligned}$$

See Fig. 3.9 on page 78.

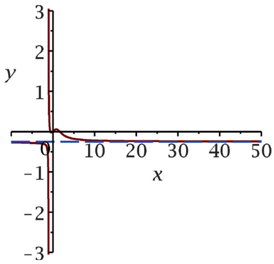


Fig. 3.9 $h(z) = (z^2 + z - z^3)/(4z^3 + 5z + 10)$

The other situation considered here is when the values of a function $f(x)$ grow without bound as $x \to a$ or as $x \to \pm\infty$. In fact, this has already been defined in this section, see Definition 12. *These limits are not limits in the same sense as before since $f(x)$ does not approach a value. We are simply noting the fact that the function values grow without bound in either the positive or negative direction.* The following definition makes this more formal.

Definition 14 (Infinite limits). Let a be a real number or let $a = \pm\infty$. The limit of $f(x)$ as $x \to a$ is ∞,

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

if for all sequences $a_n \to a$,

$$\lim_{n\to\infty} f(a_n) = \infty.$$

Similarly, the limit of $f(x)$ as $x \to a$ is $-\infty$,

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

if for all sequences $a_n \to a$,

$$\lim_{n\to\infty} f(a_n) = -\infty.$$

There are a couple of classic examples that give some idea of what can happen as the value of f grows without bound.

Example 105. Let $f(x) = 1/x^2$ and consider $\lim_{x\to 0} f(x)$, see Fig. 3.10 on page 78.

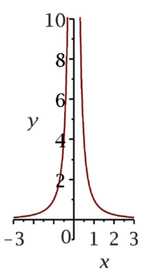


Fig. 3.10 $f(x) = 1/x^2$

For any M and any sequence $a_n \to 0$ we need to show that for some N, if $n > N$ then $f(a_n) > M$. If $M < 0$, replace M by 1. Thus, we can assume that $M > 0$. There is an N such that if $n > N$, then $|a_n| < 1/\sqrt{M}$. For such an a_n,

$$f(a_n) = a_n^2 > \left(\sqrt{M}\right)^2 = M.$$

This means that $f(a_n) \to \infty$ and that $\lim_{x\to 0} f(x) = \infty$.

Example 106. Let $f(x) = 1/x$ and consider $\lim_{x\to 0} f(x)$. The graph Fig. 3.11 on page 79 indicates that no limit exists since $f(x) < -1$ if $x \in (-1,0)$ and $f(x) > 1$ if $x \in (0,1)$.

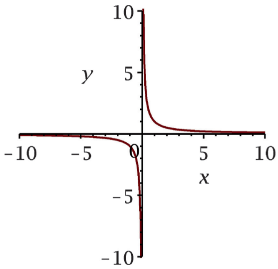


Fig. 3.11 $f(x) = 1/x$

For any $M > 0$ and any sequence $a_n \to 0^+$, there is an N such that if $n > N$ one has $0 < a_n < 1/M$. This means that as $a_n \to 0^+$, $f(a_n) \to \infty$. Similarly, for any sequence $a_n \to 0^-$, $f(a_n) \to -\infty$. The limit cannot exist.

The rules for calculating limits involving ∞ are the same as for finite limiting values except for the forms $\infty - \infty$, ∞/∞ and $\infty/0$. These are indeterminate forms, they have no fixed value, and are nontrivial to evaluate when taking limits. They are considered in Sect. 7.4.

Theorem 29 (Rules for infinite limits). *Let $f(x)$ and $g(x)$ be functions from $\mathbb{R}$ to $\mathbb{R}$ with limits K and L as x goes to a, and let $h(x)$ be a function from $\mathbb{R}$ to $\mathbb{R}$ with limit M as x goes to a. Here a can be a real number or $\pm\infty$ and K and L can be real numbers, one can be infinite or both can be infinite with the same sign. Then the following hold:*

(i) $\lim_{x\to a} (f+g)(x) = K + L$

(ii) $\lim_{x\to a} h(x) f(x) = MK$

(iii) $\lim_{x\to a} (f\,g)(x) = KL$

(iv) *If $M \neq 0$, then* $\lim_{x\to a} \dfrac{f(x)}{h(x)} = \dfrac{K}{M}$

The last example leads to the next definition of one sided limits equaling ∞.

Definition 15. The limit as $x \to a^+$ $(x \to a^-)$ is infinity if for all sequences $a_n \to a^+$ and all $M > 0$, there is an N such that if $n > N$, $f(a_n) > M$ $(f(a_n) > M)$. This is denoted by

$$\lim_{x\to a^+} f(x) = \infty \qquad \left(\lim_{x\to a^-} f(x) = \infty\right).$$

The limit as $x \to a^+$ $(x \to a^-)$ is negative infinity if for all sequences $a_n \to a^+$ $(a_n \to a^-)$ and all $M > 0$, there is an N such that if $n > N$, $f(a_n) < -M$ $(f(a_n) < -M)$. This is denoted by

$$\lim_{x\to a^+} f(x) = -\infty \qquad \left(\lim_{x\to a^-} f(x) = -\infty\right).$$

Example 106 on page 79 is a simple example of this concept.

Definition 16 (Infinite limits at infinity). Let $f : \mathbb{R} \to \mathbb{R}$. The limit as $x \to \infty$ of $f(x)$ is ∞ ($-\infty$), $\lim_{x\to\infty} f(x) = \infty$ ($\lim_{x\to\infty} f(x) = -\infty$), if for every sequence $a_n \to \infty$ and every $M > 0$ there is an N such that if $n > N$ then $a_n > M$ and $f(a_n) > M$ ($f(a_n) < -M$).

The limit as $x \to -\infty$ of $f(x)$ is ∞ ($-\infty$), $\lim_{x\to-\infty} f(x) = \infty$ ($\lim_{x\to-\infty} f(x) = -\infty$), if for every sequence $a_n \to -\infty$ and every $M > 0$ there is an N such that if $n > N$ then $a_n < -M$ and $f(a_n) > M$ ($f(a_n) < -M$).

Example 107. Let $f(x) = x$. Then $\lim_{x\to\infty} f(x) = \infty$. Let $a_n \to \infty$ and $M > 0$. by definition, after some N, $a_n > M$. Since $f(a_n) = a_n$, we also have $f(a_n) > M$. This means that $\lim_{x\to\infty} f(x) = \infty$.

Example 108. Let $f(x) = \ln(x)$. Let $a_n \to \infty$ and fix $M > 0$. Since $\ln(2) > 0$, there is a $K > 0$ such that $K\ln(2) > M$. Choose an N such that if $n > N$ then $a_n > 2^K$. In addition, $\ln(a_n) > \ln(2^K) > K\ln(2) > M$. This means that $\lim_{x\to\infty} \ln(x) = \infty$.

The theorem on limits of compositions, Theorem 17, generalizes here. The next result is used in Sect. 7.4.

Theorem 30 (Limits of compositions involving infinity).

(I) Assume that $f(x)$ has a limit L as $x \to \pm\infty$ and that $g(y)$ is continuous at L. Then the limit as $x \to \pm\infty$ of $g \circ f(x)$ exists and equals $g(L)$.
(II) Assume that $\lim_{x\to\infty} f(x) = \infty$ and that $\lim_{y\to\infty} g(y) = \infty$. Then $\lim_{x\to\infty} g \circ f(x) = \infty$. This also holds when one replaces infinity by negative infinity in any way that makes sense.

A couple examples show how Theorem 30 is used.

Example 109. Let $f(x) = 1 - \ln(x)/x$ and let $g(x) = e^x$. Then $\lim_{x\to\infty} f(x) = 1$ and $\lim_{x\to\infty} g \circ f(x) = e^1$.

Example 110. Let $f(x) = -x^2 + x$ and $g(y) = \ln(|y|)$. Then $\lim_{x\to\infty} f(x) = -\infty$ and $\lim_{x\to\infty} g \circ f(x) = -\infty$.

Exercises

1. Do the following sequences go to infinity?
   (a) $a_n = n^2$
   (b) $a_n = \sqrt{n+10}$
   (c) $a_n = \frac{n^2+4}{n-n^2}$
   (d) $a_n = \frac{n^3+n+3}{n^2-4n+6}$
   (e) $b_n = \frac{n^2-4n+6}{n^3+n+3}$
   (f) $b_n = \tan\left(\frac{\pi}{2} - \frac{1}{n}\right)$
   (g) $b_n = \tan\left(\frac{\pi}{2} + \frac{1}{n}\right)$
   (h) $b_n = \ln\left(\frac{1}{n}\right)$
2. Find the following limits, if they exist.
   (a) $\lim\limits_{x\to\infty} \sin\left(\frac{1}{x}\right)$
   (b) $\lim\limits_{x\to\infty} \sin\left(x^2\right)$
   (c) $\lim\limits_{x\to-\infty} \frac{1+x^2}{3x^2-1}$
   (d) $\lim\limits_{z\to\infty} \ln(z)$
   (e) $\lim\limits_{z\to\infty} \ln\left(1+\frac{1}{z}\right)$
   (f) $\lim\limits_{z\to-\infty} \frac{\sqrt{1-z}}{z}$
   (g) $\lim\limits_{z\to-\infty} e^{z^2-z}$
   (h) $\lim\limits_{z\to\infty} e^{-z+50}$

3. Find the following limits, if they exist.

(a) $\displaystyle\lim_{x\to 2}\frac{x^2-4}{(x-2)^3}$

(b) $\displaystyle\lim_{x\to 2}\frac{x^2-4}{(x-2)^2}$

(c) $\displaystyle\lim_{x\to -1}\ln(|x+1|)$

(d) $\displaystyle\lim_{z\to \pi/2}\tan(z)$

(e) $\displaystyle\lim_{z\to 0}\csc(z)$

(f) $\displaystyle\lim_{z\to 2}\frac{z^2-4z+4}{z^3-12z^2+32z-64}$

(g) $\displaystyle\lim_{w\to -3}\frac{w^3-9w^2+27w-27}{w^3-9}$

(h) $\displaystyle\lim_{w\to 0}\csc^2(w)$

4. Find both of the one sided limits for the following functions at the given point, if they exist.

(a) $f(x)=\dfrac{x^2-4}{(x-2)^3}$, $x=2$

(b) $f(x)=\dfrac{x^2-4}{(x-2)^2}$, $x=2$

(c) $g(x)=\ln(|x+1|)$, $x=-1$

(d) $g(z)=\tan(z)$, $z=\pi/2$

(e) $h(z)=\csc(z)$, $z=0$

(f) $h(z)=\dfrac{z^2-4z+4}{z^3-12z^2+32z-64}$, $z=2$

(g) $s(w)=\dfrac{w^3-9w^2+27w-27}{w^3-9}$, $w=3$

(h) $s(w)=\csc^2(w)$, $w=0$

5. Find the one sided limits for the following functions at the given point, if they exist.

(a) $\displaystyle\lim_{x\to 0^+}\ln(x)$

(b) $\displaystyle\lim_{x\to 0^-}\frac{1}{\sqrt{-x}}$

(c) $\displaystyle\lim_{x\to 1^+}\frac{\sqrt{x-1}}{x-1}$

(d) $\displaystyle\lim_{x\to 1^+}\frac{x-1}{\sqrt{x-1}}$

6. Show that the piecewise defined function

$$w(y)=\begin{cases} y^3 & \text{if } y<0\\ y^2 & \text{if } y\ge 0\end{cases}$$

has a derivative at $y=0$. You should use two one-sided limits.

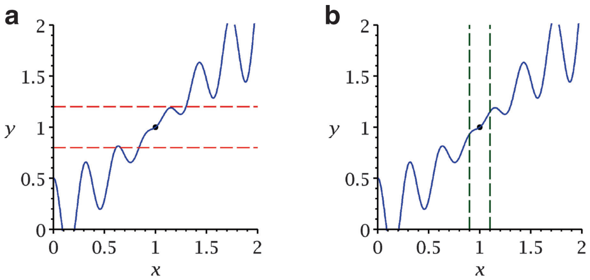
7. Show that the piecewise defined function

$$w(y)=\begin{cases} y^3+y & \text{if } y<0\\ y^2 & \text{if } y\ge 0\end{cases}$$

does not have a derivative at $y=0$. You should use two one-sided limits.

3.3 Limits and the Epsilon-Delta Definition

In most calculus books limits are defined indirectly using the idea that when we want $\mathbf{f}(x)$ to be "close" to $\mathbf{L}$ when x is "close enough" to a. We can be view this as a statement about sets in the domain of $\mathbf{f}(x)$ of the form $\{x : |||\mathbf{f}(x)-\mathbf{L}\| < \varepsilon\}$. Since many students are not well versed in the behavior of an "inverse" of $\mathbf{f}(x)$ that may not be single valued, this is difficult for many students to understand. The definition following this idea is precise and is the way that mathematicians



A horizontal band around $y = L$ A vertical band around $x = a$

Fig. 3.12

often approach the concept of a limit. It is equivalent to the definition used in this book, so there is no loss of generality in using either definition.

Here we want to capture the idea that $\mathbf{f}(x)$ approaches $\mathbf{L}$ as $x \to a$. The idea of being close to $\mathbf{L}$ is captured by saying that $\|\mathbf{f}(x) - \mathbf{L}\|$ is small. In mathematical terms, $\|\mathbf{f}(x) - \mathbf{L}\| < \varepsilon$ for some $\varepsilon > 0$. This is illustrated by Fig. 3.12a on page 82.

The idea of x being close to a is mathematically expressed as $|x - a| < \delta$ for a $\delta > 0$. This is illustrated by Fig. 3.12b on page 82.

When we combine these two pictures, the idea becomes that the portion of the graph of the function $\mathbf{f}(x)$ in the "vertical" stripe must be contained in the "horizontal" stripe. This is shown in Fig. 3.13 on page 82.

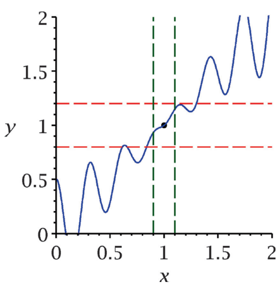


Fig. 3.13 Matched horizontal and vertical bands

In this picture, the graph of $f(x)$ is between the horizontal lines whenever the x value is between the vertical lines.

With that motivation we come to the ε-δ definition of a limit. Again, the value of $\mathbf{f}(x)$ at a is not considered.

Definition 17. Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ be defined on an interval $I = (b, c)$, excluding the point $a \in (b, c)$. We say that the limit as x goes to a of $\mathbf{f}(x)$ is $\mathbf{L}$,

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

if for all $\varepsilon > 0$ there is a $\delta > 0$ such that if $0 < |x - a| < \delta$ and $x \in I$ we have

$$\|\mathbf{f}(x) - \mathbf{L}\| < \varepsilon.$$

The results on limits from earlier sections can be proven using ε-δ proofs that are similar to those using the sequence based definition. *The two definitions of the limit in this book are* *equivalent.* This means that if a limit exists under one of the definitions, it exists under the other definition and has the same value. The next example shows how we demonstrate that

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

using the $\varepsilon - \delta$ definition. This example shows all of the work required to get a δ for a specific ε.

Example 111. Let $f(x) = x^2$ and consider $\lim_{x\to 2} f(x)$. Assume that this limit is 4 and let $x = 2+h$. Then, if we assume that $|h| < 4$,

$$\begin{aligned} |f(x) - 4| &= |(2+h)^2 - 4| \\ &= |4h + h^2| \\ &\le |4h| + h^2 \\ &\le 8|h|. \end{aligned}$$

This means that for any $\varepsilon > 0$, if we take

$$|x-2| = |h| < \delta = \frac{\varepsilon}{8}$$

we have

$$|f(x) - 4| \le 8|h| < 8\delta = \varepsilon.$$

The conclusion is that $f(x) = x^2$ is continuous at $x = 2$.

For completeness of the presentation, the equivalence of the two definitions of the limits of functions is stated and proved.

Theorem 31. *Let $\mathbf{f}(x)$ be a function defined on an interval $I = (b,c)$, except possibly a point $a \in (b,c)$. Then,*

$$\lim_{n\to\infty} \mathbf{f}(a_n) = \mathbf{L}$$

for every sequence $a_n \to a$, $a_n \neq a$, if and only if for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $0 < |x-a| < \delta$ then

$$\|\mathbf{f}(x) - \mathbf{L}\| < \varepsilon.$$

Proof. First it is shown that if a limit exists using the ε-δ definition, then it exists under the sequence definition with the same limit value. Assume that the limit as $x \to a$ of $\mathbf{f}(x)$ is $\mathbf{L}$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $0 < |x-a| < \delta$ then

$$\|\mathbf{f}(x) - \mathbf{L}\| < \varepsilon.$$

Let $a_n \to a$, $a_n \neq a$, and fix any $r > 0$. Then there is a $\delta > 0$ such that if $0 < |x-a| < \delta$ then

$$\|\mathbf{f}(x) - \mathbf{L}\| < r.$$

Since $a_n \to a$, after some $N > 0$, $|a_n - a| < \delta$. This means that for any sequence $a_n \to a$, $a_n \neq a$, and any $r > 0$ there is an $N > 0$ such that if $n > N$,

$$\|\mathbf{f}(a_n) - \mathbf{L}\| < r.$$

This gives us that the sequence definition of convergence is fulfilled and completes half of the equivalence.

Now assume that the limit of $\mathbf{f}$ exists at a under the sequence definition of limit and equals $\mathbf{L}$. Then for any sequence $a_n \to a$, $a_n \neq a$, and any $r > 0$ there is an $N > 0$ such that if $n > N$,

$$\|\mathbf{f}(a_n) - \mathbf{L}\| < r.$$

Also assume that for some $\varepsilon > 0$ and all $\delta > 0$, there is an x_δ with $0 < |x_\delta - a| < \delta$ and

$$\|\mathbf{f}(x_\delta) - \mathbf{L}\| \geq \varepsilon.$$

(The previous statement is a contradiction to the limit existing with the ε-δ definition. If this leads to an impossible situation, the ε-δ definition must hold.) This means that for every integer n there is an a_n with $0 < |a_n - a| < \delta = 1/n$ and $\|\mathbf{f}(a_n) - \mathbf{L}\| \geq \varepsilon$. Since $a_n \to a$, this contradicts the assumption that $\mathbf{f}(a_n) \to \mathbf{L}$ for every sequence $a_n \to a$, $a_n \neq a$. Since the contradicts the assumption that the sequence definition of the limit holds, the $\varepsilon - \delta$ definition must hold and the proof is complete.

Exercises

1. For each of the following functions $f(x)$, points a, values L, and $\varepsilon > 0$ find a $\delta > 0$ such that if $|x - a| < \delta$ then

$$|f(x) - L| < \varepsilon.$$

(a) $f(x) = x$, $a = 2$, $L = 2$, and $\varepsilon = 0.1$
(b) $f(x) = -x$, $a = 2$, $L = -2$, and $\varepsilon = 0.1$
(c) $f(x) = x + 1$, $a = 2$, $L = 3$, and $\varepsilon = 0.2$
(d) $f(x) = -x + 1$, $a = -2$, $L = 3$, and $\varepsilon = 0.2$
(e) $f(x) = 2x$, $a = 1$, $L = 2$, and $\varepsilon = 0.2$
(f) $f(x) = 2x - 1$, $a = 2$, $L = 3$, and $\varepsilon = 0.1$
(g) $f(x) = x^2 + 1$, $a = 0$, $L = 1$, and $\varepsilon = 0.01$
(h) $f(x) = x^3$, $a = 0$, $L = 0$, and $\varepsilon = 0.008$

2. For each of the following functions $\mathbf{f}(x)$, points a, values $\mathbf{L}$, and $\varepsilon > 0$ find a $\delta > 0$ such that if $|x - a| < \delta$ then

$$\|\mathbf{f}(x) - \mathbf{L}\| < \varepsilon.$$

(a) $\mathbf{f}(x) = (1, x)$, $a = 2$, $\mathbf{L} = (1, 2)$, and $\varepsilon = 0.1$
(b) $\mathbf{f}(x) = (-x, 2, 1)$, $a = 2$, $\mathbf{L} = (-2, 2, 1)$, and $\varepsilon = 0.1$
(c) $\mathbf{f}(x) = (x, x)$, $a = 1$, $\mathbf{L} = (1, 1)$, and $\varepsilon = 0.1$
(d) $\mathbf{f}(x) = (-x, 4, x)$, $a = -1$, $\mathbf{L} = (1, 4, -1)$, and $\varepsilon = 0.1$
(e) $\mathbf{f}(x) = (2x, x, -x)$, $a = 5$, $\mathbf{L} = (10, 5, -5)$, and $\varepsilon = 0.05$
(f) $\mathbf{f}(x) = (x^2, -1, 0)$, $a = 0$, $\mathbf{L} = (0, -1, 0)$, and $\varepsilon = 0.01$
(g) $\mathbf{f}(x) = (-x^2, x, -1)$, $a = -2$, $\mathbf{L} = (-4, -2, 1)$, and $\varepsilon = 0.02$
(h) $\mathbf{f}(x) = (x^3, x^3, 1)$, $a = 1$, $\mathbf{L} = (1, 1, 1)$, and $\varepsilon = 0.008$

3. Use the ε-δ definition of the limit to show that $f(x) = \lfloor x \rfloor$ is not continuous at $x = 0$.

4. Use the ε-δ definition of the limit to show that $g(y) = 1/(y^2 - 1)$ does not have a limit at $y = 1$.
5. Use the ε-δ definition of the limit to show that

$$h(z) = \begin{cases} \sin\left(\frac{1}{z}\right) & \text{if } z \neq 0 \\ 0 & \text{if } z = 0 \ . \end{cases}$$

is not continuous at $z = 0$.

Chapter 4
Rules for Finding Derivatives

The idea of a derivative, instantaneous rate of change, is difficult to use in many settings unless one can easily find the derivative of a function. This chapter is devoted to rules, formulas, for finding derivatives. Examples of how these rules can be applied to functions are included throughout the chapter.

Without these rules, it is very difficult to solve many simple problems in science and engineering. Although derivatives are often approximated numerically, you will still need these derivative rules to work with both known and unknown functions.

4.1 Elementary Rules

This section is devoted to some of the most basic rules, formulas, for finding derivatives. These rules are used to show that all polynomials have derivatives everywhere and are used throughout science and engineering.

The first two rules are for very simple functions.

Theorem 32. *The derivative of any constant function,* $\mathbf{f}(x) = \mathbf{c}$*, is* $\mathbf{f}\,'(x) = \mathbf{0}$.

Proof. We simply calculate using the definition.

$$\begin{aligned}
\frac{d\,\mathbf{f}(x)}{d\,x} &= \lim_{h\to 0} \frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} \\
&= \lim_{h\to 0} \frac{\mathbf{c} - \mathbf{c}}{h} \\
&= \lim_{h\to 0} \mathbf{0} \\
&= \mathbf{0}.
\end{aligned}$$

Theorem 33. *The derivative of* $f(x) = x$ *is* $f'(x) = 1$.

Proof. This is a simple computation.

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{(x+h) - x}{h} \\
&= \lim_{h\to 0} \frac{h}{h} \\
&= 1.
\end{aligned}$$

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_4

Example 112. The above results mean that the derivative of $g(x) = 2$ is $g'(x) = 0$.

Consider the positions of two cars going west on a road from New York. The distance of car B from New York is always twice the distance of car A from New York. If the speeds of the cars are both constant, the speed of car B must be twice the speed of car A. The following result can be interpreted as saying that the same conclusion is true for any speeds.

Theorem 34 (Scalar Multiple). *Assume that* $\mathbf{f}(x)$ *is differentiable at* $x = a$ *and that* s *is a constant. The derivative of* $\mathbf{g}(x) = s\mathbf{f}(x)$ *at* $x = a$ *is given by*

$$\mathbf{g}'(a) = s\mathbf{f}'(a).$$

Proof. This is a calculation using the definition of the derivative.

$$\begin{aligned}
\mathbf{g}'(a) &= \lim_{h\to 0} \frac{\mathbf{g}(a+h) - \mathbf{g}(a)}{h} \\
&= \lim_{h\to 0} \frac{s\mathbf{f}(a+h) - s\mathbf{f}(a)}{h} \\
&= s \lim_{h\to 0} \frac{\mathbf{f}(a+h) - \mathbf{f}(a)}{h} \\
&= s\mathbf{f}'(a).
\end{aligned}$$

Example 113. Let $\mathbf{f}(x) = 6\,(x,4)$. As we will show later, $\mathbf{g}(x) = (x,4)$ is differentiable at every x with derivative $\mathbf{g}'(x) = (1,0)$. This means we have

$$\mathbf{f}'(x) = 6\mathbf{g}'(x) = 6\ (1,0).$$

Example 114. Let $g(t) = -32\cos(t)$. Since $\cos(t)$ is differentiable everywhere,

$$g'(t) = -32\frac{d\cos(t)}{dt} = 32\,\sin(t).$$

Example 115. Consider an object going around a circle of radius 5 cm centered at the origin at constant speed such that it takes 2π seconds to go around the circle. This means that the speed of the object is 5 m/s. The same geometry that is used in the beginning of Sect. 2.4 holds. First, this gives the parametrization $\mathbf{r}(t) = 5\big(\cos(t), \sin(t)\big)$ of the circle. Second, the velocity $\mathbf{v}(t)$ is in the direction $\big(-\sin(t), \cos(t)\big)$ when the object is at $5\big(\cos(t), \sin(t)\big)$. Combining these facts tells us that $\mathbf{v}(t) = 5\big(-\sin(t), \cos(t)\big)$.

On the other had, using the scalar multiple rule, we have

$$\begin{aligned}
\mathbf{v}(t) &= \frac{d}{dt}\mathbf{r}(t) \\
&= 5\left(\frac{d}{dt}\cos(t), \frac{d}{dt}\sin(t)\right) \\
&= 5\big(-\sin(t), \cos(t)\big).
\end{aligned}$$

Now consider a train that is a mile long is traveling south at 20 mph and a person on the train who is walking at 2 mph from the back of the train toward the front of the train. How far will the person have traveled after a half hour? The back of the train will have gone 10 mi and the person will have gone from the back to the front, a total of 11 mi. Since the speeds are all constant, we can conclude that the speed of the person is $20 + 2 = 22$ mph. This example may convey the idea behind the sum rule, the rate of change of a sum of functions is the sum of their rates of change.

Theorem 35 (Sum Rule). *Let* $\mathbf{f}(x)$ *and* $\mathbf{g}(x)$ *be functions that have derivatives at a. Then*

$$(\mathbf{f}+\mathbf{g})'(a) = \mathbf{f}'(a) + \mathbf{g}'(a).$$

Proof. Assume that $\mathbf{f}(x)$ and $\mathbf{g}(x)$ are differentiable at a. Then

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

Example 116. Let $f(x) = 4x + 5$. Then

$$\begin{aligned}
\frac{d\,f(x)}{dx} &= \frac{d\,(4x)}{dx} + \frac{d\,(5)}{dx} \\
&= 4\frac{d\,(x)}{dx} + 0 \\
&= 4 \cdot 1 = 4.
\end{aligned}$$

The last abstract rule for this section is the product rule. It is not easy to motivate, but is extremely important. We will use it extensively throughout the rest of the book.

Theorem 36 (Product Rule). *Let* f *and* g *be functions that are differentiable at a. Then*

$$(f\,g)'(a) = f'(a)\,g(a) + f(a)\,g'(a).$$

Proof. We only need to do a careful calculation.

$$\begin{aligned}
(f\,g)'(a) &= \lim_{h\to 0} \frac{f(a+h)\,g(a+h) - f(a)\,g(a)}{h} \\
&= \lim_{h\to 0} \frac{f(a+h)\,g(a+h) - f(a)\,g(a+h) + f(a)\,g(a+h) - f(a)\,g(a)}{h} \\
&= \lim_{h\to 0} \left(\frac{f(a+h) - f(a)}{h}\; g(a+h) \right) + \lim_{h\to 0} \left(f(a)\; \frac{g(a+h) - g(a)}{h} \right) \\
&= \left(\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} \right) \left(\lim_{h\to 0} g(a+h) \right) \\
&\qquad + \left(\lim_{h\to 0} f(a) \right) \left(\lim_{h\to 0} \frac{g(a+h) - g(a)}{h} \right) \\
&= f'(a)\,g(a) + f(a)\,g'(a).
\end{aligned}$$

We can use this to calculate the derivative of $f(x) = x^n$ is n is a positive integer. For example, since we know that the derivative of $g(x) = x$ is 1, then

$$\begin{aligned}
\frac{d}{dx} x^2 &= \frac{d}{dx}(xx) \\
&= \left(\frac{d}{dx}(x) \right) x + x \left(\frac{d}{dx}(x) \right) \\
&= x + x \\
&= 2x.
\end{aligned}$$

Using this reasoning we can, by induction, prove the following.

Theorem 37 (Power Rule). *Let n be a nonnegative integer. The derivative of $g(x) = x^n$ is given by*

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

Here $0 \cdot x^{-1}$ is interpreted as 0 for all x.

Example 117. The derivative of $g(x) = x^{101}$ is $g'(x) = 101x^{100}$.

Example 118. Combining the sum, constant multiple, and power rules allows the following calculation.

$$\begin{aligned}\frac{d}{dz}\left(4z^8 - 3z^3 - 4\right) &= \frac{d}{dz}\left(4z^8\right) + \frac{d}{dz}\left(-3z^3\right) + \frac{d}{dz}(-4)\\ &= 4\frac{d}{dz}\left(z^8\right) - 3\frac{d}{dz}\left(z^3\right) + 0\\ &= 32z^7 - 9z^2\end{aligned}$$

Using the technique in the last example and induction we can prove the following rule for finding the derivatives of polynomials, Theorem 25 on page 61.

Theorem 38. *Let $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ be any polynomial of degree n. The derivative of $p(x)$ is given by*

$$\frac{d\,p(x)}{dx} = a_1 + 2a_2x + 3a_3x^2 + \cdots + ma_mx^{m-1} + \cdots + na_nx^{n-1}.$$

Example 119. The derivative of $p(w) = -16w^{10} + 12w^8 + 3w^5 - 9w^2 + 2w + 16$ is given by

$$\begin{aligned}\frac{d\,p(w)}{dw} &= -16 \cdot 10 \cdot w^{10-1} + 12 \cdot 8 \cdot w^{8-1} + 3 \cdot 5 \cdot w^{5-1} - 9 \cdot 2 \cdot w^{2-1} + 2\\ &= -160w^9 + 96w^7 + 15w^4 - 18w + 2.\end{aligned}$$

Example 120. In elementary physics it is common to give the motion of a projectile shot from a canon with horizontal position $x(t) = v_x t$ m and vertical position $y(t) = -4.9t^2 + v_y t + y_0$ m. In vector form this is

$$\mathbf{r}(t) = \left(v_x t,\, -4.9t^2 + v_y t + y_0\right) \text{ m}.$$

Using Theorem 38 on page 90 and Theorem 24 on page 55 we have

$$\begin{aligned}\mathbf{r}' &= \left(\frac{d}{dt}(v_x t),\, \frac{d}{dt}\left(-4.9t^2 + v_y t + y_0\right)\right) \text{ m/s}\\ &= (v_x,\, -9.8t + v_y) \text{ m/s}.\end{aligned}$$

The rules for derivatives of products involving vector valued functions are similar to those for scalar valued functions. The only real change is that the products are different. Proving this result is almost identical to proving the product rule, Theorem 36 on page 89. The major difference is that the computations are done coordinate by coordinate.

Theorem 39. *Let $\mathbf{f}(t)$, $\mathbf{g}(t)$, and $h(t)$ be functions that have derivatives at $t = t_0$. Then*

(1) $\frac{d}{dt}(\mathbf{f} \cdot \mathbf{g})(t_0) = \mathbf{f}'(t_0) \cdot \mathbf{g}(t_0) + \mathbf{f}(t_0) \cdot \mathbf{g}'(t_0)$*, and*

(2) $\frac{d}{dt}(h\mathbf{g})(t_0) = h'(t_0) \cdot \mathbf{g}(t_0) + h(t_0) \cdot \mathbf{g}'(t_0)$.

Example 121. Let $\mathbf{f}(t) = (t^2, t^3, 4)$, $\mathbf{g}(t) = (1-t, 3-t, t^2)$, and $h(t) = t^2$. Then

$$\begin{aligned}
\frac{d}{dt}(\mathbf{f}\cdot\mathbf{g})(t) &= \left(2t, 3t^2, 0\right)\cdot\left(1-t, 3-t, t^2\right) + \left(t^2, t^3, 4\right)\cdot(-1,-1,2t) \\
&= \left(2t - 2t^2 + 9t^2 - 3t^3 + 0\right) + \left(-t^2 - t^3 + 8t\right) \\
&= -4t^3 + 6t^2 + 9t.
\end{aligned}$$

Also

$$\begin{aligned}
\frac{d}{dt}(h\mathbf{g})(t) &= 2t\,\left(1-t, 3-t, t^2\right) + t^2\,(-1,-1,2t) \\
&= \left(2t - 3t^2, 6t - 3t^2, 0\right).
\end{aligned}$$

Exercises

1. Find the derivatives of the following functions.

(a) $f(x) = 10$
(b) $g(y) = -\frac{1}{23}$
(c) $h(z) = x + y$
(d) $f(x) = x^5$
(e) $g(y) = x^{37}$
(f) $h(z) = 15z^2$
(g) $f(x) = 3x^5$
(h) $g(y) = 10 + 5y - y^2 + 3y^3$
(i) $h(z) = 11z^5 - 2z^4 + z - 9$
(j) $f(x) = x\sin(x)$
(k) $g(y) = e^y\cos(y)$
(l) $h(z) = 24ze^z$
(m) $f(x) = 2x^2\cos(x) - x\sin(x)$
(n) $g(y) = y\cos(y)\,e^y$
(o) $h(z) = z^{15}\sin(z)\ln(z)$

2. Find the following derivatives, if they exist.

(a) $\mathbf{f}(x) = (3xe^x, x^2\cos(x), 3x^5\ln(x))$
(b) $\mathbf{g}(y) = (10y^3\sin(y), 10 + 5y - y^2e^y + 3y^3)$
(c) $\mathbf{h}(z) = (10z^2\ln(z), 6e^z\sin(z), \cos(z)\sin(z))$
(d) $\mathbf{f}(x) = (3, x^2\cos(x), 3x^5\sin(x))$
(e) $\mathbf{g}(y) = (e^y\sin(y), e^y\ln(y))$
(f) $\mathbf{h}(z) = (10z^2e^z\sin(z), z\cos(z)\ln(z))$

3. Find the derivatives of the following functions.

(a) $\mathbf{f}(t) = t^2\,(\sin(t), \cos(t))$
(b) $\mathbf{r}(t) = e^t\left(t^2, 3t^5\right)$
(c) $\mathbf{f}(t) = \cos(t)\,(\ln(t), \sin(t))$
(d) $\mathbf{r}(t) = \ln(t)\left(te^t, t^{16}\right)$

4. Use the product rule to find derivatives of the following functions.

(a) $f(x) = e^{2x}$
(b) $f(x) = \cos(2x)$
(c) $g(x) = \sin(2x)$
(d) $g(z) = \ln\left(z^z\right)$

5. Find the slope-intercept form of the tangent line to the graph of the function at the given x value.

(a) $f(x) = xe^x$, $x = 0$

(b) $f(x) = x\ln(x)$, $x = 1$

(c) $g(x) = x^2\sin(x)$, $x = \frac{\pi}{2}$

(d) $g(x) = x^3 - 3x^2 + 2x - 1$, $x = -1$

6. Find a parametrization for the tangent line to the image of the function at the given t value.

(a) $\mathbf{r}(t) = (5t + \cos(t), t\sin(t), e^t)$, $t = 0$

(b) $\mathbf{r}(t) = (t^2 - 2t, t\ln(t^2))$, $t = 2$

(c) $\mathbf{g}(t) = (t\sqrt{t}, \sin(t)\,e^t)$, $t = 0$

(d) $\mathbf{g}(t) = (6t^2 - 5t, t^3 - 6, 4t^2 + 6t^6)$, $t = -1$

7. In the following $\mathbf{r}(t)$ is the position of an object as a function of time. Find a unit vector in the direction of travel of the object at the given t.

(a) $\mathbf{r}(t) = (5t + \cos(t), t\sin(t), e^t)$, $t = 0$

(b) $\mathbf{r}(t) = (t^2 - 2t, t\ln(t^2))$, $t = 2$

(c) $\mathbf{r}(t) = (5\sqrt{t} - t, \sin(\pi t)\,e^t)$, $t = 0$

(d) $\mathbf{r}(t) = (t\,e^t, 4t^2 - 4t)$, $t = -1$

8. If $\mathbf{r}(1) = (3,-1,2)$, $\mathbf{x}(1) = (-1,1,-1)$, $\mathbf{r}'(1) = (0,-2,2)$, and $\mathbf{x}' = (-1,-1,0)$ find

$$\frac{d}{dt}\big(\mathbf{r}(t)\cdot\mathbf{x}(t)\big)$$

at $t = 1$.

9. If $\mathbf{r}(1) = (2,-4,1)$, $\mathbf{x}(1) = (2,1,3)$, $\mathbf{r}'(1) = (-2,2,1)$, and $\mathbf{x}'(1) = (-1,-2,3)$ find

$$\frac{d}{dt}\big(\mathbf{r}(t)\cdot\mathbf{x}(t)\big)$$

at $t = 1$.

10. A person is walking around an elliptical track that can be parameterized by $\mathbf{s}(\theta) = \big(4\cos(\theta), \sin(\theta)\big)$ km. The person is walking counterclockwise around the track. The person's position and speed at $t = 2$ h are $\mathbf{s}\left(\frac{\pi}{4}\right)$ and 5 km/h. Approximate the person's position at $t = 2.05$ h. (Hint: Draw a picture.)

11. What is the derivative of the product of three scalar valued functions: $f(x)$, $g(x)$, and $h(x)$? (**Hint**: Consider the product of $f(x)$ and $k(x) = g(x)\cdot h(x)$.)

12. Complete the induction proof of the power rule, Theorem 37 on page 90.
The idea of mathematical induction is to prove a statement P that depends on integers n is true for all $n \geq n_0$ for a fixed n_0. To prove a mathematical result by induction we must show that the result is true for the base case n_0, the first n where we want the result to be true. We then must prove that if the result is true for an arbitrary $n \geq n_0$, then the result is true for the $n+1$st case. This shows that the result is true for all $n \geq n_0$.

4.2 The Quotient Rule

This section is devoted to the quotient rule. It is a method commonly used to find the derivative of the quotient of two functions. Many students find it difficult to remember the formula properly. As you will see in this section, there are multiple ways of actually taking the derivative of a quotient of functions.

The first thing that you need to realize is that the derivative of a function $h(x) = \dfrac{f(x)}{g(x)}$ cannot exist at a point a where $g(a) = 0$ since h is not defined at a. The assumptions that are made for the quotient are that $f(x)$ and $g(x)$ are differentiable at a and that $g(a) \neq 0$. Since $g(a) \neq 0$, $g(x)$ is not zero on some interval $(a-\delta, a+\delta)$, see Theorem 21 on page 46. This justifies the assumption, in what follows, that $g(x)$ is never 0 at the points we consider.

Before considering the quotient rule, the following useful result is proven.

Lemma 1. *Assume that $g(x)$ is differentiable at a and that $g(a) \neq 0$. Then*

$$\frac{d}{dx}\frac{1}{g(a)} = -\frac{g'(a)}{(g(a))^2}.$$

Proof. The proof of this result is not extremely obvious. The main part is rewriting numerator of the difference quotient as a single fraction.

$$\begin{aligned}
\frac{d}{dx}\frac{1}{g(a)} &= \lim_{h\to 0}\frac{\frac{1}{g(a+h)} - \frac{1}{g(a)}}{h} \\
&= \lim_{h\to 0}\frac{g(a) - g(a+h)}{h\,g(a)\,g(a+h)} \\
&= -\lim_{h\to 0}\frac{g(a+h) - g(a)}{h}\lim_{h\to 0}\frac{1}{g(a)\,g(a+h)} \\
&= -g'(a)\,\frac{1}{(g(a))^2}
\end{aligned}$$

This can be used to find the derivatives of secant, cosecant and x^{-n} for positive integers n.

Example 122. Since $\sec(x) = 1/\cos(x)$, we have

$$\begin{aligned}
\frac{d}{dx}\sec(x) &= \frac{d}{dx}\frac{1}{\cos(x)} \\
&= \frac{-1}{(\cos(x))^2}\frac{d\cos(x)}{dx} \\
&= \frac{\sin(x)}{(\cos(x))^2} \\
&= \sec(x)\tan(x).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{d}{dx}\csc(x) &= \frac{d}{dx}\frac{1}{\sin(x)} \\
&= \frac{-1}{(\sin(x))^2}\frac{d\sin(x)}{dx} \\
&= \frac{-\cos(x)}{(\sin(x))^2} \\
&= -\csc(x)\cot(x).
\end{aligned}$$

Example 123. Using the fact that $x^{-n} = 1/x^n$ we have

$$\begin{aligned}
\frac{d}{dx}\left(x^{-n}\right) &= \frac{d}{dx}\left(\frac{1}{x^n}\right) \\
&= \frac{-1}{x^{2n}}\frac{d}{dx}x^n \\
&= \frac{-1}{x^{2n}}nx^{n-1} \\
&= \frac{-n}{x^{n+1}} \\
&= -nx^{-n-1}
\end{aligned}$$

On its own the proposition above is very useful. It can be used to prove the quotient rule.

Theorem 40 (Quotient Rule). *Let* $\mathbf{f}(x) : \mathbb{R} \to \mathbb{R}^n$ *and* $g(x) : \mathbb{R} \to \mathbb{R}$ *be differentiable functions at* $x = a$ *and assume that* $g(a) \neq 0$*. Then*

$$\frac{d}{dx}\left(\frac{\mathbf{f}}{g}\right)(a) = \frac{\mathbf{f}'(a)g(a) - \mathbf{f}(a)g'(a)}{(g(a))^2}.$$

Proof. This uses the product rule and Proposition 1 on page 93. Assume that $\mathbf{f}(x)$ and $g(x)$ satisfy the hypotheses of the theorem. Then

$$\begin{aligned}
\frac{d}{dx}\left(\frac{\mathbf{f}(x)}{g(x)}\right)(a) &= \frac{d}{dx}\left(\mathbf{f}(x)\frac{1}{g(x)}\right)(a) \\
&= \frac{\mathbf{f}'(a)}{g(a)} + \mathbf{f}(a)\frac{-g'(a)}{(g(a))^2} \\
&= \frac{\mathbf{f}'(a)\,g(a) - \mathbf{f}(a)\,g'(a)}{(g(a))^2}
\end{aligned}$$

This can be used to find the derivatives of $\tan(\theta)$ and $\cot(\theta)$.

Example 124. The derivative of $\tan(\theta)$ is $\sec^2(\theta)$ and the derivative of $\cot(\theta)$ is $-\csc^2(\theta)$.

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

and

$$\begin{aligned}
\frac{d}{d\theta}\cot(\theta) &= \frac{d}{d\theta}\frac{\cos(\theta)}{\sin(\theta)} \\
&= \frac{-\sin(\theta)\sin(\theta) - \cos(\theta)\cos(\theta)}{(\sin(\theta))^2} \\
&= \frac{-1}{(\sin(\theta))^2} \\
&= -(\csc(\theta))^2.
\end{aligned}$$

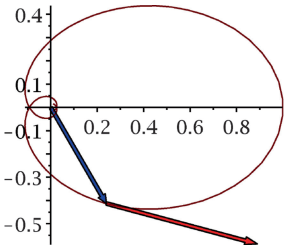


Fig. 4.1 Location and derivative vectors for $\mathbf{h}(s)$ as in Example 125

This rule can also be used to find the derivative of $\mathbf{h}(s) = \dfrac{(\cos(s), \sin(s))}{s^2+1}$. The graph of this function is in Fig. 4.1. The derivative at $s = -\pi/3$ is shown in red.

Example 125. The derivative of $\mathbf{h}(s) = \dfrac{(\cos(s), \sin(s))}{s^2+1}$ is found by the following computation.

$$\begin{aligned}\frac{d}{ds}\left(\frac{(\cos(s), \sin(s))}{s^2+1}\right) &= \frac{\big((-\sin(s), \cos(s))\,(s^2+1) - 2s\,(\cos(s), \sin(s))\big)}{(s^2+1)^2} \\ &= \frac{\big(-(s^2+1)\,\sin(s) - 2s\,\cos(s), (s^2+1)\,\cos(s) - 2s\,\sin(s)\big)}{(s^2+1)^2}.\end{aligned}$$

The derivative at $s = -\pi/3$ is

$$\begin{aligned}\frac{d\mathbf{h}(s)}{ds}\left(-\frac{\pi}{3}\right) &= \frac{\left((\frac{\sqrt{3}}{2}, \frac{1}{2})\,((\frac{\pi}{3})^2+1) - 2\frac{\pi}{3}\,\left(\frac{1}{2}, -\frac{\sqrt{3}}{3}\right)\right)}{\left((\frac{\pi}{3})^2+1\right)^2} \\ &\approx (0.651, \ -.174).\end{aligned}$$

(See Fig. 4.1.)

Exercises

1. Find the derivatives of the following functions.

(a) $f(x) = \dfrac{x}{x+1}$

(b) $g(y) = -\dfrac{\cos(y)}{y}$

(c) $h(z) = \dfrac{z^3}{\sin(z)}$

(d) $f(x) = \dfrac{\sin(x)}{\ln(x)}$

(e) $g(y) = \dfrac{e^y}{y}$

(f) $h(z) = \dfrac{\sqrt{z}}{z^2+3z-2}$

(g) $f(x) = \dfrac{x\sin(x)}{x^2+4}$

(h) $g(y) = \dfrac{e^y}{y\cos(y)}$

(i) $h(z) = \dfrac{z\sqrt{z}}{z^4-4z}$

2. Find the derivatives of the following functions.

(a) $f(x) = \sec(x)\csc(x)$

(b) $g(y) = \dfrac{1}{\sqrt{y}}$

(c) $h(z) = \dfrac{\sin(z)}{\sqrt{z}}$

(d) $f(x) = \dfrac{2+x^2}{x+\sqrt{x}}$

3. Find the equations for the tangent lines to the graph of the function at the given point.

(a) $f(x) = \dfrac{6x-2}{x^2+1}$ at $x = 2$

(b) $g(y) = \dfrac{1}{\sqrt{y}}$ at $y = 3$

(c) $h(z) = \dfrac{\sin(z)}{\sqrt{z}}$ at $z = 1$

(d) $f(x) = \dfrac{2+x^2}{x+\sqrt{x}}$ at $x = 4$

4. Find the derivatives of the following functions.

(a) $\mathbf{f}(x) = \dfrac{(\cos(x), x^2)}{\sin(x)}$

(b) $\mathbf{g}(t) = \dfrac{(t, t^2, 3\sqrt{t})}{t+t^2}$

(c) $\mathbf{h}(s) = \dfrac{(\cos(s), \sin(s))}{1+\exp(-s)}$

(d) $\mathbf{x}(t) = \dfrac{(t, \exp(t))}{\exp(t)+\exp(-t)}$

4.3 The Chain Rule

This section is devoted to the chain rule. It is the method commonly used to find the derivative of the composition of two functions. The idea behind this can be interpreted using a simple example.

Consider two wheels on an axle. The first wheel has radius 1 m and the other wheel has radius r m. If there is a belt around the wheel of radius 1 m that is traveling at $s(t)$ m/s, a point on the surface of the other wheel is traveling r times as fast.

The idea is that the rates of change multiply. In this section we use the notation Df to denote the derivative of $f(x)$. For example $D(x^2) = 2x$. It makes many of the results easier to read.

Theorem 41 (Chain Rule). *Assume that $f(y)$ and $g(x)$ are functions from $\mathbb{R}$ to $\mathbb{R}$ such that $g(x)$ is differentiable at a and $f(y)$ is differentiable at $g(a)$. Then*

$$D(f \circ g)(a) = (Df)(g(a))\,Dg(a).$$

Proof. Assume that $f(y)$ is differentiable at $g(a)$ and $g(x)$ is differentiable at a. There are two cases to consider, when $Dg(a) \neq 0$ and when $Dg(a) = 0$. The second case is more delicate and is omitted here.

Assume that $Dg(a) \neq 0$. Then, for h near 0, $g(a+h) \neq g(a)$. With this we can write,

$$\begin{aligned} D(f \circ g)(a) &= \lim_{h\to 0} \frac{f \circ g(a+h) - f \circ g(a)}{h} \\ &= \lim_{h\to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \, \frac{g(a+h) - g(a)}{h} \\ &= \lim_{h\to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \, \lim_{h\to 0} \frac{g(a+h) - g(a)}{h} \\ &= (Df)(g(a))\,Dg(a) \end{aligned}$$

since $g(a+h) \to g(a)$ as $h \to 0$.

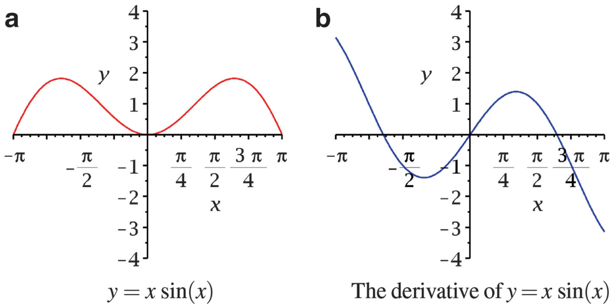


Fig. 4.2

This has some consequences that are fairly obvious, but they show the usefulness of the chain rule. First, if a function is shifted s units to the right, the rate of change graph is also shifted s units to the right.

Example 126. Assume that $f(x)$ is differentiable, then

$$\begin{aligned}\frac{d}{dx}f(x-s) &= Df(x-s)\,\frac{d\,(x-s)}{dx}\\ &= Df(x-s)\,1\\ &= f'(x-s).\end{aligned}$$

This is illustrated here with the graphs of $f(x) = x\sin(x)$ in Fig. 4.2 and $g(x) = (x-\pi/4)\sin(x-\pi/4)$ in Fig. 4.3.

In a similar way, if we multiply the input of $f(x)$ by a constant s, the function $f(sx)$ varies s times as fast. The idea here is that if we pedal a bicycle twice as fast in the same gear, we go twice as fast.

Example 127. Consider the functions $f(\theta) = \cos(\theta)$ and $g(\theta) = \cos(3\theta)$. Then

$$f'(\theta) = -\sin(\theta)$$

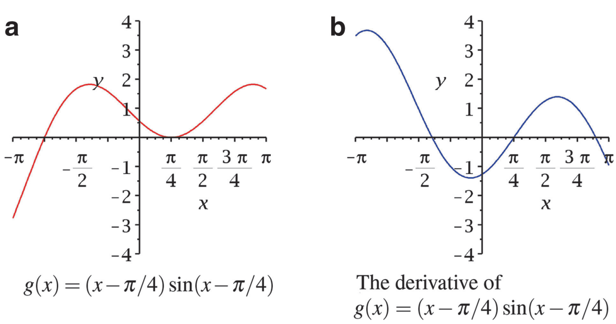
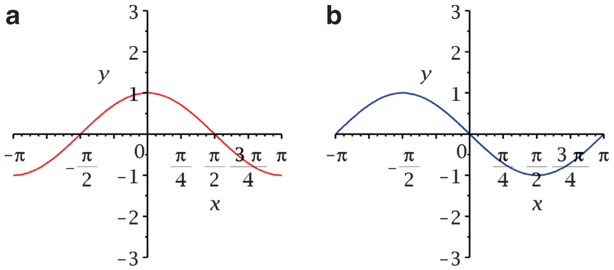
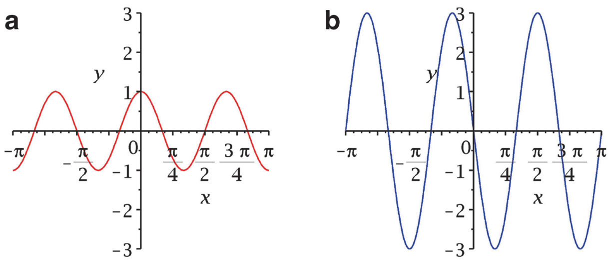


Fig. 4.3



$y = \cos(\theta)$ The derivative of $y = \cos(\theta)$

Fig. 4.4



$y = \cos(3\theta)$ The derivative of $y = \cos(3\theta)$

Fig. 4.5

and

$$
\begin{aligned}
g'(\theta) &= \frac{d}{d\theta}\cos(3\theta) \\
&= -\sin(3\theta)\,\frac{d\,(3\theta)}{d\theta} \\
&= -\sin(3\theta)\,3 \\
&= 3\,f'(3\theta)
\end{aligned}
$$

See Figs. 4.4 on page 98 and 4.5 on page 98

Some additional examples will demonstrate how the chain rule is used. The first is the exponential of a function.

Example 128. Let $f(x) = e^{x^2}$. Then, since $\dfrac{d\,e^u}{du} = e^u$,

$$
\begin{aligned}
\frac{d f(x)}{dx} &= e^{x^2}\,\frac{d\,x^2}{dx} \\
&= e^{x^2}\,2x.
\end{aligned}
$$

Next is the composition of two trigonometric functions.

Example 129. Let $g(z) = \cos(\sin(z))$. Then, since $D(\cos(x)) = -\sin(x)$,

$$\begin{aligned}\frac{d\,g(z)}{dz} &= -\sin(\sin(z))\,\frac{d\,\sin(z)}{dz}\\ &= -\sin(\sin(z))\,\cos(z).\end{aligned}$$

Next we have an example with a composition of three functions.

Example 130. Let $h(\theta) = \sin^2(3\theta)$. Then, using the fact that $D(x^2) = 2x$ and $D(\sin(z)) = \cos(z)$, we can calculate

$$\begin{aligned}\frac{d\,h(\theta)}{d\theta} &= \frac{d\,(\sin(3\theta))^2}{d\theta}\\ &= 2\sin(3\theta)\frac{d\,\sin(3\theta)}{d\theta}\\ &= 2\sin(3\theta)\,\cos(3\theta)\,\frac{d\,(3\theta)}{d\theta}\\ &= 6\sin(3\theta)\,\cos(3\theta).\end{aligned}$$

Using the fact that $a^x = e^{x\ln(a)}$ the chain rule allows us to take the derivative of a^x,

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

Example 131. The derivative of 10^x is given by

$$\frac{d\,10^x}{dx} = \ln(10)\,10^x.$$

The chain rule can easily be extended to the case when the outer function is vector valued.

Theorem 42. *Let $\mathbf{f}(y)$ be a function from $\mathbb{R}$ to $\mathbb{R}^n$ that is differentiable at $g(a)$ and let $g(x)$ be a continuous function from $\mathbb{R}$ to $\mathbb{R}$ that is differentiable at a. Then*

$$D(\mathbf{f}\circ g)(a) = D(\mathbf{f})(g(a))\,D(g)(a).$$

Proof. First note that $\mathbf{f}\circ g(x) = \big(f_1(g(x)), f_2(g(x)), \ldots, f_n(g(x))\big)$. Since

$$D(f_i\circ g)(a) = D(f_i)(g(a))\,D(g)(a)$$

for each i, differentiating coordinate by coordinate we have

$$\begin{aligned}D(\mathbf{f}\circ g)(a) &= \big(D(f_1\circ g)(a), D(f_2\circ g)(a), \ldots, D(f_n\circ g)(a)\big)\\ &= \big(D(f_1)(g(a))\,D(g)(a), D(f_2)(g(a))\,D(g)(a), \ldots,\\ &\qquad D(f_n)(g(a))\,D(g)(a)\big)\\ &= \big(D(f_1)(g(a)), D(f_2)(g(a)), \ldots, D(f_n)(g(a))\big)\,D(g)(a)\\ &= D(\mathbf{f})(g(a))\,D(g)(a).\end{aligned}$$

Example 132. Let $\mathbf{T}(\theta) = (\sin(\theta), \cos(\theta))$ and let $g(t) = 3t+2$. Then $D(\mathbf{T})(\theta) = (\cos(\theta), -\sin(\theta))$ and $D(g)(t) = 3$. Thus

$$\begin{aligned} D\big(\sin(3t+2), \cos(3t+2)\big) &= D(\mathbf{T}\circ g)(t) \\ &= (D\mathbf{T})(3t+2)\, D(g)(t) \\ &= \big(\cos(\theta), -\sin(\theta)\big)\Big|_{\theta=3t+2} 3 \\ &= 3\big(\cos(3t+2), -\sin(3t+2)\big). \end{aligned}$$

In many cases differential notation is used. This can simplify things. For example if we assume that f is a function of u and u is a function of x we can write

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

Example 133. Let $f(u) = \ln(u)$ and let $u(x) = x^3 + \sin(x)$. Here we have $\dfrac{df}{du} = 1/u$ and

$$\frac{du}{dx} = 3x^2 + \cos(x).$$

Using the chain rule we get

$$\begin{aligned} \frac{d}{dx} f\circ u(x) &= \frac{df}{du}\bigg|_{u=x^3+\sin(x)} \frac{du}{dx} \\ &= \frac{1}{u}\bigg|_{u=x^3+\sin(x)} \left(3x^2+\cos(x)\right) \\ &= \frac{1}{x^3+\sin(x)} \left(3x^2+\cos(x)\right). \end{aligned}$$

Exercises

1. Find the derivatives of the following functions.

(a) $f(x) = \sin(2x)$
(b) $g(y) = \tan(3y)$
(c) $h(z) = e^{-7z}$
(d) $f(x) = \ln(4x)$
(e) $g(y) = 5\cos(4-3y)$
(f) $h(z) = \sec(5z-10)$
(g) $f(x) = \cot(2x-3)$
(h) $g(y) = 5\cos(2\pi y)$
(i) $h(z) = \csc\big(\pi(z-10)\big)$
(j) $f(x) = \ln(e^2x-2)$
(k) $g(s) = 2^s$
(l) $h(w) = 4^w$

2. Find the derivatives of the following functions.

(a) $f(x) = \left(x^2-4x+10\right)^{51}$
(b) $g(y) = e^{y^3-4y}$
(c) $h(z) = e^{\cos(z)}$
(d) $f(w) = (\sin(w)+\cos(w))^{-10}$
(e) $h(z) = 3^{z^2+4z}$
(f) $r(t) = 3^{3^t}$

3. Find the derivatives of the following functions.

(a) $f(x) = \sec\left(e^{x^2+2x}\right)$

(b) $g(y) = e^{\sec(\ln(y))}$

(c) $h(z) = \sqrt{z^2 + \cos^2(z)}$

(d) $f(x) = \cos\left(x^2 - 2x - 2x^2 + 4\right)$

(e) $g(y) = \ln\left(4 + e^{y^3}\right)$

(f) $h(z) = \tan\left(z^2 + \sin^2(6z+2)\right)$

(g) $r(s) = \cos\left(2^{s^2+1} - 2\pi s\right)$

(h) $f(x) = 2^{\cos(3\pi x)}$

4. Find the derivatives of the following functions using the chain rule.

(a) $\mathbf{r}(s) = \left(\sec\left(\pi x^2\right), \csc\left(\pi x^2\right)\right)$

(b) $\mathbf{f}(t) = \left(\exp\left(\sin(3x)\right), \ln\left(\sin(3x)\right)\right)$

(c) $\mathbf{r}(s) = \left(2^{s^2-s}, 3^{s^2-s}, 5^{s^2-s}\right)$

(d) $\mathbf{f}(t) = \left(\sin\left(t^3\right), \cos\left(t^3\right), \tan\left(t^3\right)\right)$

5. Find the tangent lines to the graphs of the following functions at the given a's.

(a) $f(x) = \sin(4x)$ at $a = \frac{\pi}{12}$.

(b) $g(y) = \exp(y^2)$ at $a = -2$.

(c) $h(z) = \sqrt[4]{\sin(\pi x) + 2x}$ at $a = \frac{1}{4}$.

(d) $\mathbf{r}(s) = \left(\sin\left(\frac{\pi x^2}{4}\right), \cos\left(\frac{\pi x^2}{4}\right)\right)$ at $s = 1$.

(e) $\mathbf{f}(t) = \left(\exp\left(\ln(3x) + 2\right), \ln\left(\ln(3x) + 2\right)\right)$ at $a = e$.

6. The volume of a sphere of radius r is given by $V = \frac{4}{3}\pi r^3$. If the radius of a sphere is 20 cm when the radius of the sphere is decreasing at $\frac{1}{3}$ cm/s, how fast is the volume of the sphere changing?

7. The volume of a sphere of radius r is given by $V = \frac{4}{3}\pi r^3$. If the radius of a sphere is 20 cm when the volume of the sphere is decreasing at $2\,\mathrm{cm}^3/\mathrm{s}$, how fast is the radius of the sphere changing? (Hint: Use the chain rule when differentiating the volume equation. Then solve for $r'(t)$.)

4.4 Implicit Differentiation and Inverse Functions

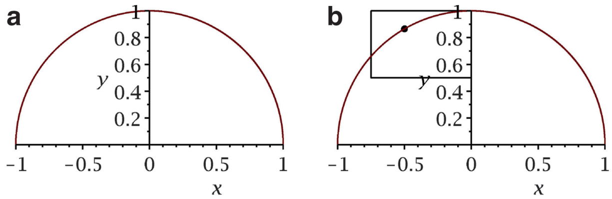
In the previous sections a dependent variable y was always written as a function of an independent variable x. In many cases this is not possible. A simple example where y is not always defined as a function of x is the unit circle,

$$x^2 + y^2 = 1.$$

Here, for each $x \in (-1,1)$ there are two y values for each x value, y is not a function of x.

Instead of requiring that an equation in two variables defines one variable as a function of the other, the idea is to consider equations that define one variable as a function of the other around a point (x_0, y_0). As an example consider the unit circle centered at $(0,0)$. If we take $x \in (-1,1)$ and $y > 0$, there is a unique y value for each x value. See Fig. 4.6a on page 102.

If we look at any point $(x, \sqrt{1-x^2})$ with $x \in (-1,1)$, there is a box around that point such that for each x there is only one y with (x,y) on the curve within the box. For example, consider the point $\left(-1/2, \sqrt{3}/2\right)$ and the box with $x \in [-3/4, 0]$ and $y \in [1/2, 1]$. See Fig. 4.6b on



The unit upper half circle

A point where y is a function of x on the unit upper half circle

Fig. 4.6

page 102. Within this box, for every x there is a unique y on the unit circle. In this case we say that $x^2 + y^2 = 1$ *implicitly* defines y as a function of x around $\left(-1/2, \sqrt{3}/2\right)$.

In this special case we can solve for y as a function of x when we take $y > 0$. Solving for y in terms of x gives

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

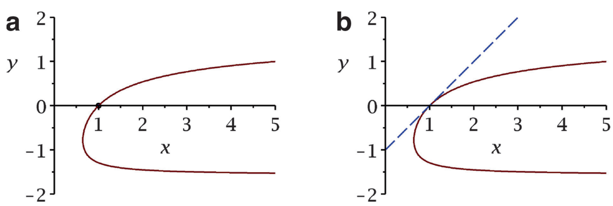
This functions is differentiable if $x \in (-1, 1)$. If we have a point (x_0, y_0) on the circle where $y_0 > 0$, we can find the derivative of y with respect to x at (x_0, y_0). See Fig. 4.6b on page 102. For example, if we take the point $(\sqrt{2}/2, \sqrt{2}/2)$, the derivative of y at $x = 1/2$ is

$$\begin{aligned} \frac{d}{dx}\sqrt{1-x^2}\bigg|_{x=\frac{\sqrt{2}}{2}} &= \frac{-x}{\sqrt{1-x^2}}\bigg|_{x=\frac{\sqrt{2}}{2}} \\ &= -1. \end{aligned}$$

Note that there are no boxes around $(1, 0)$ and $(-1, 0)$ where we have only one y for each x. If we look at a box around $(1, 0)$, say $x \in (1-\delta, 1+\delta)$ and $y \in (-\omega, \omega)$, there is an $x_0 \in (1-\delta, 1)$ such that $\sqrt{1-x_0^2} < \omega$. For this x_0 both of the points $(x_0, \sqrt{1-x_0^2})$ and $(x_0, \sqrt{1-x_0^2})$ are in the box. This means that the portion of the unit circle in any box around $(1, 0)$ does not define y as a function of x. (Add figure)

The above approach for finding the derivative of y with respect to x relies on *explicitly* solving for y in terms of x. This is not always possible. Therefore another approach is needed to handle cases when we cannot explicitly solve for y.

Consider the equation $x\cos(y) = e^y$ and the point $(x, y) = (1, 0)$ that satisfies the equation. (See Fig. 4.7a on page 102.) Ignoring the lower part of the graph, it appears that we should be able to assume that the equation defines y as a function of x near $(1, 0)$. We cannot, however, explicitly solve this equation for y in terms of x.



The curve defined by $x\cos(y) = e^y$

A tangent line to the curve defined by $x\cos(y) = e^y$

Fig. 4.7

If we assume that $y(x)$ is a differentiable function of x, we can differentiate both sides of the equation

$$x\cos(y(x)) = e^{y(x)}$$

and maintain the equality. Using the chain rule, this gives

$$\cos(y(x)) - x\sin(y(x))\,y'(x) = e^{y(x)}\,y'(x).$$

Solving for $y'(x)$ yields

$$y'(x) = \frac{\cos(y)}{e^y + x\sin(y)}.$$

Substituting in $x = 1$ and $y = 0$ into the equation gives

$$y'(0) = 1.$$

Plotting the tangent line given by this derivative, $y = x - 1$ and the graph of the equation together gives credence to the conclusion that this derivative is correct. (See Fig. 4.7b on page 102.)

This process is call *implicit differentiation*. It is usually applied when looking for the derivative of a function $y(x)$ when we only have an equation defining the relationship between y and x. This means that we need to know a pair (x_0, y_0) satisfying the relationship.

If we look closely at this process, we find that the equation to solve for $y'(x)$ is always linear in $y'(x)$, our equation is of the form $h(x,y)\cdot y'(x) + g(x,y) = k(x,y)\cdot y'(x) + \ell(x,y)$ for some functions $h(x,y)$, $g(x,y)$, $k(x,y)$ and $\ell(x,y)$. This means that, at least formally, we can always go through the process of solving for $y'(x)$. Assuming that the equation defines y as a function of x near (x_0, y_0), the only problem with the process is that we must avoid division by zero.

Division by zero in this process can mean there is a vertical tangent line. Returning to the unit circle

$$x^2 + y^2 = 1,$$

we can show that division by zero occurs at $(-1,0)$ and $(1,0)$. As was noted earlier in this section, at both of these points y is not defined as a function of x. In this case there are vertical tangents to the unit circle at $(1,0)$ and $(-1,0)$.

Example 134. Consider the relationship between x and y given by $x^2 = y^2$ and the point $(0,0)$ that satisfies this relationship. (See Fig. 4.8 on page 103.)

All of the functions $f_1(x) = x$, $f_2(x) = -x$, $f_3(x) = |x|$, and $f_4(x) = -|x|$ satisfy the relationship $x^2 = y^2$. This means that the relationship does not uniquely define y as a function of x around $(0,0)$. We cannot use implicit differentiation in this case.

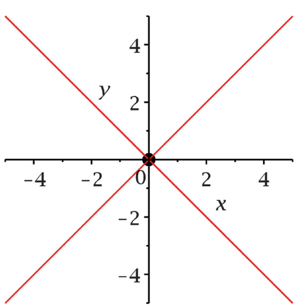


Fig. 4.8 The set defined by $x^2 = y^2$ does not define y as a function of x around $(0,0)$

As the next example demonstrates, even though the computations may be messy, implicit differentiation is usually a straight forward process.

Example 135. Assume that $3y^5+2x^2y^3+x^3y^2-6y=0$ defines y as a function of x around $(1,1)$. See Fig. 4.9a on page 104.

Replacing y with $y(x)$ gives

$$3y(x)^5+2x^2y(x)^3+x^3y(x)^2-6y(x)=0.$$

Differentiating both sides yields

$$15y(x)^4y'(x)+4xy(x)^3+6x^2y(x)^2y'(x)+3x^2y(x)+2x^3y(x)y'(x)-6y'(x)=0.$$

Solving for $y'(x)$ gives

$$y'(x)=-\frac{4xy(x)^3+3x^2y(x)}{15y(x)^4+6x^2y(x)^2+2x^3y(x)-6}.$$

The derivative of y at $(1,1)$ is

$$y'(1)=-\frac{7}{17}.$$

This is illustrated in Fig. 4.9b on page 104.

We can also use implicit differentiation to find the derivatives of rational powers of x. Finding derivatives of arbitrary powers of x involves using the natural exponential and is left for later in this book.

Consider the function $y(x)=x^{\frac{n}{m}}$ where n and m are integers. Then $y^m=\left(x^{\frac{n}{m}}\right)^m=x^n$. Taking derivatives of both sides gives

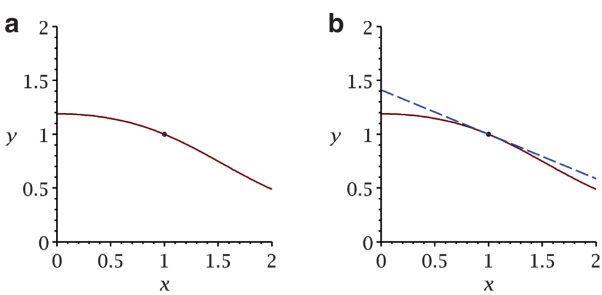
$$my^{m-1}y'(x)=nx^{n-1}.$$

or, since $y(x)=x^{n/m}$,

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

Simplifying the exponent of x on the right side gives the desired result,

$$\frac{d}{dx}x^{\frac{n}{m}}=\frac{n}{m}x^{\frac{n}{m}-1}.$$



The set defined by $3y^5+2x^2y^3+x^3y^2-6y=0$

A tangent line to the set

Fig. 4.9

Example 136. The derivative of $f(x) = x^{13/16}$ is

$$f'(x) = \frac{13}{16}x^{\frac{13}{16}-1} = \frac{13}{16}x^{-\frac{3}{16}}.$$

4.4.1 Derivatives of Inverse Functions

A very important use of implicit differentiation is finding the derivatives of inverse functions. The idea is very simple. We differentiate the equation

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

This yields

$$\left(Df^{-1}\right)(f(x))Df(x) = 1,$$

or

$$\left(Df^{-1}\right)(f(x)) = \frac{1}{Df(x)}. \tag{4.1}$$

The problem of finding the derivative of an inverse function becomes rewriting the above equation in the form

$$D\left(f^{-1}\right)(y) = g(y).$$

Example 137. Let $f(x) = e^x$ and assume that, as indicated earlier, $\frac{d}{dx}e^x = e^x$. Putting this into Eq. (4.1) gives

$$D\left(f^{-1}\right)(e^x) = \frac{1}{e^x}.$$

Replacing e^x by y gives

$$D\left(f^{-1}\right)(y) = \frac{1}{y}.$$

Since the inverse function of e^x is $\ln(x)$, This equation says the derivative of $\ln(x)$ is $1/x$. This was assumed earlier.

The same technique can be used to find the derivative of $g(x) = \log_a(x)$ when $a > 0$. This is the inverse function for $f(x) = a^x$. Recall that $y = \log_a(x)$ if and only if $x = a^y$. We first find the derivative of $f(x) = a^x$ when $a > 0$ using the chain rule.

Since $a = e^{\ln(a)}$, $a^x = \left(e^{\ln(a)}\right)^x = e^{x\ln(a)}$. Using the chain rule we get

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

Example 138. If $h(z) = 3^{2z}$, then

$$\begin{aligned}\frac{d}{dz}3^{2z} &= \ln(3)\,3^{2z}\frac{d}{dz}(2z)\\ &= 2\ln(3)\,3^{2z}.\end{aligned}$$

We now consider the derivative of $f(x) = \log_a(x)$.

Example 139. If, for $a > 0$, $f(y) = a^y$ then the derivative of $f(y)$ is $f'(y) = \ln(a)\,a^y$. Using Eq. (4.1) we find that

$$\begin{aligned}\frac{d}{dx}\log_a(x) &= \left.\frac{1}{\frac{d}{dy}\,a^y}\right|_{y=\log_a(x)}\\ &= \left.\frac{1}{\ln(a)\,a^y}\right|_{y=\log_a(x)}\\ &= \frac{1}{\ln(a)\,x}.\end{aligned}$$

Notice that this is consistent with the derivative of the formula $\log_a(x) = \ln(x)/\ln(a)$ that is often given in classes before calculus:

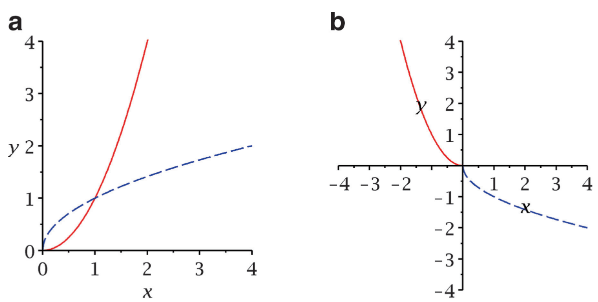
$$\frac{d}{dx}\log_a(x) = \frac{d}{dx}\frac{\ln(x)}{\ln(a)} = \frac{1}{x\ln(a)}.$$

If we consider $h(w) = \log_4(\tan(w))$ we get

$$\begin{aligned}\frac{d}{dw}h(w) &= \frac{1}{\ln(4)\,\tan(w)}\,\frac{d}{dw}\tan(w)\\ &= \frac{\sec^2 w}{\ln(4)\,\tan(w)}\\ &= \frac{1}{\ln(4)\,\sin(w)\cos(w)}.\end{aligned}$$

There are cases where a function has inverse functions only on restricted domains.

Example 140. Consider the function $f(x) = x^2$. Since $f(x) = f(-x)$ for any x, this function is not one to one on its domain. On the other hand, if $f(x)$ is restricted to $[0,\infty)$ or $(-\infty,0]$, $f(x)$ is one to one. On $[0,\infty)$ the inverse is $f^{-1}(x) = \sqrt{x}$ and on $(-\infty,0]$ the inverse is $f^{-1}(x) = -\sqrt{x}$. This is illustrated in Fig. 4.10a, b on page 106.



a $y = x^2$ and its inverse on $[0,\infty)$

b $y = x^2$ and its inverse on $(-\infty,0]$

Fig. 4.10

We can implicitly find the derivative of the inverse of $f(x) = x^2$ at $(-1,1)$ Again, using $f^{-1}(x^2) = x$, we have

$$\left(Df^{-1}\right)(x^2)\,2x = 1,$$

or

$$\left(Df^{-1}\right)(x^2) = \frac{1}{2x}.$$

At the point $(-1,1)$ this gives

$$\left(Df^{-1}\right)(1) = -\frac{1}{2}.$$

(Check that this is the correct value for the derivative.)

Exercises

1. In the following the equation defines y as a function of x around the given point $\mathbf{a}$. Find $\dfrac{dy}{dx}$ at $\mathbf{a}$ using implicit differentiation.

(a) $x^2y - y^3 + 3x = 8$, $\mathbf{a} = (0,-2)$
(b) $e^{xy} - y^2 + x = 0$, $\mathbf{a} = (0,1)$
(c) $x = \tan(y)$, $\mathbf{a} = (1, 5\pi/4)$
(d) $x = y\cos(y)$, $\mathbf{a} = (0, \pi/2)$
(e) $x^5y^3 - 2x^2y^2 + x = 0$, $\mathbf{a} = (0,-1)$
(f) $x^2y^{1/3} - xy^2 + 20x^{1/3} = -42$, $\mathbf{a} = (1,8)$
(g) $\ln(xy+4) - y^2 + x^2 = -8$, $\mathbf{a} = (-1,3)$
(h) $y^{2/3} - x^2 + xy = 1$, $\mathbf{a} = (1,1)$

2. Find the derivatives of the inverses of the following functions at the given point.

(a) $f(x) = \sin(2x) + x^3$, $x = -2$
(b) $g(y) = e^{y^3+4y}$, $y = 5$,
(c) $h(z) = \tan(z/2)$, $z = \frac{\pi}{2}$
(d) $f(x) = \sin(\sqrt{x})$, $x = \frac{\pi^2}{4}$

3. Explain why implicit differentiation cannot be used to find $\frac{dy}{dx}$ with the following equations at the given point.

(a) $x^2 + y^2 = 1$, $\mathbf{a} = (1,0)$
(b) $e^{xy} + y^2 - x^2 = 0$, $\mathbf{a} = (0,0)$
(c) $x^4 - 4x^3 + 2x^2 + 2x^2y^2 - 4y^2x + 2y^2 + y^4 + 4x = 3$, $\mathbf{a} = (1,0)$

4. Find the derivatives of the following functions.

(a) $f(x) = x^{3/2}$
(b) $g(y) = y^{-4/5}$
(c) $h(z) = (6x+10)^{1/3}$
(d) $f(x) = \sqrt[7]{\sin(x)}$
(e) $g(y) = \sqrt[5]{(3x-2)^{17}}$
(f) $h(z) = 5^{x/3}$
(g) $f(x) = 7^{\sin(x)}$
(h) $g(y) = \left(\frac{1}{2}\right)^{3x-2}$
(i) $h(z) = \tan\left(x^{2/3} + x^{3/2}\right)$
(j) $g(y) = \log_3(y)$
(k) $h(z) = \log_{10}(\sin(3z))$
(l) $f(x) = \log_5\left(x^3 - 4x\right)$
(m) $g(y) = 10^{\left(\log_{10}(y) + y^2\right)}$
(n) $h(t) = \dfrac{x^3+1}{\sqrt[3]{\sin(2x)}}$
(o) $w(z) = \sec(z)^{\exp(\sin(z))}$
(p) $h(t) = \log_4\left(3^t\right)$

5. The following functions have inverses on intervals around the given $x = a$ values. Find the derivatives of $f^{-1}(y)$ at $y = f(a)$.

(a) $f(x) = x^2 + \sin(\pi x)$ at $a = 3$.
(b) $f(x) = \tan(x) + \sin(x)$ at $a = \frac{\pi}{3}$.
(c) $f(x) = e^x - e^{-x} + x$ at $a = 0$.
(d) $f(x) = \ln(x) + x^3$ at $a = 2$.
(e) $f(x) = \sqrt{x^2+3} + x^7$ at $a = 1$.
(f) $f(x) = e^x + \cos(x^2) - x$ at $a = 2$.

6. Assume that $y = f(x)$ has an inverse on an interval around $x = a$ and that $f'(a) = 0$. Why does $f^{-1}(y)$ not have a derivative at $y = f(a)$?

4.5 Inverse Trigonometric Functions

The trigonometric functions are used extensively in many areas of application. Because they are so important, this section is devoted to the inverses of the trigonometric functions and the derivatives of these functions. The most important fact about inverse functions here is that the original function must be one to one on the set in the domain used to define an inverse.

This makes it clear that we must restrict the domain of all trigonometric functions to define inverses. The "best" choices are the largest intervals around $x = 0$ where the trigonometric function is defined and is one-to-one. The choice for the sine function is the interval $[-\pi/2, \pi/2]$, see Fig. 4.11a on page 108.

When this portion of the graph of $\sin(x)$ is reflected across $y = x$, exchanging the x and y values for each point in the graph, we get the graph in Fig. 4.11b on page 108 for the inverse of $\sin(x)$, $\arcsin(x)$. This function is also denoted by $\sin^{-1}(x)$. For most computer programs and programming languages $\arcsin(x)$ is used for the inverse of $\sin(x)$.

Example 141. Since $\sin(\pi/6) = 1/2$, the arcsin of $1/2$ is $\pi/6$.

Using the technique demonstrated in Sect. 4.4.1, we have

$$\left(D\sin^{-1}\right)(\sin(x))\cos(x) = 1.$$

Since $\cos(x) = \sqrt{1-\sin^2(x)}$ in the range of x values, $[-1,1]$, formally

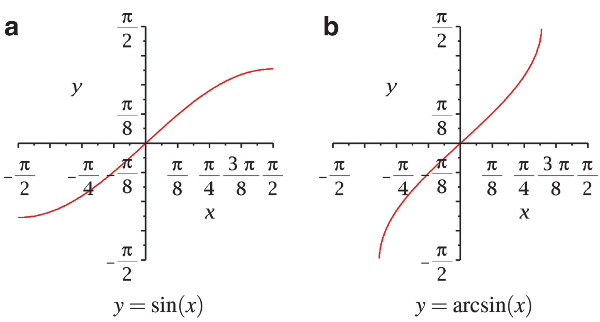


Fig. 4.11

$$\left(D\sin^{-1}\right)(y) = \frac{1}{\sqrt{1-y^2}}.$$

The domain of the derivative must exclude -1 and 1 since we would be dividing by zero. Therefore the domain of $\left(D\sin^{-1}\right)$ is $(-1,1)$.

Example 142. The derivative of $\arcsin(x)$ at $x = 1/2$ is

$$D\arcsin\left(\frac{1}{2}\right) = \frac{1}{\sqrt{1-\left(\frac{1}{2}\right)^2}} = \frac{2}{\sqrt{3}}.$$

Example 143. The derivative of $f(x) = \arcsin(\sqrt{x})$ is

$$D\arcsin(\sqrt{x}) = \frac{1}{\sqrt{1-(\sqrt{x})^2}}D\sqrt{x} = \frac{1}{\sqrt{1-x}}\frac{1}{2\sqrt{x}} = \frac{1}{2\sqrt{x-x^2}}.$$

Note that the implied domain of $\sqrt{x}$ is $x \geq 0$ and the domain of this derivative is $x \in (0,1)$.

The other inverse trigonometric functions are defined in a similar manner and their derivatives are found using the same methods. Consider $\cos(x)$. In this case the common choice for the domain of $\cos(x)$ is $[0,\pi]$, see Fig. 4.12a on page 109. The domain of the inverse of $\cos(x)$, $\arccos(x)$ or $\cos^{-1}(x)$, is $[-1,1]$ and the graph is in Fig. 4.12b on page 109. The derivative is

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

on $(-1,1)$.

Example 144. Since the cos of $\pi/3$ is $1/2$, the *arccos* of $1/2$ is $\pi/3$.

Example 145. The derivative of $\arccos(x)$ at $x = 1/4$ is

$$D\arccos\left(\frac{1}{4}\right) = \frac{-1}{\sqrt{1-\left(\frac{1}{4}\right)^2}} = \frac{-4}{\sqrt{15}}.$$

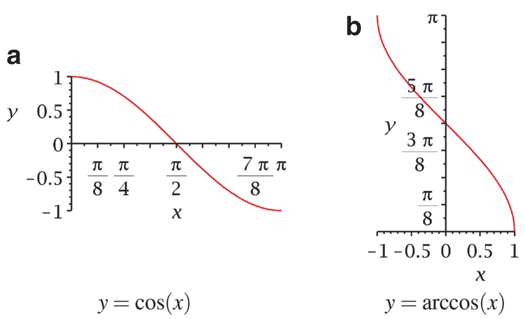


Fig. 4.12

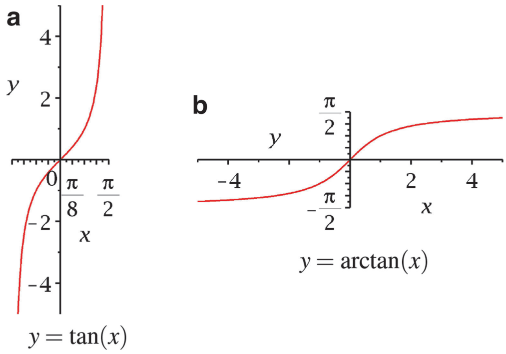


Fig. 4.13

When making the choices of domain for $\tan(x)$ and $\csc(x)$ to define the inverse functions we use the same domain as is chosen for defining $\sin^{-1}(x)$, excluding the points where $\cos(x) = 0$. (Other choices can be made, but this is a common choice.) The graphs of $\tan(x)$ and $\arctan(x)$ are in Fig. 4.13 on page 110 and the graphs of $\csc(x)$ and $\csc^{-1}(x)$ are the in Fig. 4.14 on page 111.

Their derivatives are

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2} \quad \text{on} \quad (-\infty,\infty) \qquad \text{and}$$
$$\frac{d}{dx}\csc^{-1}(x) = \frac{-1}{x^2\sqrt{1-\frac{1}{x^2}}} \quad \text{on} \quad (-\infty,-1)\cup(1,\infty).$$

Example 146. Since $\tan(\pi/4) = 1$ and $\sec(\pi/4) = \sqrt{2}$, $\arctan(1) = \pi/4$ and $\csc^{-1}(1) = \sqrt{2}$.

Example 147. At $x = 2$ we have

$$\begin{aligned} D\arctan(1) &= \frac{1}{1+2^2} \\ &= \frac{1}{5} \end{aligned}$$

and

$$\begin{aligned} D\csc^{-1}(2) &= \frac{-1}{2^2\cdot\sqrt{1-\frac{1}{2^2}}} \\ &= \frac{-1}{2\sqrt{3}}. \end{aligned}$$

Example 148. The derivative of $f(x) = x\arctan(x^2)$ can be calculated as

$$\begin{aligned} \frac{d\,f(x)}{x} &= \left(\frac{d}{dx}x\right)\arctan\left(x^2\right) + x\frac{d}{dx} \\ &= \arctan\left(x^2\right) + x\frac{1}{1+\left(x^2\right)^2}\frac{d}{dx}x^2 \\ &= \arctan\left(x^2\right) + \frac{2x^2}{1+x^4}. \end{aligned}$$

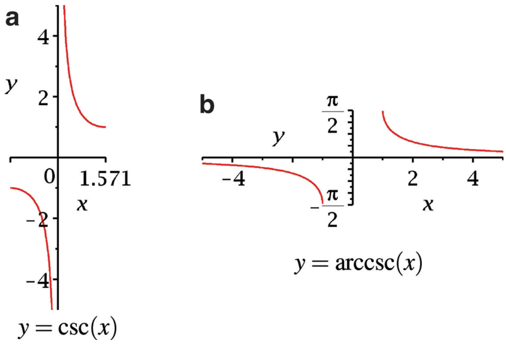


Fig. 4.14

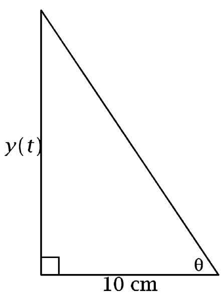


Fig. 4.15 A triangle for Example 149

Example 149. Consider the situation of a right triangle with one of the angles, θ that is not the right angle, has an adjacent side of length 10 cm. The opposite side has length $y(t) = 2t + 3$ cm. What is the derivative of θ with respect to t when $t = 4$? See Fig. 4.15 on page 111.

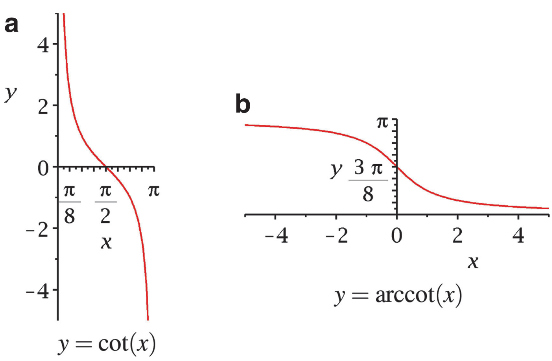


Fig. 4.16

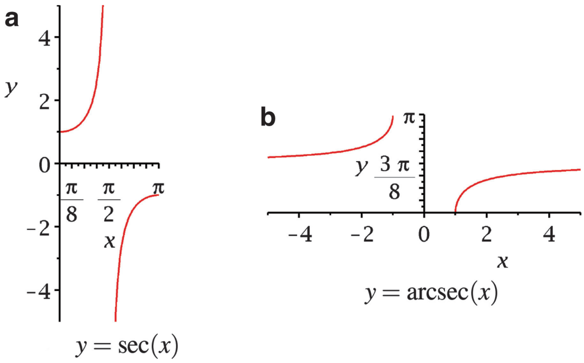


Fig. 4.17

Since $\theta = \arctan\left((2t+3)/10\right)$, we have

$$\begin{aligned}\frac{d\theta}{dt} &= \frac{d}{dt}\arctan\left(\frac{2t+3\,\mathrm{cm}}{10\,\mathrm{cm}}\right)\\ &= \frac{1}{1+\left(\frac{2t+3}{10}\right)^2}\frac{d}{dt}\left(\frac{2t+3}{10}\right)\\ &= \frac{1}{1+\left(\frac{2t+3}{10}\right)^2}\frac{2}{10}\\ &= \frac{2}{100+(2t+3)^2}.\end{aligned}$$

At $t = 4$ we have

$$\frac{d\theta}{dt} = \frac{2}{100+11^2} = \frac{2}{221}.$$

When making the domain choices for $\cot(x)$ and $\sec(x)$ to define the inverse functions we use the same domain as is chosen for defining $\cos^{-1}(x)$ excluding the points of the interval where $\sin(x) = 0$. (Other choices can be made, but this is a common choice.) The graphs of $\cot(x)$ and $\cot^{-1}(x)$ are in Fig. 4.16 on page 111 and the graphs of $\sec(x)$ and $\sec^{-1}(x)$ are in Fig. 4.17 on page 112. Their derivatives are

$$\begin{aligned}\frac{d}{dx}\cot^{-1}(x) &= \frac{-1}{1+x^2} \quad \text{on} \quad (-\infty,\infty) \qquad \text{and}\\ \frac{d}{dx}\sec^{-1}(x) &= \frac{1}{x^2\sqrt{1-\frac{1}{x^2}}} \quad \text{on} \quad (-\infty,-1)\cup(1,\infty).\end{aligned}$$

Example 150. Since $\cot\left(\frac{3\pi}{4}\right) = -1$ and $\sec\left(\frac{3\pi}{4}\right) = -\sqrt{2}$, $\operatorname{arccot}(-1) = \frac{3\pi}{4}$ and $\sec^{-1}(\sqrt{2}) = \frac{3\pi}{4}$.

$f(x)$	$f'(x)$
$\sin^{-1}(x)$	$\dfrac{1}{\sqrt{1-x^2}}$
$\cos^{-1}(x)$	$\dfrac{-1}{\sqrt{1-x^2}}$
$\tan^{-1}(x)$	$\dfrac{1}{1+x^2}$
$\cot^{-1}(x)$	$\dfrac{-1}{1+x^2}$
$\sec^{-1}(x)$	$\dfrac{1}{x^2\sqrt{1-\frac{1}{x^2}}}$
$\csc^{-1}(x)$	$\dfrac{-1}{x^2\sqrt{1-\frac{1}{x^2}}}$

Table 4.1 The derivatives of the inverse trigonometric functions

Example 151. The derivative of $\sec^{-1}\big(\exp(x)\big)$ is calculated as

$$\begin{aligned} D\sec^{-1}\big(\exp(x)\big) &= \frac{1}{e^{2x}\sqrt{1-\frac{1}{e^{2x}}}}\,e^x \\ &= \frac{1}{\sqrt{e^{2x}-1}}. \end{aligned}$$

Table 4.1 on page 113 summarizes the derivatives of the inverse trigonometric functions.

Exercises

1. Evaluate the following.

(a) $\sin^{-1}(-\sqrt{2}/2)$
(b) $\cos^{-1}(1/2)$
(c) $\cos^{-1}(-1/2)$
(d) $\sin^{-1}(\sqrt{3}/2)$
(e) $\sec^{-1}(-2)$
(f) $\arctan(1)$
(g) $\cot^{-1}(-\sqrt{3})$
(h) $\csc^{-1}(2)$

2. Find the derivatives of the following functions.

(a) $f(x) = \arccos(3x-2)$
(b) $g(y) = \tan^{-1}(y^2+\sqrt{y})$
(c) $f(x) = \sec^{-1}(6x)$

(d) $g(y) = \csc^{-1}(7y-2)$

(e) $f(x) = \arcsin\left(\dfrac{x}{1+x^2}\right)$

(f) $g(y) = \cot^{-1}(y^3)$

(g) $f(x) = \tan^{-1}(e^x)$

(h) $g(y) = \sin^{-1}(\cos(y))$

(i) $f(x) = \cos^{-1}\left(\tan\left(x^2\right)\right)$

(j) $g(y) = \sec^{-1}\left(e^y + e^{-y}\right)$

(k) $f(x) = \tan^{-1}(\cos(x) + \sin(x))$

(l) $g(y) = \cot^{-1}\left(\ln\left(y^2\right)\right)$

(m) $\mathbf{r}(t) = \left(\cot^{-1}\left(t^2\right), \exp(\arcsin(t))\right)$

(n) $\mathbf{s}(t) = \left(\sin^{-1}(4t-7), \arccos(4t)^2\right)$

3. Find the domains and ranges of the following functions.

(a) $f(x) = \arccos(3x-2)$

(b) $g(y) = \tan^{-1}(y^2)$

(c) $f(x) = \sec^{-1}(6x+5)$

(d) $g(y) = \arcsin(y^3)$

(e) $f(x) = \operatorname{arccsc}(\sin(2y))$

(f) $g(y) = \sin(\operatorname{arccot}(2y))$

4. Let θ be an angle that is not a right angle in a right triangle. If the ratio of the length of the adjacent side to the opposite side is increasing at a rate of $\frac{1}{3}\,\mathrm{s}^{-1}$ when the lengths of the two sides are respectively 4 and 5 ft, what is the rate of change of θ with respect to time at that time.

4.6 Higher Order Derivatives

There are many situations where more information is needed about how a function changes besides the values of its derivative. Additional information can be obtained through derivatives of derivatives, called higher order derivatives.

A simple example is the behavior of a quadratic polynomial. The derivative of a quadratic $q(x) = c_2x^2 + c_1x + c_0$ is $r(x) = q'(x) = 2c_2x + c_1$. At $x = 0$ we have $q'(0) = c_1$ and the linear approximation at $x = 0$ is $p_1(x) = c_1x + c_0$. The derivative of $r(x)$ is $r'(x) = 2c_2$. Using the fact that $q(0) = c_0$ this means that $p_2(x) = q(0) + q'(0)x + r'(0)x^2/2 = c_0 + c_1x + c_2x^2$, the original function $q(x)$. All of the information needed to recover this polynomial is contained in $q'(x)$ and $r'(x)$. This idea can be generalized to higher degree polynomials and used for approximations of *nice* functions that are not polynomials.

In order to consider these approximations, the concept of *higher order derivatives* is required. The second derivative of a function $f(x)$, if it exists, is the derivative of the derivative of $f(x)$. This is denoted by

$$f''(x), \quad \frac{d^2 f(x)}{dx^2}, \quad D^2 f(x), \quad \text{or} \quad f^{(2)}(x).$$

Example 152. Consider the function $f(x) = \cos(x)$. Its first derivative is $f'(x) = -\sin(x)$ and the second derivative of f is $f''(x) = -\cos(x)$.

Example 153. Consider the function $f(x) = 1/(1+x^2)$. Its first derivative is

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

and the second derivative of f is calculated as

$$\begin{aligned}\frac{d^2 f(x)}{dx^2} &= -2\,\frac{\left(1+x^2\right)^2 - x\cdot 2\left(1+x^2\right)\cdot 2x}{(1+x^2)^4} \\ &= \frac{6x^2-2}{(1+x^2)^3}\end{aligned}$$

Example 154. If $\mathbf{r}(t)$ is the position of an object in two or three space, the first derivative is the velocity, $\mathbf{v}(t) = \frac{d}{dt}\mathbf{r}(t)$ and the derivative of the velocity, the second derivative of position, is the acceleration, $\mathbf{a}(t) = \frac{d}{dt}\mathbf{v}(t) = \frac{d^2}{dt^2}\mathbf{r}(t)$.

If the position of an object is given by $\mathbf{r}(t) = \big(15\cos(t)+\cos(10t), 15\sin(t)+\cos(10t)\big)$, then the velocity is $\mathbf{v}(t) = \big(-15\sin(t) - 10\sin(10t), 15\cos(t) - 10\sin(10t)\big)$ and the acceleration is $\mathbf{a}(t) = \big(-15\cos(t) - 100\cos(t), -15\sin(t) - 100\cos(t)\big)$.

Example 155. Consider circular motion around the origin at constant speed,

$$\mathbf{r}(t) = A\left(\cos(\omega t), \sin(\omega t)\right).$$

In this case we have

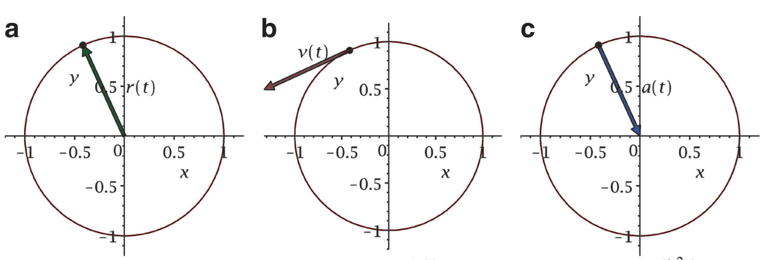
$$\mathbf{v}(t) = \frac{d}{dt}\mathbf{r}(t) = \omega A\big(-\sin(\omega t), \cos(\omega t)\big)$$

and

$$\mathbf{a}(t) = \frac{d}{dt}\mathbf{v}(t) = \omega^2 A\big(-\cos(\omega t), -\sin(\omega t)\big).$$

For this case we have $\mathbf{r}(t)\cdot\mathbf{v}(t) = 0$, $\mathbf{v}(t)\cdot\mathbf{a}(t) = 0$ and $\mathbf{a}(t) = -\omega^2\mathbf{r}(t)$. See Fig. 4.18 on page 115 where $A = 1$ and $\omega = 1$. We will not demonstrate it here, but these relationships, where $\mathbf{r}(t)$ is the position vector relative to the center of the circle, characterize circular motion.

The first and second derivatives can be used to approximate a function around a point y using a quadratic polynomial as was done for $q(x) = ax^2 + bx + c$. The idea is to choose a quadratic polynomial $q(x) = c_2x^2 + c_1x + c_0$ which matches the function value and first and second derivatives of $f(x)$ at z. This means that



$\mathbf{r}(t) = (\cos(\omega t), \sin(\omega t))$ $\mathbf{v}(t) = \frac{d\mathbf{r}(t)}{dt}$ $\mathbf{a}(t) = \frac{d^2\mathbf{r}(t)}{dt^2}$

Fig. 4.18

$$\begin{aligned} f(z) &= c_2x^2 + c_1x + c_0 \\ f'(y) &= 2c_2x + c_1 \quad \text{and} \\ f''(y) &= 2c_2. \end{aligned}$$

This system of equations can be solved to give the quadratic approximation

$$f(x) \approx f(z) + f'(z)(x-z) + \frac{f''(z)}{2}(x-z)^2.$$

The first two terms of this polynomial are simply the linear approximation to $f(x)$ centered at z.

Example 156. Consider $f(x) = \cos(x)$ around $z = 0$. From Example 152, $f(0) = 1$, $f'(0) = 0$ and $f''(0) = -1$. The quadratic approximation to $\cos(x)$ around $x = 0$ is

$$\cos(x) \approx 1 + 0(x-0) - \frac{1}{2}(x-0)^2 = 1 - \frac{x^2}{2}.$$

The graphs in Fig. 4.19 on page 116 show that this quadratic is a good approximation for $\cos(x)$ near 0.

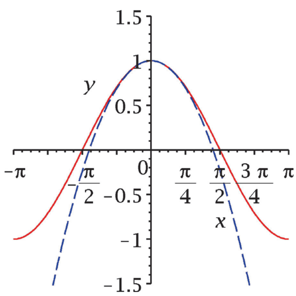


Fig. 4.19 $y = \cos(x)$ and $y = 1 - x^2/2$

Higher order derivatives are defined recursively in the same way that the second derivative is defined using the first derivative. This recurrence relation is

$$\frac{d^{n+1} f(x)}{dx^{n+1}} = \frac{d}{dx}\left(\frac{d^n f(x)}{dx^n}\right).$$

This is very simple to use.

Example 157. The first five derivatives of $\sin(x)$ are

$$\begin{aligned} \frac{d\sin(x)}{dx} &= \cos(x), \\ \frac{d^2\sin(x)}{dx^2} &= \frac{d\cos(x)}{dx} = -\sin(x), \\ \frac{d^3\sin(x)}{dx^3} &= \frac{d(-\sin(x))}{dx} = -\cos(x), \\ \frac{d^4\sin(x)}{dx^4} &= \frac{d(-\cos(x))}{dx} = \sin(x), \quad \text{and} \\ \frac{d^5\sin(x)}{dx^5} &= \frac{d\sin(x)}{dx} = \cos(x). \end{aligned}$$

Note that the derivatives of $\sin(x)$ and $\cos(x)$ return to the original function every fourth derivative.

Example 158. The first four derivatives of $g(x) = 1/x$ are

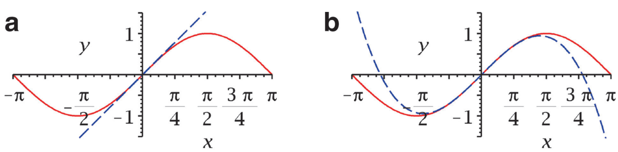
$$\begin{aligned}
\frac{d\,g(x)}{dx} &= -\frac{1}{x^2}.\\
\frac{d^2\,g(x)}{dx^2} &= \frac{d}{dx}\left(-\frac{1}{x^2}\right) = \frac{2}{x^3}.\\
\frac{d^3\,g(x)}{dx^3} &= \frac{d}{dx}\left(\frac{2}{x^3}\right) = -\frac{6}{x^4}. \quad \text{and}\\
\frac{d^4\,g(x)}{dx^4} &= \frac{d}{dx}\left(-\frac{6}{x^4}\right) = \frac{24}{x^5}.
\end{aligned}$$

The polynomial of degree n that matches the value of $f(x)$ at a and the first n derivatives of $f(x)$ at a is called the *Taylor polynomial* of degree n for $f(x)$ centered at a. These can be very good approximations for the function $f(x)$. With a little work we can show that the formula for an nth degree Taylor polynomial for $f(x)$ centered at a is

$$f(x) \approx f(a) + f'(a)(x-a) + \cdots + \frac{f^{(i)}(a)}{i!}(x-a)^i + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The next example shows how this can be used.

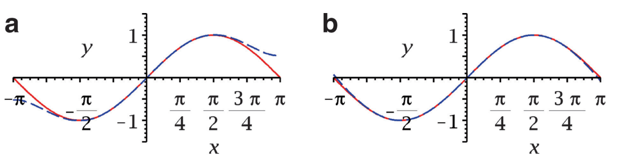
Example 159. Consider the function $f(x) = \sin(x)$ and let $a = 0$. The first four odd numbered Taylor polynomials are



$y = \sin(x)$ and $y = x$

$y = \sin(x)$ and $y = x - x^3/6$

Fig. 4.20



$y = \sin(x)$ and $y = x - x^3/6 + x^5/120$

$y = \sin(x)$ and $y = x - x^3/6 + x^5/120 - x^7/7!$

Fig. 4.21

$$\begin{aligned}
T_1(x) &= T_2(x) = x\\
T_3(x) &= T_4(x) = x - \frac{x^3}{6}\\
T_5(x) &= T_6(x) = x - \frac{x^3}{6} + \frac{x^5}{120} \quad \text{and}\\
T_7(x) &= T_8(x) = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{7!}.
\end{aligned}$$

The graphs of the Taylor polynomials along with $\sin(x)$ are shown in Figs. 4.20 on page 117 and 4.21 on page 117.

These graphs indicate that the higher the degree of the Taylor polynomial for $\sin(x)$, the better the approximation.

While it is often true, it is not always true that increasing the degree of the Taylor polynomial increases the accuracy of the approximation for a function.

Example 160. The function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Has derivatives of all orders at every point with

$$\frac{d^n f}{dx^n}(0) = 0$$

for all integers $n > 0$. (This is not trivial and is not demonstrated here.) This means that all of the Taylor polynomials for f centered at $x = 0$ are $p_n(x) = 0$. These polynomials do not approximate $f(x)$ very well since $f(x) \neq 0$ for all $x \neq 0$.

Exercises

1. Find the first five derivatives of each of the following functions.

(a) $f(x) = x^3 - 6x^2 + 4x - 5$
(b) $g(y) = y^6 - 6y^4 + 10y^3 - y^2 + 7y + 15$
(c) $h(z) = z^5 + \dfrac{10}{z^5}$
(d) $f(x) = \tan^{-1}(x)$
(e) $g(y) = \sin(y^3)$
(f) $h(z) = e^{5z+1}$
(g) $f(x) = \cos(\ln(x))$
(h) $g(y) = \ln(y)$
(i) $h(z) = e^{-\frac{1}{z^2}}$
(j) $f(x) = \ln\left(x^2 + x\right)$
(k) $g(y) = \sin\left(\cos^2(y)\right)$
(l) $h(z) = \sqrt{3z+5}$
(m) $f(x) = 2^x$
(n) $g(y) = \log_3\left(y^2\right)$

2. Find a quadratic approximation to the function centered at the given point.

(a) $f(x) = x^3 - 6x^2 + 4x - 5$, $x = 0$
(b) $g(y) = y^6 - 6y^4 + 10y^3 - y^2 + 7y + 15$, $y = 0$
(c) $h(z) = z^5 + \dfrac{10}{z^5}$, $z = 1$
(d) $f(x) = \tan^{-1}(x)$, $x = 1$

(e) $g(y) = \sin(y^3)$, $y = \sqrt[3]{\pi/4}$
(f) $h(z) = e^{5z+1}$, $z = 0$
(g) $f(x) = \cos(\ln(x))$, $x = 1$
(h) $g(y) = \ln(y)$, $y = \ln(2)$

3. Find the cubic approximations for the following functions centered at 0.

(a) $f(x) = x^3 + 6x - 3$
(b) $g(y) = \tan(x)$
(c) $f(x) = \sin^{-1}\left(\frac{x}{2}\right)$
(d) $g(y) = \arctan(y)$
(e) $f(x) = x^4 - 6x^2 + 2$
(f) $g(y) = \sin(x^3)$
(g) $f(x) = \sec(x)$
(h) $g(y) = e^y$

4. For each of the following functions $f(x)$ and points a and b find the first four Taylor polynomial approximations to $f(x)$ centered at a. (This will be polynomials of degrees 0,1,2, and 3.) Then find the error for each of the polynomials at b, $f(b) - p(b)$.

(a) $f(x) = \cos(x)$, $a = 0$, $b = \frac{\pi}{6}$
(b) $f(x) = \sin(x)$, $a = 0$, $b = -\frac{\pi}{6}$
(c) $f(x) = e^x$, $a = 1$, $b = \frac{7}{8}$
(d) $f(x) = e^x$, $a = -1$, $b = -\frac{9}{8}$
(e) $f(x) = \ln(x)$, $a = 1$, $b = 1.05$
(f) $f(x) = x\cos(x)$, $a = 0$, $b = \frac{\pi}{12}$

4.7 Derivative Practice Problems

This section is simply a collection of problems to practice taking derivatives. Unless stated otherwise either take the indicated derivative or take the derivative of the given function.

1. $f(x) = \text{arcsec}\,(x-5)$
2. $f(x) = \tan(\tan(x))$
3. $g(w) = \dfrac{6w^2 + 8w + 7}{9w^2 + 2w + 12}$
4. $h(\theta) = \sin^{18}(\theta)$
5. $z(t) = 2^{2t}$
6. $g(z) = 7z^2 + 3\ln(z) - 6$
7. Find $y'(x)$ if $e^{yx^7} = y^2x^2$ defines y as a function of x.
8. $f(x) = \sin^{-1}\left(\dfrac{x}{3} - \dfrac{1}{2}\right)$
9. $g(w) = \dfrac{w^7}{\tan(w)}$
10. $\dfrac{d}{ds}\left(2s^7\right)$
11. $\dfrac{d}{dx}\left(-12x^{\frac{48}{17}}\right)$
12. $g(w) = 3\cos(w)\sin(w)$
13. $f(t) = \exp(t)$
14. $L(w) = 4^{\sin(w)}$
15. $g(z) = \dfrac{e^z}{8 + 3z - 5z^2}$
16. $f(\theta) = \cos(\theta)$
17. Find the second derivative of $h(x) = -8x^4 - 4x^3 + 3x^2 + 9x + 6$.
18. $g(w) = -3w^12\sin(w) + 7w^{13}$
19. Find $y'(x)$ if $y^6 - 9y = -3x^2y^2$ defines y as a function of x.
20. Find the second derivative of $r(x) = \tan(14x)$
21. $g(y) = \tan^{-1}\left(\dfrac{y}{8} + \dfrac{1}{10}\right)$
22. $f(x) = \sin(\ln(3x))$
23. $r(s) = \log_7(\ln(3s))$
24. $h(x) = \sec(x)$
25. $g(w) = 11e^w w^15$

26. $g(z) = -z^2 - 6\sin(z)$

27. $s(r) = \dfrac{\sin(3r)}{\sqrt[3]{x^4 - 3x^3 + x}}$

28. $f(x) = x^{23}\sin(3x)\,e^{x^2}$

29. Find the third derivative of $h(z) = \arccos(2z)$.

30. $z(t) = t^{t^2}$

31. $f(x) = 82 + 7x^5 - 26x^4 + 21x^3 - 28x^2 - 90x$

Chapter 5
Applications of Limits and Derivatives

5.1 The Intermediate Value Theorem

The Intermediate Value Theorem (IVT) is the first application considered that uses concepts from the last three chapters. It is often considered an "obvious" result that has some important uses. The hypotheses in the IVT are very important and must be properly stated. The idea of the IVT is that a continuous function $f(x)$ on an interval $[a,b]$ cannot skip any values between $f(a)$ and $f(b)$, intermediate values. We consider one application of the IVT, the bisection method for finding zeroes of a function.

Theorem 43 (Intermediate Value Theorem). *Let $f:[a,b]\to\mathbb{R}$ be continuous on $[a,b]$. Then, for every s between $f(a)$ and $f(b)$ there is a point $c\in(a,b)$ such that $f(c)=s$.*

Proof. This is an outline to give an idea of what is happening. *It is not rigorous.*

Assume that $s=0$ and that $f(a)>0>f(b)$. There is a largest $c\in(a,b)$ such that $f(c)\geq 0$. This implies that $f(x)<0$ if $x\in(c,b)$. If $f(c)>0$, then, because $f(x)$ is continuous at c, $f(x)>0$ on $(c-\delta,c+\delta)$ for some $\delta>0$. This contradicts the assumption that $f(x)<0$ if $x\in(c,b]$. Thus $f(c)=0$.

Example 161. Let $f(x)=cos(x)-x$. Since $\cos(x)$ and x are continuous, so is $f(x)$. The values of $f(x)$ at $x=0$ and $x=\frac{\pi}{2}$ are 1 and $-\frac{\pi}{2}$. The Intermediate Value Theorem says that for each of the values $y=-1,-\frac{1}{2},0,\frac{1}{2}$ there is a $c\in\left(0,\frac{\pi}{2}\right)$ such that $f(c)=y$. See Fig. 5.1a on page 122.

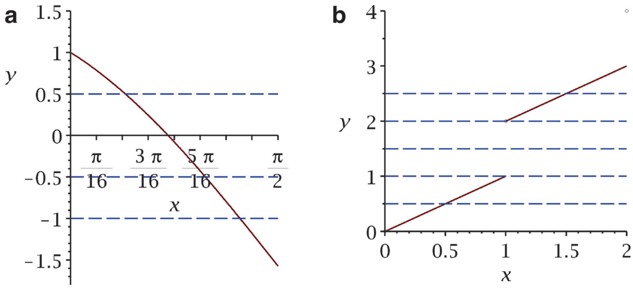
It is necessary that $f(x)$ be continuous for the Intermediate Value Theorem to hold. The next example shows that if $f(x)$ is not continuous $f(x)$ may or may not attain certain values on an interval.

Example 162. Consider the function $f(x)=x+\lfloor x\rfloor$ on $[0,2]$ where the $\lfloor x\rfloor$ is the floor function, Example 92 on page 70 as in Fig. 5.1b on page 122. This function is not continuous on $[0,2]$ When one considers the values $y=\frac{1}{2},1,\frac{3}{2},2,\frac{5}{2}$ one finds that $f(x)$ takes on the values $y=\frac{1}{2},2,\frac{5}{2}$ and skips the values $y=1,\frac{3}{2}$.

Remark 5. The graph in the last example shows the gap in the graph of the function. Most calculators and many computer programs graph a function by generating points $(x,f(x))$ and connecting those points. They often do not test to see if the function is actually continuous. Take care by *not* assuming that a computer generated graph of a function demonstrates continuity.

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_5) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_5



$f(x) = cos(x) - x$ for Example 161 $f(x) = x + \lfloor x \rfloor$ for Example 162

Fig. 5.1

The intermediate value theorem does not say anything about possible function values not between $f(a)$ and $f(b)$. The following example illustrates this.

Example 163. Let $f(x) = \sin(x)$ on the interval $I = \left[-\frac{\pi}{4}, \frac{9\pi}{4}\right]$. The IVT tells one that $\sin(x)$ takes on every value in $[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$ on I. One the other hand $\sin\left(\frac{\pi}{2}\right) = 1$ and $\sin\left(\frac{3\pi}{2}\right) = -1$. These values are outside the interval for the IVT, see Fig. 5.2 on page 122.

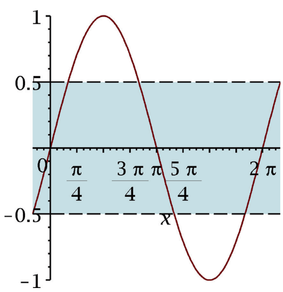


Fig. 5.2 Where $f(x) = \sin(x)$ is between $y = -\frac{1}{2}$ and $y = \frac{1}{2}$

From a computational point of view the problem with the Intermediate Value Theorem is that it only says that a point exists. The IVT does not say how to find an appropriate point. In many cases, as in solving $\cos(x) - x = 0$, one cannot explicitly solve for x. Because the IVT says, in a loose sense, where a root of a function must be, it can be used to narrow the interval where a zero must lie.

Consider the continuous function $f(x) = 2\cos(10x) - x$ on the interval $[0, \pi/2]$, see Fig. 5.3a on page 123. Since $f(0) > 0$ and $f(\pi/2) < 0$, there must be a point c in $[0, \pi/2]$ where $f(x)$ is zero. If one takes a guess for c and $f(c)$ is not 0, one has a smaller interval that contains a root of $f(x)$. The common choice for the guess at c is $\frac{a+b}{2}$. In this case the point is $\frac{\pi}{4}$ and $f(\pi/4) = -\frac{\pi}{4}$, see Fig. 5.3b on page 123. The fact that $f(0) \cdot f\left(\frac{\pi}{4}\right) < 0$ tells us that $f(c) = 0$ for some $c \in \left(0, \frac{\pi}{4}\right)$. If we keep repeating this procedure we get the *bisection* or *interval halving* method.

The bisection, interval halving, method is used for finding roots of a function. One starts the method with a continuous function $f(x)$ on a closed interval $[a, b]$ such that $f(a)\,f(b) < 0$ and an $\varepsilon > 0$ that tells how close the point returned by the algorithm should be to a root of $f(x)$. The assumption that $f(a)\,f(b) < 0$ guarantees that $f(x)$ has opposite signs at a and b. Since 0

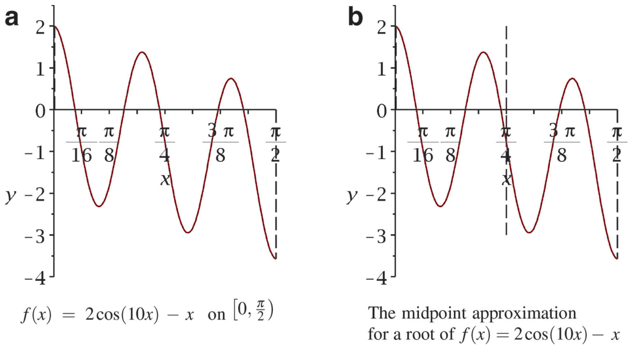


Fig. 5.3

is an intermediate value, this means that there is a root of $f(x)$ in (a,b). The goal is to find a point x such that x is within ε of a zero of $f(x)$.

At each step on has an interval $[a_i,b_i]$ such that $f(a_i)\,f(b_i) < 0$. One takes the midpoint of that interval, $c = \frac{a_i+b_i}{2}$, as an approximation to a root of the function. After n steps of doing this, the interval containing a root of the function has length $\frac{b-a}{2^n}$. This means that the midpoint of that interval is at most $\frac{b-a}{2^{n+1}}$ away from a root of the function. This is used to give the number of steps taken in the following algorithm. Note that $\rceil x\rceil$ is the smallest integer greater than or equal to x.

1. Set $n = \left\lceil \log_2\left(\dfrac{b-a}{\varepsilon}\right)\right\rceil$.
2. Repeat the following $n-1$ times.
   (a) Set $c = \frac{a+b}{2}$.
   (b) If $f(c) = 0$, return c and stop. A zero of $f(x)$ has been found.
   (c) If $f(a)\,f(c) > 0$ then replace a with c. Otherwise, replace b with c.
3. Return $\frac{a+b}{2}$.

This algorithm either returns an exact zero of $f(x)$ or it returns the midpoint of an interval with length $\frac{b-a}{2^{n-1}}$. Since the midpoint of the interval with length $\frac{b-a}{2^{n-1}}$ is less than $\frac{b-a}{2^n}$ away from any point inside the interval, this point is within ε of a zero of $f(x)$.

Example 164. Again consider the function $f(x) = 2\cos(10x) - x$ on the interval $[0,2]$. Table 5.1 gives the iterations of the bisection method to find a root of $f(x)$ to within 0.001. Since $\frac{2-0}{2^{11}} <$ 0.001, a total of 10 iterations are completed.

The final approximation for a zero of $f(x)$ is $x = 0.74707$. A plot of $f(x)$ on $[0.746, 0.7481]$ indicates there is a root in this interval. (See Fig. 5.4 on page 124)

Iteration	a	$f(a)$	b	$f(b)$	c	$f(c)$
0	0	1	2	−1.1838	1	−2.67814
1	0.	2.	1.0000	−2.6781	0.50000	0.067324
2	0.50000	0.067324	1.0000	−2.6781	0.75000	−0.056729
3	0.50000	0.067324	0.75000	−0.056729	0.62500	1.3739
4	0.62500	1.3739	0.75000	−0.056729	0.68750	0.97236
5	0.68750	0.97236	0.75000	−0.056729	0.71875	0.51770
6	0.71875	0.51770	0.75000	−0.056729	0.73438	0.24238
7	0.73438	0.24238	0.75000	−0.056729	0.74219	0.095382
8	0.74219	0.095382	0.75000	−0.056729	0.74609	0.019911
9	0.74609	0.019911	0.75000	−0.056729	0.74805	−0.018270
10	0.74609	0.019911	0.74805	−0.018270	0.74707	

Table 5.1 Ten steps of the bisection method for $f(x) = 2\cos(10x) - x$

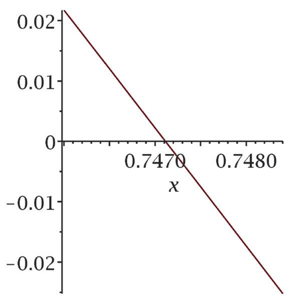


Fig. 5.4 A closeup view of the root of $f(x) = 2\cos(10x) - x$

Exercises

1. Do the following functions take on the given value in the given interval? Explain why the IVT applies or why it does not apply.

(a) $f(x) = 6x^2 - 10$, $f(c) = 2$, $[0,2]$

(b) $g(y) = \sin(y) + 2y$, $g(c) = 4$, $[1,4]$

(c) $h(z) = 2z + e^z$, $h(c) = 3.5$, $[0,3]$

(d) $f(x) = \tan(x)$, $f(c) = 1.4$, $[0, \frac{2\pi}{3}]$

(e) $g(y) = e^y - y$, $g(c) = 0$, $[-1,2]$

(f) $h(z) = 2z + \lfloor z \rfloor$, $h(c) = 3.5$, $[0,3]$

(g) $f(x) = 2x + \lfloor x \rfloor$, $f(c) = 2.5$, $[0,3]$

(h) $g(y) = e^y - 6y$, $g(c) = 0$, $[-1,2]$

2. Use the interval halving method to approximate a point where the function takes on the given value within the given interval. Do at least 5 steps.

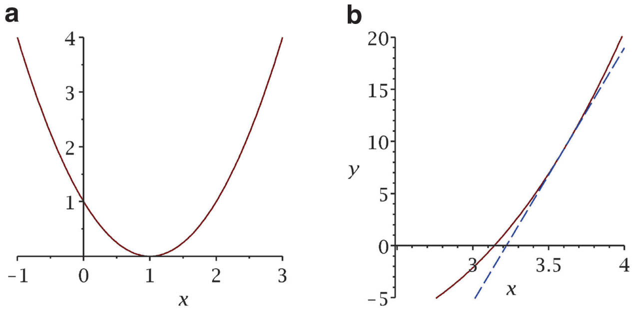
(a) $f(x) = x^3 + x^2 + 4x - 5$, $f(c) = 0$, $[0,2]$

(b) $g(y) = y^6 - 6y^4 + 10y^3 - y^2 + 7y + 15$, $g(c) = 18$, $[0,1]$

(c) $h(z) = z^5 - \dfrac{10}{z^5}$, $h(c) = 2$, $[1,2]$

(d) $f(x) = \tan^{-1}(x)$, $f(c) = 1.4$, $[0,10]$

(e) $g(y) = y\sin(y)$, $g(c) = 1$, $[0, \frac{\pi}{2}]$



The graph of $f(x) = x^2 - 2x + 1$ for Example 165

The root of a tangent line approximating the root of a function

Fig. 5.5

(f) $h(z) = e^{5z+1}$, $h(c) = 6$, $[0,1]$

(g) $f(x) = \cos(x) - x^2 + x$, $f(c) = 0$, $[0,2]$

(h) $g(y) = \sec y + y - 1$, $g(c) = 1$, $[0,1]$

3. Explain why the hypotheses of the bisection algorithm do not apply for finding a zero of the given function over the given interval. Each of these functions does have a zero in the given interval.

(a) $f(x) = x^2 + 2x + 1$, $[a,b] = [-2,0]$

(b) $g(y) = \sec(y) - 1$, $[a,b] = [-2,1]$

(c) $f(x) = \lfloor x \rfloor + x - \frac{5}{2}$, $[a,b] = [0,2]$

(d) $g(y) = \tan(y) - y$, $[a,b] = [1,6]$

5.2 Newton's Method

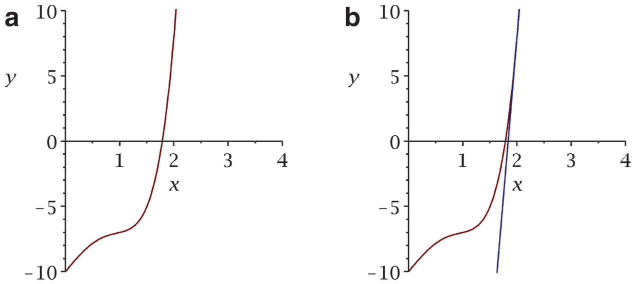
In the last section the bisection method was introduced as a way of finding zeros of a function. The advantages of the bisection method are that, once it is started, it is always possible to get as close to a zero as desired. The biggest problem with bisection is that it may be difficult to find an interval where $f(x)$ changes sign. A simple example shows that a function can have a root without changing sign.

Example 165. Let $f(x) = x^2 - 2x + 1$. This function has a zero at $x = 1$, as in Fig. 5.5a on page 125. However, since $f(x) = (x-1)^2$, the function is always nonnegative.

A different method based on using tangent line approximations to functions is Newton's method. The idea behind this method is very simple, find an approximation to the zero of a function by finding the zero of a tangent line approximation to $f(x)$. As Fig. 5.5b on page 125 shows, if the point where one is taking the tangent line approximation is close to the actual zero of the function, the new approximation is often much closer to the root. Hopefully, repeating this process will get one closer and closer to a zero of the function.

Before writing out the algorithm we need to find the zero of the tangent line to $f(x)$. The equation of a tangent line to the graph of $f(x)$ at x_0 is $y = f(x_0) + f'(x_0)(x - x_0)$. Setting $y = 0$, giving a zero of this line, and solving for x gives

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}. \tag{5.1}$$



a The graph of $f(x)$ for Example 166

b A first Newton step for Example 166

Fig. 5.6

Example 166. Consider the function $f(x) = x^5 - 3x^3 + 5x - 10$ in Fig. 5.6a on page 126 From the graph there is a zero between 0 and 2. Taking $x = 2$ as an initial guess for the root of this function, one can find a better approximation using Eq. (5.1).

Since $f'(x) = 5x^4 - 9x^2 + 5$, our new approximation is

$$\begin{aligned} x_1 &= 2 - \frac{f(2)}{f'(2)} \\ &= 2 - \frac{8}{49} \\ &\approx 1.837. \end{aligned}$$

This is shown in Fig. 5.6b on page 126

The actual value of the root is $x = 1.785585$ to six decimal places.

As with the bisection method, this procedure is repeated until the x value is close to the zero of the function. This is often judged by taking two consecutive values of x to be within some $\varepsilon > 0$ of each other or by taking the value of $f(x)$ to be within some $\delta > 0$ of 0. Since a sequence of points generated by Newton's method may fail to converge to a root of the function, the maximum number of iterations allowed is often set to some $N > 0$.

Assume that one has an initial point x_0, a differentiable function $f(x)$, the derivative $f'(x)$ of $f(x)$, a maximum number of iteration N, and two tolerances $\varepsilon, \delta > 0$. The algorithm is expressed below.

(1) Start with an initial x_0 for the sequence to be constructed.

(2) Set

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

and set $n = 1$.

(3) Repeat the following until $|x_n - x_{n-1}| < \varepsilon$, $|f(x_n)| < \delta$, or $n \geq N$.

(A) Set

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

(B) Set $n = n + 1$.

(4) Return x_n.

Here are a few steps of the Newton's method for $f(x) = x^5 - 3x^3 + 5x - 10$.

Example 167. Let $f(x) = x^5 - 3x^3 + 5x - 10$ and $x_0 = 2$. Using 10 digit arithmetic one gets the values in Table 5.2 on page 127.

n	x	$f(x)$
0	2	8
1	1.83673694	1.49861258
2	1.789225066	0.09930111
3	1.785604969	0.00053822
4	1.785585133	0.00000002
5	1.785585132	−0.00000001

Table 5.2 A table of five Newton steps for Example 167

Figure 5.7 on page 127 gives a graphical representation of the first two steps of Newton's method.

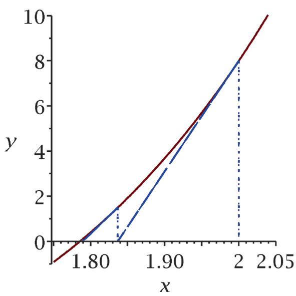


Fig. 5.7 The first two Newton steps for Example 167

As one can see in the previous example, Newton's method can converge very quickly. The following result, stated without proof, tells one how fast Newton's method converges.

Theorem 44. *Let $f(x)$ be a function that has continuous first and second derivatives. Also assume that $f(a) = 0$ and $f'(a) \neq 0$. Given an x_0, define a sequence by*

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

If x_0 is close enough to a, then $|x_{n+1} - a| < c|x_n - a|^2$ for some $c > 0$.

To get an idea what this means, assume that this holds for a sequence $\{x_n\}_{n=0}^{\infty}$ with a $c = 1$. If $|x_k - a| < 0.01$, then $|x_{k+1} - a| < 0.0001$. The basic conclusion is that one doubles the number of digits of accuracy with each step of Newton's method when the derivative of $f(x)$ is not zero at the zero of $f(x)$ and one is close enough to the zero.

When comparing Newton's method and bisection there are several things to consider. If one has a function and its derivative, starting Newton's method only requires choosing a point.

n	x	$f(x)$
0	0.5	0.35
1	0.8500000000	0.1225000000
2	1.258333333	0.1667361109
3	0.9356182792	0.1041450060
4	1.744427087	0.6541716879
5	1.305047798	0.1930541591
6	0.9886151596	0.1001296146
7	5.386112630	19.33798400
8	3.181656699	4.859625952
9	2.067909987	1.240431740
10	1.487134568	0.3373000873

Table 5.3 Ten Newton' steps for Example 168. The x values bounce around the local minimum

Choosing a good point may be a problem, but starting Newton's method is rather simple. Starting bisection is more difficult. To start bisection we must have an interval where the function changes sign. If the values of the function are predominantly of one sign and the region where the function takes on the other sign is very small, finding a starting interval can be quite difficult.

Once we have an interval where the function changes sign, bisection is guaranteed to converge. Unfortunately this convergence is quite slow. It take approximately 10 steps to bisection to gain three decimal digits of accuracy. On the other hand, as was noted earlier, if Newton's method is converging to a root c of the function and the derivative of the function is not zero at c, Newton's will converge very quickly. Newton's method is often chosen since it is easy to start and often converges very quickly.

Other behavior is possible for Newton's method. The points in the sequence can go off to infinity or bounce around. The following two examples show two ways that this can happen.

Example 168. This example show how Newton's method can bounce around a point that is almost a zero of a function. Let $f(x) = (x-1)^2 + 0.01$ and take $x_0 = 0.5$. The first ten elements of the sequence generated by Newton's method are in Table 5.3 on page 128.

The Fig. 5.8 on page 129 shows how one moves from one side of the minimum to the other side. This is repeated until one goes toward another zero of the function or the derivative at x_n is zero, causing an error. It may simple go back and forth around a point forever.

The initial point of the sequence generated by Newton's method may affect the convergence of Newton's method.

Example 169. Let $f(x) = \arctan(x)$. If one chooses an initial point near 0, the sequence converges to 0, the only zero of $f(x)$. The sequence for $x_0 = 1$ is in Table 5.4 on page 129. Figure 5.9a on page 129 shows the first two steps.

If one starts at $x_0 = 1.5$, the size of the x_n's goes to infinity. Part of the sequence for $x_0 = 1.5$ is in Table 5.5 on page 130.

The first few steps are illustrated in Fig. 5.9b on page 129.

Although Newton's method is very powerful when it works, one must be careful. To avoid the problems of the sequence going off toward infinity or cycling around a point, one often limits the number of iterations of the method. There can also be problems with calculating

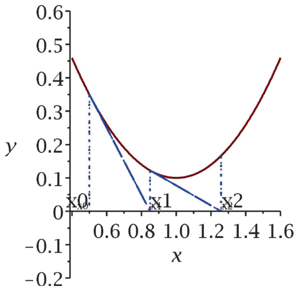
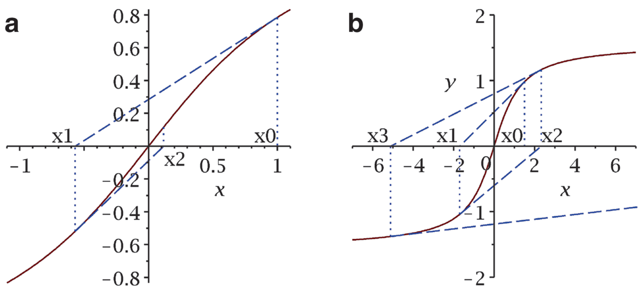


Fig. 5.8 Moving across a minimum with Newton's method as in Example 168

n	x	$f(x)$
0	1.	0.7853981634
1	−0.570796327	−0.5186693694
2	0.1168599041	0.1163322652
3	−0.0010610221	−0.001061021702
4	$7.97\,10^{-10}$	$7.970000000\,10^{-10}$

Table 5.4 Four Newton steps for $\arctan(x)$ starting at $x = 1$



Newton's method converging to 0 for $\arctan(x)$

Newton's method diverging for $\arctan(x)$

Fig. 5.9

function values and values of the derivative. Because of this, Newton's method is often halted if the step size is very small, the function values are small, or the function values are large.

Exercises

1. Find a zero of the given function to 5 decimal digits starting at the given point.

(a) $f(x) = 6x^2 - 10$, $x_0 = 2$

(b) $g(y) = \sin(y) - 2y + 1$, $y_0 = 2$

(c) $h(z) = 2z + e^z - 2$, $z_0 = 3$

(d) $f(x) = \tan(x) - x$, $x_0 = 4.7$

(e) $g(y) = e^y - 4y$, $y_0 = 0$

(f) $h(z) = z^2 - 6z + 9$, $z_0 = 1$

2. Explain the behavior of Newton's method for the following functions starting at the given point.

n	x	$f(x)$
0	1.5	0.9827937232
1	−1.694079600	−1.037546359
2	2.321126960	1.164002042
3	−5.114087826	−1.377694528
4	32.29568375	1.539842327
5	−1575.316935	−1.570161534

Table 5.5 Four Newton steps for $\arctan(x)$ starting at $x = 1.5$

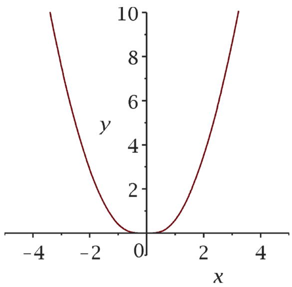


Fig. 5.10 Graph of $h(x)$ for Problem 3

(a) $f(x) = x^2 - 2x + 1$, $x_0 = 5$

(b) $g(y) = y^2 - 1.9y + 1$, $y_0 = 5$

(c) $h(z) = 0.9 - e^{-z^2}$, $z_0 = 1.6$

(d) $f(x) = 0.9 - e^{-x^2}$, $x_0 = 1.4$

(e) $g(y) = 0.9 - e^{-y^2}$, $y_0 = 1.1$

(f) $h(z) = cos(5z) - z + 3$, $z_0 = 0.2$

3. Consider the function

$$h(x) = \frac{(100x+1)^2\,(100x-1)^2}{10{,}000\,(1+x^2)}$$

with graph as in Fig. 5.10.

(a) Explain what happens when Newton's method is started at $z = 1$.

(b) Explain what happens when Newton's method is started at $z = -1$.

(c) Explain why it is hard to start the interval halving method for this function.

5.3 Related Rates

There are many situations where there are two quantities that are related and where the rate of change of one of the quantities is known. What we often want to know is how the second quantity is changing. A simple example is the relationship between the radius and circumference of a circle. If C is the circumference and R is the radius their relationship is

$$C = 2\pi R.$$

Assuming that both C and R are functions of t we have

$$C(t) = 2\pi R(t).$$

Using the following proposition, we can differentiate both sides and maintain the equality.

Theorem 45 (Uniqueness of the derivative). *If* $\mathbf{f}$ *and* $\mathbf{g}$ *are functions on some open interval* $\mathscr{I} \subset \mathbb{R}$ *such that* $\mathbf{f}(t) = \mathbf{g}(t)$ *for all* $t \in \mathscr{I}$ *and* $\mathbf{f}$ *is differentiable at* $t_0 \in \mathscr{I}$*, then* $\mathbf{g}$ *is differentiable at* t_0 *and* $\mathbf{g}'(t_0) = \mathbf{f}'(t_0)$.

Proof. Simply note that the difference quotients for $\mathbf{f}$ and $\mathbf{g}$ centered at t_0 are the same for any h with $t_0 + h \in \mathscr{I}$. Since the limit of the difference quotients exists for $\mathbf{f}$ as $h \to 0$, the limit of the difference quotients for $\mathbf{g}$ must also exist and equal the limit for $\mathbf{f}$.

Returning to the relationship between the radius and circumference of a circle, Proposition 45 on page 131 tells us that

$$\frac{dC(t)}{dt} = 2\pi \frac{dR(t)}{dt}.$$

This tells us that if the radius of a circle is changing at 2 in/min, then the circumference of the circle is change at 4π in/min .

There are simple examples where more derivatives involved. The following is one such example.

Example 170. A north-south highway and an east-west highway intersect at the origin $(0,0)$. Assume that at time $t = 0$ a woman starts 2 km south of the origin and walks north at 4 km/h Also at time $t = 0$ a man starts 3 km east of the origin and walks west at 3 km/h. What is the rate of change of the distance between the two people at $t = 4$? See Fig. 5.11 on page 132.

In order to illustrate the ideas here, the solution is given using general functions for the positions and the specific functions are put in at the end. Assume that the position of the woman is $(0, y(t))$ m and the position of the man is $(x(t), 0)$ m. The distance function is

$$s(t) = \sqrt{x(t)^2 + y(t)^2}.$$

Using the chain rule we get

$$s'(t) = \frac{x(t)x'(t) + y(t)y'(t)}{\sqrt{x(t)^2 + y(t)^2}}.$$

Since the walking speeds are constant $x'(t) = -3$ km/h and $y'(t) = 4$ km/h and the positions are $(0, y(4)) = (0, -2 + 4 \cdot 4) = (0, 14)$ and $(x(4), 0) = (3 - 3 \cdot 4, 0) = (-9, 0)$, at $t = 4$,

$$\begin{aligned} s'(4) &= \frac{-9 \cdot (-3) + 14 \cdot 4}{\sqrt{9^2 + 14^2}} \\ &= \frac{83}{\sqrt{277}} \, \mathrm{km/h}. \end{aligned}$$

Another example often used in calculus texts is that of a ladder with one end on the ground and the other end sliding down a wall.

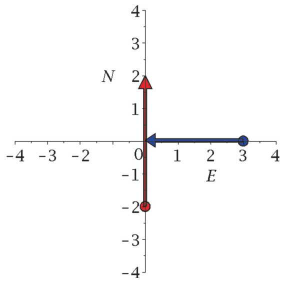


Fig. 5.11 Directions of travel of a woman and a man

Example 171. Consider a ladder that is 7 m long. One end is sliding away from a wall and the other end is against the wall. If the bottom of the ladder is moving away from the wall at 1/2 m/s when the bottom is 2 m from the wall, how fast is the top of the ladder moving at that time? See Fig. 5.12 on page 133.

If the top of the ladder is at $(0, y(t))$ and the bottom is at $(x(t), 0)$, we have

$$7 = \sqrt{x(t)^2 + y(t)^2}.$$

Taking the derivative of both sides with respect to t yields

$$0 = \frac{2y(t)y'(t) + 2x(t)x'(t)}{2\sqrt{x(t)^2 + y(t)^2}}.$$

Solving for $y'(t)$ and noting that if $x = 2$ then $y = \sqrt{45}$ gives

$$y'(t) = -\frac{x(t)}{y(t)}x'(t),$$

or

$$y'(t) = \frac{1}{\sqrt{45}}\,\mathrm{m/s}.$$

There are many other situations where related rates are used. In fact, the technique is a staple in science and engineering.

Example 172. Consider a conical tank with its point toward the ground. This tank is 10 m high and its radius at the top is 5 m. See Fig. 5.13a on page 133. Water is being pumped into the tank at $1/2\,\mathrm{m}^3$/min. The problem is to find how fast the water is rising when the height of the water is 6 m.

The volume of a right circular cone with height h and radius r is $V = \pi r^2 h/3$. In this case we always have $r = h/2$ using similar triangles as in the side on view of the tank in Fig. 5.13b on page 133.

The relationship between V and h is now

$$V = \frac{1}{12}\pi h^3.$$

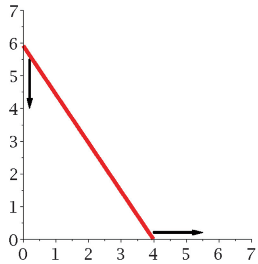


Fig. 5.12 A ladder sliding down a wall

Differentiating both sides with respect to time, t, gives

$$\frac{dV}{dt} = \frac{1}{4}\pi h^2 \frac{dh}{dt}.$$

Substituting in the given values yields

$$\frac{1}{2} = \frac{1}{4}\pi 6^2 \frac{dh}{dt}.$$

This gives

$$\frac{dh}{dt} = \frac{9}{2\pi}\,\mathrm{m/min}.$$

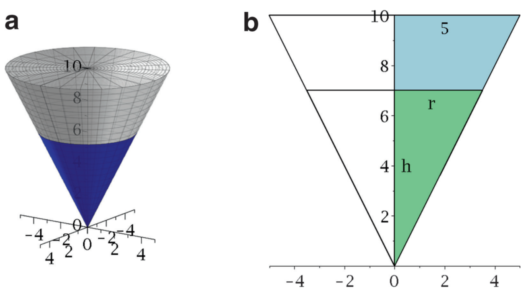


Fig. 5.13 A conical tank partially filled with water

As a final application of related rates the rate of change of the angle between the direction of travel of one object and the direction from that object to another object is considered. This might be the angle between the forward direction of one airplane and the direction to another airplane. If the position of the first object is given by $\mathbf{r}(t)$ and the position of the second object is given by $\mathbf{s}(t)$, the angle under consideration is given by

$$\theta = \cos^{-1}\left(\frac{\mathbf{r}'(t)\cdot(\mathbf{s}(t)-\mathbf{r}(t))}{\|\mathbf{r}'(t)\|\,\|\mathbf{s}(t)-\mathbf{r}(t)\|}\right).$$

(See Eq. 1.14.)

To make the computations simpler set $f = \mathbf{r}'(t)\cdot(\mathbf{s}(t)-\mathbf{r}(t))$ and $g = \|\mathbf{r}'(t)\|\,\|\mathbf{s}(t)-\mathbf{r}(t)\|$. Taking derivatives with respect to t we have

$$f' = \mathbf{r}''(t)\cdot(\mathbf{s}(t)-\mathbf{r}(t)) + \mathbf{r}'(t)\cdot(\mathbf{s}'(t)-\mathbf{r}'(t))$$

and

$$g' = \frac{\mathbf{r}''(t)\cdot\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}\,\|\mathbf{s}(t)-\mathbf{r}(t)\| + \frac{(\mathbf{s}'(t)-\mathbf{r}'(t))\cdot(\mathbf{s}(t)-\mathbf{r}(t))}{\|\mathbf{s}(t)-\mathbf{r}(t)\|}\,\|\mathbf{r}'(t)\|.$$

Differentiating the equation for θ with respect to t gives

$$\begin{aligned}\frac{d\theta}{dt} &= \frac{d}{dt}\cos^{-1}\left(\frac{f}{g}\right)\\ &= \frac{-1}{\sqrt{1-\frac{f^2}{g^2}}}\,\frac{f'g-fg'}{g^2}\\ &= \frac{-1}{g\sqrt{g^2-f^2}}\,\left(f'g-fg'\right)\end{aligned}$$

To show how this works, an example that simplifies the computations is considered. Assume that the motion of the first object is circular in the xy-plane,

$$\mathbf{r}(t) = (\cos(t),\sin(t),0)$$

and that the motion of the second object is linear

$$\mathbf{s}(t) = (t-1,1-2t,-t).$$

This means that $\|\mathbf{r}'(t)\| = 1$, $\mathbf{r}'(t)\cdot\mathbf{r}'(t) = 1$, $\mathbf{r}''(t)\cdot\mathbf{r}(t) = -1$, and $\mathbf{r}'(t)\cdot\mathbf{r}(t) = 0$. The derivatives for f and g now reduce to

$$f' = \mathbf{r}''(t)\cdot\mathbf{s}(t) + \mathbf{r}'(t)\cdot\mathbf{s}'(t)$$

and

$$g' = \frac{(\mathbf{s}'(t)-\mathbf{r}'(t))\cdot(\mathbf{s}(t)-\mathbf{r}(t))}{\|\mathbf{s}(t)-\mathbf{r}(t)\|}.$$

At $t=0$ we have $\mathbf{r}(0) = (1,0,0)$, $\mathbf{r}'(0) = (0,1,0)$, $\mathbf{r}''(0) = (-1,0,0)$, $\mathbf{s}(0) = (-1,1,0)$ and $\mathbf{s}'(0) = (1,-2,-1)$. From this we have $\mathbf{s}(0)-\mathbf{r}(0) = (-2,1,0)$, $f(0) = (0,1,0)\cdot(-2,1,0) = 1$, and $g(0) = \|(0,1,0)\|\,\|(-2,1,0)\| = \sqrt{5}$.

This means that

$$f'(0) = (-1,0,0)\cdot(-1,1,0) + (0,1,0)\cdot(1,-2,-1) = -1$$

and

$$g'(0) = \frac{((1,-2,-1)-(0,1,0))\cdot(-2,1,0)}{\|(-2,1,0)\|} = \frac{-5}{\sqrt{5}} = -\sqrt{5}.$$

Finally

$$\begin{aligned}\frac{d\theta}{dt} &= \frac{1}{\sqrt{5}\sqrt{5+1}}\,\Big((-1)\cdot\sqrt{5} - 1\cdot(-\sqrt{5})\Big)\\ &= 0.\end{aligned}$$

Another similar problem is finding the rate of change of the distance between two objects. It would seem easy to simply write the distance between the objects as a function of time and then take the derivative. It turns out that this is very simple using what has been presented about vector valued functions.

Example 173. If two masses have positions $\mathbf{r}(t) = \big(\cos(\pi t), 3\sin(\pi t)\big)$ and $\mathbf{w}(t) = \big(t^2+1, 2t-3\big)$, how fast is the distance between the objects changing when $t = 1/2$? In vector notation the distance between the objects is

$$s(t) = \sqrt{\big(\mathbf{r}(t) - \mathbf{w}(t)\big) \cdot \big(\mathbf{r}(t) - \mathbf{w}(t)\big)}.$$

Taking the derivative of this expression with respect to t gives

$$s'(t) = \frac{\big(\mathbf{r}(t) - \mathbf{w}(t)\big) \cdot \big(\mathbf{r}'(t) - \mathbf{w}'(t)\big)}{\sqrt{\big(\mathbf{r}(t) - \mathbf{w}(t)\big) \cdot \big(\mathbf{r}(t) - \mathbf{w}(t)\big)}}. \tag{5.2}$$

Calculating gives

$$\mathbf{z}(t) = \mathbf{r}(t) - \mathbf{w}(t) = \big(\cos(\pi t) - t^2 - 1,\ 3\sin(\pi t) - 2t + 3\big)$$

and

$$\mathbf{z}'(t) = (-\pi \sin(\pi t) - 2t,\ 3\pi \cos(\pi t) - 2)\,.$$

At $t = 1/2$ this gives

$$\mathbf{z}\left(\frac{1}{2}\right) = \left(-\frac{5}{4},\, 6\right)$$

and

$$\mathbf{z}'\left(\frac{1}{2}\right) = (-\pi - 1,\ -2)\,.$$

Plugging these into the expression for $s'(t)$ yields

$$s'\left(\frac{1}{2}\right) = \frac{\left(-\frac{5}{4},\, 6\right) \cdot (-\pi - 1,\ -2)}{\sqrt{\left(-\frac{5}{4},\, 6\right) \cdot \left(-\frac{5}{4},\, 6\right)}} = \frac{5\pi - 43}{\sqrt{601}}.$$

Exercises

1. Water is draining from a right cylindrical tank at 5 l/s. The tank has a radius of 4 m and is 15 m tall. How fast is the height of the water changing when the height of the water in the tank is 7 m?
2. A man, who is 2 m tall, is walking away from a lamppost with a lamp that is 7 m above the ground. If the man is walking at 0.5 m/s, how fast is the length of his shadow changing when he is 10 m from the lamppost?
3. A man, who is 2 m tall, is walking along a path that, at its closest point, is 8 m away from a lamppost with a lamp that is 7 m above the ground. If the man is walking at 0.5 m/s and moving away from the lamppost, how fast is the length of his shadow changing when he is 10 m from the closest point on the path to the lamppost?

4. One person is driving north on a north-south highway at 50 mph and another person is traveling east on an east-west highway at 70 mph. How fast is the distance between the people changing when the first person is 20 mi north of the intersection of the highways and the second person is 40 mi west of the intersection of the highways?
5. A hot air balloon passes over a person on the ground at a height of 400 ft. The balloon is traveling at 10 mph. The person is tracking the balloon with their camera. How fast is the angle between the lens of the camera and the ground changing 10 min after the balloon has passed over the person on the ground?
6. A point moves along the curve $y = x^3 - x$. All distances are measured in meters. Assume the rate of change of the x-coordinate of the point is 3 m/s. How fast is the distance from the point to $(4, -2)$ m changing when the point is at $(2, 6)$ m?
7. Sand is being dumped onto the top of a pile of sand. The pile of sand is a conical pile whose height is always equal to the diameter of the base. If the sand is falling at a rate of 3 m^3/min, how fast is the height of the pile increasing when the height of the pile is 7 m?
8. A balloon has the shape of a right circular cylinder topped by a hemisphere. The height of the cylinder is 1.6 times the diameter of the base. If air is being pumped in at 1 m^3/min, how fast is the diameter of the cylinder increasing when the diameter is 4 m?
9. A pulsar is 10,000 light-years from earth and rotates 650 times per second. How fast does the beam of photons from this pulsar travel across the surface of the earth? (Set this up as a related rates problem. This is an approximation for the pulsar PSR B1937+21.) Compare this speed with the speed of light.
10. A drawbridge has two spans that are 25 m long that rotate up (Fig. 5.14 on page 136). If the angle that the spans make with level is increasing at $1/4$ rad/min when the spans are $\frac{\pi}{6}$ radians from level, how fast is the distance between the ends of the spans increasing?

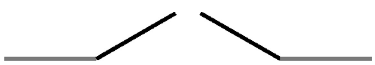


Fig. 5.14 A sketch of a draw bridge

11. If a spherical drop of water is picking up water at a rate proportional to its surface area, show that the radius is changing at a constant rate.
12. Consider the triangular trough with a triangular cross section in Fig. 5.15 on page 136. The lengths of the sides of the triangular cross section are 2, 2, and 3 ft. The trough is 8 ft long. If water is being pumped into the tank at 5 ft^3/min, how fast is the water rising when the depth of the water is 1 ft?

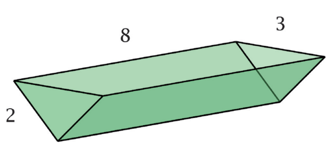


Fig. 5.15 The trough for Problem 12

13. A boat is being pulled into a dock from the top of the dock using a rope. The rope is being pulled in at a rate of 16 ft/min and the top of the dock is 10 ft above where the rope attaches to the boat. How fast is the distance from the boat to the dock changing when the boat is 20 ft from the dock?

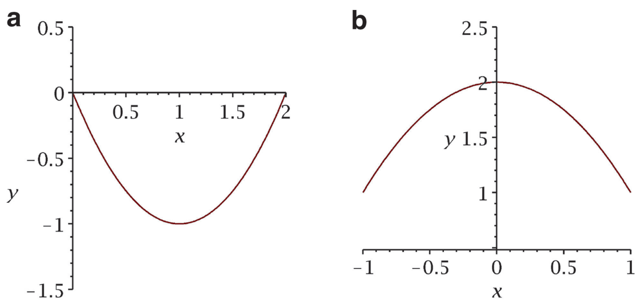
14. At $t = t_0$ an object is at $(0,0,1)$ with velocity $(1,2,-1)$ and at the same time a second object is at $(1,1,0)$ with velocity $(3,-1,2)$. What is the rate of change of the distance between the two objects at $t = t_0$?

15. A crane has a 200 ft boom that is set at an angle of $60°$ from horizontal. The crane is raising a beam at 10 ft/s. How fast is the distance from the base of the crane to the beam changing when the beam is 40 ft above the ground?

16. A crane has a 200 ft boom that is set at an angle of $60°$ from horizontal. The crane is raising a beam at 10 ft/s. How fast is the angle between the boom of the crane and the line from the base of the crane to the beam changing when the beam is 40 ft above the ground?

17. A new species of animal was introduced into a circular area. The region inhabited by the animals remains circular. Assume that the density of animals remains constant. If the population of animals increases at 10 % per year, how fast is the radius of the inhabited region increasing when the radius of the region is 10 mi? (Think about the relationship between the area and the population.)

18. A 100 m long incline has its west edge fixed and its east edge is being raised at 1 m/min. If a 10 kg mass is on the incline, what is the rate of change of the force due to gravity that is pushing the mass down the incline when the east edge of the incline is 4 m above the west edge of the incline? (Remember that force is a vector quantity.)



The graph of $f(x) = x^2 - 2x$ The graph of $g(x) = 2 - x^2$

Fig. 5.16 Graphs for Example 174

5.4 Extreme Values of Functions

A major area of application of calculus is finding the largest or smallest value of a function. The examples in this course are the first steps in learning about this. In physics a major concept is that things will happen in a way that uses the least energy possible. As an example, this idea is used to find the best shapes for ships and planes. Applications such as these are well beyond this course.

This section works only with the simplest case, minimizing or maximizing nice functions from $\mathbb{R}$ to $\mathbb{R}$. In order to do this we start with some basic definitions.

Definition 18. Let $\mathscr{C}$ be a subset of the real numbers. A function $f : \mathscr{C} \to \mathbb{R}$ has a *local maximum (or minimum)* at $c \in \mathscr{C}$ if for some $r > 0$, we have $f(x) \leq f(c)$ (or $f(x) \geq f(c)$) for all $x \in (c-r, c+r) \cap \mathscr{C}$.

Let $\mathscr{C}$ be a subset of the real numbers. A function $f : \mathscr{C} \to \mathbb{R}$ has a *global maximum (or minimum)* at $c \in \mathscr{C}$ if for all $x \in \mathscr{C}$, $f(x) \le f(c)$ (or $f(x) \ge f(c)$).

A point where a function has a global or local maximum or minimum is called an *extreme point*.

A few examples will make this definition clearer. The first is a simple example with simple quadratic functions.

Example 174. The function $f(x) = x^2 - 2x$ has a local minimum at $x = 1$ and $g(x) = 2 - x^2$ has a local maximum at $x = 0$. See Fig. 5.16 on page 137.

These are also global extreme points.

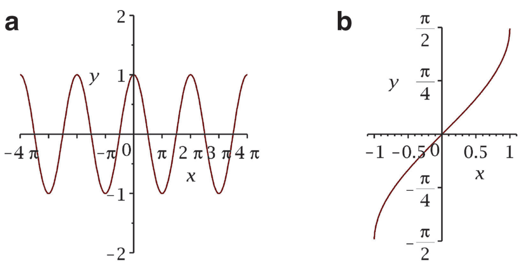
The following example shows that a function may have many maxima and minima.

Example 175. The function $\cos(\theta)$ has local maxima at the points $\theta = 2n\pi$, $n = 0, \pm1, \pm2, \ldots$ and local minima at the points $\theta = (2n+1)\pi$, $n = 0, \pm1, \pm2, \ldots$. At these points $\cos(\theta) = \pm1$. See Fig. 5.17a on page 138. In fact, all of the local maxima and minima are global maxima and minima.

Next is an example where the function has its maximum and minimum at the endpoints of its domain.

Example 176. The function $\arcsin(x)$ has a global maximum at $x = 1$ and a global minimum at $x = -1$. See Fig. 5.17b on page 138

Here is an example where a function only has a local minimum and a local maximum.



A graph of $\cos(\theta)$ for Example 175

A graph of $\arcsin(x)$ for Example 176

Fig. 5.17

Example 177. Consider the function $f(x) = x^3 - 3x$. From the graph in Fig. 5.18 on page 139 we can see that $f(x)$ has a local maximum at $x \approx -1$ with $f(-1) = 2$ and a local minimum at $x \approx 1$ with $f(1) = -2$. Since $\lim_{x\to-\infty} f(x) = -\infty$ and $\lim_{x\to\infty} f(x) = \infty$, these are not global extreme points.

If we want to find the maxima and minima of a function the first problem is finding the extreme points of the function. In order to accomplish this task we often use the following result.

Theorem 46 (First order optimality condition). *Assume that $f(x)$ is a function that has an extreme point c and that $f(x)$ is defined on an interval (a,b) that contains c. If $f'(c)$ exists, then $f'(c) = 0$.*

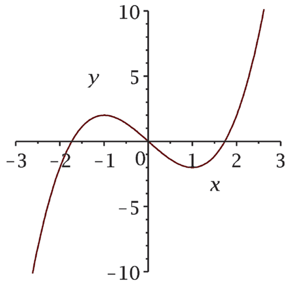


Fig. 5.18 The graph of $f(x) = x^3 - 3x$ for Example 177

Proof. The two possibilities, a local maximum and a local minimum, are almost identical. Because of this, it is assumed that $f(x)$ has a local maximum at c. We have, since c is a local maximum, for any $x < c$ in (a,b)

$$\frac{f(x)-f(c)}{x-c} \geq 0.$$

Since $f(x)$ is differentiable at c,

$$f'(c) = \lim_{x\to c^-} \frac{f(x)-f(c)}{x-c} \geq 0.$$

Similarly, if $x > c$ and $x \in (a,b)$

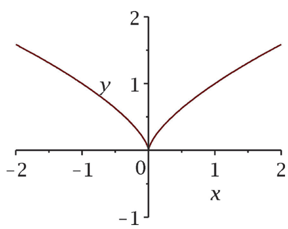


Fig. 5.19 The graph of $f(x) = x^{2/3}$ for Example 179

$$\frac{f(x)-f(c)}{x-c} \leq 0.$$

Since $f(x)$ is differentiable at c,

$$f'(c) = \lim_{x\to c^+} \frac{f(x)-f(c)}{x-c} \leq 0.$$

The only possible value for $f'(c)$ is 0 and the result is true.

In some cases this is easy to apply.

Example 178. Consider, again, $f(x) = x^3 - 3x$. This function has derivative $f'(x) = 3(x^2 - 1)$. Since the derivative is 0 at $x = \pm 1$, the candidates for maxima and minima are $x = \pm 1$. The graph in Fig. 5.18 on page 139 indicates that $x = -1$ is a local maximum and that $x = 1$ is a local minimum.

There are, however, cases where a function has an extreme point where it does not have a derivative.

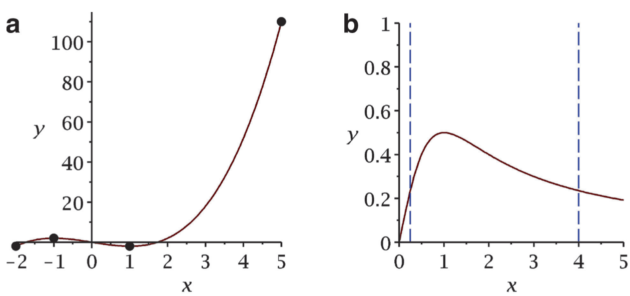
Example 179. Consider the function $f(x) = x^{2/3}$, see Fig. 5.19 on page 139. This function has a global minimum at $x = 0$ since $f(x) \geq 0$ for all x. However, it does not have a derivative at $x = 0$.

It is necessary to use the following result in the rest of this section. The conclusion of this theorem is that a continuous function on a closed bounded interval $[a, b]$ *must* attain both a maximum and a minimum value on the interval. It is an extremely important result whose proof is beyond this class. Because of that, the proof is omitted.

Theorem 47 (Maximum-Minimum Theorem). *Assume that $f(x)$ is a continuous function on a closed bounded interval $[a, b]$. Then there are points c and d in $[a, b]$ such that $f(c) \geq f(x)$ for all $x \in [a, b]$ and $f(d) \leq f(x)$ for all $x \in [a, b]$. These are global extrema for $f(x)$ on $[a, b]$.*

This result leads to a basic general procedure for finding the extreme points of a continuous function on a closed bounded interval $\mathscr{I} \subset \mathbb{R}$ with endpoints a and b. First we find the points where $f'(x) = 0$ and where $f'(x)$ does not exist (DNE). These are the *critical points* of $f(x)$. We then add the endpoints of the interval to this set of points. The extreme points can be found by comparing the values of $f(x)$ at the set of points we have found.

If the endpoints of the interval are not included in the interval, we must consider the limits $\lim_{x \to a^+} f(x)$ and $\lim_{x \to b^-} f(x)$ to determine if there are any global extrema for the function.



The graph of $f(x) = x^3 - 3x$ for Example 180

The graph of $f(x) = x/(1 + x^2)$ for Example 181

Fig. 5.20

Example 180. Returning to $f(x) = x^3 - 3x$ with a restricted domain of $\mathscr{I} = [-2, 5]$. The critical points are $x = \pm 1$ and the endpoints are -2 and 5. Calculating the values of $f(x)$ at $[-2, -1, 1, 5]$ gives the values $[-2, 2, -2, 110]$. The conclusion is that the maximum value of $f(x)$ on $\mathscr{I}$ is 110 at $x = 5$ and the minimum value of $f(x)$ is -2 at $x = -2$ and at $x = 1$. See Fig. 5.20a on page 140.

Here is another example of finding the extrema of a function on a closed bounded interval.

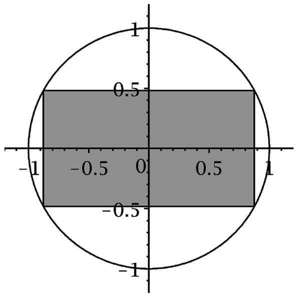


Fig. 5.21 A rectangle inscribed in a unit circle

Example 181. Find the minimum and maximum of $f(x) = x/(1+x^2)$ on $[1/4, 4]$, see Fig. 5.20b on page 140.

The derivative of $f(x)$ is

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

Since this is 0 at $x = \pm 1$, the three points where the optima can be are $x = 1/4, 1,$ and 4. Since the values of $f(x)$ at the points are $f(1/4) = 4/17$, $f(1) = 1/2$, and $f(4) = 4/17$. The maximum value is $1/2$ at $x = 1$ and the minimum value is $4/17$ at both $x = 1/4$ and $x = 4$.

There are some geometric examples that show how finding extrema can be useful.

Example 182. A rectangle is to be inscribed in a circle of radius r. The objective is to find the dimensions of the largest rectangle that can be inscribed in the circle. We can assume that the circle is centered at the origin, two sides are parallel to the x-axis, two sides are parallel to the y-axis, and all four corners are on the circle. (Why can we make these assumptions?) Figure 5.21 on page 141 illustrates this with $r = 1$.

Fig. 5.22 A right circular cone as in Example 183

The corners of the rectangle are now at the points, going counter clockwise from the first quadrant: $\left(x, \sqrt{r^2-x^2}\right)$, $\left(-x, \sqrt{r^2-x^2}\right)$, $\left(-x, -\sqrt{r^2-x^2}\right)$, and $\left(x, -\sqrt{r^2-x^2}\right)$. Since the lengths of the sides of the rectangle are $2x$ and $2\sqrt{r^2-x^2}$, the area of the rectangle as a function of x is

$$A(x) = 4x\sqrt{r^2-x^2}.$$

The domain of this function is $[0, r]$. (Why?)

Differentiating $A(x)$ with respect to x gives

$$\begin{aligned}\frac{dA(x)}{dx} &= 4\sqrt{r^2-x^2} - \frac{4x^2}{\sqrt{r^2-x^2}}\\ &= 4\,\frac{r^2-2x^2}{\sqrt{r^2-x^2}}.\end{aligned}$$

This is zero only when $x^2 = r^2/2$ or $x = \sqrt{2}\,r/2$.

In this case the y coordinate of the corner in the first quadrant is $y = \sqrt{2}\,r/2$ and the figure is a square with area $2r^2$.

This section concludes with another geometric example.

Example 183. The volume of a right circular cone is given by $V = \pi hr^2/3$ where h is the height of the cone and r is the radius of the base. (See Fig. 5.22 on page 141.) If the sum of r and h is 10, what dimensions will give the largest volume?

Since $h = 10 - r$, the volume can be written as

$$V(r) = \frac{1}{3}\,\pi(10-r)\,r^2.$$

Both r and h must be non-negative. Hence, the domain of $V(r)$ is $[0,10]$.

Differentiating $V(r)$ gives

$$\frac{dV(r)}{dr} = \frac{1}{3}\,\pi(20r-3r^3).$$

To find the critical points set $\frac{dV(r)}{dr}$ equal to 0. This gives

$$0 = \frac{1}{3}\,\pi(20r-3r^3).$$

The critical points are $r = 0$ and $r = \pm\sqrt{20/3}$. Since $-\sqrt{20/3}$ is outside of the domain of $V(r)$, the points to check are $r = 0$, $\sqrt{20/3}$, and 10. The function values at these points are 0, $20\pi/9\left(10-\sqrt{20/3}\right)$ and 0. This means that the maximum volume is $20\pi\left(10-\sqrt{20/3}\right)/9$ when $r = \sqrt{20/3}$ and $h = 10-\sqrt{20/3}$.

Exercises

1. Find all critical points for the following functions.

(a) $f(x) = x^2 - 4x + 6$

(b) $g(y) = 3y^3 + 3y^2 - 12y + 4$

(c) $f(x) = x + \dfrac{1}{x}$

(d) $g(y) = \dfrac{y}{1+4y^2}$

(e) $f(x) = \dfrac{x^2}{16+x^4}$

(f) $g(y) = \ln(y) - y^2 + 3y$

(g) $f(x) = 12x + 2x^3 + 9x^2 - 3$

(h) $g(y) = \ln(y^2+y) - \dfrac{y^2}{2} - y$

(i) $h(w) = \left|w^2 - 5w + 6\right|$

(j) $g(y) = \left|y^3 - 12y\right|$

2. Find all local and global maxima and minima for the following functions on the given intervals.

(a) $f(x) = x^2 - 4x + 6$ on $[1,4]$

(b) $g(y) = 3y^3 + 3y^2 - 12y + 4$ on $[-4,4]$

(c) $f(x) = x + \dfrac{1}{x}$ on $\left[\dfrac{1}{2}, 4\right]$

(d) $g(y) = \dfrac{y}{1+y^2}$ on $[-4,3]$

(e) $f(x) = \dfrac{x^2}{16+x^4}$ on $[-2,3]$

(f) $g(y) = \ln(y) - y^2 + 3y$ on $[1,4]$

(g) $f(x) = 12x + 2x^3 + 9x^2 - 3$ on $[-2,2]$

(h) $g(y) = \ln(y^2 + y) - \dfrac{y^2}{2} - y$ on $\left[\dfrac{1}{2}, 3\right]$

(i) $h(w) = \left|w^2 - 5w + 6\right|$ on $[-4,3]$

(j) $g(y) = \left|y^3 - 12y\right|$ on $[-2,4]$

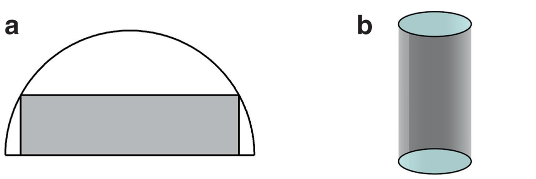
3. A rectangle is to be inscribed in a half circle of radius 1 ft. If the rectangle is as in the Fig. 5.23a on page 143, what dimensions will give the largest area?

4. A cylinder with caps on both ends has a volume of 1. What dimensions will minimize the area of the cylinder? See Fig. 5.23b on page 143.

5. If the position of a mass is given by $\mathbf{r}(t) = (\cos(t), 4\sin(t))$ m and time is in seconds, find the time(s) when the speed of the mass is greatest.

5.5 The Mean Value Theorem

The Mean Value Theorem, like the results on linear approximations, relates the values of a function with the derivative of the function. In this case we are comparing the slope of the secant line between $(a, f(a))$ and $(b, f(b))$ with the derivative of $f(x)$ in (a,b). Many uses of the Mean Value Theorem are beyond this class. Most of the applications involve estimating values of functions.



a A rectangle inscribed in a half circle of radius 1 ft

b A cylinder with caps on both ends

Fig. 5.23

Before stating the Mean Value Theorem, a special case called Rolle's theorem is considered. In this case the slope of the secant line is assumed to be 0.

Theorem 48 (Rolle's Theorem). *Let $f(x)$ be a function that is continuous on $[a,b]$ and differentiable on (a,b). Assume that $f(a) = f(b) = 0$. Then there is a $c \in (a,b)$ such that $f'(c) = 0$.*

Proof. There are three cases to consider: when $f(x) = 0$ on $[a,b]$, when $f(x) > 0$ for some $x \in (a,b)$, and when $f(x) < 0$ for some $x \in (a,b)$. The last two cases are almost identical, hence only the first of them will be considered.

First assume that $f(x) = 0$ on $[a,b]$. The result holds since $f'(c) = 0 = (f(b) - f(a))/(b-a)$ for all $c \in (a,b)$.

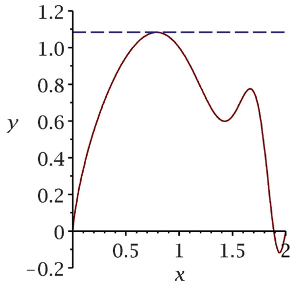


Fig. 5.24 A point where a function has a horizontal tangent as in Rolle's Theorem

Now assume that $f(x) > 0$ for some $x \in (a,b)$. See Fig. 5.24 on page 144.

There is a point $c \in (a,b)$ such that $f(x) \leq f(c)$ for all $x \in [a,c]$, since $f(x)$ attains a maximum on the closed bounded interval. (Recall Theorem 47 on page 140.) Since $f(x)$ is differentiable at c, by Theorem 46 on page 138, $f'(c) = 0$. This proves the result if there is a point where $f(x) > 0$.

The reasoning if there is a point where $f(x) < 0$ is similar and is left to the reader.

Remark 6. The hypothesis that $f(a) = f(b) = 0$ can be replaced by the assumption that $f(a) = f(b)$. If $f(a) = f(b) \neq 0$ we can replace $f(x)$ by $g(x) = f(x) - f(a)$ where $g(a) = g(b) = 0$ and $g'(x) = f'(x)$ for all $x \in (a,b)$.

We can now state and prove the Mean Value Theorem.

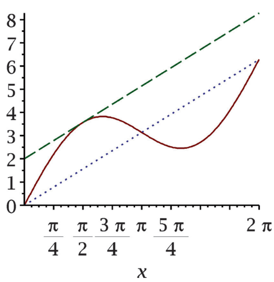


Fig. 5.25 A secant line and tangent line as in the Mean Value Theorem

Theorem 49 (Mean Value Theorem). *Let $[a,b]$ be a closed interval in $\mathbb{R}$ and let $f(x)$ be a function that is continuous on $[a,b]$ and differentiable on (a,b). Then there is a point $c \in (a,b)$ such that*

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

or, equivalently,

$$f(b) = f(a) + f'(c)\,(b - a).$$

The graph in Fig. 5.25 on page 144 illustrates that the tangent line to the graph, the dashed line, is parallel to the secant line, the line of dots.

Proof. The proof is an application of Rolle's Theorem to the function $f(x)$ minus the secant line for $f(x)$ between $(a, f(a))$ and $(b, f(b))$. Writing out the secant line gives the function $g(x) = f(a) + (f(b) - f(a))/(b-a)\,(x-a)$. Subtracting g from f gives

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b-a}\,(x-a).$$

We have $h(a) = h(b) = 0$ and hence, by Rolle's Theorem, Theorem 48 on page 143, there is a $c \in (a, b)$ such that

$$h'(c) = 0 = f'(c) - \frac{f(b) - f(a)}{b-a}.$$

Rewriting this equation gives the two forms of the Mean Value Theorem.

The remainder of this section is devoted to two applications of the Mean Value Theorem. The first involves the behavior of functions on intervals and the second illustrates how the Mean Value Theorem can be used for estimation.

Consider a function $h(x)$ that is defined and differentiable on an interval $\mathscr{I}$. The Mean Value Theorem tells us that for any x and y in $\mathscr{I}$ where $y > x$ we have

$$h(y) = h(x) + h'(c)\,(y-x).$$

If $h'(c) > 0$, we get that $h(y) > h(x)$. The following puts this in a concise form. The definition gives common terminology used for functions that increase or decrease on an interval.

Definition 19 (Increasing and decreasing functions). A function $f(x)$ is *increasing* (or *decreasing*) on an interval $\mathscr{I}$ if for all x and y with $y > x$ we have $f(y) \geq f(x)$ (or $f(y) \leq f(x)$).

A function $f(x)$ is *strictly increasing* (or *strictly decreasing*) on an interval $\mathscr{I}$ if for all x and y with $y > x$ we have $f(y) > f(x)$ (or $f(y) < f(x)$).

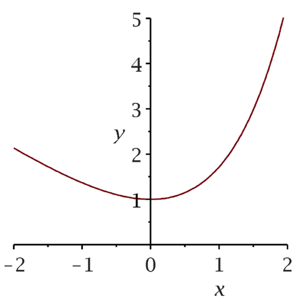


Fig. 5.26 A graph of $f(x) = e^x - x$ for Example 186

Theorem 50 (Derivatives and increasing or decreasing functions). *Let $f(x)$ be a function that is differentiable on an interval $\mathscr{I}$. If the derivative of $f(x)$ is nonnegative (or positive) on $\mathscr{I}$, then $f(x)$ is increasing (or strictly increasing) on $\mathscr{I}$.*

Let $f(x)$ be a function that is differentiable on an interval $\mathscr{I}$. If the derivative of $f(x)$ is nonpositive (or negative) on $\mathscr{I}$, then $f(x)$ is decreasing (or strictly decreasing) on $\mathscr{I}$.

There are many uses for this result.

Example 184. The function $f(x) = e^x$ is strictly increasing on $\mathbb{R}$. Since $f'(x) = e^x > 0$ for all $x \in \mathbb{R}$ the above theorem says that $f(x)$ is strictly increasing on $\mathbb{R}$.

Example 185. Consider $f(x) = x^3$. This function is strictly increasing on $\mathbb{R}$, but the theorem above cannot be applied directly since $f'(x) = 3x^2$ is 0 at $x = 0$. We must divide the proof into three cases: when $x < y \leq 0$, when $0 \leq x < y$, and when $x < 0 < y$. Note that $f'(c) = 3c^2 > 0$ if $c \neq 0$.

In the first and second cases we have $f(y) = f(x) + f'(c)(y-x) > f(x)$ since $f'(c) > 0$ and $y - x > 0$. In the third case we have, for some $c_1 \in (x, 0)$ and some $c_2 \in (0, y)$,

$$\begin{aligned} f(y) - f(x) &= f(y) - f(0) + f(0) - f(x) \\ &= f'(c_2)(y-0) + f'(c_1)(0-x) \\ &> 0. \end{aligned}$$

This shows that $f(x) = x^3$ is strictly increasing on $\mathbb{R}$.

This can also help when looking at critical points to find maxima and minima.

Assume that $f(x)$ is a function that is continuous on (a,b) and is differentiable on the intervals (a,c) and (c,b). If $f'(x) < 0$ on (a,c), then for all $x \in (a,c)$ we have $f(x) > f(c)$. If, in addition, $f'(x) > 0$ on (c,b), then for all $x \in (c,b)$ we have $f(x) > f(c)$. Combining these facts shows that $f(x)$ has a local minimum at c.

This can be stated, in the case when $f'(x)$ is a continuous function, as the following.

Lemma 2. *Assume that $f(x)$ has a continuous derivative on (a,b) with $f'(c) = 0$ for some $c \in (a,b)$. If $f'(x)$ changes from negative to positive at c, then $f(x)$ has a local minimum at c. If $f'(x)$ changes from positive to negative at c, then $f(x)$ has a local maximum at c.*

Example 186. Let $f(x) = e^x - x$. Then $f'(x) = e^x - 1$, which is 0 at $x = 0$. In addition, if $x < 0$ we have $e^x - 1 < 1 - 1 = 0$ and if $x > 0$ we have $e^x - 1 > 1 - 1 = 0$. Thus $f(x)$ has a local minimum at $x = 0$. In fact, it is a global minimum. (Why?) See Fig. 5.26 on page 145.

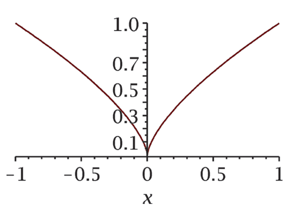


Fig. 5.27 A graph of $f(x) = x^{2/3}$ for Example 187

Remark 7. In Proposition 2 the assumption, for a maximum, that $f'(x)$ is continuous on (a,b) can be replaced with the assumptions that $f(x)$ is continuous on (a,b), c is a critical point of $f(x)$, $f'(x)$ exists and is greater than 0 on (a,c), and $f'(x)$ exists and is less than 0 on (c,b). Similar changes can be made to the assumptions for a minimum.

Example 187. Let $f(x) = x^{2/3}$. Then $f'(x) = 2x^{-1/3}/3$ and $x = 0$ is the only critical point of $f(x)$. Since $f'(x) < 0$ if $x < 0$ and $f'(x) > 0$ if $x > 0$, the point $x = 0$ is a local minimizer for $f(x)$. See Fig. 5.27 on page 146.

A simple but useful consequence of the Mean Value Theorem is the following result. Students are asked to prove the result in Exercise 9 on page 148.

Theorem 51. *Assume that a function $f(x)$ is continuous on an interval $[a,b]$ and that $f'(x) = 0$ on (a,b). Then $f(x)$ is constant on $[a,b]$.*

The converse of this statement, a constant function is continuous and has a zero derivative, is also true and is easily proven.

Example 188. If a function $h(z)$ is continuous on $[1,3]$, $h(2) = 4$ and $h'(x) = 0$ on $(1,3)$, then $h(z) = 4$ for all $z \in [1,3]$.

The Mean Value Theorem can also be used for estimations. Consider a function $f(x)$ that is differentiable on an interval $[a,b]$. If $f'(x)$ is between ℓ and u on $[a,b]$ then the Mean Value Theorem tells us that

$$f(a) + \ell\,(b-a) \le f(b) \le f(a) + u\,(b-a).$$

This follows directly from MVT, for some c_1 and c_2 in (a,b),

$$f(b) = f(a) + f'(c_1)\,(b-a) \ge \ell\,(b-a),$$

and

$$f(b) = f(a) + f'(c_2)\,(b-a) \le u\,(b-a).$$

Example 189. Let $f(x) = \sin(x)$ and let $a = 0$ and $b \in (0, \pi/3]$. Since the derivative of $\sin(x)$, $\cos(x)$, is between $1/2$ and 1 on $[0, \pi/3]$, we have

$$\frac{x}{2} \le \sin(x) \le x$$

for $x \in [0, \pi/3]$. See Fig. 5.28 on page 147.

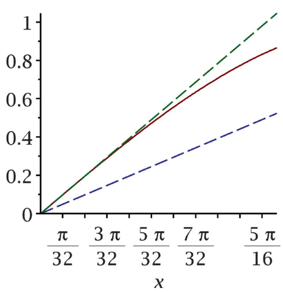


Fig. 5.28 The range of values allowed by the Mean Value Theorem in Example 189

Exercises

1. Find where each of the following functions is increasing and where it is decreasing.

(a) $f(x) = x^2 - 3x + 7$
(b) $g(x) = 2x^3 + 3x^2 - 36x - 18$
(c) $f(y) = y^3 + 3y^2 + 3y - 3$
(d) $g(y) = \tan(y)$
(e) $f(x) = \cos(2x)$
(f) $g(x) = \sec x$
(g) $f(y) = e^{y^2-3y+2}$
(h) $g(y) = \tan^{-1}\left(y^2 - 1\right)$

2. In each part of this problem a value for a function $f(x)$ at a is given, a range for the derivative of $f(x)$ on (a,b) is given, and a possible value for $f(b)$ is given. Assuming that $f(a)$ is correct and the range for $f'(x)$ is correct, can $f(b)$ be correct?

(a) $f(1) = 2$, $f'(x) \in [-1,1]$, $f(5) = 0$
(b) $f(0) = -1$, $f'(x) \in [0,2]$, $f(4) = -2$
(c) $f(-1) = 5$, $f'(x) \in [-2,0]$, $f(5) = 6$
(d) $f(e) = 2$, $f'(x) \in [0,1]$, $f(\pi) = 2.3$

3. Can the Mean Value Theorem be applied to the following functions on the given intervals? Justify your answer.

(a) $f(x) = |x|$ on $[-1,1]$.
(b) $f(x) = |x|$ on $[0,5]$.
(c) $g(x) = \lfloor x \rfloor$ on $[2,6]$.
(d) $h(z) = z^{15} - 34z^2 + 2z - 3$ on $[-2,10]$.
(e) $r(w) = w^3\cos(w)$ on $[-\pi, 2\pi]$.
(f) $s(t) = t^2|t|$ on $[-1,1]$.

4. Does a change of sign of the derivative show that the critical points of the following functions are local maxima or local minima?

(a) $f(x) = x^3$ on $[-1,1]$.
(b) $f(x) = x^4$ on $[-1,1]$.
(c) $g(x) = x^5 - 4x^3$ on $[-3,3]$.
(d) $h(z) = \cos(z) + z$ on $[-10,10]$.
(e) $r(w) = \sin(w) - \frac{w}{2}$ on $[-2\pi, 2\pi]$.
(f) $s(t) = |t|$ on $[-1,1]$.

5. Show that the point c between a and b guaranteed by the Mean Value Theorem is always $\frac{a+b}{2}$ for any a, b, and any quadratic function.

6. A car enters a toll road segment going at the speed limit of 70 mph through a “pay pass” toll station and then leaves the 70 mi segment going 70 mph through a “pay pass” toll station 1 h later. What can you say about the speed of the car over the hour?

7. A car stops at a toll booth and enters a toll road segment that is 70 mi long. The car exits the segment of the road 1 h later by stopping at a second toll booth. What can you say about the speed of the car over the hour?

8. The Mean Value Theorem does not hold for vector valued functions. Consider $\mathbf{r}(t) = \left(t^2, t^3\right)$ on the interval $[-1,1]$. Show that there is only one point c_1 where the first component function $r_1(t) = x^2$ satisfies $r_1(1) - r_1(-1) = r_1'(c_1)\left(1 - (-1)\right)$ and show that there are two points c_2 where $r_2(1) - r_2(-1) = r_2'(c_2)\left(1 - (-1)\right)$ with $r_2(t) = t^3$. Note that these three points are distinct to conclude that the Mean Value Theorem does not hold for vector valued functions.

9. Use the Mean Value Theorem to prove that if $h(z)$ is continuous on a closed interval $[a,b]$ with $h'(z) = 0$ for all $z \in (a,b)$, then $h(x) = h(y)$ for any $x, y \in [a,b]$. Use this to conclude that Theorem 51 on page 147 is true.

10. Use Theorem 51 on page 147 and Eq. 5.2 in Example 173 on page 135 to show that the following is true. If the position vector of motion for an object $\mathbf{r}(t)$ is always perpendicular to the velocity of the object, then the motion of the object is restricted to a circle in $\mathbb{R}^2$ or to a sphere in $\mathbb{R}^3$ centered at the origin.

11. Assume that $f(z)$ is continuous on $[a,b]$ and differentiable on (a,b) with $b > a$. Explain how and why the critical points of

$$h(z) = f(z) - f(a) - \frac{f(b) - f(a)}{b - a}(z - a)$$

in (a,b) are related to the points guaranteed to exist for $f(z)$ by the Mean Value Theorem.

5.6 Concavity and Extrema

The previous sections in this book have been concerned with how a function changes as described by the derivative. The ideas of functions increasing or decreasing on an interval and critical points have been demonstrated to be useful. However, these concepts only give us a very rough idea about the shape of the graph of a function. The following two examples give a better idea of how different the graphs of two functions can be with the same derivative sign information

Example 190. Let $f(x) = e^x$. Since $f'(x) = e^x$, the derivative is always increasing and the graph shows $f(x)$ increasing at an increasing rate. See Fig. 5.29a on page 149.

The next example is a function that is also always increasing on its domain It is, however, quite different.

Example 191. Let $h(x) = (x^2 - 1)/x^2$. Since $h'(x) = 2/x^3$, the derivative is always positive if $x > 0$ and the graph shows $f(x)$ is increasing toward the horizontal asymptote $y = 1$. See Fig. 5.29a on page 149.

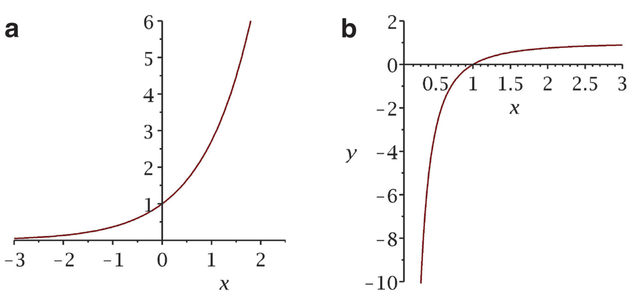
The derivative of $f'(x)$ from Example 190 is $f''(x) = e^x > 0$. This means that the slopes of tangent lines are increasing. See Fig. 5.30a on page 150.

On the other hand, the derivative of $h(x)$ from Example 191 is $h''(x) = -6/x^4 < 0$ on $(0, \infty)$. Here the slopes of the tangent lines are decreasing, See Fig. 5.30b on page 150.

Using the functions as above examples leads to the following definition.

Definition 20. A function $f(x)$ is *concave up* (or *concave down*) on an interval (a,b) if $f''(x) \geq 0$ (or $f''(x) \leq 0$) on (a,b).

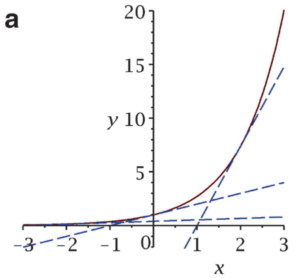
Remark 8. This is not the only definition used for the concepts of concave up, also known as convex, and concave down, also known as simply as concave. The other definitions are beyond this course and will not be considered.



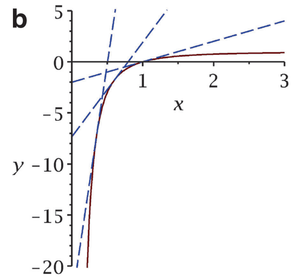
$f(x) = e^x$ and $f'(x) = e^x$ are both increasing

$h(x) = (x^2 - 1)/x^2$ is increasing and $h'(x) = 2/x^3$ is decreasing

Fig. 5.29



$f''(x) = e^x > 0$ and the slopes of the tangent lines are increasing



$h''(x) = -6/x^4 < 0$ and the slopes of the tangent lines are decreasing

Fig. 5.30

Example 192. The function $f(x) = x^3 - x$ is concave down on $(-\infty, 0)$ and concave up on $(0, \infty)$. See in Fig. 5.31a on page 151. Since $f''(x) = 6x$, we have $f''(x) < 0$ if $x < 0$ and $f''(x) < 0$ if $x > 0$.

A point (x, y) where a function changes concavity is called an *inflection point*. In this book we assume that the curve $\mathbf{g}(t) = (t, f(t))$ must have a tangent line at $t = x$ in order for $(x, f(x))$ to be an inflection point.

Example 193. Consider the function $f(x) = \sqrt[3]{x}$ in Fig. 5.31b on page 151. This curve can be parameterized either as $\mathbf{g}(t) = (t, \sqrt[3]{t})$ or as $\mathbf{h}(t) = (t^3, t)$. Since the second parametrization has a nonzero derivative everywhere, the curve has a tangent line at $(0, 0)$.

In addition, $\dfrac{d^2 f(x)}{dx^2} = -2/\left(9x^{5/3}\right)$. This means that $f(x)$ is concave down if $x > 0$ and concave up if $x < 0$. Since the graph of $f(x)$ has a tangent line at $(0,0)$, $(0,0)$ is an inflection point.

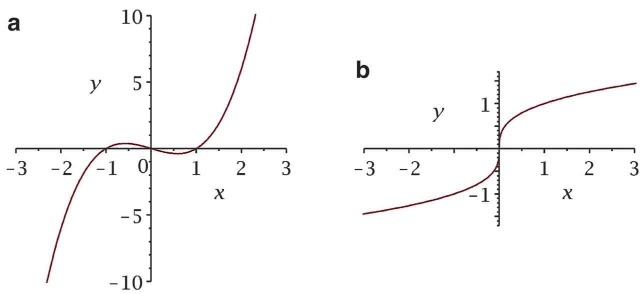
One of the main uses of concavity is for second order optimality conditions. What was shown earlier in Sect. 5.4 is that $\dfrac{df}{dx}(c) = 0$ is required (necessary) when $f(x)$ is differentiable and has a local extreme point at c. Assume that, in addition to $f'(c) = 0$, $f''(x) < 0$ around c. This means that $f(x)$ is concave down around c and $f'(x)$ changes from positive to negative at c. The plots in Fig. 5.32 on page 151 illustrates this situation. The conclusion is that $f(x)$ increase up to c and the decreases after c. Because of this, $f(x)$ has a local maximum at c. With a little more work we can prove the following result.

Theorem 52 (Second order optimality conditions). *Assume that $f(x)$ is differentiable on (a,b), $f'(c) = 0$ for some $c \in (a,b)$, and $f''(c)$ exists.*

(i) If $f''(c) > 0$, then $f(x)$ has a local minimum at c.

(ii) If $f''(c) < 0$, then $f(x)$ has a local maximum at c.

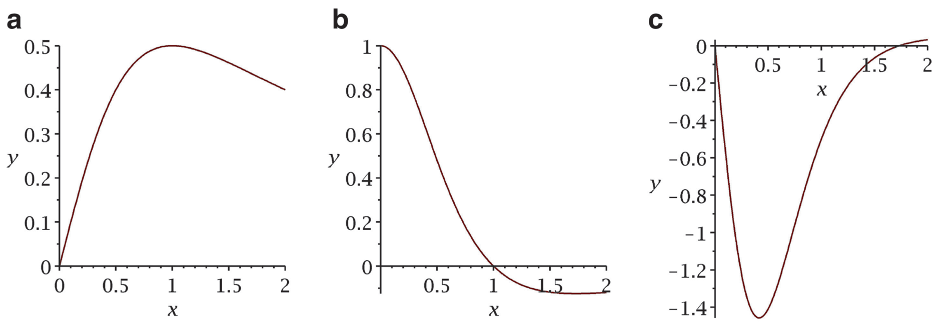
Proof. The proofs of the two cases are almost identical. Because of this, only the first case is considered.



The inflection point for $f(x) = x^3 - x$ at $x = 0$

The inflection point for $f(x) = x^{1/3}$ at $x = 0$

Fig. 5.31



A concave down $f(x)$ with a maximum at $x = 1$

The derivative of $f(x)$ with a zero at $x = 1$

The second derivative of $f(x)$ is negative around $x = 1$

Fig. 5.32

If $f'(c) = 0$ and $f''(c) > 0$, then, from the definition of the second derivative at c, there is an $a_1 \in (a,c)$ such that $f'(x) < 0$ on (a_1,c). There is also a $b_1 \in (c,b)$ such that $f'(x) > 0$ on (c,b_1). By the Mean Value Theorem we have $f(x) > f(c)$ on both (a_1,c) and (c,b_1). This means that $f(x)$ has a local minimum at c.

Example 194. Consider the function $f(x) = x/\left(1+x^2\right)$ (Fig. 5.33 on page 151).

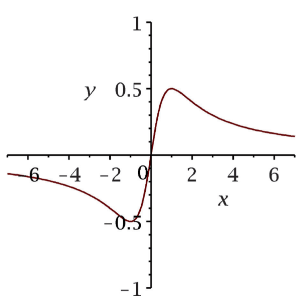


Fig. 5.33 A plot of $f(x) = x/\left(1+x^2\right)$ for Example 194

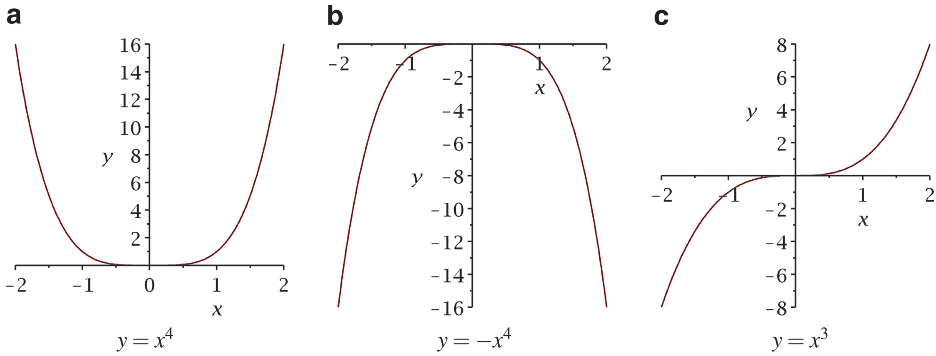


Fig. 5.34

This function has first derivative

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

and second derivative

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

The first derivative has zeros at $x = -1$ and $x = 1$. The second derivative has zeros at $x = -\sqrt{3}$, $x = 0$, and $x = \sqrt{3}$.

Since $f''(-1) = 1/2$ and $f''(x)$ is continuous at $x = -1$, $f(x)$ has a local minimum at $x = -1$. Similarly, since $f''(1) = -1/2$, $f(x)$ has a local maximum at $x = 1$.

The function $f''(x)$ is a continuous function with zeros at $x = -\sqrt{3}$, $x = 0$, and $x = \sqrt{3}$. Since $f''(x)$ is continuous it has constant sign on $(-\infty, -\sqrt{3})$, $(-\sqrt{3}, 0)$, $(0, \sqrt{3})$, and $(\sqrt{3}, \infty)$. Because $f''(-2) = -4/125$, $f''(x) < 0$ on $(-\infty, -\sqrt{3})$ and $f(x)$ is concave down on $(-\infty, -\sqrt{3})$. Similarly $f(x)$ is concave up on $(-\sqrt{3}, 0)$ and $(\sqrt{3}, \infty)$ and $f(x)$ is concave down on $(0, \sqrt{3})$. The points $\left(-\sqrt{3}, -\sqrt{3}/4\right)$, $(0,0)$, and $\left(\sqrt{3}, \sqrt{3}/4\right)$ are inflection points of $f(x)$.

Example 195. The case when $f'(a) = 0$ and $f'' = 0$ is not covered in Theorem 52 since we can have a local maximum, a local minimum or a point that is neither at such a point. The functions $f(x) = x^4$, $g(x) = -x^4$ and $h(x) = x^3$ all have first and second derivatives equal to 0 at $x = 0$. However, $f(x)$ has a minimum at $x = 0$, $g(x)$ has a maximum at $x = 0$ and $h(x)$ is always increasing and has neither a maximum nor a minimum at $x = 0$. The graphs of these functions are in Fig. 5.34 on page 152.

Exercises

1. Find where each of the following functions is concave up and where it is concave down. Where are the inflection points of the functions?

(a) $f(x) = x^2 - 3x + 7$

(b) $g(x) = 2x^3 + 3x^2 - 36x - 18$

(c) $f(y) = y^3 + 3y^2 + 3y - 3$

(d) $g(y) = \tan(y)$

(e) $f(x) = \cos(2x)$

(f) $g(x) = \sec x$

(g) $f(y) = e^{y^2 - 3y + 2}$

(h) $g(y) = \tan^{-1}\left(y^2 - 1\right)$

(i) $f(x) = \dfrac{x^3 - 1}{x^2 + 3x - 10}$

(j) $g(x) = \dfrac{x^2 - x - 6}{x^3 - 3x^2 + 4}$

(k) $f(x) = \left(\dfrac{x}{1+x^2}\right)^{\frac{2}{3}}$

(l) $h(w) = \tan\left(\sin(w) - \dfrac{1}{2}\right)$

2. Use the second derivative test to classify the critical points of the following functions, if possible.

(a) $g(x) = 2x^3 + 3x^2 - 36x - 18$

(b) $f(y) = y^3 + 3y^2 + 3y - 3$

(c) $g(x) = \dfrac{x^3 - 1}{x^2 + 3x - 10}$

(d) $f(y) = \dfrac{y^2 - y - 6}{y^3 - 3y^2 + 4}$

(e) $g(x) = |x^2 - 2x - 3|$

(f) $f(y) = \cos(2y)$

(g) $h(w) = \lfloor |w| \rfloor - w^2$

3. Assume that a function $f(x)$ is defined around $x = c$ and that $f'(c) = 0$, $f''(c) = 0$, and $f^{(3)}(c) \neq 0$. If $f^{(3)}(c)$ is continuous around $x = c$, explain why $f(x)$ has neither a local maximum nor a local minimum at $x = c$. (Hint: Look at $h(x) = x^3$ around $x = 0$.)

5.7 Calculus and Graphs of Functions

Using the information from the first and second derivatives it is easier to understand the graph of a function that has two (continuous) derivatives. The first derivative tells us where a function is increasing and where it is decreasing. From the second derivative we can tell where a function is concave up or concave down, It also gives inflection points. Given this information, we can get a good picture of the behavior of a function.

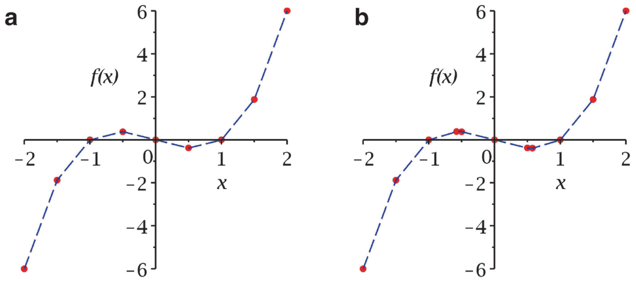
Each piece of information we obtain gives us a better idea of the appearance of the graph. The process of sketching the graph of a function is a step by step process of obtaining information. Typically, we start with the basic information about a function obtained without using calculus. The information that we find may includes the zeroes of the function, the values of the function at several points, and whether the function is odd or even.

Example 196. Consider the familiar function $f(x) = x^3 - x$. This function has zeroes at $x = -1, 0,$ and 1. In addition, this is an odd function. We can also get the values of f at several points as in Table 5.6 on page 153. With this information we can plot the points and get a possible graph by connecting the points with line segments. See Fig. 5.35a on page 154.

The first derivative gives intervals where a function is increasing or decreasing and the critical points. This is used to refine the graph.

x	-2	$-3/2$	-1	$-1/2$	0	$1/2$	1	$3/2$	2
$f(x)$	-6	$15/8$	0	$3/8$	0	$-3/8$	0	$15/8$	6

Table 5.6 Some function values for $f(x) = x^3 - x$.



a A possible graph for $f(x) = x^3 - x$ from a few points

b A possible graph for $f(x) = x^3 - x$ from a few points with the critical points added

Fig. 5.35

x	-2	$-3/2$	-1	$-1/\sqrt{3}$	$-1/2$	0	$1/2$	$1/\sqrt{3}$	1	$3/2$	2
$f(x)$	-6	$15/8$	0	$2/9\sqrt{3}$	$3/8$	0	$-3/8$	$-2/9\sqrt{3}$	0	$15/8$	6

Table 5.7 The points for plotting $f(x) = x^3 - x$ with the critical points added

Example 197. Continuing with the function $f(x) = x^3 - x$, we have $f'(x) = 3x^2 - 1$. Since this is continuous everywhere and is zero at $x = \pm 1/\sqrt{3}$, we can calculate that $f'(x) > 0$ on $\left(-\infty, -1/\sqrt{3}\right)$ and $\left(1/\sqrt{3}, \infty\right)$. Also, $f'(x) < 0$ on $\left(-1/\sqrt{3}, 1/\sqrt{3}\right)$. From this information we can conclude that $f(x)$ is increasing on $\left(-\infty, -1/\sqrt{3}\right)$ and $\left(1/\sqrt{3}, \infty\right)$. and it is decreasing on $\left(-1/\sqrt{3}, 1/\sqrt{3}\right)$. This means that $f(x)$ has a local maximum at $x = -1/\sqrt{3}$ and a local minimum at $x = 1/\sqrt{3}$.

Adding the two critical point to the list of points used from plotting gives Table 5.7 on page 154. See Fig. 5.35b on page 154.

The information from the first derivative excludes the graph in the Fig. 5.36a on page 155 since there are only two critical point.

It does not, however, exclude the shape in Fig. 5.36b on page 155. In this graph there are only two critical points and the regions where $f(x)$ is increasing and decreasing match the first derivative information.

Now we can add information obtained from the second derivative.

Example 198. Continuing with the function $f(x) = x^3 - x$, we have $f''(x) = 6x$. Since this is continuous everywhere and is zero at $x = 0$, we can calculate that $f''(x) < 0$ if $x < 0$ and $f''(x) > 0$ if $x > 0$. From this information we can conclude that $f(x)$ is concave down on $(-\infty, 0)$ and concave up on $(0, \infty)$. There is also an inflection point at $x = 0$.

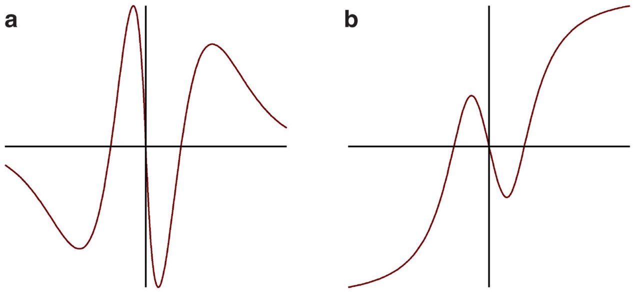
This information leads us to Fig. 5.37 on page 155.

The main thing that is missing from the pictures here are asymptotes. There are two types of asymptotes that are commonly used for graphing, horizontal and vertical asymptotes. Horizontal asymptotes tell us about the behavior of functions $f(x)$ as $x \to \pm\infty$.

Definition 21. Let $f(x)$ be a function from $\mathbb{R}$ to $\mathbb{R}$.

(I) A horizontal line $y = a$ is a *horizontal asymptote* for $f(x)$ if

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



Not a possible graph for $f(x) = x^3 - x$ since it has 4 critical points

A possible graph for $f(x) = x^3 - x$ without concavity information

Fig. 5.36

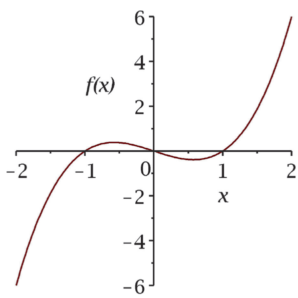


Fig. 5.37 The graph of $f(x) = x^3 - x$ using full information

(II) A vertical line $x = b$ is a *vertical asymptote* for $f(x)$ if

$$\lim_{x\to b^-} f(x) = \pm\infty \qquad \text{or} \qquad \lim_{x\to b^+} f(x) = \pm\infty.$$

The next example gives the ideas of these definitions.

Example 199. Consider the function $g(x) = x^2/(x^2-4)$. It has a horizontal asymptote $y = 1$ since

$$\lim_{x\to\infty} \frac{x^2}{x^2-4} = \lim_{x\to\infty} \frac{1}{1-\frac{4}{x^2}}$$
$$= 1.$$

The limit as $x \to -\infty$ is also 1.

The function also has vertical asymptotes at $x = -2$ and $x = 2$ since there are infinite limits for $f(x)$ as x approaches ± 2 from either side. First note that $f > 0$ if $x \in (-\infty, -2) \cup (2, \infty)$ and $f(x) < 0$ if $x \in (-2, 2)$. Since

$$\lim_{x\to 2^+} x^2 = 4 \qquad \text{and} \qquad \lim_{x\to 2^+} x^2 - 4 = 0,$$

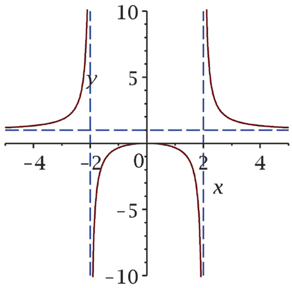


Fig. 5.38 The graph of $g(x) = x^2/(x^2 - 4)$ with asymptotes

we have

$$\lim_{x \to 2^+} f(x) = \infty.$$

Using similar reasoning we can show that

$$\lim_{x \to 2^-} f(x) = -\infty, \quad \lim_{x \to -2^+} f(x) = -\infty \quad \text{and} \quad \lim_{x \to -2^-} f(x) = \infty.$$

See Fig. 5.38 on page 156.

Remark 9. Rational functions have asymptotes where there is a zero of the denominator of the reduced form of the rational function. For example

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

does not have an asymptote at $x = 1$ since it is the same as $g(x) = x + 1$ if $x \neq 1$.

On the other hand, the function

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

has an asymptote at $x = 1$ since it is the same as $k(x) = 1/(x - 1)$ except at $x = -1$.

Combining all of the basic information we get using the above examples and definitions, we can get a good understanding of the graph of a function. The following set of steps can be used to understand the graphs of functions. Often not all of the steps are necessary, or even useful.

Steps for Understanding Graphs of Functions.

(I) Find all zeroes and y intercepts for the function.

(II) Find all asymptotes.

(III) Find the first derivative.

- (i) Find the zeroes of the derivative and all other critical points.
- (ii) Find all regions where the function is increasing and decreasing.

x	$-8.$	$-6.$	-4.11	$-2.$	$-.500$	$.500$	1.50	$2.$	3.50	$5.$	$7.$	$9.$
$f(x)$	$.926$	$.914$	$.899$	$.889$	$.926$	1.29	$-.200$	$0.$	-1.69	2.25	1.39	1.22

Table 5.8 All of the points to plot

(IV) Find the second derivative and all points where the second derivative is 0.

(i) Find the regions where the function is concave up and concave down.

(ii) Find all inflection points.

(V) Classify all critical points.

Another example will illustrate the whole process.

Example 200. Let $f(x) = (x^2-4x+4)/(x^2-5x+4)$. Using the quotient rule gives a derivative of

$$f'(x) = -\frac{x^2-4}{(x^2-5x+4)^2}.$$

This means that $f(x)$ has two critical points, $x = \pm 2$.

Since $f(x)$ is a rational function that is not defined at $x = 1$ and $x = 4$, $f'(x)$ can only change signs at $x = -2, 1, 2,$ and 4. Checking one value of the derivative in each of the intervals $(-\infty,-2)$, $(-2,1)$, $(1,2)$, $(2,4)$, and $(4,\infty)$ will indicate where $f(x)$ is increasing and decreasing. The values of $f'(x)$ at $x = -3, 0, 3/2, 3,$ and 5 are $-5/784, 1/4, 28/25, -5/4, -21/16$. This means that $f(x)$ is increasing on $(-2,1)$ and $(1,2)$ and $f(x)$ is decreasing on $(-\infty,-2)$, $(2,4)$ and $(4,\infty)$. This gives basic shape information and tells us that $f(x)$ has a local minimum at $x = -2$ and a local maximum at $x = 1$

The second derivative of $f(x)$ is

$$\frac{d^2 f(x)}{dx^2} = \frac{2\,(x^3-12x+20)}{(x^2-5x+4)^3}.$$

This has a single root at $x_0 = -2^{4/3} - 2^{2/3}$. Since $f''(0) = 5/8$, this means that $f''(x)$ is negative if $x < x_0$ and $f''(x)$ is positive if $x > x_0$. From this we conclude that $f(x)$ is concave down on $(-\infty, x_0)$, $f(x)$ is concave up on $(x_0, 1)$, and $f(x)$ has an inflection point at $x_0 \approx 4.10724$. In addition, since $f''(2) = -8/(2^3) = -1 < 0$, $f(x)$ is concave down on $(1,4)$. And $f''f(5) = 17/800 > 0$ means that $f(x)$ is concave up on $(5,\infty)$.

There is a horizontal asymptote at $y = 1$ since $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = 1$. Also, there is a vertical asymptote at $x = 1$ since for the numerator of $f(x)$, we have $\lim_{x\to 1} x^2-4x+4 = 1$ and for the denominator of $f(x)$ on has $\lim_{x\to 1} x^2-5x+4 = 0$. In a similar manner, $f(x)$ also has a vertical asymptote at $x = 4$.

We also want to plot the points corresponding to the inflection and critical points. In addition, plotting the function at some other points will help. The Table 5.8 on page 157 gives the points to plot accurate to three decimal places.

All of the information is confirmed by the graph of $f(x)$ on $[-5,5]$, Fig. 5.39 on page 158.

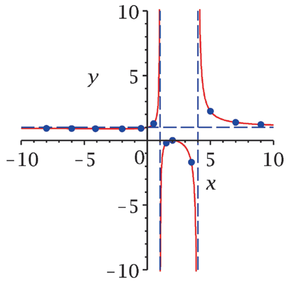


Fig. 5.39 The graph of $f(x) = (x^2 - 4x + 4)/(x^2 - 5x + 4)$ as plotted using calculus

Exercises

1. Find all of the relevant information and sketch the graph of each function by hand.

(a) $f(x) = x^2 - 3x + 7$

(b) $g(x) = 2x^3 + 3x^2 - 36x - 18$

(c) $f(y) = y^3 + 3y^2 + 3y - 3$

(d) $g(y) = \tan(y)$

(e) $f(x) = \cos(2x)$

(f) $g(x) = \sec x$

(g) $f(y) = e^{y^2 - 3y + 2}$

(h) $g(y) = \tan^{-1}\left(y^2 - 1\right)$

(i) $f(x) = \dfrac{x}{x^2 - 4}$

(j) $g(x) = \dfrac{x^2}{x^2 - 1}$

(k) $f(x) = \dfrac{x^3 - 1}{x^2 + 3x - 10}$

(l) $g(x) = \dfrac{x^2 - x - 6}{x^3 - 3x^2 + 4}$

(m) $f(y) = \dfrac{y^3 - y}{y^2 - 4}$

(n) $g(y) = y^4 + 2y^3 - 7y^2 - 8y + 12$

(o) $f(x) = x^2 + \dfrac{4}{1 + x^2}$

(p) $g(x) = \dfrac{1}{x^2} - \dfrac{16}{1 + x^2}$

(q) $f(x) = \exp\left(-x^2\right)$

(r) $g(x) = \ln\left(1 + x^2\right)$

2. Consider the function $f(x) = (x^3 - x) \cdot \left(1 - \frac{x}{1{,}000}\right) \cdot \left(1 + \frac{x}{1{,}000}\right)$. Collect the information using algebra and calculus that you would use to graph this function. (You will need to find numerical approximations for many of the points of interest.) Explain why trying to plot a graph of this function that illustrates all of the information you have cannot be done on a 1920 by 1080 pixel screen.

3. Give an example showing that a function can be concave up on its entire domain and yet never achieve a minimum value.

5.8 Applied Optimization

This section is devoted to uses of the derivative for finding maximum and minimum values of functions from applications. This is often called optimization. In order to use calculus in a situation where we want to find the best of something, the optimal value, we first model the situation using a function that represents the thing to be optimized. This function should have as its domain the variable we can control. The output is the quantity to be maximized or minimized. We can, if the situation is nice, use calculus to find the optimal point. The ideas in Sect. 5.4 are used to find maxima and minima for applied problems.

Recall Theorem 47 on page 140.

Theorem 53. *Assume that $f(x)$ is a continuous function on a closed bounded interval $[a,b]$. Then there are points c and d in $[a,b]$ such that $f(c) \geq f(x)$ for all $x \in [a,b]$ and $f(d) \leq f(x)$ for all $x \in [a,b]$.*

What was shown before was that if a function $f(x)$ is defined on an interval, c is a point inside the interval, and $f(x)$ has a local maximum or minimum at c, then c is a critical point of $f(x)$. Combining this with the theorem above gives the following result. There are many situations where we have limits on the resources available. The mathematical translation is that the collection of possible inputs to a function to be optimized is bounded and, often, closed.

Theorem 54 (Extreme point theorem). *Let $f(x)$ be a continuous function on a closed bounded interval $[a,b]$. The maximizing and minimizing points for $f(x)$ on $[a,b]$ are either critical points or endpoints of the interval.*

Proof. The proof is very simple given the information we have. First, there is a point c where $f(x)$ takes on its maximum (minimum) value. If this point is not an endpoint of the interval, it must be a local maximum (minimum). This means it is a critical point.

Theoretically, it is easy to use this theorem. We should simply find the critical points of the function on (a,b), compute the values of $f(x)$ at the critical points and the endpoints of the interval, and chose the maximum and minimum. The examples from Sect. 5.4 illustrate how this works.

An example of an application of optimization is the problem of enclosing an area inside a perimeter with a fixed length. This will be seen in a number of different problems. The following is one of the simplest examples.

Example 201. Find the dimensions the largest rectangular area that can be enclosed in a perimeter of length 30 cm, see Fig. 5.40 on page 160.

Let the length of the region be denoted by h and denote the width of the region by w. Since the length of the perimeter is 30 cm, the relationship between the width and the length is $30 = 2h + 2w$. Note that w and h must both be nonnegative and neither can be larger than 15 cm. This means that both h and w must be in the interval $[0,15]$.

The area of the region is $A = hw$. Since all of our results on finding optimal points require a function of one variable, either the h or the w must be eliminated from the area function.

Solving for h in terms of w gives $h = 15 - w$. Substituting this into the area expression gives

$$A(w) = 15w - w^2.$$

The derivative of $A(w)$ is $A'(w) = 15 - 2w$. The critical point is where $A'(w) = 0$. Solving gives $x = 7.5$. Since $A(0) = A(15) = 0$, the maximum area is at the critical point. This means $w = 7.5$ and $h = 15 - 7.5 = 7.5$. The largest area is when the region is a square.

Another common example is finding the dimensions of a box that maximize the volume.

Example 202. Assume that a box without a top is to be made from a sheet of cardboard that measures 75 cm by 60 cm. This box is to be made by cutting and folding the cardboard. What are the dimensions of the box that will maximize the volume of the box? See Fig. 5.41 on page 160.

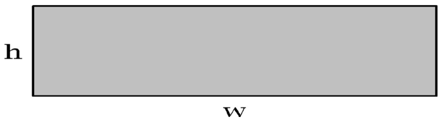


Fig. 5.40

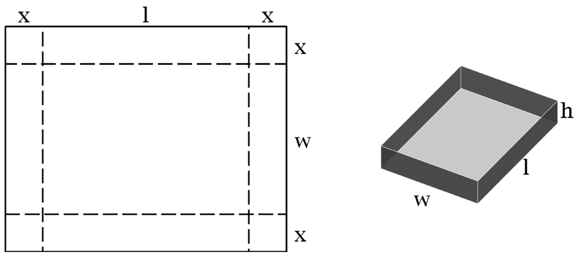


Fig. 5.41

To make the problem easier it is assumed that the cardboard does not have any thickness. It is also assumed that the folds will be parallel to the sides of the cardboard. (Can you explain why this assumption is appropriate?) It is also assumed that four folds are made, all the same distance away from an edge of the sheet of cardboard. (Again, can you explain why this assumption is appropriate?) Often this problem is stated so that the squares at each corner are removed. In the manufacture of many boxes these squares are used as gluing tabs, so here they remain attached.

If x is the distance of the fold lines from the edges, the length and width of the base of the box are $\ell = 75 - 2x$ and $w = 60 - 2x$. For the rest of the problem units will be omitted from the expressions, except for the final answers. The height of the box is $h = x$. The volume of the box is

$$V(x) = x(75 - 2x)(60 - 2x) = 4x^3 - 270x^2 + 4{,}500x.$$

To find the critical points we set

$$\frac{dV(x)}{dx} = 12x^2 - 540x + 4{,}500 = 0.$$

This gives

$$x = \frac{45 + 5\sqrt{21}}{2} \approx 34, \frac{45 - 5\sqrt{21}}{2} \approx 11.$$

The choice of $x = \frac{45+5\sqrt{21}}{2}$ is physically impossible, so $x = \frac{45-5\sqrt{21}}{2}$. (Why is $x = (45 + 5\sqrt{21})/2$ cm physically impossible?). The dimensions of the box are $\ell = 30 + 5\sqrt{21}$ cm, $w = 15 + 5\sqrt{21}$ cm, and $h = (45 - 5\sqrt{21})/2$ cm.

The plot of $V(x)$ in Fig. 5.42 on page 161 shows that this is a reasonable x value for maximizing V.

Remark 10. This working of the problem is a common way for students to approach the problem. It is better if we first use the information from the problem to find the restrictions on the variables x, l, and w. All of these variables are nonnegative and these variables must satisfy

$$0 \le 2x + w \le 60 \quad \text{and} \quad 0 \le 2x + l \le 75.$$

In particular, these conditions imply that $0 \le x \le 30$. Thus we should maximize the function $V(x) = x(75 - 2x)(60 - 2x)$ on $[0, 30]$.

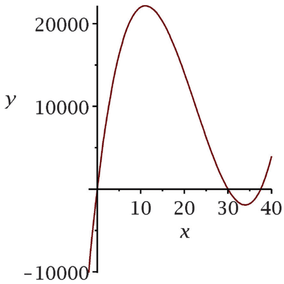


Fig. 5.42

Optimization problems arise in a large number of applications. One of those areas is economics. A common goal in economics is to try and make the largest profit possible. Businesses often keep track of their income and their costs. What we often try to maximize the is profit, the income minus the costs.

In economics the rate of change of income with respect to the amount produced is called the *marginal income* (marginal revenue) and the rate of change of cost is called the *margin cost*. Assume a company is producing a product, p. If the income as a function of the amount of product produced is $I(p)$ and the cost of production as a function of the amount of product produced is $C(p)$, the profit as a function of the amount of product produced is $P(p) = I(p) - C(p)$. Assuming that the functions are all differentiable, the maximum profit occurs when

$$0 = \frac{dP(p)}{dp} = \frac{dI(p)}{dp} - \frac{dC(p)}{dp}.$$

This says that the maximum is when

$$\frac{dI(p)}{dp} = \frac{dC(p)}{dp},$$

or when the marginal cost equals the marginal income.

Example 203. Assume that the selling price of a product is given by $100 - p$ if p units are sold. This gives an income of $I(p) = p(100 - p)$ if p units of the product are produced. Assume that the cost of producing the product is $p + 100/p$ per unit of p produced when p units are produced. This gives a total cost $C(p) = p(p + 100/p)$ for producing p units of the product.

To find the amount of the product that maximizes profit we set the marginal cost equal to the marginal income,

$$100 - 2p = 2p.$$

This means that the maximum profit is when $p = 25$ units. The profit at that point is

$$\begin{aligned} P(25) &= I(25) - C(25) \\ &= 25 \cdot 75 - 25 \cdot (25 + 4) \\ &= 900. \end{aligned}$$

The next example is another common example. For many people the answer we get goes against their intuition.

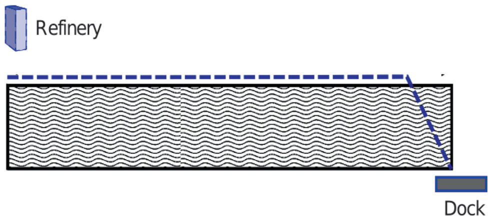


Fig. 5.43

Example 204. Consider the problem of building a pipeline from an oil port along a canal to a refinery 10 mi away. It is assumed that the port and the refinery are on opposite sides of the canal that is $1/2$ mi wide. The pipeline can be laid along the edge of the canal or under the water. The pipeline will run along a straight line under the canal and then along the straight edge of the canal of the refinery, see Fig. 5.43 on page 162. If it costs $1\frac{1}{2}$ times as much to run the pipeline under the water as on land, where should the pipeline come out of the canal?

Solution: Assume the cost of the pipeline on land is 1 unit per mile. Let x be the distance down the canal from the oil port where the pipeline comes out of the water. The cost of building the pipeline is

$$C(x) = (10-x) + 1.5\sqrt{(1/2)^2 + x^2}.$$

The derivative of this function is

$$\frac{dC(x)}{dx} = -1 + 1.5\frac{x}{\sqrt{(1/2)^2 + x^2}}.$$

Since this is differentiable everywhere and we want $x \in [0,10]$ (Why?), we only need to check for any critical points where $x \in (0,10)$ and the endpoints of the interval, $x = 0$, and $x = 10$.

Setting the derivative equal to zero and doing a little simplification gives

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

Squaring both sides and collecting all x^2 terms on one side yields

$$\frac{5x^2}{4} = \frac{1}{4}.$$

Since x is nonnegative we get the critical point

$$x = \frac{\sqrt{5}}{5}.$$

Plugging the x values into $C(x)$ gives

$$\begin{aligned} C(0) &= 10\frac{3}{4}, \\ C\left(\frac{\sqrt{5}}{5}\right) &= 10 - \sqrt{5}5 + \frac{9\sqrt{5}}{20} \approx 10.56, \quad \text{and} \\ C(10) &= \frac{3}{2}\sqrt{\frac{401}{4}} \approx 15.02. \end{aligned}$$

From this it is clear that the pipeline should come out of the water $\sqrt{5}/5$ mi down from the oil port.

We can also find the minimum distance between two people who are moving.

Example 205. Two roads meet at the origin. One is north-south and the other is east-west. A woman starts walking north at $t = 0$ from 2 km south of the origin going north at 2 km/h and a man starts walking east at $t = 0$ from 5 km west of the origin going east at 3 km/h. What is the smallest distance between the two people and when are they that distance apart?

A frequent, but incorrect, assumption is that the minimum distance occurs when one of the people is at the origin. Here is what happens when we make that assumption. If the position of the woman is given by $\mathbf{w}(t) = (0, y(t))$ and the position of the man is given by $\mathbf{m}(t) = (x(t), 0)$ the distance between them is given by

$$s(t) = \sqrt{x(t)^2 + y(t)^2}.$$

Taking the derivative with respect to time yields

$$\frac{ds}{dt} = \frac{x(t)x'(t) + y(t)y'(t)}{\sqrt{x(t)^2 + y(t)^2}}. \tag{5.3}$$

If, for example, $y(t_0) = 0$ and the distance between the people is at a minimum, we must have $x(t_0) = 0$, $x'(t_0) = 0$, or $\sqrt{x(t)^2 + y(t)^2} = 0$. These are the only conditions that make t_0 a critical point for $s(t)$. Since, in our case, none of the three conditions are satisfied when $y(t_0) = 0$, the distance is not minimized when the woman is at the origin.

Now for the correct solution. In this situation $y(t) = -2 + 2t$ and $x(t) = -5 + 3t$. Plugging these into Eq. (5.3) gives

$$\begin{aligned}\frac{ds}{dt} &= \frac{3(3t-5) + 2(2t-2)}{\sqrt{(3t-5)^2 + (2t-2)^2}} \\ &= \frac{13t - 19}{\sqrt{13t^2 - 38t + 29}}.\end{aligned}$$

Since the distance between the people $s(t) = \sqrt{13t^2 - 38t + 29}$ is never 0, the only critical point occurs when $13t - 19 = 0$ or when $t = 19/13$. The distance at this time is $s(19/13) = 4\sqrt{13}/13$.

The last problem is an easier specific case of a more general problem, finding the minimum distance between two objects that are moving along different paths. This distance problem was considered in Example 173 in the section on related rates. We can use the derivative from that example to find an optimality condition for the problem of finding the minimum distance between two objects with positions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, where $\mathbf{x}_i(t)$ is differentiable. Let $\mathbf{r}(x)$ be the position of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$, $\mathbf{r}(t) = \mathbf{x}_2(t) - \mathbf{x}_1(t)$.

The distance between the objects is $s(t) = \sqrt{\mathbf{r}(t) \cdot \mathbf{r}(t)}$. The derivative of this function is

$$s'(t) = \frac{\mathbf{r}(t) \cdot \mathbf{r}'(t)}{\sqrt{\mathbf{r}(t) \cdot \mathbf{r}(t)}}.$$

This means that, in order for the distance to be minimized, we must have $\mathbf{r}(t) \cdot \mathbf{r}'(t) = 0$, or the velocity of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$ is perpendicular to the radius vector from $\mathbf{x}_1$ to $\mathbf{x}_2$. The idea is illustrated in Fig. 5.44 on page 164.

As the next example shows, using basic calculus we can sometimes handle much more complex problems. In this situation we must divide the problem up into different phases to use

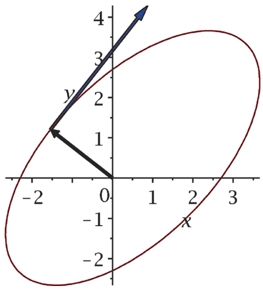


Fig. 5.44

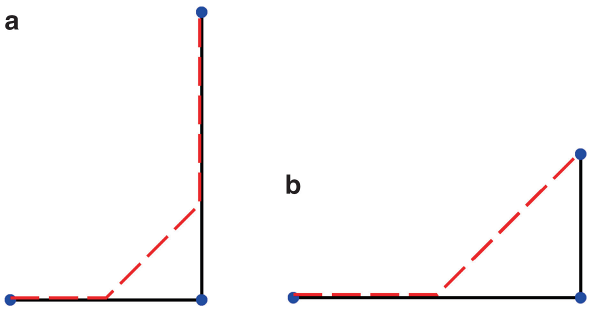


Fig. 5.45

basic calculus, Here there are two input variables x and y. We can find the optimal y for a fixed x in terms of x. We can then use that information to create a function that only depends on x to optimize. Although contrived, this example should give the flavor of these problems.

Example 206. A farmer is on a north-south road that intersects with an east-west road 3 km south of where the farmer has run out of fuel. The farmer's barn is 2 km west of the north-south road on the east-west road. The farmer must walk to the barn and wants to do it as quickly as possible. The farmer can walk at 5 km/h on the north-south road, 5 km/h on the east-west road and 3 km/h across the field. The farmer will only walk along straight lines. What route should the farmer take? See Fig. 5.45a on page 164.

First consider the case where the farmer is x miles north of the intersection and will first head across the field to the east-west road and then go to the barn. After finding the fastest way to do this for a fixed x, the time to get to x km north of the intersection is added to this quantity to get the time as a function of x. Minimizing this function gives the minimum time required. See Fig. 5.45b on page 164.

The time required to get from x km north of the intersection across the field to a point y km west of the intersection and then down the east-west road to the barn is

$$C_1(y) = \frac{2-y}{5} + \frac{\sqrt{x^2+y^2}}{3}.$$

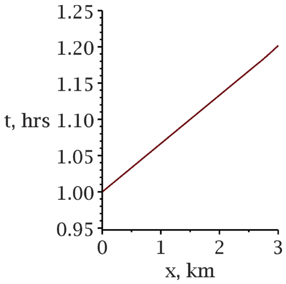


Fig. 5.46

The derivative of this function is

$$\frac{dC_1(y)}{dy} = -\frac{1}{5} + \frac{y}{3\sqrt{x^2+y^2}}.$$

A first simplification gives

$$x^2 + y^2 = \frac{25y^2}{9}.$$

Remembering that $y \in [0,2]$ and solving for y gives

$$y = \frac{3x}{4}$$

when $x \in [0, 8/3]$. This is the minimum for these x values. The minimum is at $y = 2$ when $x \in [8/3, 3]$.

From this we get that the best time for a given x value is

$$\begin{aligned} C(x) &= \frac{8-3x}{20} + \frac{5x}{12} + \frac{3-x}{5} \\ &= 1 + \frac{x}{15} \end{aligned}$$

when $x \in [0, \frac{8}{3}]$ and

$$C(x) = \frac{3-x}{5} + \frac{\sqrt{4+x^2}}{3}$$

when $x \in [8/3, 3]$ as illustrated in Fig. 5.46.

Since the derivatives of both parts of the $C(x)$ function are positive and $C(x)$ is continuous, the minimum of $C(x)$ occurs when $x = 0$ and the minimum time is 1 h.

Exercises

1. You are to make a box without a lid from a 100 cm by 35 cm piece of cardboard. The box is to be made in the same manner as Example 202 from this section, Sect. 5.4. What are the dimensions that will maximize the volume? What is the maximum volume?

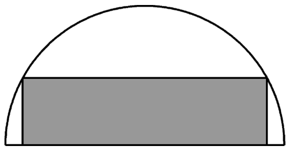


Fig. 5.47 A rectangle inscribed in a half circle

Fig. 5.48 The cylinder for Exercise 3

2. A rectangle is to be inscribed in a half circle of radius 1 ft. If the rectangle is as in Fig. 5.47 what dimensions will give the largest area?
3. A cylindrical container with caps on both ends and a volume of 1 l cm^3 is to be made from material that costs \$ 0.02 cm^{-2}. What is the minimum cost to make the container? See Fig. 5.48 on page 166.
4. A cylindrical container with caps on both ends and a volume of 1 l cm^3 is to be made from material that costs \$ 0.02 cm^{-2} for the cylinder and \$ 0.03 cm^{-2} for the top and bottom. What is the minimum cost to make the container? See Fig. 5.48 on page 166.
5. Redo the previous problem assuming that the disks for the caps are cut from squares and that one must pay for all of the material in the squares.
6. There is a 6 km long canal that is 1 km wide. At one end on the south shore there is an oil terminal. At the other end of the canal on the north side there is a refinery. The objective is to build a pipeline from the oil terminal to the refinery that has the smallest cost. The pipeline will go on a straight line from the oil terminal across the canal and then along the straight north shore of the canal to the refinery. It costs twice as much to put pipe under the water as to lay pipe on land. What path should the pipeline take? See Fig. 5.49 on page 166.

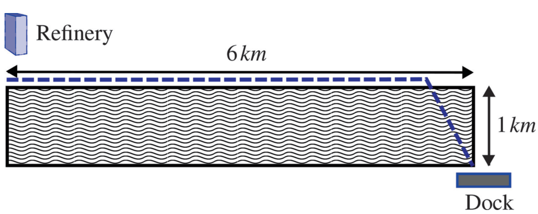


Fig. 5.49 The canal from Exercise 6

7. A 1,000 kg mass is between two masses of 350,000 and 600,000 kg whose centers are 100,000 m apart. At what distance from the center of the 600,000 kg mass will the sum of the norms of the gravitational forces from the 350,000 and 600,000 kg masses acting on the 1,000 kg mass be minimized? See Fig. 5.50 on page 167.
8. A holder is needed for a cylindrical glass lamp shade. The holder will consist of two wire rings, for the top and the bottom, and five equally spaced vertical wires connecting the two

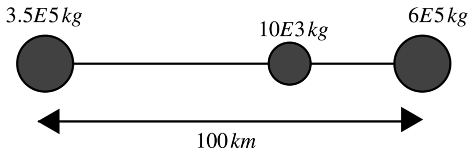


Fig. 5.50 The mass arrangement for Exercise 7

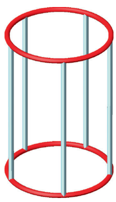


Fig. 5.51 Lamp shade holder for Exercise 8

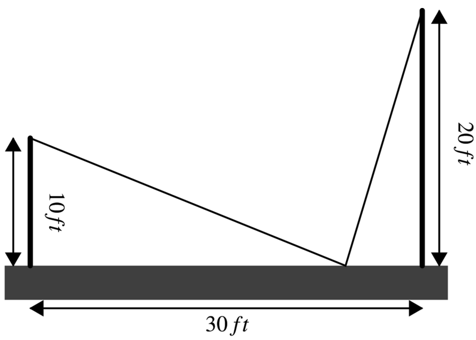
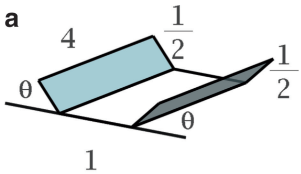


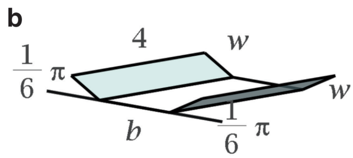
Fig. 5.52 The wire between poles for Exercise 9

rings. If there is 2 m of wire, what is the largest volume that can be enclosed by the frame? See Fig. 5.51 on page 167.

9. A wire is to run from the top of a 10 ft pole to the ground and then to the top of a 20 ft pole that is 30 ft from the first pole. Assuming that the wire is stretched tight so that the lengths are straight lines, what is the minimum amount of wire required? See Fig. 5.52 on page 167.
10. According to Fermat's principle a beam of light travels along the path that minimizes the time it takes to go from A to B. Light is traveling from a source A to a point B by reflecting off a mirror. Assume that A is 15 in. from the mirror, B is 10 in. from the mirror, and the closest points on the mirror to A and B are 10 in. apart. What path does the light take from A to B. (Hint: Draw a picture.)
11. A poster is needed that has 100 in.^{2} of printed area with 3 in. margins on the top and bottom and 2.0 in margins on the left and right. What dimensions will minimize the total area of the poster?
12. A trough is to be made by bending a 2 m by 4 m sheet of metal. The two bends will be parallel to the long sides and $\frac{1}{2}$ m from the long sides. What angle θ will give the trough with the largest volume? See Fig. 5.53a on page 168.
13. A trough is to be made by bending a 2 m by 4 m sheet of metal. The two bends will be parallel to the long sides and the same distance from the long sides. if the bending angle is $\frac{\pi}{6}$ what distance from the long sides will give the largest volume? See Fig. 5.53b on page 168.



The trough for exercise 12



The trough for exercise 13

Fig. 5.53

14. The intensity of light from a light source is proportional to the intensity of the light source divided by the square of the distance from the source. There are two light sources on an east-west line that are 10 m apart. If the intensity of the western light source is four times the intensity of the eastern light source, where on the line between the two light sources is the light intensity weakest?
15. A company wants to design a tin without a top and with a volume of 2 l. This is to be a rectangular tin with a square bottom. If the material for the bottom of the tin costs twice as much as the material for the sides of the tin, what dimensions for the tin will minimize the cost of materials for the tin?
16. You are to cut a beam from a log with an elliptical cross section. The length of the semi-major axis is 1 m and the length of the semi-minor axis is 3/4 m. If the stiffness of a beam is proportional to the breadth of the beam times the cube of the depth of the beam, what are dimensions of the stiffest beam you can cut from the log?

Chapter 6
Integration

6.1 Antiderivatives and Differential Equations

The first chapters of this text are devoted to the derivative. This is one of two fundamental concepts in calculus, the other is the antiderivative or integral. Many models of phenomena in the world are written in terms of the relationship between a function and its derivatives. The most basic and earliest use of this type of models is simple Newtonian mechanics.

Newtonian mechanics rests on the relationships between position, velocity, and acceleration. The first two relationships are that velocity is the derivative of position with respect to time,

$$\frac{d\,\mathbf{r}(t)}{dt} = \mathbf{v}(t),$$

and that acceleration is the derivative of velocity with respect to time,

$$\frac{d\,\mathbf{v}(t)}{dt} = \mathbf{a}(t),$$

The third relationship considered here is Newton's third law, the total force acting on a mass is the mass of the object times the acceleration,

$$\mathbf{F} = m\mathbf{a}.$$

This last relationship can be rewritten, using the fact that acceleration is the derivative of the derivative of position as

$$\mathbf{F} = m\,\frac{d^2\mathbf{r}(t)}{dt^2}.$$

These relationships are differential equations.

Definition 22. A *differential equation* is an equation that involves an independent variable, say x; a function of the variable, say $\mathbf{y}(x)$; and derivatives of the function with respect to the independent variable, say $\mathbf{y}'(x), \mathbf{y}''(x), \ldots, \mathbf{y}^{(n)}(x)$. A common form for writing a differential equation when $\mathbf{y}$ is a function of x is

$$\mathbf{f}\Big(x, \mathbf{y}, \mathbf{y}'(x), \mathbf{y}''(x), \ldots, \mathbf{y}^{(n)}(x)\Big) = \mathbf{0}.$$

Given a differential equation

$$\mathbf{f}\Big(x, \mathbf{y}, \mathbf{y}'(x), \mathbf{y}''(x), \ldots, \mathbf{y}^{(n)}(x)\Big) = \mathbf{0}.$$

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_6

a function $\mathbf{g}(x)$ is a *solution* to the differential equation if

$$\mathbf{f}\left(x, \mathbf{g}(x), \mathbf{g}'(x), \mathbf{g}''(x), \ldots, \mathbf{g}^{(n)}(x)\right) = \mathbf{0}$$

on some interval (a,b).

Example 207. Consider the differential equation

$$\frac{d\,y(t)}{dt} = k\,y(t)$$

and the function $g(t) = Ce^{kt}$ where C is an arbitrary constant. Since $g'(t) = kCe^{kt} = k\,g(t)$ for all t, $g(t)$ is a solution to the differential equation.

Returning to the basic Newtonian mechanics system we have the differential equation relating velocity and position,

$$\frac{d\,\mathbf{r}(t)}{dt} = \mathbf{v}(t).$$

This is a very special type of differential equation. One is looking for a function whose derivative is $\mathbf{v}(t)$. Such a function $\mathbf{r}(t)$ is called *an antiderivative* of $\mathbf{v}(t)$ if $\mathbf{r}'(t) = \mathbf{v}(t)$.

Definition 23. The *antiderivative* (family) of a function $\mathbf{f}(x)$ defined on an open interval (a,b) is the collection of all functions $\mathbf{F}(x)$ such that $\mathbf{F}'(x) = \mathbf{f}(x)$. The antiderivative of $\mathbf{f}(x)$ is denoted by

$$\mathbf{F}(x) = \int \mathbf{f}(x)\,dx.$$

Example 208. The antiderivative family of $f(x) = x$ is the collection of all functions $F(x) = x^2/2 + C$ where C is an arbitrary constant. This is true since $\frac{dx^2/2}{dx} = x$ and $\frac{dC}{dx} = 0$.

Example 209. Since the derivative of $\mathbf{r}(t) = (t^2, 3t, \sin(4t))$ is $\mathbf{v}(t) = (2t, 3, 4\cos(4t))$ *an antiderivative* of $\mathbf{v}(t) = (2t, 3, 4\cos(4t))$ is $\mathbf{r}(t) = (t^2, 3t, \sin(4t))$. Because $\frac{d}{dt}(c_1, c_2, c_3) = (0,0,0)$, all of the functions $\mathbf{F}(t) = (t^2 + c_1, 3t + c_2, \sin(4t) + c_3)$ are also antiderivatives of $\mathbf{r}(t)$. In fact, they are all of the antiderivatives of $\mathbf{v}(t)$,

$$\int (2t, 3, 4\cos(4t))\,dt = (t^2 + c_1, 3t + c_2, \sin(4t) + c_3) = (t^2, 3t, \sin(4t)) + \mathbf{c}.$$

We must be careful about the domain of a function when we consider which functions are antiderivatives. If we extend the definition to allow functions whose domains are collections of disjoint intervals, we can choose different constants on each interval.

Example 210. Consider the function $f(x) = 1/x$. This function's domain is the union of two disjoint open intervals. The antiderivative of $1/x$ is often written as

$$\begin{aligned} \int \frac{1}{x}\,dx &= \ln|x| + C \\ &= \begin{cases} \ln(-x) + C & x < 0 \\ \ln(x) + C & x > 0 \end{cases}. \end{aligned}$$

In fact, the function

$$h(x) = \begin{cases} \ln(-x) + 7 & x < 0 \\ \ln(x) - 10 & x > 0 \end{cases}$$

has derivative $f(x) = 1/x$ if $x \neq 0$ and is an antiderivative of $f(x) = 1/x$.

We can reverse the process of taking derivatives for all of the elementary functions you have seen before. This gives some functions to work with throughout this chapter.

Example 211. Since $\sin(x) = -\frac{d}{dx}\cos(x)$, the antiderivative family of $\sin(x)$ is

$$\int \sin(x)\,dx = -\cos(x) + C.$$

Here C is an arbitrary real number.

Example 212. Since

$$\frac{d\tan^{-1}(x)}{dx} = \frac{1}{1+x^2},$$

the antiderivative family of $1/(1+x^2)$ is

$$\int \frac{1}{1+x^2}\,dx = \tan^{-1}(x) + C.$$

The following are basic results for finding antiderivatives of functions. They are simply differentiation rules rewritten in term of antiderivatives.

Theorem 55. *Assume that $\mathbf{f}(x)$ and $\mathbf{g}(x)$ are functions with antiderivatives and that c is a constant. Then*

(i) $\displaystyle\int (\mathbf{f}(x) + \mathbf{g}(x))\,dx = \int \mathbf{f}(x)\,dx + \int \mathbf{g}(x)\,dx,$ *and*

(ii) $\displaystyle\int c\,\mathbf{f}(x)\,dx = c\int \mathbf{f}(x)\,dx.$

Proof. Only the first result is proven. Assume that $\mathbf{f}(x)$ and $\mathbf{g}(x)$ have antiderivatives $\mathbf{F}(x)$ and $\mathbf{G}(x)$. Then

$$\begin{aligned}\frac{d}{dx}(\mathbf{F}(x) + \mathbf{G}(x)) &= \frac{d}{dx}\mathbf{F}(x) + \frac{d}{dx}\mathbf{G}(x)\\ &= \mathbf{f}(x) + \mathbf{g}(x).\end{aligned}$$

The second result is the power rule. It will be used for an example involving acceleration, velocity, and position. Instead of multiplying by the old power and subtracting 1 from the exponent as in the power rule for derivatives, here we add one to the exponent and then divide x to the new exponent by the new exponent.

Theorem 56 (Integral Power Rule). *Let q be a real number not equal to -1. Then*

$$\int x^q\,dx = \frac{1}{q+1}x^{q+1} + C.$$

Proof. Using the power rule for derivatives we have

$$\begin{aligned}\frac{d}{dx}\left(\frac{1}{q+1}x^{q+1} + C\right) &= \frac{1}{q+1}(q+1)x^{(q+1)-1} + 0\\ &= x^q.\end{aligned}$$

These results can be used to find the antiderivatives of polynomials.

Example 213. The antiderivative of $p(x) = 4x^5 - 3x + 2$ is given by

$$\begin{aligned}
\int p(x)\,dx &= \int \left(4x^5 - 3x + 2\right)\,dx \\
&= 4\int x^5\,dx - 3\int x\,dx + \int 2\,dx \\
&= 4\,\frac{x^{5+1}}{5+1} - 3\,\frac{x^{1+1}}{1+1} + 2x + C \\
&= \frac{4}{6}x^6 - \frac{3}{2}x^2 + 2x + C.
\end{aligned}$$

This can be applied to a standard problem in elementary physics, projectile motion. A shell of mass M kilograms is shot with an initial velocity $\mathbf{v}_0 = (v_x, v_y)$. It is assumed that the only force acting on the mass is gravity. It is also assumed that the mass starts at a position (x_0, y_0) and the second coordinate is the altitude above the ground.

The acceleration is the constant gravitational acceleration,

$$\frac{d\,\mathbf{v}}{dt} = \mathbf{a}(t) = (0, -9.8)\,\mathrm{m/s}^2.$$

The velocity is an antiderivative of $(0, -9.8)$, or

$$\mathbf{v}(t) = \int (0, -9.8)\,dt = (0, -9.8t) + \mathbf{c}_0.$$

The condition that $\mathbf{v}(0)$ is (v_x, v_y) is called an *initial condition* . Using this gives

$$\mathbf{v}(0) = (v_x, v_y) = (0, -9.8 \cdot 0) + \mathbf{c}_0.$$

From this we conclude that $\mathbf{c}_0 = (v_x, v_y)$ and hence that

$$\mathbf{v}(t) = (v_x, -9.8t + v_y).$$

Taking the antiderivative again gives

$$\begin{aligned}
\mathbf{r}(t) &= \int \mathbf{v}(t)\,dt \\
&= \int (v_x, -9.8t + v_y)\,dt \\
&= (v_x t, -9.8\frac{t^2}{2} + v_y t) + \mathbf{c}_1.
\end{aligned}$$

Since $\mathbf{r}(0)$ was assumed to be (x_0, y_0),

$$\mathbf{r}(0) = (x_0, y_0) = (v_x 0, -4.9 \cdot 0^2 + v_y 0) + \mathbf{c}_1 = \mathbf{c}_1.$$

We finally get that

$$\mathbf{r}(t) = (v_x t + x_0, -4.9t^2 + v_y t + y_0). \tag{6.1}$$

Example 214. Assume a shell is fired at 60° above horizontal with an initial speed of 50 m/s from a gun barrel that is 7 m above ground level. Where and when does the shell hit the ground?

The initial position can be set to $\mathbf{r}_0 = (0, 7)$ and the initial velocity $\mathbf{v}_0 = 50\left(1/2, \sqrt{3}/2\right)$. Plugging this into Eq. (6.1) gives the position of the shell as a function of time,

$$\mathbf{r}(t) = \left(25t, -4.9t^2 + 25\sqrt{3}t + 7\right). \tag{6.2}$$

The shell lands when the second component is zero,

$$-4.9t^2 + 100\sqrt{3}t + 7 = 0.$$

The roots of this quadratic are

$$t = \frac{25\sqrt{3} \pm \sqrt{1875 + 137.2}}{9.8}.$$

Since the only time of interest is when $t > 0$, we get $t \approx 9$ s. Putting this time back into Eq. 6.2 shows that shell lands approximately 450 m from where it started.

Exercises

1. Find the antiderivatives of the following functions.

(a) $f(x) = \cos(x)$
(b) $g(x) = -\sin(x)$
(c) $f(y) = \sec^2(y)$
(d) $g(y) = \tan(y)\sec(y)$
(e) $f(x) = -\csc^2(x)$
(f) $g(x) = -\cot(x)\csc(x)$
(g) $f(y) = \dfrac{1}{y}$
(h) $g(y) = 2y$
(i) $f(x) = 3x^2$
(j) $g(x) = e^x$
(k) $f(x) = \dfrac{1}{1+x^2}$
(l) $g(x) = \dfrac{1}{\sqrt{1-x^2}}$
(m) $f(y) = \dfrac{1}{y^2\sqrt{1-\frac{1}{y^2}}}$
(n) $g(y) = 1$

2. Find the antiderivatives of the following functions.

(a) $f(x) = \cos(x) + \sin(x)$
(b) $g(x) = 3x^2 - \sin(x)$
(c) $f(y) = 1 - \sec^2(y)$
(d) $g(y) = \tan(y)\sec(y) + e^y$
(e) $f(x) = \dfrac{1}{x} - \csc^2(x)$
(f) $g(x) = 2x - \cot(x)\csc(x)$
(g) $f(y) = \dfrac{1}{y} - y^4$
(h) $g(y) = \dfrac{1}{2\sqrt{y}} + 3$
(i) $f(x) = 3x^2 + \sin(x)$
(j) $g(x) = e^x + x^7$
(k) $f(x) = \dfrac{1}{1+x^2} + \dfrac{3}{x}$
(l) $g(x) = \cos(x) - \dfrac{4}{\sqrt{1-x^2}}$
(m) $f(y) = \dfrac{1}{y^2\sqrt{1-\frac{1}{y^2}}} + \sqrt{y}$
(n) $g(y) = 1 - y^3 + y^4 + 12y^5$
(o) $f(x) = x^3 + 6x^5 + 1$
(p) $g(x) = 4x^{17} - 36x^2 + 7x$
(q) $f(y) = 3y^7 + 4y^4 - 2y + 5$
(r) $g(y) = 63y^{63} - 41y^{40} + 3y^8$

3. Assume a shell is fired at 30° above horizontal with an initial speed of 60 m/s from a gun barrel that is 3 m above ground level. Where and when does the shell hit the ground?

4. Assume a shell is fired at 70° above horizontal with an initial speed of 100 m/s from a gun barrel that is 2 m above ground level. Where and when does the shell hit the ground?

5. Assume a shell is fired at 65° above horizontal with an initial speed of 85 m/s from a gun barrel that is 3 m above ground level. Where and when does the shell hit the ground?

6. Show that the following functions are solutions to the associated differential equations.

(a) $P(t) = Ce^{4t}, \quad \dfrac{dP}{dt} = 4P$

(b) $y(x) = C_1 \cos(2t) + C_2 \sin(2t), \quad \dfrac{d^2y(x)}{dx^2} + 4y(x) = 0$

(c) $z(x) = \tan(x), \quad \dfrac{dz(x)}{dx} = 1 + z(x)^2$

(d) $\mathbf{w}(t) = \big(\ln(4t), \exp(2t)\big), \quad \dfrac{dw(t)}{dt} = \left(\dfrac{1}{t}, 2\exp(2t)\right)$

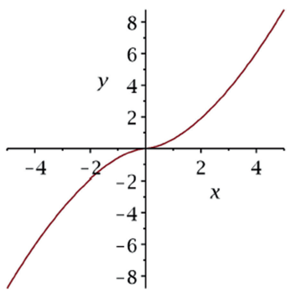
7. Match the function with the graph of its antiderivative.

(a) $f(x) = x\,\mathrm{arccot}(x)$

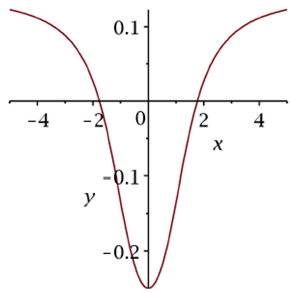
(b) $g(x) = -x^2 + x$

(c) $h(x) = \sqrt[3]{x^2}$

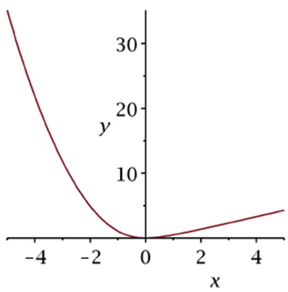
(d) $r(x) = \dfrac{x}{4+x^4}$



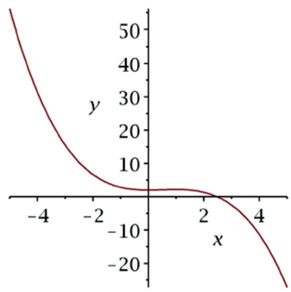
a



c



b



d

6.2 Area and Riemann Sums

In Sect. 6.4 of this chapter the derivative will be related to what is called the definite integral. The definite integral is defined in this section and the next section without the benefit of the derivative. This is done through approximating the area "under" the graph of a function of one variable. Although many of the technical details of this process are difficult, some of them are presented without proofs.

The process used to find areas is by approximation with rectangular areas. Assuming that a function $f(x)$ is nonnegative and continuous on an interval $[a,b]$, the value of $f(x)$ will be almost constant over a very small interval $[x_i, x_i + h]$ contained in $[a,b]$. This means that the

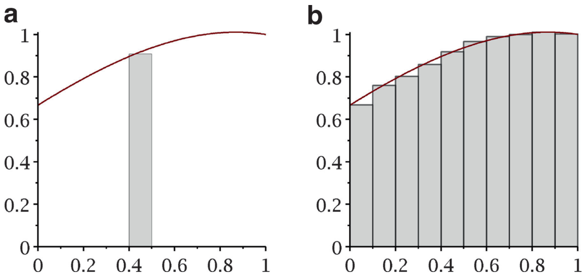


Fig. 6.1

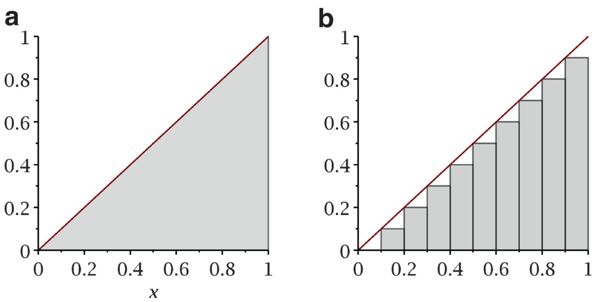


Fig. 6.2

area under the graph of $f(x)$ over $[x_i, x_i+h]$ is approximately $f(\xi_i)h$ for any ξ_i in $[x_i, x_i+h]$, see Fig. 6.1a on page 175.

Summing the areas of a collection of these rectangles from a to b should approximate the area under the graph of the function from a to b, see Fig. 6.1b on page 175.

Example 215. Consider the function $f(x) = x$ on the interval $[0,1]$. The area under the graph is $\frac{1}{2}$ since the area is a triangle with base and height 1, see Fig. 6.2a on page 175.

Dividing the interval into ten equal subintervals and using the left-hand endpoints for evaluating the function give

$$\begin{aligned}\text{Area} &\approx 0\frac{1}{10} + \frac{1}{10}\frac{1}{10} + \frac{2}{10}\frac{1}{10} + \frac{3}{10}\frac{1}{10} + \frac{4}{10}\frac{1}{10} \\ &\quad + \frac{5}{10}\frac{1}{10} + \frac{6}{10}\frac{1}{10} + \frac{7}{10}\frac{1}{10} + \frac{8}{10}\frac{1}{10} + \frac{9}{10}\frac{1}{10} \\ &\approx 0.45.\end{aligned}$$

This is a fairly good approximation as illustrated in Fig. 6.2b on page 175.

If we take the right endpoints of the subintervals we get

$$\begin{aligned}\text{Area} &\approx \frac{1}{10}\frac{1}{10} + \frac{2}{10}\frac{1}{10} + \frac{3}{10}\frac{1}{10} + \frac{4}{10}\frac{1}{10} + \frac{5}{10}\frac{1}{10} + \frac{6}{10}\frac{1}{10} \\ &\quad + \frac{7}{10}\frac{1}{10} + \frac{8}{10}\frac{1}{10} + \frac{9}{10}\frac{1}{10} + \frac{10}{10}\frac{1}{10} \\ &\approx 0.55.\end{aligned}$$

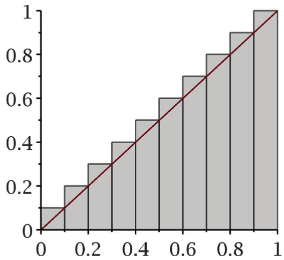


Fig. 6.3

Again, as illustrated in Fig. 6.3 on page 176, this is a fairly good approximation.

Remark 11. Using this scheme with function values times the widths of the rectangles we say that the area between the function $y(x) = -x$ and the x-axis on the interval $[0,1]$, using a sum with evaluation at right endpoints, is approximately

$$\begin{aligned}\text{Area} \approx\ & \frac{-1}{10}\frac{1}{10} + \frac{-2}{10}\frac{1}{10} + \frac{-3}{10}\frac{1}{10} + \frac{-4}{10}\frac{1}{10} + \frac{-5}{10}\frac{1}{10}\\ & + \frac{-6}{10}\frac{1}{10} + \frac{-7}{10}\frac{1}{10} + \frac{-8}{10}\frac{1}{10} + \frac{-9}{10}\frac{1}{10} + \frac{-10}{10}\frac{1}{10}\\ \approx\ & -0.55.\end{aligned}$$

This is shown in Fig. 6.4 on page 176.

For this and other reasons *the area between a curve below the x-axis and the x-axis is taken to be negative*.

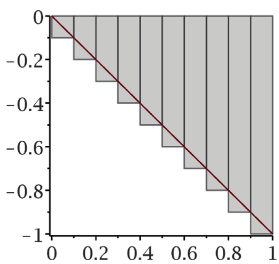


Fig. 6.4

It should be clear from the long sums in the above examples that better notation is needed for sums. The notation used here is called *summation notation.* Let $a_1, a_2, \ldots, a_n$ be a sequence of n numbers or expressions. We write the sum of these n terms as

$$a_1 + a_2 + \cdots + a_n = \sum_{i=1}^{n} a_i.$$

We will use this to simplify the notation of this section.

Example 216. To find the sum of the first 11 positive integers we can write

$$\begin{aligned}\sum_{i=1}^{11} i &= 1+2+\cdots+11 \\ &= 66.\end{aligned}$$

There are a few general formulas for sums that will be used in this section. They are used to illustrate the theoretical process for finding areas under curves. The proofs of these formulas involve mathematical induction. Since this is not a course on proofs, the proof that is given is not essential.

Lemma 3. *Assume that n is a positive integer, then the following hold.*

(i) $$\sum_{i=1}^{n} i = \frac{n(n+1)}{2},$$

(ii) $$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}, \quad \textit{and}$$

(iii) $$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}.$$

Proof. Only the first formula is considered. The idea behind mathematical induction is that if a statement is true for a fixed integer m and if we can show that when the statement is true for some n, it is also true for $n+1$, the statement is true for all $n \geq m$. In this case the statement for $n = 1$ is

$$\sum_{i=1}^{1} i = 1 = \frac{1(1+1)}{2},$$

a true statement.

Now assume that for some $n \geq 1$

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Then

$$\begin{aligned}\sum_{i=1}^{n+1} i &= \left(\sum_{i=1}^{n} i\right) + (n+1) \\ &= \frac{n(n+1)}{2} + \frac{2(n+1)}{2} \\ &= \frac{n^2+3n+2}{2} \\ &= \frac{(n+1)(n+2)}{2}.\end{aligned}$$

Since this is the desired formula for $n+1$, the formula is true for all $n \geq 1$.

Now that the notation and formulas are available, it is time to return to the sums representing areas "under" curves. Let $f(x)$ be a function on an interval $[a,b]$. A division of an interval $[a,b]$ into n subintervals $[a_0,a_1],[a_1,a_2],\ldots,[a_{n-1},a_n]$ with $a_0 = a$, $a_n = b$ and $a_i < a_{i+1}$ is called a *partition of* $[a,b]$, $\mathscr{P} = a_0, a_1, a_2, \ldots, a_n$ with $a_k < a_{k+1}$. The idea is that as the "size" of the partitions gets smaller, goes to 0, the area corresponding to the partitions goes to a number, the definite integral of $f(x)$ over $[a,b]$.

To be consistent we need to insist that a number of technical conditions are satisfied. These include that the points where the functions are evaluated in each subinterval must be arbitrary. We also must insist that the endpoints of the partition intervals can be arbitrary with the *mesh*, $\max_{i=1\ldots n}(a_i - a_{i-1})$, becoming arbitrarily small. For our purposes, it is also assumed that the function $f(x)$ is *piecewise continuous*, continuous on a finite set of subintervals of $[a,b]$, $\{[c_0,c_1],[c_1,c_2],\ldots,[c_{m-1},c_m]\}$, the limits $\lim_{x\to c_j^+} f(x)$ exist and are finite for $j = 0,1\ldots,m-1$ and the limits $\lim_{x\to c_j^-} f(x)$ exist and are finite for $j = 1,2\ldots,m$. For the rest of the section it is assumed that the technical conditions are satisfied except when examples are given where the conditions do not hold. It is important to note that all of the technicalities are satisfied if the function is continuous on the closed bounded interval $[a,b]$

The following definition puts it all together. What we will actually use are the results after the definition.

Definition 24. Assume that $f(x)$ is a function from $[a,b]$ into $\mathbb{R}$. If the limit, taken over all partitions $\mathscr{P}$, $a = a_0 < a_1 < \cdots < a_n = b$, of $[a,b]$ and arbitrary $\xi_i \in [a_{i-1}, a_i]$,

$$\lim_{\max(a_i - a_{i-1})\to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1}) = A$$

exists, then the *area between* the graph of $f(x)$ and the x-axis is A.

The sums

$$\sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1})$$

are called *Riemann sums*.

The next result is not proven here as the proof is well beyond the level of this class. It is, however, essential to the rest of this calculus class.

Theorem 57. *Assume that $f(x)$ is continuous on $[a,b]$, then the limit over all partitions $\mathscr{P}$ of $[a,b]$,*

$$\lim_{\max(a_i - a_{i-1})\to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1})$$

exists.

Finally comes the result actually used in this section for approximating areas. It allows us to choose a sequence of partitions for approximating the limit much as knowing that the limit $\lim_{x\to a} f(x)$ exists implies that we can use any sequence to approximate the limit value. Again, the proof is omitted.

Theorem 58. *Assume that the limit*

$$\lim_{\max(a_i - a_{i-1})\to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1}) \tag{6.3}$$

exists. Then for any sequence of partitions P_k with $\max\,(a_i - a_{i-1}) \to 0$, *$a_i$ and a_{i-1} are in $\mathscr{P}_k$, and any points $\xi_i \in [a_{i-1}, a_i]$, the limit as $k \to \infty$ of the sums (6.3) exists and is the area under the graph of the function.*

Using these results we can approximate the areas under the graphs of several functions.

Example 217. Let $f(x) = x$ on the interval $[0,1]$. As was noted above, the area under this curve is $\frac{1}{2}$. Since $f(x)$ is continuous the limit

$$\lim_{\max(a_i - a_{i-1}) \to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1})$$

exits. Because of this, we can choose partitions that have $n+1$ equally spaced points, $0, 1/n, \ldots, (n-1)/n, 1$ and choose the left endpoints of the intervals given by the partitions. The lengths of the intervals are all $(a_i - a_{i-1}) = 1/n$ and the function value for the ith interval is $(i-1)/n$.

With these choices the Riemann sums become

$$\sum_{i=1}^{n} \frac{i-1}{n} \frac{1}{n}.$$

Using Lemma 3, we can rewrite the sums as

$$\begin{aligned} \sum_{i=1}^{n} \frac{i-1}{n} \frac{1}{n} &= \frac{1}{n^2} \frac{n(n-1)}{2} \\ &= \frac{1}{2} \frac{n^2 - n}{n^2}. \end{aligned}$$

The area under the curve is

$$\begin{aligned} A &= \lim_{n \to \infty} \frac{1}{2} \frac{n^2 - n}{n^2} \\ &= \frac{1}{2}. \end{aligned}$$

We can do a similar approximation and limit process with $f(x) = x^2$.

Example 218. Let $f(x) = x^2$ and let the interval under consideration be $[0,2]$. An appropriate choice is equal interval lengths of $2/n$ and appropriate points are the right endpoints of the intervals. The points for function evaluation are $2/n, 4/n, \ldots, (2n-2)/n, 2$. (See Fig. 6.5 on page 179.)

The Riemann sum in this case is

$$\sum_{i=1}^{n} \left(\frac{2i}{n}\right)^2 \frac{2}{n}.$$

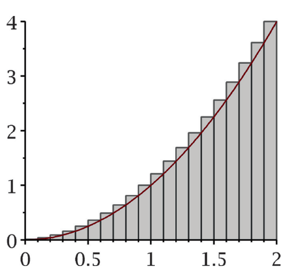


Fig. 6.5

Rearranging and simplifying using Lemma 3 gives

$$\begin{aligned}\sum_{i=1}^{n}\left(\frac{2i}{n}\right)^2\frac{2}{n} &= \frac{8}{n^3}\sum_{i=1}^{n} i^2\\ &= \frac{8}{n^3}\,\frac{n(n+1)(2n+1)}{6}\\ &= \frac{4}{3}\,\frac{2n^3+3n^2+n}{n^3}.\end{aligned}$$

Taking the limit as $n \to \infty$ yields an area

$$\begin{aligned}A &= \lim_{n\to\infty}\frac{4}{3}\left(2+\frac{3}{n}+\frac{1}{n^2}\right)\\ &= \frac{8}{3}.\end{aligned}$$

The area under $f(x) = x^2$ from 0 to 2 is $8/3$.

There are many functions for which we can only approximate the area under the curve. Unlike the functions in the rest of this chapter, they do not have antiderivatives that are common simple functions. In these cases, we can use Riemann sums for approximating the area under the curve. However, Riemann sums are a poor method for finding approximations. The following example illustrates this.

Example 219. Let $f(x) = e^x$ and consider the area under this curve from $x = 1$ to $x = 3$, see Fig. 6.6 on page 180.

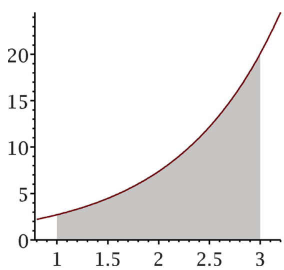


Fig. 6.6

The Riemann sums are done with 2^k equal intervals and right endpoints. This gives the largest possible error for these partitions. Figure 6.7 on page 181 shows the graph and the rectangles for 4 equal intervals.

Table 6.1 on page 181 is a table of the values of the Riemann sums for a number of different partitions of $[1,3]$. The error is the actual value minus the approximation from the Riemann sum.

The size of the error decreases by approximately a factor of 2 for each doubling of the number of intervals. This is what theory, derived in numerical analysis classes, tells us how the errors should behave. This means that to get one more digit of accuracy, we must use approximately ten times as many intervals. There are methods that do much better than this. Several better techniques are discussed in second semester calculus courses.

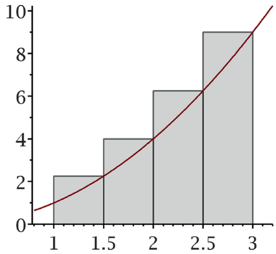


Fig. 6.7

Intervals	Sum value	Error
4	10.75000000	−2.083333333
8	9.687500000	−1.020833333
16	9.171875000	−0.505208333
32	8.917968750	−0.251302083
64	8.791992194	−0.125325527
128	8.729248048	−0.062581381

Table 6.1 Riemann sum values for different numbers of intervals

Exercises

1. Evaluate the following sums exactly.

(a) $\sum_{n=1}^{10} \frac{1}{n}$

(b) $\sum_{k=1}^{7} k^2$

(c) $\sum_{m=1}^{5} m!$

(d) $\sum_{n=1}^{7} \frac{1}{k^2}$

2. Approximate the following sums to five decimal places.

(a) $\sum_{n=1}^{10} \sqrt{n}$

(b) $\sum_{k=1}^{15} \frac{k}{k^2+3}$

(c) $\sum_{m=1}^{35} \frac{(-1)^m}{5+\sqrt{m}}$

(d) $\sum_{n=3}^{19} \frac{2}{k^3-8}$

3. Find expressions for the Riemann sums for the following functions on the given intervals with the given number of equal intervals.

(a) $f(x) = 5 - x$, $[1,3]$, $n = 2$

(b) $g(x) = x^2 - x$, $[-1,1]$, $n = 4$

(c) $f(y) = \sin(y)$, $[0, \frac{\pi}{2}]$, $n = 3$

(d) $g(y) = \tan^{-1}(y)$, $[-1,1]$, $n = 4$

4. Use a calculator or computer to compute the left and right endpoint Riemann sums for the given function over the given interval with the specified number of equal intervals.

(a) $f(x) = x + 1$, $[-2,3]$, $n = 25$
(b) $g(x) = x^3 - x^2$, $[1,4]$, $n = 24$
(c) $f(y) = \sin(y)$, $[0, \frac{\pi}{2}]$, $n = 15$
(d) $g(y) = \tan^{-1}(y)$, $[-1,1]$, $n = 15$
(e) $f(x) = \cos(x) + \sin(x)$, $[0,\pi]$, $n = 20$
(f) $g(x) = 3x^2 - \sin(x)$, $[1,11]$, $n = 50$
(g) $f(y) = 1 - \tan^2(y)$, $[0,1.3]$, $n = 30$
(h) $g(y) = e^{-y^2}$, $[-2,1]$, $n = 14$

5. Use Riemann sums and limits to find the area under the given function on the given interval.

(a) $f(x) = x + 1$, $[0,2]$
(b) $g(x) = x^2 + 1$, $[0,2]$
(c) $f(y) = y^2$, $[1,2]$
(d) $g(y) = y^2 + y$, $[2,4]$
(e) $f(x) = x^3$, $[1,2]$
(f) $g(x) = x^3 - x$, $[1,3]$

6. Use Riemann sums and limits to find the area under the given function on the given interval.

(a) $f(x) = x$, $[0,b]$
(b) $g(x) = x$, $[a,0]$
(c) $f(y) = y^2$, $[0,b]$
(d) $g(y) = y^2$, $[a,0]$
(e) $f(x) = x^3$, $[0,b]$
(f) $g(x) = x^3$, $[a,0]$

6.3 The Definite Integral

In the last section on Riemann sums the idea of defining areas under curves using limits of sums of rectangular approximations was defined. We use this to define the definite integral. *Throughout this section, unless stated otherwise, it is assume that the limits of Riemann sums exits for the functions under consideration.* The main emphasis of this section is on the properties of the definite integral.

Definition 25. The *definite integral*, *integral*, of $f(x)$ from a to b,

$$\int_a^b f(x)\,dx,$$

is the limit, taken over all partitions $a = a_0 < a_1 < \cdots < a_n = b$ of $[a,b]$ and arbitrary $\xi_i \in [a_{i-1}, a_i]$,

$$\lim_{\max\,(a_i - a_{i-1}) \to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1}).$$

A function that has an integral on $[a,b]$ is said to be *integrable* on $[a,b]$.

The notation in this definition uses the same notation as was used in approximations using the derivative in Eq. 2.4, using g for the function,

$$dg = g'(a)\,dx\,.$$

As in that case, the dx represents an infinitely small length. In this setting the integral can be interpreted as the sum of an infinite number of "areas" with height $f(x)$ and with width dx. This is frequently used when setting up integrals in applications.

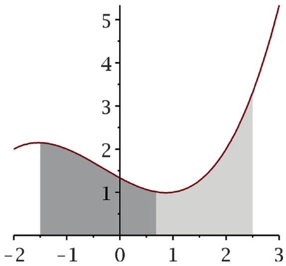


Fig. 6.8

We can also view dx as a "small" length and then we have, as in differential notation, a small area $f(x)\Delta x = f(x)\,dx$. Either interpretation is reasonable. Use the one with which you feel comfortable.

As was noted in Theorem 57 on page 178, all continuous functions on an interval $[a,b]$ are integrable on $[a,b]$. The next two results tells us somewhat more about integrability. The proof of the first is omitted and an idea of the proof of the second is provided.

Theorem 59. *Assume that $f(x)$ is integrable on $[a,b]$ and that $[c,d] \subset [a,b]$. Then $f(x)$ is integrable on $[c,d]$.*

Theorem 60. *Assume that $f(x)$ is integrable on $[a,c]$ and $[c,b]$. Then $f(x)$ is integrable on $[a,b]$ and*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Idea of this result is an illustrated in Fig. 6.8 on page 183. It is the graph of a function of x. The area under the curve from -1.5 to 2.5 is the sum of the areas under the curve from -1.5 to c and from c to 2.5.

Proof. (Idea only for the case when $c \in (a,b)$.) Let $a = a_0, a_1 < \cdots < a_k = c$ be a partition of $[a,c]$ and let $c = b_0 < b_1 < \cdots < b_m = b$ be a partition of $[c,b]$. Since the limits

$$\lim_{\max\,(a_i - a_{i-1}) \to 0} \sum_{i=1}^{k} f(\xi_i)(a_i - a_{i-1})$$

and

$$\lim_{\max\,(b_j - b_{j-1}) \to 0} \sum_{j=1}^{m} f(\mu_j)(b_j - b_{j-1})$$

both exist, the limit

$$\lim_{\max\,(a_i - a_{i-1}) \to 0,\ \max\,(b_j - b_{j-1}) \to 0} \left(\sum_{i=1}^{k} f(\xi_i)(a_i - a_{i-1}) \right.$$
$$\left. + \sum_{i=1}^{m} f(\mu_j)(b_j - b_{j-1}) \right)$$

exists. This limit is the definite integral of $f(x)$ on $[a,b]$.

The justification of the last statement is omitted. It is very technical and beyond the scope of this class. However, the idea is contained in the work above.

Remark 12. This result tells us that a piecewise continuous function on $[a,b]$ with a finite number of pieces is integrable on $[a,b]$. Simple examples of this are that $\lfloor x \rfloor$ is integrable on any $[a,b]$ and the signum function,

$$sgn(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases},$$

is integrable on any $[a,b]$

The same basic properties, Theorem 55 on page 171, of antiderivatives hold for the definite integral.

Theorem 61. *Assume that $f(x)$ and $g(x)$ are integrable on $[a,b]$ and that c is a real number. Then*

(i) $\displaystyle\int_a^b f(x)+g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx,$ *and*

(ii) $\displaystyle\int_a^b c\,f(x)\,dx = c\int_a^b f(x)\,dx\,.$

Proof. Only the second result is proven.

Let $a = a_0 < a_1 < \cdots < a_n = b$ be a partition of $[a,b]$ with corresponding points ξ_i. Then

$$\sum_{i=1}^{n} c\,f(\xi_i)(a_i - a_{i-1}) = c\sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1}).$$

Hence

$$\begin{aligned}\int_a^b c\,f(x)\,dx &= \lim_{\max\,(a_i - a_{i-1})\to 0} \sum_{i=1}^{n} c\,f(\xi_i)(a_i - a_{i-1}) \\ &= c \lim_{\max\,(a_i - a_{i-1})\to 0} \sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1}) \\ &= c\int_a^b f(x)\,dx.\end{aligned}$$

There are also some properties that are defined to make the definite integral consistent. The first definition can be viewed as saying that a rectangle with no width has no area. The second definition can be viewed in terms of accumulation distance traveled in terms of time.

If we integrate velocity forward in time from $t = a$ to $t = b$, we add that displacement to the position at $t = a$ to get the position at $t = b$. On the other hand, if we integrate the velocity backward in time from $t = b$ to $t = a$, we should be able to add this to the position at $t = b$ to get the position at $t = a$. This is the negative of the displacement from $t = a$ to $t = b$. The integral from $t = b$ to $t = a$ should be the negative of the integral from $t = a$ to $t = b$.

Definition 26. Assume that $f(x)$ is defined at a. Then

$$\int_a^a f(x)\,dx = 0.$$

Definition 27. Assume that $f(x)$ is integrable on $[a,b]$. Then

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

Some simple examples give an idea of how to evaluate definite integrals. The starting example is constant functions.

Theorem 62. *Any constant function $f(x) = c$ is integrable on any $[a,b]$ and*

$$\int_a^b c\,dx = c\,(b-a).$$

In particular

$$\int_a^b 0\,dx = 0.$$

The next result gives a way to compare or estimate definite integrals.

Theorem 63. *Assume that $f(x)$ and $g(x)$ are integrable on $[a,b]$ with $f(x) \geq g(x)$ for all $x \in [a,b]$, then*

$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$

In particular, if $f(x)$ is an integrable function on $[a,b]$ with $f(x) > 0$ for all $x \in [a,b]$ and $b > a$, then

$$\int_a^b f(x)\,dx > 0.$$

Proof. The only case considered here is showing that if $h(x) \geq 0$ on $[a,b]$ then

$$\int_a^b h(x)\,dx \geq 0.$$

The main result follows immediately from this inequality by letting $h(x) = f(x) - g(x)$ and noting that

$$\int_a^b f(x) - g(x)\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx.$$

Let $a_0 < a_1 < \cdots < a_n$ be any partition of $[a,b]$ and let ξ_i be corresponding points. Since $h(\xi_i) \geq 0$ for all i,

$$\sum_{i=1}^{n} h(\xi_i)\;(a_i - a_{i-1}) \geq 0.$$

In the limit in definition of the definite integral all of the sums are nonnegative, thus the limit must be nonnegative. Therefore

$$\int_a^b h(x)\,dx \geq 0.$$

The final result in this section is a version of the Mean Value Theorem, Theorem 49 on page 144. This time it is for integrals. It is a geometrically interesting result and is also used in the proof of the Fundamental Theorem of Calculus, Theorem 65 on page 189.

Theorem 64 (Mean Value Theorem for Integrals). *Assume that $f(x)$ is continuous on $[a,b]$. Then there is a $c \in [a,b]$ such that*

$$\int_a^b f(x)\,dx = f(c)\;(b-a).$$

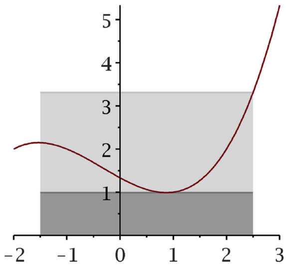


Fig. 6.9

Proof. Since $f(x)$ is continuous on $[a,b]$ it attains a maximum value M at $z_1 \in [a,b]$ and it attains a minimum value m at $z_2 \in [a,b]$. See Fig. 6.9 on page 186 The Riemann sum for any partition $a = a_0 < a_1 < \cdots < a_n = b$ and points ξ_i in the intervals satisfies

$$\begin{aligned}\sum_{i=1}^{n} f(\xi_i)\,(a_i - a_{i-1}) &\geq \sum_{i=1}^{n} m\,(a_i - a_{i-1}) \\ &\geq m\,(b-a).\end{aligned}$$

Similarly

$$\begin{aligned}\sum_{i=1}^{n} f(\xi_i)\,(a_i - a_{i-1}) &\leq \sum_{i=1}^{n} M\,(a_i - a_{i-1}) \\ &\leq M\,(b-a).\end{aligned}$$

This implies that the limit of the Riemann sums, $\int_a^b f(x)\,dx$, must be between $m\,(b-a)$ and $M\,(b-a)$. The value is $r\,(b-a)$ for some $r \in [m,M]$. By the Intermediate Value Theorem, Theorem 43 on page 121, there is a $c \in [a,b]$ such that $f(c) = r$. This proves the result.

An important use of the definite integral is finding the area between two curves. As Fig. 6.10 on page 187 shows. The area between the graphs of two functions can be approximated by a sum of areas of rectangles with heights $f(\xi_i) - g(\xi_i)$ and widths Δx_i, assuming $f(x) \geq g(x)$ on the interval of interest, $[a,b]$. Writing this as a Riemann sum gives

$$\begin{aligned}\text{Area} &\approx \sum_{i=1}^{n} \left(f(\xi_i) - g(\xi_i)\right)\,\Delta x_i \\ &\approx \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i - \sum_{i=1}^{n} g(\xi_i)\,\Delta x_i.\end{aligned}$$

Taking the limit of the right side of the last equation as the maximum of Δx_i goes to zero and assuming that both $f(x)$ and $g(x)$ are integrable on $[a,b]$ gives

$$\begin{aligned}\text{Area} &= \int_a^b f(x)\,dx - \int_a^b g(x)\,dx \\ &= \int_a^b (f(x) - g(x))\,dx.\end{aligned}$$

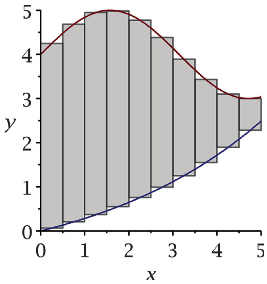


Fig. 6.10

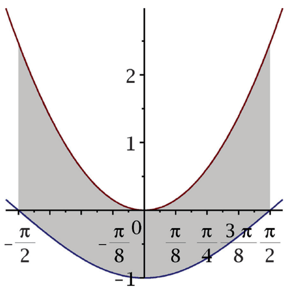


Fig. 6.11

As an example, consider the area between the curves $f(x) = x^2$ and $g(x) = -\cos(x)$ on the interval $[-\pi/2, \pi/2]$, see Fig. 6.11 on page 187. By the formula above we have

$$\begin{aligned}\text{Area} &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(x^2 + \cos(x)\right)\,dx \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\,dx + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(x)\,dx.\end{aligned}$$

In Sect. 6.4 of this chapter you will learn how to find the values of the preceding integrals. The values are

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\,dx = \frac{\pi^3}{12} \qquad \text{and} \qquad \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(x)\,dx = 2.$$

This means that the area between the curves is $2 + \pi^3/12$.

Exercises

1. Use information from the problems in the last section to evaluate the following.

(a) $\displaystyle\int_a^b 1\,dx$

(b) $\displaystyle\int_a^b x\,dx$

(c) $\displaystyle\int_a^b x^2\,dx$

(d) $\displaystyle\int_a^b x^3\,dx$

2. Use the results from the first problem in this section to evaluate the following.

(a) $\int_{-1}^{2} (x+2)\,dx$

(b) $\int_{0}^{3} x^2 - 2\,dx$

(c) $\int_{1}^{4} \left(x^3 - x\right)\,dx$

(d) $\int_{1}^{-2} \left(4 - 3x^2 - 2x\right)\,dx$

(e) $\int_{3}^{0} x^2 - 2\,dx$

(f) $\int_{4}^{1} \left(x^3 - x\right)\,dx$

(g) $\int_{-7}^{-7} \left(x^2 + x + 1\right)\,dx$

(h) $\int_{-2}^{1} \left(4 - 3x^2 - 2x\right)\,dx$

(i) $\int_{-10}^{3} \left(4x^3 - 2x^2 + 3\right)\,dx$

(j) $\int_{4}^{6} 5 - x\,dx$

3. Assume that $f(x)$ is a odd function that is integrable on $[-a,a]$ where $a > 0$. Why is

$$\int_{-b}^{b} f(x)\,dx = 0$$

for any $b \in (0,a]$?

4. Assume that $f(x)$ is an even function that is integrable on $[-a,a]$ where $a > 0$, Why is

$$\int_{-b}^{b} f(x)\,dx = 2\int_{0}^{b} f(x)\,dx$$

for any $b \in (0,a]$?

5. Why is

$$\int_{a}^{a+2\pi} \cos(x)\,dx = 0$$

for any $a \in \mathbb{R}$?

6. Why is

$$\int_{a}^{a+2\pi} \sin(x)\,dx = 0$$

for any $a \in \mathbb{R}$?

7. If $m > n$, how does $\int_{0}^{1} \frac{1}{1+y^n}\,dy$ relate to $\int_{0}^{1} \frac{1}{1+t^m}\,dt$?

8. If $\int_1^2 f(x)\,dx = 3$ and $\int_2^5 f(x)\,dx = -1$, what is $\int_1^5 f(x)\,dx$?

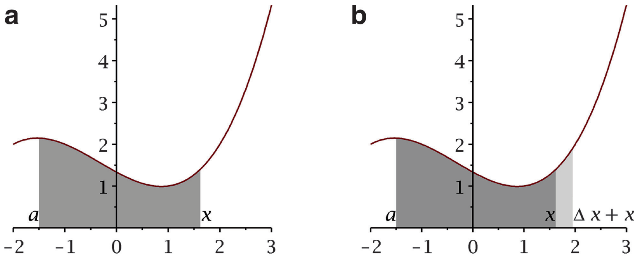
9. If $\int_{-1}^3 f(x)\,dx = 3$ and $\int_2^3 f(x)\,dx = -1$, what is $\int_{-1}^2 f(x)\,dx$?

10. If $\int_{-5}^0 f(x)\,dx = 3$ and $\int_0^7 f(x)\,dx = 61$, what is $\int_{-5}^7 f(x)\,dx$?

11. If $\int_3^{10} g(t)\,dt = 6$ and $\int_8^{10} g(x)\,dx = -1$, what is $\int_3^8 g(w)\,dw$?

12. If $\int_3^5 g(t)\,dt = 4$, what is $\int_5^3 g(w)\,dw$?

13. If $\int_{-2}^2 g(t)\,dt = -3$, what is $\int_2^{-2} g(w)\,dw$?



a The integral of f from a to x

b The integral of f from a to $x+\Delta x$

Fig. 6.12

6.4 The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus is the capstone of the first semester of calculus. It ties the integral (antiderivative) and the derivative together and shows how they can be used together.

The idea is illustrated by considering a continuous function $f(x)$ and the definite integral of that function, $F(x)$, see Fig. 6.12a on page 189,

$$F(x) = \int_a^x f(\tau)\, d\tau.$$

The difference

$$F(x+\Delta x) - F(x)$$

can be rewritten, using the Mean Value Theorem for Integrals, Theorem 64 on page 185 as

$$\begin{aligned} F(x+\Delta x) - F(x) &= \int_a^{x+\Delta x} f(x)\, dx - \int_a^x f(x)\, dx \\ &= \int_x^{x+\Delta x} f(x)\, dx \\ &= f(c)\,\Delta x \end{aligned}$$

for some c between x and $x+\Delta x$. Thus the area under $f(x)$ from x to $x+\Delta x$ is $f(c)\Delta x$. (See Fig. 6.12b on page 189.)

Since c is between x and $x+\Delta x$ and $\lim_{\Delta x\to 0} c = x$,

$$\begin{aligned} F'(x) &= \lim_{\Delta x\to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} \\ &= \lim_{\Delta x\to 0} \frac{f(c)\,\Delta x}{\Delta x} \\ &= f(x). \end{aligned}$$

This is the first half of the Fundamental Theorem of Calculus.

Theorem 65 (Fundamental Theorem of Calculus). *Let $f(x)$ be continuous on an interval $[a,b]$ and assume that $y \in [a,b]$ and $x \in (a,b)$. Then*

$$\frac{d}{dx}\int_y^x f(\tau)\, d\tau = f(x).$$

Assume that $f(x)$ is integrable on $[a,b]$ and that $F(x)$ is an antiderivative of $f(x)$ on $[a,b]$. Then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Proof. The first part was derived before the theorem was stated. Thus only the second statement is considered.

Let

$$G(x) = \int_a^x f(\tau)\,d\tau$$

and let $F(x)$ be any antiderivative of $f(x)$. Then $F(x) = G(x) + C$ and $G(a) = 0$. Combining these statements gives

$$\begin{aligned} F(b) - F(a) &= G(b) - G(a) \\ &= G(b) \\ &= \int_a^b f(x)\,dx. \end{aligned}$$

This is the desired result.

Remark 13. Since antiderivatives of vector valued functions are done coordinate by coordinate, the Fundamental Theorem of Calculus also holds for vector valued functions.

Remark 14. In the rest of this book the following notation is used for evaluating definite integrals. Here $F(x)$ is usually an antiderivative of some function.

$$F(x)\Big|_{x=a}^{b} = F(b) - F(a).$$

If we know the antiderivative of a function the Fundamental Theorem of Calculus, FTC, is easy to use.

Example 220. Since an antiderivative of $f(x) = x$ is $F(x) = x^2/2$ we have

$$\begin{aligned} \int_1^5 5x\,dx &= 5\int_1^x x\,dx \\ &= 5\,\frac{x^2}{2}\Big|_{x=1}^{5} \\ &= 5\left(\frac{5^2}{2} - \frac{1^2}{2}\right) \\ &= 60. \end{aligned}$$

Example 221. Consider the acceleration of an object given by $\mathbf{a}(t) = (t, 1/t, -1)$. The change in velocity of the object from $t = 1$ to $t = 3$ is given by

$$\begin{aligned} \mathbf{v}(3) - \mathbf{v}(0) &= \int_1^3 \mathbf{a}(t)\,dt \\ &= \int_1^3 (t, 1/t, -1)\,dt \\ &= \left(\frac{x^2}{2}, \ln|x|, -x\right)\Big|_1^3 \end{aligned}$$

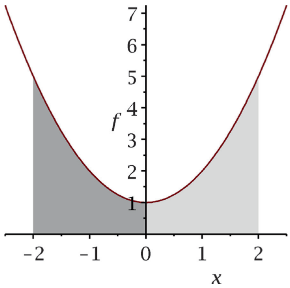


Fig. 6.13 Equal areas for a symmetric and an even function

$$\begin{aligned}
&= \left(\frac{3^2}{2}, \ln(3), -3\right) - \left(\frac{1^2}{2}, \ln(1), -1\right) \\
&= (4, \ln(3), -2).
\end{aligned}$$

The FTC also makes it easy to find the area under the graph of some functions.

Example 222. The area under the curve $f(x) = 1 + x^2$ from $x = -2$ to $x = 2$, see Fig. 6.13 on page 191, is given by

$$A = \int_{-2}^{2} 1 + x^2\, dx.$$

Since $f(x)$ is an even function, the integrals from $x = -2$ to $x = 0$ and from $x = 0$ t $x = 2$ are the same. This means that

$$\begin{aligned}
A &= 2 \int_0^2 x^2 + 1\, dx \\
&= 2\left(\frac{x^3}{3} + x\right)\Bigg|_{x=0}^{2} \\
&= 2\left(\frac{2^3}{3} + 2 - \left(\frac{0^3}{3} + 0\right)\right) \\
&= \frac{28}{3}.
\end{aligned}$$

Example 223. The question here is what is the are between the curves $f(x) = x$ and $f(x) = x^3$ for x between -1 and 1. See Fig. 6.14 on page 192. Here $x^3 \geq x$ if $x \in [-1, 0]$ and $x \geq x^3$ if $x \in [0, 1]$. Note that the areas between the curves from $x = -1$ to $x = 0$ and from $x = 0$ to $x = 1$ are the same by symmetry. This implies that the area is

$$\begin{aligned}
Area &= 2 \int_0^1 (x - x^3)\, dx \\
&= 2\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{1} \\
&= 2\left(\frac{1^2}{2} - \frac{1^4}{4}\right) - 2\left(\frac{0^2}{2} - \frac{0^4}{4}\right) \\
&= \frac{1}{2}.
\end{aligned}$$

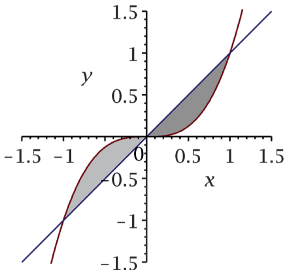


Fig. 6.14 The area between $y = x$ and $y = x^3$

Exercises

1. Find the derivatives of the following.

(a) $f(x) = \int_2^x \tan(y)\,dy$

(b) $g(x) = \int_{-1}^x \sin(2\tau)\,d\tau$

(c) $f(y) = \int_y^{-3} \left(x^2 - 6x + 2\right)\,dx$

(d) $g(y) = \int_{-10}^y \tan^{-1}(z)\,dz$

2. Evaluate the following definite integrals.

(a) $\int_0^2 2\,dx$

(b) $\int_{-3}^{10} 0\,dx$

(c) $\int_2^5 -1\,dx$

(d) $\int_{-2}^2 \sin(3x)\,dx$

(e) $\int_{-1}^1 \left(x^2 + 2\right)\,dx$

(f) $\int_0^2 x^3 - x\,dx$

(g) $\int_{-3}^0 \left(x^3 + x^2\right)\,dx$

(h) $\int_0^\pi \cos(\theta)\,d\theta$

(i) $\int_0^\pi \sin(w)\,dw$

(j) $\int_0^1 \frac{1}{1+t^2}\,dt$

3. Find the following functions.

(a) $\int_0^x \cos(t)\,dt$

(b) $\int_x^0 \cos(t)\,dt$

(c) $\int_{\frac{x}{2}}^x y^2\,dy$

(d) $\int_{-x}^x \left(t^4 + t^2 + 1\right)\,dt$

(e) $\int_{\frac{\pi}{4}}^\theta \sin(t)\,dt$

(f) $\int_{-\frac{\pi}{6}}^\theta \sec(\tau)\tan(\tau)\,d\tau$

4. Find the area between the given functions on the given interval.

(a) $f(x) = x - 4$, $g(x) = \cos(x)$ and $[-\pi, \pi]$

(b) $h(y) = -y$, $g(y) = y$ and $[-2, 2]$

(c) $f(\theta) = \cos(\theta)$, $s(\theta) = \sin(\theta)$ and $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

(d) $h(y) = \sqrt{y}$, $r(y) = \frac{y}{3}$ and $[1, 2]$

(e) $f(x) = x^2$, $g(x) = -x^2$ and $[-1, 1]$

5. Explain why the area between the curves in Example 223 is not given by

$$\int_{-1}^{1} x - x^3\, dx.$$

6. Consider the function

$$f(z) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}.$$

(a) Show that

$$h(z) = \begin{cases} x+1 & x > 0 \\ -x-1 & x < 0 \end{cases}.$$

and

$$g(z) = |z|$$

are both antiderivatives of $f(z)$.

(b) Does the second statement in the Fundamental Theorem of Calculus hold for either $h(z)$ or $g(z)$?

(c) Explain why this does or does not contradict the Fundamental Theorem of Calculus.

6.5 Integration by Substitution

The idea of the simple antiderivatives is to reverse the operation of differentiation. The first method beyond simple integration in these notes is the substitution method. It is basically reversing the chain rule, Theorem 41 on page 96,

$$\frac{d}{dx}(f \circ g)(x) = f'(g(x))\, g'(x).$$

Reversing this says that if an integral can be framed as

$$\int f'(g(x))\, g'(x)\, dx,$$

the integral is then

$$\int f'(g(x))\, g'(x)\, dx = (f \circ g)(x) + C. \tag{6.4}$$

Example 224. Consider the integral

$$\int 2\cos(2x+1)\, dx.$$

If we set $f(y) = \sin(y)$ and $g(x) = 2x+1$, we have

$$f'(y) = \cos(y) \qquad \text{and} \qquad g'(x) = 2.$$

The integral now has the form in Eq. (6.4) on page 193 and we get

$$\begin{aligned}\int 2\cos(2x+1)\,dx &= \int f'(g(x))\,g'(x)\;dx\\ &= f(g(x))+C\\ &= \sin(2x+1)+C.\end{aligned}$$

This method is rather convoluted and is seldom used. The technique that is used does the same thing, but with easier notation. It is called *substitution*. Instead of using the functions $f(y)$ and $g(x)$ where we use the functions $f'(x)$ and $g'(x)$. We often set $g(x)=u$.

The idea of the technique is to rewrite the integral as

$$\int f'(u)\,du$$

where $f'(u)$ has a simple antiderivative $f(u)+C$. The function $g(x)$ then replaces u to get the integral of the original function. The next example repeats Example 224 using this technique.

Example 225. Consider, again, the integral

$$\int 2\cos(2x+1)\,dx.$$

Setting $u=2x+1$ and taking the differentials of both sides gives

$$du = 2\,dx.$$

The integral now becomes, changing from x to u,

$$\begin{aligned}\int 2\cos(2x+1)\,dx &= \int \cos(2x+1)\;2\,dx\\ &= \int \cos(u)\,du.\end{aligned}$$

Since the integral of $\cos(u)$ is $\sin(u)+C$ we get

$$\int \cos(u)\,du = \sin(u)+C.$$

Substituting $2x+1$ back in for u gives the final result

$$\int 2\cos(2x+1)\,dx = \sin(2x+1)+C.$$

Remark 15. It is relatively easy to check if the integral we get is correct by differentiating the result. For Example 225 we have

$$\begin{aligned}\frac{d}{dx}\left(\sin(2x+1)+C\right) &= \cos(2x+1)\,\frac{d}{dx}\,(2x+1)+0\\ &= 2\,\sin(2x+1).\end{aligned}$$

This verifies that the integral obtained in Example 225 is correct.

This technique can get very complicated, so a couple more examples are in order here.

Example 226. Consider

$$\int (2y+1)e^{2y^2+2y-5}\, dy.$$

Since the only composition in the integral that is obvious is e^{y^2+y-5}, the substitution to try is

$$w = 2y^2 + 2y - 5.$$

We have

$$dw = (4y+2)\, dy = 2\,(2y+1)\, dy.$$

We can rewrite this as

$$\frac{dw}{2} = (2y+1)\, dy.$$

and use this to replace the expression $(2y+1)\, dy$.

Applying this substitution yields

$$\begin{aligned} \int (2y+1)e^{2y^2+2y-5}\, dy &= \int e^w\, \frac{dw}{2} \\ &= \frac{e^w}{2} + C \\ &= \frac{e^{2y^2+2y-5}}{2} + C. \end{aligned}$$

Remark 16. In Example 226 we could also rewrite the integral as

$$\int (2y+1)e^{2y^2+2y-5}\, dy = \frac{1}{2} \int (4y+2)e^{2y^2+2y-5}\, dy.$$

This is multiplying by $2/2 = 1$. We can then use

$$dw = (4y+2)\, dy$$

directly. This is equivalent to what is done in Example 226.

There are cases where the choice of the u function is not as clear.

Example 227. Consider the integral

$$\int \sin(x)\cos(x)\, dx.$$

If we take $u = \sin(x)$ and $du = \cos(x)\, dx$ we get

$$\begin{aligned} \int \sin(x)\cos(x)\, dx &= \int u\, du \\ &= \frac{u^2}{2} + C \\ &= \frac{(\sin(x))^2}{2} + C. \end{aligned}$$

We could also use $u = \cos(x)$. (Check and see what you get.)

In either case we can interpret the original integral as

$$\int (\sin(x))^1 \cos(x)\,dx.$$

to see the composition that is being used.

A case where there is a substitution that does not use any of the functions directly in the integrand is the antiderivative of $\sec(x)$.

Example 228. Consider

$$\int \sec(x)\,dx.$$

A clever way of evaluating this integral is to multiply $\sec(x)$ by $(sec(x)+\tan(x))/(\sec(x)+\tan(x))$ and use the fact that

$$\frac{d}{dx}\big(\sec(x)+\tan(x)\big) = \sec(x)\,\tan(x)+\sec^2(x)$$

in the substitution $u=\sec(x)+\tan(x)$.

The integral is easy to do with the multiplication and substitution above.

$$\begin{aligned}
\int \sec(x)\,dx &= \int \frac{\sec(x)\big(\sec(x)+\tan(x)\big)}{\sec(x)+\tan(x)}\,dx \\
&= \int \frac{\sec^2(x)+\sec(x)\,\tan(x)}{\sec(x)+\tan(x)}\,dx \\
&= \int \frac{du}{u} \\
&= \ln|u|+C \\
&= \ln\big|\sec(x)+\tan(x)\big|+C.
\end{aligned}$$

Using substitution for definite integrals in a little bit more complicated. The difficulty arises in the limits of integration. If the limits of integration do not change with the variable, we will usually have the wrong value somewhere in the problem.

Example 229. Consider the integral

$$\int_0^{\pi} \sin(2x)\,dx.$$

Since $F(x) = -\cos(2x)/2$ is an antiderivative of $\sin(2x)$,

$$\begin{aligned}
\int_0^{\pi} \sin(2x)\,dx &= -\left.\frac{\cos(2x)}{2}\right|_0^{\pi} \\
&= -\frac{\cos(2\pi)}{2} - \left(-\frac{0}{2}\right) \\
&= 0.
\end{aligned}$$

This is the correct answer since $\sin(x)$ goes through a half cycle of positive and negative values from $x=0$ to $x=\pi$.

If we use the substitution $u = 2x$ and $du/2 = dx$ we have

$$\int \sin(2x)\,dx = \int \frac{\sin(u)}{2}\,du.$$

If the limits of integration are not changed we get

$$\begin{aligned}\int_0^{\pi} \frac{\sin(u)}{2}\,du &= -\left.\frac{\cos(u)}{2}\right|_0^{\pi}\\ &= -\frac{\cos(\pi)}{2} + \frac{\cos(0)}{2}\\ &= 1.\end{aligned}$$

This is not the correct answer.

If we note that when $x = 0$ and $x = \pi$ we get $u = 0$ and $u = 2\pi$, the limits for u are different from the limits for x. Using these limits yields

$$\begin{aligned}\int_0^{2\pi} \frac{\sin(u)}{2}\,du &= -\left.\frac{\cos(u)}{2}\right|_0^{2\pi}\\ &= -\frac{\cos(2\pi)}{2} + \frac{\cos(0)}{2}\\ &= 0.\end{aligned}$$

This is the correct answer.

There are three basic ways to do a substitution for a definite integral. The first is to find an antiderivative for the function and then evaluate it between the limits. The second way is equivalent. It is to carry the limits in terms of the original variable through the substitution and reversion to the original variables. The last method is to change the limits of integration to limits in terms of the new variable. Since the first two are equivalent, only the second is done. The next example shows how to do both of the last two techniques.

Example 230. In this example it is demonstrated how we can evaluate definite integrals when using substitution,

$$\int_{-1}^{4} \frac{x\ln\left(x^2+1\right)}{x^2+1}\,dx$$

A substitution that works for this integral is

$$u = \ln\left(x^2+1\right) \quad \text{with} \quad du = \frac{2x}{x^2+1}\,dx.$$

First the substitution is done with a change of limits. Note that if $x = -1$ then $u = \ln(2)$ and if $x = 4$ then $u = \ln(17)$. Evaluating the integral gives

$$\begin{aligned}\int_{-1}^{4} \frac{x\ln\left(x^2+1\right)}{x^2+1}\,dx &= \int_{\ln(3)}^{\ln(17)} \frac{u}{2}\,du\\ &= \left.\frac{u^2}{4}\right|_{u=\ln(2)}^{\ln(17)}\\ &= \frac{(\ln(17))^2 - (\ln(2))^2}{4}.\end{aligned}$$

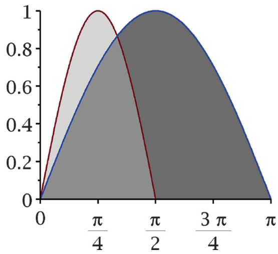


Fig. 6.15 The integral of $\sin(x)$ and $\sin(u)$, $u = 2x$

If we leave the limits in terms of x we get

$$\begin{aligned}\int_{-1}^{4} \frac{x\ln\left(x^2+1\right)}{x^2+1}\,dx &= \int_{x=-1}^{x=4} \frac{u}{2}\,du \\ &= \left.\frac{u^2}{4}\right|_{x=-1}^{x=4} \\ &= \left.\frac{\ln\left(x^2+1\right)^2}{4}\right|_{x=-1}^{x=4} \\ &= \frac{(\ln(17))^2-(\ln(2))^2}{4}.\end{aligned}$$

To finish this section a couple examples are provided that help illustrate the geometry of integration by substitution. The first example is very simple and the second is slightly more complex.

Example 231. Consider the integral

$$\int_0^{\pi/2} \sin(2x)\,dx.$$

The substitution used is $u = 2x$ with limits $u = 0$ when $x = 0$ and $u = \pi$ when $x = \pi/2$. The graphs of the two functions $\sin(2x)$ and $\sin(u)$ over the appropriate intervals are in Fig. 6.15 on page 198.

This transformation takes any interval $[x, x+h]$ in x to the interval $[2x, 2x+2h]$ in u. It doubles the length of each interval and hence doubles the area. See Fig. 6.16 on page 199. In terms of Riemann sums, a sum approximating the integral in terms of x, and $h_i = a_i - a_{i-1}$,

$$S_x = \sum_{i=1}^{n} \sin(2\xi_i)\, h_i$$

becomes, multiplying all partition points by 2,

$$S_u = \sum_{i=1}^{n} \sin(2\xi_i)\, 2h_i = 2\,S_x.$$

Since this is true for any Riemann sum, the area is doubled.

The simple transformation in the previous example does not give a very complete picture of the geometry. The next example gives a case where the lengths of the u intervals are not simply multiples of the lengths of the x intervals. In order to keep the example simple, u is a differentiable, strictly increasing function of x.

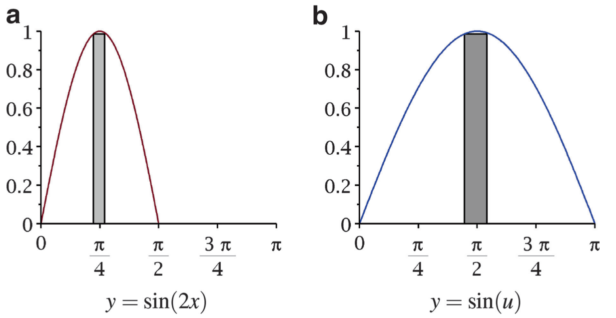


Fig. 6.16

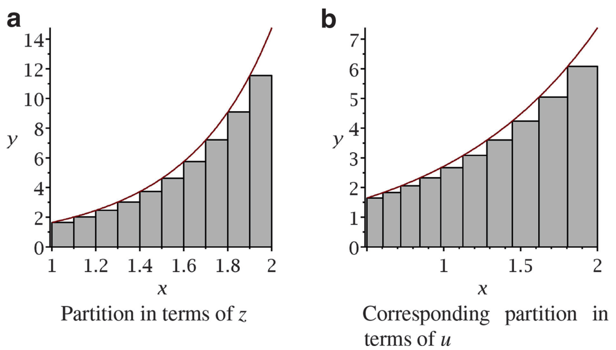


Fig. 6.17

Example 232. Consider the integral

$$\int_1^2 z e^{z^2/2}\,dz.$$

The substitution used here is $u = z^2/2$ and $du = z\,dz$. Consider a partition of $[1,2]$ into 10 equal intervals, $z_0 = 1, z_1 = 1.1, z_2 = 1.2, \ldots, z_{10} = 2$ using left endpoints as illustrated in Fig. 6.17a on page 199. In terms of u these points are $u_0 = 1, u_1 = (1.1)^2, u_2 = (1.2)^2, \ldots, u_{10} = 4$. See Fig. 6.17b on page 199

Writing out the Riemann sums in the first case with $h = 1/10$ gives

$$A \approx \sum_{i=1}^{10} (1+(i-1)h)e^{(1+(i-1)h)^2/2}\,h.$$

In the second case we get

$$\begin{aligned} A_1 &\approx \sum_{i=1}^{10} e^{(1+(i-1)h)^2/2}\left(\frac{(1+ih)^2}{2} - \frac{(1+(i-1)h)^2}{2}\right) \\ &\approx \sum_{i=1}^{10} \frac{e^{(1+(i-1)h)^2/2}}{2}\left(2(1+(i-1)h)h - h^2\right) \end{aligned}$$

Since h^2 is "very small" and does not contribute to the limit as h goes to zero, we can write

$$A_1 \approx \sum_{i=1}^{10} e^{(1+(i-1)h)^2/2}\,(1+(i-1)h)\,h.$$

Since the approximate areas are the same for a given number of intervals, the definite integrals are the same.

$$\begin{aligned}\int_1^2 z e^{\frac{z^2}{2}}\, dz &= \int_{\frac{1}{2}}^2 e^u\, du \\ &= e^u \Big|_{\frac{1}{2}}^2 \\ &= e^2 - e^{\frac{1}{2}}.\end{aligned}$$

Substitution can also be used when we have the composition of a vector valued function, $\mathbf{r}(w)$ and a scalar valued function $s(t)$, $\mathbf{r} \circ s(t)$. If the integral under consideration is of the form

$$\int \mathbf{r}'(s(t))\, s'(t)\, dt,$$

then each of the component integrals is of the form

$$\int r_i'(s(t))\, s'(t)\, dt = r_i(s(t)) + C_i,$$

This means that

$$\int \mathbf{r}'(s(t))\, s'(t)\, dt = \mathbf{r}(s(t)) + \mathbf{C},$$

Example 233. Consider the function

$$\mathbf{r}(t) = 2t \left(\sin(t^2), \cos(t^2), \frac{1}{t^2} \right).$$

Setting $\mathbf{r}'(w) = \left(\sin(w), \cos(w), \frac{1}{w}\right)$, $s(t) = t^2$ and $s'(t) = 2t$ we have

$$\int \mathbf{r}'(s(t))\, s'(t)\, dt = \left(-\cos\left(t^2\right), \sin\left(t^2\right), \ln\left(t^2\right)\right) + \mathbf{C}.$$

Exercises

1. Evaluate the following indefinite integrals.

(a) $\int 4 e^{4y}\, dy$

(b) $\int 2 \sin(2\tau)\, d\tau$

(c) $\int 6 \cos(6x + 3)\, dx$

(d) $\int 4 e^{4z-3}\, dz$

(e) $\int \frac{1}{3x - 4}\, dx$

(f) $\int \csc(\theta)\, d\theta$

(g) $\int \cot(\phi)\, d\phi$

2. Evaluate the following indefinite integrals.

(a) $\int 4z e^{2z^2-3}\, dz$

(b) $\int 6x^2 \sin(2x^3 + 3)\, dx$

(c) $\int (4z - 3)\, e^{2z^2-3z}\, dz$

(d) $\int (6x^2 + 1) \sec^2(2x^3 + x + 1)\, dx$

(e) $\displaystyle\int \frac{4z+1}{2z^2+z+2}\,dz$

(f) $\displaystyle\int \frac{\sin(x)}{\cos^4(x)}\,dx$

3. Evaluate the following indefinite integrals.

(a) $\displaystyle\int \cos(z)\,e^{\sin(z)}\,dz$

(b) $\displaystyle\int e^x \sin(e^x)\,dx$

(c) $\displaystyle\int \frac{\cos(\ln(z))}{z}\,dz$

(d) $\displaystyle\int \sin(x)\sec^2(\cos(x))\,dx$

4. Evaluate the following definite integrals.

(a) $\displaystyle\int_0^1 4\,e^{4y}\,dy$

(b) $\displaystyle\int_{-1}^1 \sin(\pi x)\,dx$

(c) $\displaystyle\int_2^5 e^{3z-2}\,dz$

(d) $\displaystyle\int_{-\pi}^{\pi} \sin(y)\cos(y)\,dy$

(e) $\displaystyle\int_{-2}^0 x\sqrt{x^2+3}\,dx$

(f) $\displaystyle\int_{-2}^3 (2z-2)e^{z^2-2z+1}\,dz$

(g) $\displaystyle\int_{-4}^4 we^{-w^2}\,dw$

(h) $\displaystyle\int_0^2 \frac{w+1}{\sqrt{w^2+2w+3}}\,dw$

(i) $\displaystyle\int_{-2}^4 \frac{z}{1+z^4}\,dz$

5. Show that Definition 27 on page 184 is consistent with what we get when using the substitution $u = b+a-x$ in the integral

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

Chapter 7
The Cross Product and L'Hôpital's Rule

7.1 Determinants and the Cross Product

The cross product is a product defined on two vectors in three dimensions. Two applications of the cross product in science and engineering involve rotational motion and the magnetic field generated by current through a wire. Because of this, the cross product is important. The uses of the cross product when working with planes and the volume of a parallelepiped are more mathematical, but are also quite useful.

The cross product can be defined geometrically, just as the dot product can be defined geometrically. The cross product is only defined in $\mathbb{R}^3$. We can define the cross product in $\mathbb{R}^2$ by adding a zero third coordinate to all vectors. The following definition is the geometric definition of the cross product.

Definition 28. The *cross product* of two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$ is the *unique* vector $\mathbf{v}$ such that $\mathbf{v}$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$, such that $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{v}$ form a right hand system, and such that the length of $\mathbf{v}$ is $\|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin(\theta)$ where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. (See Fig. 7.1 on page 204.)

The cross product is used for finding torque, the equivalent to force in rotational motion, for finding normals to planes, for finding volumes, and for finding a vector perpendicular to a pair of vectors. It is also used in relating current to magnetic fields.

It is nontrivial to calculate the cross product from first principles. We can do this by solving the system of linear equations $\mathbf{x}\cdot\mathbf{a} = 0$ and $\mathbf{x}\cdot\mathbf{b} = 0$. Since this system of equations is two linear equations in 3 unknowns, it either has one solution $\mathbf{0}$ or an infinite number of solutions. The correct choice gives the cross product.

In these notes the cross product will be computed using determinants of matrices. The determinant of a 2×2 array, matrix, is defined by

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

This is easy to compute.

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_7) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_7

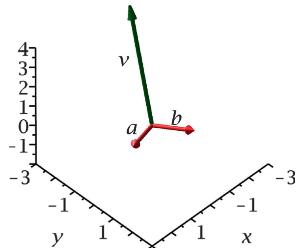


Fig. 7.1 The vector **v** is the cross product of the vectors **a** and **b**

Remark 17. The notation here can be a problem. A matrix can be denoted by either of the following:

$$\begin{pmatrix}1 & 2\\4 & 7\end{pmatrix}, \qquad \begin{bmatrix}1 & 2\\4 & 7\end{bmatrix}.$$

Straight vertical bars denote the determinant,

$$\begin{vmatrix}1 & 2\\4 & 7\end{vmatrix}.$$

Since parentheses can be hard to distinguish from vertical bars, in this book a matrix is always written using square brackets.

Example 234. The determinant of $\begin{bmatrix}2 & 1\\3 & 4\end{bmatrix}$ is

$$\begin{vmatrix}2 & 1\\3 & 4\end{vmatrix} = 2\cdot 4 - 1\cdot 3 = 5.$$

Example 235. The determinant of $\begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}$ is

$$\begin{aligned}\begin{vmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{vmatrix} &= \cos(\theta)\cdot\cos(\theta) - (-\sin(\theta))\cdot\sin(\theta)\\ &= 1.\end{aligned}$$

With the determinant of a 2×2 matrix we can define the cross product of two vectors in the xy-plane. It is a vector in $\mathbb{R}^3$,

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= \left(0, 0, \begin{vmatrix}a_1 & a_2\\ b_1 & b_2\end{vmatrix}\right)\\ &= (0, 0, a_1b_2 - a_2b_1).\end{aligned}$$

or, if we want to stay in $\mathbb{R}^2$, the cross product can be considered a scalar quantity since it is always a multiple of $(0,0,1)$,

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= \begin{vmatrix}a_1 & a_2\\ b_1 & b_2\end{vmatrix}\\ &= a_1b_2 - a_2b_1.\end{aligned}$$

To calculate the cross product in $\mathbb{R}^3$ we can use the determinant of a 3×3 matrix. The determinant of a 3×3 matrix can be calculated using the expressions

$$\begin{aligned}\det\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix} &= \begin{vmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}\\ &= a_1\begin{vmatrix}b_2 & b_3\\ c_2 & c_3\end{vmatrix} - a_2\begin{vmatrix}b_1 & b_3\\ c_1 & c_3\end{vmatrix} + a_3\begin{vmatrix}b_1 & b_2\\ c_1 & c_2\end{vmatrix}\\ &= a_1b_2c_3 - a_1b_3c_2 - a_2b_1c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1.\end{aligned}$$

Example 236. The determinant of $\begin{bmatrix}1 & 3 & 0\\ 2 & 4 & -2\\ -1 & 5 & 6\end{bmatrix}$ is given by

$$\begin{aligned}\begin{vmatrix}1 & 3 & 0\\ 2 & 4 & -2\\ -1 & 5 & 6\end{vmatrix} &= 1\begin{vmatrix}4 & -1\\ 5 & 6\end{vmatrix} - 3\begin{vmatrix}2 & -2\\ -1 & 6\end{vmatrix} + 0\begin{vmatrix}2 & 4\\ -1 & 5\end{vmatrix}\\ &= 1(24+5) - 3(12-2) + 0(10+4)\\ &= 29 - 30 + 0\\ &= -1.\end{aligned}$$

Letting $\hat{\imath} = (1,0,0)$, $\hat{\jmath} = (0,1,0)$, and $\hat{k} = (0,0,1)$ the cross product of $\mathbf{a} = (a_1,a_2,a_3)$ and $\mathbf{b} = (b_1,b_2,b_3)$ is given by

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= \begin{vmatrix}\hat{\imath} & \hat{\jmath} & \hat{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}\\ &= \hat{\imath}\begin{vmatrix}a_2 & a_3\\ b_2 & b_3\end{vmatrix} - \hat{\jmath}\begin{vmatrix}a_1 & a_3\\ b_1 & b_3\end{vmatrix} + \hat{k}\begin{vmatrix}a_1 & a_2\\ b_1 & b_2\end{vmatrix}\\ &= \hat{\imath}(a_2b_3 - a_3b_2) - \hat{\jmath}(a_1b_3 - a_3b_1) + \hat{k}(a_1b_2 - a_2b_1)\\ &= \big(a_2b_3 - a_3b_2,\, a_3b_1 - a_1b_3,\, a_1b_2 - a_2b_1\big). \qquad (7.1)\end{aligned}$$

This is the formula we get for the cross product no matter how we define it. It is reasonable to simply memorize one of the expressions on the right side in the above list to calculate the cross product.

Example 237. Consider the three vectors $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$. All of the vectors are at right angles to each other and in the given order they form a right hand system. This means that $\hat{\imath}\times\hat{\jmath}$ must be in the positive $\hat{k}$ direction with length $\|\hat{\imath}\|\,\|\hat{\jmath}\|\,\sin(\pi/2) = 1$. In other words, $\hat{\imath}\times\hat{\jmath} = \hat{k}$.

If we calculate this using the matrix format we get

$$\begin{aligned}\hat{\imath}\times\hat{\jmath} &= \begin{vmatrix}\hat{\imath} & \hat{\jmath} & \hat{k}\\ 1 & 0 & 0\\ 0 & 1 & 0\end{vmatrix}\\ &= \hat{\imath}\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix} - \hat{\jmath}\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix} + \hat{k}\begin{vmatrix}1 & 0\\ 0 & 1\end{vmatrix}\\ &= \hat{k}.\end{aligned}$$

In a similar manner we have $\hat{\jmath}\times\hat{k} = \hat{\imath}$ and $\hat{k}\times\hat{\imath} = \hat{\jmath}$. Also $\hat{\jmath}\times\hat{\imath} = -\hat{k}$.

Example 238. The cross product of $\mathbf{u} = (2,3,1)$ and $\mathbf{v}(-2,1,-1)$ is

$$\begin{aligned}
\mathbf{u}\times\mathbf{v} &= \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 2 & 3 & 1 \\ -2 & 1 & -1 \end{vmatrix} \\
&= \hat{\imath}\begin{vmatrix} 3 & 1 \\ 1 & -1 \end{vmatrix} - \hat{\jmath}\begin{vmatrix} 2 & 1 \\ -2 & -1 \end{vmatrix} + \hat{k}\begin{vmatrix} 2 & 3 \\ -2 & 1 \end{vmatrix} \\
&= -4\hat{\imath} + 0\hat{\jmath} + 8\hat{k} \\
&= (-4,0,8).
\end{aligned}$$

Before continuing with the geometry of the cross product, we take some time to look at some rules used for computing with cross products. These rules are basically the same as those for the dot product, except that the cross product is not commutative.

Theorem 66. *Let* $\mathbf{a}$*,* $\mathbf{b}$*, and* $\mathbf{c}$ *be vectors in* $\mathbb{R}^3$ *and the* α *be a constant. Then the following hold:*

(1) $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$

(2) $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$ *and* $(\mathbf{a}+\mathbf{b})\times\mathbf{c} = \mathbf{a}\times\mathbf{c}+\mathbf{b}\times\mathbf{c}$

(3) $\alpha(\mathbf{a}\times\mathbf{b}) = (\alpha\mathbf{a})\times\mathbf{b} = \mathbf{a}\times(\alpha\mathbf{b})$

(4) $\dfrac{d}{dt}(\mathbf{a}(t)\times\mathbf{b}(t)) = \left(\dfrac{d}{dt}\mathbf{a}(t)\right)\times\mathbf{b}(t)+\mathbf{a}\times\left(\dfrac{d}{dt}\mathbf{b}(t)\right)$

Proof. All of these results can be done by calculating from the definitions. Because of this, we only consider the first formula. We use the explicit formula for the cross product, Eq. 7.1 on page 205. Through a short sequence of equations we transform the left side of (1) in Theorem 66 to the right side of the equation.

$$\begin{aligned}
\mathbf{a}\times\mathbf{b} &= \big(a_2b_3 - a_3b_2,\, a_3b_1 - a_1b_3,\, a_1b_2 - a_2b_1\big) \\
&= -\big(b_2a_3 - b_3a_2,\, b_3a_1 - b_1a_3,\, b_1a_2 - b_2a_1\big) \\
&= -\mathbf{b}\times\mathbf{a}.
\end{aligned}$$

The following are examples of how the results work.

Example 239. Let $\mathbf{v} = (1,2,-1)$ and $\mathbf{w} = (3,1,1)$, then

$$\mathbf{v}\times\mathbf{w} = \left(\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix}, -\begin{vmatrix} 1 & -1 \\ 3 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix}\right) = (3,4,-5)$$

and

$$\mathbf{w}\times\mathbf{v} = \left(\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix}, -\begin{vmatrix} 3 & 1 \\ 1 & -1 \end{vmatrix}, \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix}\right) = (-3,-4,5).$$

Example 240. Let $\mathbf{a}(t) = \big(t,t^2,t^3\big)$, let $\mathbf{b}(t) = (\cos(\pi t),\sin(\pi t),4t)$ and let $\mathbf{r}(t) = \mathbf{a}(t)\times\mathbf{b}(t)$. Then $\mathbf{a}'(t) = \big(1,2t,3t^2\big)$ and $\mathbf{b}'(t) = (-\pi\sin(\pi t),\pi\cos(\pi t),4)$. At $t=1$ we have $\mathbf{a}(1) = (1,1,1)$, $\mathbf{a}'(1) = (1,2,3)$, $\mathbf{b}(1) = (-1,0,4)$ and $\mathbf{b}'(1) = (0,\pi,4)$.

Combining the information above gives

$$\begin{aligned}
\frac{d}{dt}(\mathbf{a}(t)\times\mathbf{b}(t))\big|_{t=1} &= \mathbf{a}'(1)\times\mathbf{b}(1)+\mathbf{a}(1)\times\mathbf{b}'(1) \\
&= (1,2,3)\times(-1,0,4)+(1,1,1)\times(0,\pi,4) \\
&= (8,-7,2)+(4-\pi,-4,\pi) \\
&= (12-\pi,-11,2+\pi).
\end{aligned}$$

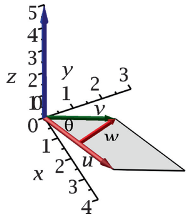


Fig. 7.2 The parallelogram spanned by **u** and **v**

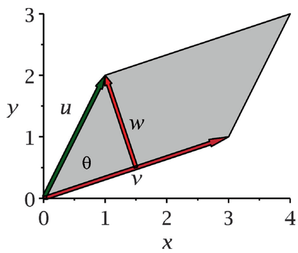


Fig. 7.3 The parallelogram spanned by **u** and **v**

One way of viewing the length of the cross product is as the area of the parallelogram spanned by two vectors. The Fig. 7.2 on age 207 shows the parallelogram spanned by **u** and **v**. The length of the base of the parallelogram $\|\mathbf{u}\|$ and the height is $\|\mathbf{w}\| = \|\mathbf{v}\| \sin(\theta)$.

Now consider the cross product of two vectors in the xy-plane, $\mathbf{u} = (u_1, u_2, 0)$ and $\mathbf{v} = (v_1, v_2, 0)$. The cross product is $(0, 0, u_1v_2 - u_2v_1)$. To support the definition of the cross product as a vector with length $\|\mathbf{u}\|\ \|\mathbf{v}\|\ \sin(\theta)$ we should have that $\|\mathbf{u}\|\ \|\mathbf{v}\|\ \sin(\theta) = |u_1v_2 - u_2v_1|$. See Fig. 7.3 on page 207. The square of the area of the parallelogram is given by

$$\begin{aligned}
Area^2 &= \|\mathbf{u}\|^2\ (\|\mathbf{v}\|\ \sin(\theta))^2 \\
&= \|\mathbf{u}\|^2\ \|\mathbf{v} - \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}\|^2 \\
&= (\mathbf{u}\cdot\mathbf{u}) \left(\mathbf{v}\cdot\mathbf{v} - 2\frac{(\mathbf{v}\cdot\mathbf{u})^2}{\mathbf{u}\cdot\mathbf{u}} + \frac{(\mathbf{v}\cdot\mathbf{u})^2}{\mathbf{u}\cdot\mathbf{u}}\right) \\
&= \|\mathbf{u}\|^2\ \|\mathbf{v}\|^2 - (\mathbf{v}\cdot\mathbf{u})^2 \\
&= (v_1u_2 - v_2u_2)^2 \\
&= \|\mathbf{u}\times\mathbf{v}\|^2.
\end{aligned}$$

Taking the square root of both sides gives the desired result.

Example 241. The area of the parallelogram spanned by the vectors $\mathbf{u} = (1, -3, 5)$ and $\mathbf{v} = (-2, 0, 3)$ is $\|\mathbf{u}\times\mathbf{v}\|$. This is

$$\begin{aligned}
\|\mathbf{u}\times\mathbf{v}\| &= \|(-9, -13, -6)\| \\
&= \sqrt{286}.
\end{aligned}$$

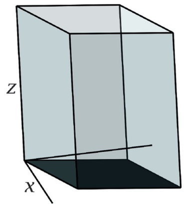


Fig. 7.4 A prism with base in the xy-plane spanned by the two vectors and with height h

Exercises

1. Find the cross products of the following pairs of vectors.

(a) $(2,0,-1)$ and $(2,-1,4)$
(b) $(0,1,2)$ and $(0,4,-2)$
(c) $(1,7,3)$ and $(1,3,-1)$
(d) $(4,-2,1)$ and $(1,-5,1)$
(e) $(3,-2,6)$ and $(-3,2,-6)$
(f) $(6,3,-2)$ and $(-1,2,0)$
(g) $(2,-1)$ and $(3,1)$
(h) $(4,1)$ and $(1,-3)$

2. Find the area of the parallelogram spanned by the following pairs of vectors.

(a) $(2,0,-1)$ and $(2,-1,4)$
(b) $(0,1,2)$ and $(0,4,-2)$
(c) $(1,7,3)$ and $(1,3,-1)$
(d) $(4,-2,1)$ and $(1,-5,1)$
(e) $(3,-2,6)$ and $(-3,2,-6)$
(f) $(6,3,-2)$ and $(-1,2,0)$
(g) $(2,-1)$ and $(3,1)$
(h) $(4,1)$ and $(1,-3)$

3. Use the cross product to find the sine of the angle between the following pairs of vectors.

(a) $(2,0,-1)$ and $(2,-1,4)$
(b) $(0,1,2)$ and $(0,4,-2)$
(c) $(1,7,3)$ and $(1,3,-1)$
(d) $(4,-2,1)$ and $(1,-5,1)$
(e) $(3,-2,6)$ and $(-3,2,-6)$
(f) $(6,3,-2)$ and $(-1,2,0)$
(g) $(2,-1)$ and $(3,1)$
(h) $(4,1)$ and $(1,-3)$

4. Find the areas of the triangles with the following triples of vertices using the cross product.

(a) $(0,0,0)$, $(2,1,3)$, and $(4,1,1)$.
(b) $(0,0,0)$, $(-1,5,-4)$, and $(1,-3,2)$.
(c) $(-1,-1,-1)$, $(2,1,3)$, and $(4,1,1)$.
(d) $(2,4,-2)$, $(5,3,1)$, and $(5,-1,2)$.

5. Find the volume of the prism with base in the xy-plane spanned by the two vectors and with height h. See Fig. 7.4 on page 208.

(a) $(1,1,0)$, $(0,2,0)$, and $h=3$.
(b) $(-1,2,0)$, $(3,2,0)$, and $h=2$.
(c) $(-2,-4,0)$, $(-1,2,0)$, and $h=6$.
(d) $(1,2,0)$, $(-3,0,0)$, and $h=7$.

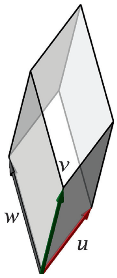


Fig. 7.5 The parallelepiped spanned by **u**, **v**, and **w**

6. Find the determinants of the following matrices.

(a) $\begin{bmatrix} 2 & 3 \\ 4 & -2 \end{bmatrix}$.

(b) $\begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$.

(c) $\begin{bmatrix} 3 & 0 \\ 2 & 0 \end{bmatrix}$.

(d) $\begin{bmatrix} \sin(t) & \cos(t) \\ \cos(t) & -\sin(t) \end{bmatrix}$.

(e) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 4 & 0 & -2 \end{bmatrix}$.

(f) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 1 \\ 1 & -1 & 1 \end{bmatrix}$.

(g) $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 0 \\ 4 & 5 & -4 \end{bmatrix}$.

(h) $\begin{bmatrix} 5 & 1 & -2 \\ 4 & -5 & 2 \\ 1 & 3 & 5 \end{bmatrix}$.

7. Show that the cross product of any two vectors **a** and **b** calculated using the formulas is orthogonal to both **a** and **b**.

7.2 Volume and Torque

This section covers two applications of the cross product, finding volumes of parallelepipeds and calculating torque. The volume of a parallelepiped is used to approximate volumes of various other figures. In rotational mechanics the concept of torque is equivalent to the idea of force in mechanics. Both of these are calculated using the cross product.

A parallelepiped is a six sided volume in three dimensions such that opposite sides are parallel parallelograms. It can also be described as the span of three vectors **u**, **v**, and **w**, see Fig. 7.5 on page 209.

The area of the base of this parallelepiped is the parallelogram spanned by **u** and **w**, see Fig. 7.6a on page 210. The height vector is the projection of **v** onto the cross product of u and w, see Fig. 7.6b on page 210. This means that the volume is given by

$$\begin{aligned} \text{Volume} &= \|\mathbf{u}\times\mathbf{w}\|\,\|\mathbf{h}\|. \\ &= \|\mathbf{u}\times\mathbf{w}\|\,\|\mathbf{v}\|\,|\cos(\phi)| \end{aligned}$$

where ϕ is the angle between $\mathbf{u}\times\mathbf{w}$ and $\mathbf{v}$. This is the absolute value of $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{w})$.

Example 242. Find the volume of the parallelepiped spanned by $\mathbf{u} = (1,-2,3)$, $\mathbf{v} = (-1,1,0)$ and $\mathbf{w} = (0,1,-1)$.

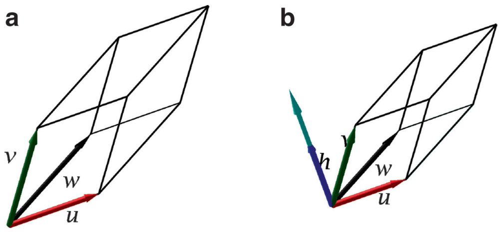


Fig. 7.6

The cross product of **u** and **v** is

$$\begin{aligned}\mathbf{u}\times\mathbf{v} &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & -2 & 3\\ -1 & 1 & 0\end{vmatrix}\\ &= (-3,-3,-1).\end{aligned}$$

The volume of the parallelepiped is

$$\begin{aligned}\text{Volume} &= |(0,1,-1)\cdot(-3,-3,-1)|\\ &= 2.\end{aligned}$$

The expression $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ is called the *triple product* of the vectors **u**, **v**, and **w**. This can be calculated directly using the determinant,

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = \begin{vmatrix}u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{vmatrix}.$$

Using this expression it is easy to calculate the volume of a parallelepiped.

Example 243. The volume of the parallelepiped spanned by $(-2,-2,1)$, $(3,1,2)$, and $(-1,2,5)$ is

$$\begin{aligned}\text{Volume} &= \left|\begin{vmatrix}-2 & -2 & 1\\ 3 & 1 & 2\\ -1 & 2 & 5\end{vmatrix}\right|\\ &= \left|-2\begin{vmatrix}1 & 2\\ 2 & 5\end{vmatrix} + 2\begin{vmatrix}3 & 2\\ -1 & 5\end{vmatrix} + 1\begin{vmatrix}3 & 1\\ -1 & 2\end{vmatrix}\right|\\ &= |-2+34+7|\\ &= 39.\end{aligned}$$

The concept of torque is very useful in the physical sciences and engineering. It is the equivalent of force in rotational motion. We can get an idea of how torque works by considering levers. Consider a lever with a 1 kg block 1 m from the fulcrum and a point at which to push down 2 m from the fulcrum on the other side from the block. If we press down with 4.9 N, the weight of a 1/2 kg mass, the 1 kg mass stays where it is. What this indicates is that the rotational force is directly proportional to the distance from the axis of rotation. See Fig. 7.7 on page 211.

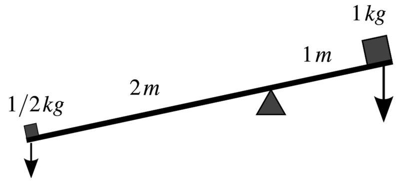


Fig. 7.7 Balancing masses

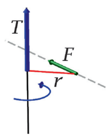


Fig. 7.8

The torque exerted by a force $\mathbf{F}$ is defined at a point $\mathbf{x}$ around an axis of rotation $\mathbf{a}+t\mathbf{v}$ is

$$\mathbf{T} = \mathbf{r} \times \mathbf{F}.$$

Here $\mathbf{r}$ is the vector from the closest point on the axis of rotation to the point $\mathbf{x}$, see Fig. 7.8 on page 211.

Example 244. Assume that a plate in the xy-plane can rotate around the z-axis. If a force $\mathbf{F} = (2,-1,0)$ is applied to the plate at $(1,1,0)$, what is the torque on the plate?

Here the torque is

$$\begin{aligned}\mathbf{T} &= (1,1,0) \times (2,-1,0) \\ &= \left(0,0,\begin{vmatrix}1 & 1 \\ 2 & -1\end{vmatrix}\right) \\ &= (0,0,-3).\end{aligned}$$

Remark 18. Here it is assumed that the force vector is perpendicular to the axis of rotation. If it is not perpendicular to the axis of rotation, two things happen. First, part of the force pushes against the axis of rotation. Second, assuming that the axis of rotation is fixed, the torque is the projection of the cross product onto the axis of rotation.

As with rectilinear forces, the total torque is the sum of the torques applied.

$$\mathbf{T} = \sum_{i=1}^{N} \mathbf{T}_i.$$

Example 245. Consider three forces $\mathbf{F}_1 = (2,3,0)$, $\mathbf{F}_2 = (1,-1,0)$, and $\mathbf{F}_3 = (3,-2,0)$ applied at $(1,1,0)$, $(2,3,0)$ and $(-1,-2,0)$ to a plate in the xy-plane that rotates around the line $x = 1$ and $y = -1$, see Fig. 7.9a on page 212. The torques are

$$\mathbf{T}_1 = \mathbf{F}_1 \times \big((1,1,0) - (1,-1,0)\big) = (2,3,0) \times (0,2,0) = (0,0,4),$$

$$\mathbf{T}_2 = \mathbf{F}_2 \times \big((2,3,0) - (1,-1,0)\big) = (1,-1,0) \times (1,4,0) = (0,0,5),$$

and

$$\mathbf{T}_3 = \mathbf{F}_3 \times \big((-1,-2,0)-(1,-1,0)\big) = (3,-2,0)\times(-2,-1,0) = (0,0,-7).$$

This means the total torque is

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 = (0,0,4)+(0,0,5)+(0,0,-7) = (0,0,2).$$

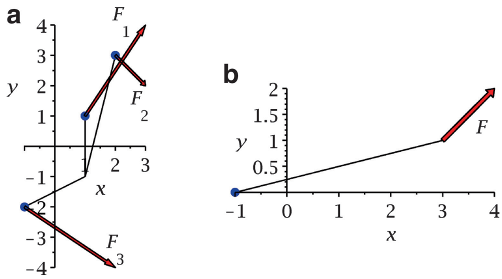


Fig. 7.9

We can also find the magnitude of a force given enough other information.

Example 246. We want to find the force in the direction of $(1,1,0)$ applied at $(3,1,0)$ to a plate in the xy-plane that is to rotate around the point $(-1,0,0)$ when the torque on the plate is $\mathbf{T} = (0,0,-20)$. See, Fig. 7.9b on page 212. The force must have the form $\mathbf{F} = \alpha(1,1,0)$ and the radial vector from the point of application of the force to the point of rotation is $\mathbf{r} = (3,1,0)-(-1,0,0) = (4,1,0)$.

The torque from the force is

$$\begin{aligned}\mathbf{T}_F &= (4,1,0)\times\alpha(1,1,0)\\ &= \alpha\begin{vmatrix}4 & 1\\ 1 & 1\end{vmatrix}\hat{k}\\ &= 3\alpha\hat{k}.\end{aligned}$$

This means that $(0,0,-20) = (0,0,3\alpha)$, or $\alpha = -20/3$. The desired force is $\mathbf{F} = -20(1,1,0)/3$.

Exercises

1. Find the volume of the parallelepiped spanned by the three vectors. Assume all distances are in meters.

(a) $(2,0,-1)$, $(1,1,0)$ and $(2,-1,4)$
(b) $(0,1,2)$, $(4,-3,2)$ and $(0,4,-2)$
(c) $(1,7,3)$, $(3,0,3)$ and $(1,3,-1)$
(d) $(4,-2,1)$, $(2,1,5)$ and $(1,-5,1)$
(e) $(2,5,1)$, $(1,1,3)$ and $(-1,4,-1)$
(f) $(6,1,1)$, $(1,6,1)$ and $(1,1,6)$
(g) $(2,0,0)$, $(0,0,3)$ and $(0,1,0)$
(h) $(0,1,3)$, $(3,0,1)$ and $(4,0,0)$

2. Find the torque generated by a force of 50 N on a plate in the xy-plane if the force is in the same direction as the given $\mathbf{v}$ at the point $\mathbf{p}$, in meters. Assume the plate rotates around the origin.

(a) $\mathbf{v}=(2,-3)$ and $\mathbf{p}=(1,1)$
(b) $\mathbf{v}=(2,-3)$ and $\mathbf{p}=(-1,-1)$
(c) $\mathbf{v}=(-2,1)$ and $\mathbf{p}=(3,0)$
(d) $\mathbf{v}=(5,-10)$ and $\mathbf{p}=(-2,2)$
(e) $\mathbf{v}=(3,1)$ and $\mathbf{p}=(1,1)$
(f) $\mathbf{v}=(6,-7)$ and $\mathbf{p}=(-1,2)$
(g) $\mathbf{v}=(-1,5)$ and $\mathbf{p}=(0,2)$
(h) $\mathbf{v}=(3,0)$ and $\mathbf{p}=(1,0)$

3. In the following you are given a point of rotation $\mathbf{p}$ in the xy-plane, a point $\mathbf{a}$ in the xy-plane where a force in the direction of $\mathbf{v}$ is applied, and a torque $\mathbf{T}$ that the force yields for rotation around the point $\mathbf{p}$. Find the force in the same direction as $\mathbf{v}$ that yields the torque. Use standard mks units.

(a) $\mathbf{p}=(0,0,0)$, $\mathbf{a}=(1,0,0)$, $\mathbf{v}=(0,-1,0)$, $T=(0,0,-4)$
(b) $\mathbf{p}=(0,0,0)$, $\mathbf{a}=(1,1,0)$, $\mathbf{v}=(-1,1,0)$, $T=(0,0,\sqrt{2})$
(c) $\mathbf{p}=(0,0,0)$, $\mathbf{a}=(1,\sqrt{3},0)$, $\mathbf{v}=(0,1,0)$, $T=(0,0,3)$
(d) $\mathbf{p}=(1,1,0)$, $\mathbf{a}=(0,0,0)$, $\mathbf{v}=(0,1,0)$, $T=(0,0,-1)$

4. A force $\mathbf{F}$ in the xy-plane is applied to a plate in the xy-plane at a point $\mathbf{p}$. Show that the same force applied at any point on the line $\ell(\alpha)=\mathbf{p}+\alpha\mathbf{F}$ gives the same torque. The line $\ell(\alpha)$ is call the *line of action* of $\mathbf{F}$.

5. A parallelepiped has one vertex at the origin and is spanned by the vectors $\mathbf{a}=(1,0,1)$ m, $\mathbf{b}=(-1,0,1)$ m, and $\mathbf{c}=\left(0,1,\frac{1}{5}\right)$ m. If $\mathbf{c}$ is moving in the direction $(-1,1,1)$ at $\frac{1}{4}$ m/s, how fast is the volume of the parallelepiped changing?

7.3 Planes

The cross product, along with the dot product, is important for working with planes in three dimensions. Before taking calculus students may have seen two different views of planes. Hopefully students are familiar with the general form of an equation for a plane in $\mathbb{R}^3$,

$$Ax+By+Cz=D$$

for constants A, B, C, and D and for coordinates x, y, and z. It is usually stated in some class before high school that three points in $\mathbb{R}^3$ define a plane. Unfortunately, students frequently do not understand how we get a plane from the three points or how the three point characterization of a plane relates to the equation of a plane.

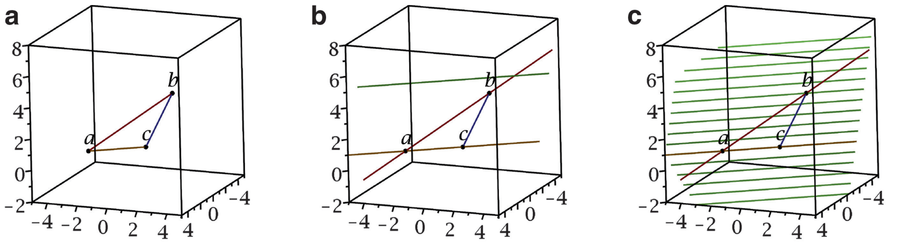
We first consider the three point characterization of a plane. Assume that three distinct non-collinear points in a plane are $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, See Fig. 7.10a on page 214. We can parametrize the lines through $\mathbf{a}$ and $\mathbf{b}$ and through $\mathbf{a}$ and $\mathbf{c}$ as

$$\ell_1(t)=\mathbf{a}+t(\mathbf{b}-\mathbf{a}) \qquad \text{and} \qquad \ell_2(s)=\mathbf{a}+s(\mathbf{c}-\mathbf{a}).$$

The second line can be replaced by any line of the form $\ell_2(s)=\mathbf{a}_1+s(\mathbf{c}-\mathbf{a})$ where $\mathbf{a}_1=\mathbf{a}+t_0(\mathbf{b}-\mathbf{a})$ is a point on the first line for a given t_0. See Fig. 7.10b on page 214.

One way of viewing this is that we are running the second line $\ell_2(s)$ along the first line $\ell_1(t)$. The result is a plane whose points are of the form

$$\mathbf{x}(s,t)=\mathbf{a}+s(\mathbf{b}-\mathbf{a})+t(\mathbf{c}-\mathbf{a}) \tag{7.2}$$



a Three points in $\mathbb{R}^3$ form a triangle **b** Three points in $\mathbb{R}^3$ define 2 lines **c** Running one line along another line builds a plane

Fig. 7.10

where t and s vary over all of $\mathbb{R}$. This process is illustrated in Fig. 7.10c on page 214. This representation is called a *parametrization* of a plane.

Example 247. To find a parametrization of the plane containing the points $\mathbf{a} = (2,-5,6)$, $\mathbf{b} = (1,6,-2)$, and $\mathbf{c} = (6,3,7)$ we find two distinct directions in the plane, say $\mathbf{b}-\mathbf{a} = (-1,11,-8)$ and $\mathbf{c}-\mathbf{a} = (4,-8,1)$. We can then write every point in the plane as

$$\mathbf{x} = (2,-5,6) + s(-1,11,-8) + t(4,-8,1).$$

Rewriting Eq. 7.2 in terms of the vectors $\mathbf{v} = \mathbf{b}-\mathbf{a}$ and $\mathbf{w} = \mathbf{c}-\mathbf{a}$ we have the general form for a parametrization of a plane

$$\mathbf{x} = \mathbf{a} + s\mathbf{w} + t\mathbf{v}.$$

This says that a plane is defined by a single point in the plane and two distinct nonzero directions that are in the plane.

Example 248. The plane containing the point $(2,-5,1)$ and the directions $\mathbf{v} = (0,1,2)$ and $\mathbf{w} = (-2,1,0)$ has a parametrization

$$\mathbf{x} = (2,-5,1) + s(0,1,2) + t(-2,1,0).$$

The relationship between the general equation of a plane, $Ax+By+Cz = D$ and a parametrization of a plane is fairly simple. If we have a plane parametrized by

$$\mathbf{x} = \mathbf{a} + s\,\mathbf{b} + t\,\mathbf{c},$$

and a nonzero vector $\mathbf{n}$ that is orthogonal to both $\mathbf{b}$ and $\mathbf{c}$, then, for any $\mathbf{x}$ in the plane,

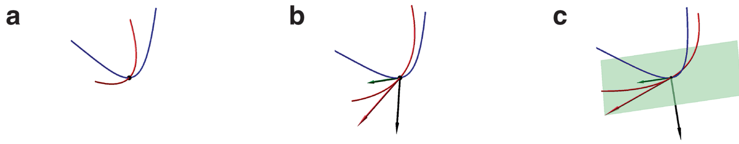
$$\begin{aligned}\mathbf{n}\cdot\mathbf{x} &= \mathbf{n}\cdot(\mathbf{a} + s\,\mathbf{b} + t\,\mathbf{c})\\ &= \mathbf{n}\cdot\mathbf{a} + s(\mathbf{n}\cdot\mathbf{b}) + t(\mathbf{n}\cdot\mathbf{c})\\ &= \mathbf{n}\cdot\mathbf{a}.\end{aligned}$$

Any such $\mathbf{n}$ is called a *normal to the plane*. Letting $\mathbf{n} = (A,B,C)$ and $\mathbf{n}\cdot\mathbf{a} = D$ means that for every $\mathbf{x} = (x,y,z)$ in the plane we have

$$Ax + By + Cz = D.$$

Example 249. Consider the plane parametrized by

$$\mathbf{x} = (2,-5,1) + s(0,1,2) + t(-2,1,0).$$



Two curves that intersect in $\mathbb{R}^3$

The tangents to the curves give a normal to both curves

The plane through the intersection with normal that is normal to both curves is tangent to both curves

Fig. 7.11

A vector perpendicular to both $(0,1,2)$ and $(-2,1,0)$ is

$$(0,1,2)\times(-2,1,0)=(-2,-4,2).$$

This means that an equation for the plane is

$$(-2,-4,2)\cdot(x,y,z)=(-2,-4,2)\cdot(2,-5,1)$$

or

$$-2x-4y+2z=18.$$

Example 250. We say that a curve $\mathbf{r}(t)$ from $\mathbb{R}$ to $\mathbb{R}^3$ is tangent to a plane at a point t_0 if $\mathbf{r}(t_0)$ is in the plane and $\mathbf{r}'(t_0)$ is a direction in the plane. The curves $\mathbf{r}(t)=(t^2,2t+2,t^2-2t+1)$ and $\mathbf{s}(w)=(e^w,4\cos(w),w^2)$ meet at the point $\mathbf{A}=(1,4,0)$ with $\mathbf{r}(1)=\mathbf{A}$ and $\mathbf{s}(0)=\mathbf{A}$ as in Fig. 7.11a on page 215.

Since the derivatives of the functions $\mathbf{r}(t)$ and $\mathbf{s}(w)$,

$$\begin{aligned}\mathbf{r}'(1)&=(2t,\,2,\,2t-2)\,\big|_{t=1}\\&=(2,\,2,\,0)\,,\end{aligned}$$

and

$$\begin{aligned}\mathbf{s}'(0)&=(e^w,\,-4\sin(w)\,2w)\,\big|_{w=0}\\&=(1,\,0,\,0)\,,\end{aligned}$$

are not in the same direction at $\mathbf{A}$, their cross product must be normal to the plane tangent to both curves at $\mathbf{A}$. (Why does such a plane exist?) See Fig. 7.11b on page 215.

The normal to the plane is

$$\begin{aligned}\mathbf{n}&=(2,2,0)\times(1,0,0)\\&=(0,0,-2).\end{aligned}$$

The equation of the plane is

$$(0,0,-2)\cdot(x,y,z)=(0,0,-2)\cdot(1,4,0),$$

or

$$z=0.$$

This is illustrated in Fig. 7.11c on page 215.

Another concept that is useful is the idea of a curve being perpendicular, normal, to a plane. This is used when looking at fluid flow and other topics in engineering and science.

Example 251. A curve $\mathbf{r}(t)$ is said to be perpendicular to a plane $Ax+By+Cz=D$ at a point $\mathbf{r}(t_0)$ in the plane if $\mathbf{r}'(t_0)$ is a normal to the plane. Consider the curve $\mathbf{r}(t)=(t^2,2t+2,t^2-2t+1)$ at $t_0=1$. The point in the plane is $\mathbf{r}(1)=(1,4,0)$ and the normal vector is $\mathbf{n}=\mathbf{r}'(1)=(2,2,0)$. Thus the normal plane is

$$2x+2y=10.$$

See Fig. 7.12 on page 216.

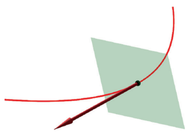


Fig. 7.12 A plane tangent to a curve

Exercises

1. Find a parametrization of the plane containing the three points **a**, **b**, and **c**.

(a) $\mathbf{a}=(2,1,1)$, $\mathbf{b}=(0,0,1)$, and $\mathbf{c}=(1,2,1)$.
(b) $\mathbf{a}=(2,2,1)$, $\mathbf{b}=(2,-1,1)$, and $\mathbf{c}=(2,3,-3)$.
(c) $\mathbf{a}=(1,1,1)$, $\mathbf{b}=(2,1,2)$, and $\mathbf{c}=(-2,1,2)$.
(d) $\mathbf{a}=(3,-1,2)$, $\mathbf{b}=(1,4,2)$, and $\mathbf{c}=(-3,3,1)$.
(e) $\mathbf{a}=(4,1,2)$, $\mathbf{b}=(-5,-6,-2)$, and $\mathbf{c}=(-2,1,-2)$.
(f) $\mathbf{a}=(3,2,5)$, $\mathbf{b}=(1,4,1)$, and $\mathbf{c}=(2,2,3)$.

2. Find an equation of the plane containing the three points **a**, **b**, and **c**.

(a) $\mathbf{a}=(2,1,1)$, $\mathbf{b}=(0,0,1)$, and $\mathbf{c}=(1,2,1)$.
(b) $\mathbf{a}=(2,2,1)$, $\mathbf{b}=(2,-1,1)$, and $\mathbf{c}=(2,3,-3)$.
(c) $\mathbf{a}=(1,1,1)$, $\mathbf{b}=(2,1,2)$, and $\mathbf{c}=(-2,1,2)$.
(d) $\mathbf{a}=(3,-1,2)$, $\mathbf{b}=(1,4,2)$, and $\mathbf{c}=(-3,3,1)$.
(e) $\mathbf{a}=(4,1,2)$, $\mathbf{b}=(-5,-6,-2)$, and $\mathbf{c}=(-2,1,-2)$.
(f) $\mathbf{a}=(3,2,5)$, $\mathbf{b}=(1,4,1)$, and $\mathbf{c}=(2,3,3)$.

3. Find the plane perpendicular to the given curve at the given t_0.

(a) $\mathbf{r}(t)=\left(t,t^2,t^3\right)$ at $t_0=1$
(b) $\mathbf{r}(t)=(\sin(t),\cos(t),t)$ at $t_0=\frac{\pi}{4}$
(c) $\mathbf{s}(t)=\left(e^t,e^{2t},6t+2\right)$ at $t_0=0$
(d) $\mathbf{s}(t)=(4t,2t-1,5t+5)$ at $t_0=-2$
(e) $\mathbf{v}(t)=\left(\ln(t),t,e^t\right)$ at $t_0=1$
(f) $\mathbf{v}(t)=(\tan(t),\cot(t),\sec(t))$ at $t_0=\frac{\pi}{4}$
(g) $\mathbf{v}(t)=\left(\tan^{-1}(t),\cot^{-1}(t),t^5\right)$ at $t_0=1$
(h) $\mathbf{v}(t)=\left(\cos(3t),\sin(t^2/\pi),\sin(2t)\right)$ at $t_0=\frac{\pi}{4}$

4. Find an equation for the plane tangent to the two curves at the given point.

(a) $\mathbf{r}(t) = \left(t^2, t, t^3\right)$, $\mathbf{s}(u) = (\cos(u), e^u, 2u+1)$, $\mathbf{a} = (1,1,1)$

(b) $\mathbf{r}(t) = (\cos(t), \sin(t), t)$, $\mathbf{s}(w) = \left(e^{(w-1)}, w^2 - 1, \ln(w)\right)$, $\mathbf{a} = (1,0,0)$

(c) $\mathbf{x}(t) = \left(\cos(2t), \ln\left(\frac{t}{\pi}\right), t\right)$, $\mathbf{y}(s) = \left(\sin(\pi s), 2s-1, 4\pi s^2\right)$, $\mathbf{a} = (1,0,\pi)$

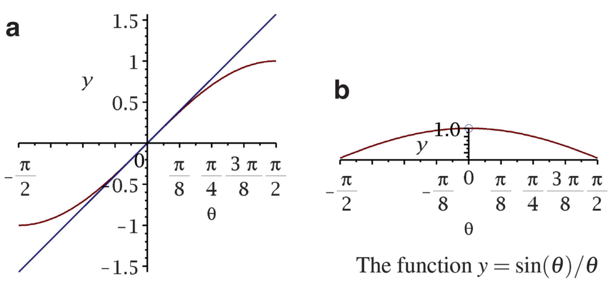
7.4 L'Hôpital's Rule

In this section we consider a tool used to find limits where the techniques we have used before do not work very well. The material in this section can be used in Chaps. 8 and 10. We consider limits that involve indeterminate forms, i.e. forms that do not have well defined values. A simple example of an indeterminate form is the limit

$$\lim_{x\to 2} \frac{x^2 - 6x + 8}{x^2 - 5x + 6}. \tag{7.3}$$

Here the limits of both the numerator and the denominator are 0 at $x = 2$. This means that the form we get when trying to evaluate the limit as a quotient of limits is $0/0$. The form $0/0$ does not have a fixed meaning since, for any number a,

$$\lim_{x\to 0} \frac{ax}{x}$$



The functions $y = \sin(\theta)$ and $y = \theta$

The function $y = \sin(\theta)/\theta$

Fig. 7.13

has the form $0/0$ at $x = 0$ but has limit a. This means that we can get any number a out of the form $0/0$.

To take the limit in Eq. 7.3 we simply cancel the factor $x - 2$ from the numerator and the denominator to get a form that is not $0/0$.

$$\begin{aligned}\lim_{x\to 2}\frac{x^2-6x+8}{x^2-5x+6} &= \lim_{x\to 2}\frac{(x-2)(x-4)}{(x-2)(x-3)}\\ &= \lim_{x\to 2}\frac{x-4}{x-3}\\ &= 2.\end{aligned}$$

The problem is that we cannot simply factor all functions that lead to an indeterminate form. A simple example is the limit

$$\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta}$$

where, if $\theta = 0$, both the numerator and denominator are 0, yielding an indeterminate form. See Fig. 7.13 on page 217.

In the case of $f(\theta) = \sin(\theta)/\theta$ both the numerator and the denominator have continuous derivatives at $\theta = 0$ with the derivative of the denominator equaling 1. This means that we can write, using the Mean Value Theorem 49,

$$\begin{aligned}f(\theta) &= \frac{\sin(\theta)}{\theta}\\ &= \frac{\sin(0)+\cos(\xi_\theta)(\theta-0)}{\theta}\\ &= \cos(\xi_\theta)\end{aligned}$$

for some ξ_θ between θ and 0. Since $\lim_{\theta\to 0}\cos(\theta) = 1$ and ξ_θ is closer to 0 than θ, $\lim_{\theta\to 0}\cos(\xi_\theta) = 1$. This gives us

$$\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1.$$

The proof of the following theorem is somewhat more complicated than the reasoning above and is left for later courses. This theorem gives us one case when the ideas above can be used.

Theorem 67 (L'Hôpital's Rule I). *Let f and g be functions that have first derivatives on an open interval containing a with $f(a) = g(a) = 0$. Assume that*

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

then

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

Remark 19. In the justification of this result a two sided limit was used. In fact, the same argument works if we take a one-sided limit. In this case with the indeterminate form $0/0$ we have, for example, if

$$\lim_{x\to a^+}\frac{f'(x)}{g'(x)} = L,$$

then

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

In the versions of L'Hôpital's rule below, the same comment holds when a is finite.

We proceed with two examples to show how this works.

Example 252. Consider the function $h(z) = \ln(z)/(z^2 - 1)$ and the limit

$$\lim_{z \to 1} h(z).$$

If we take $f(z) = \ln(z)$ and $g(x) = z^2 - 1$ then $f(1) = g(1) = 0$ and both $f(z)$ and $g(z)$ have continuous derivatives at $z = 1$. The limit of the ratio of the derivatives is

$$\lim_{z \to 1} \frac{\frac{1}{z}}{2z} = \lim_{z \to 1} \frac{1}{2z^2} = \frac{1}{2}.$$

Since the conditions for Theorem 67 are satisfied we have

$$\lim_{z \to 1} \frac{\ln(z)}{z^2 - 1} = \frac{1}{2}.$$

Remark 20. The notation and logic used in the last example is rather cumbersome. For this reason many people will use the following, or similar, notation for applying L'Hôpital's rule in the example above.

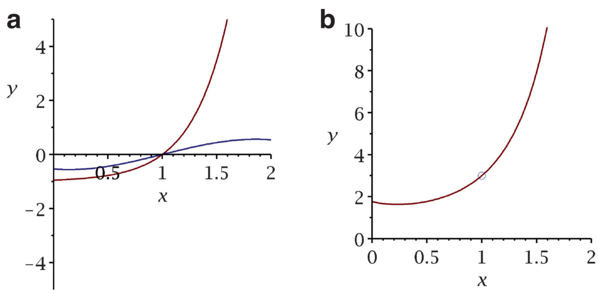
$$\begin{aligned} \lim_{z \to 1} \frac{\ln(z)}{z^2 - 1} &\overset{(L)}{=} \lim_{z \to 1} \frac{\frac{1}{z}}{2z} \\ &= \frac{1}{2}. \end{aligned}$$

Note how $\overset{(L)}{=}$ is used to indicate both that L'Hôpital's rule is being used and that the right side of the equation may not exist.

To simplify our notation, we adopt this convention.

Example 253. We must be careful to check that we have an indeterminate form before applying L'Hôpital's rule . Consider the limit

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



Two nice functions that are zero at $x = 1$

The ratio of the two nice functions

Fig. 7.14

Since the numerator is 0 at $x = -1$ and the denominator is -1 at $x = -1$, L'Hôpital's rule does not apply. The limit evaluates as $\lim_{x\to -1} (x+1)/\cos(\pi x) = 0/-1 = 0$.

If we were to apply L'Hôpital's rule to this limit by taking the limit of the quotient of the derivatives of the numerator and the denominator we would get the limit

$$\lim_{x\to -1} \frac{1}{-\pi \sin(\pi x)},$$

which does not exist.

L'Hôpital's rule can be used multiple times to find limits.

Example 254. Consider the limit as $t \to 0$ of $(1-\cos^2(t))/t^2$. This has the form $0/0$ at $t = 0$ and the ratio of the derivatives, $(\sin(t)\cos(t))/2t$, also has the form $0/0$ at $t = 0$. We can apply L'Hôpital's rule twice to find the limit.

$$\begin{aligned}\lim_{t\to 0} \frac{1-\cos^2(t)}{t^2} &\overset{(L)}{=} \lim_{t\to 0} \frac{\sin(t)\cos(t)}{2t} \\ &\overset{(L)}{=} \lim_{t\to 0} \frac{\cos^2(t)-\sin^2(t)}{2} \\ &= \frac{1}{2}.\end{aligned}$$

We can also justify L'Hôpital's rule for the form $0/0$ using Taylor polynomials in a simple case. If we use Taylor polynomials to approximate functions $f(x)$ and $g(x)$ around a zero, c, of both functions where both functions have non-zero derivatives at c we can write

$$\begin{aligned} f(x) &= f'(c)\,(x-c) + f''(\xi_1)\frac{(x-c)^2}{2} \quad \text{and} \\ g(x) &= g'(c)\,(x-c) + g''(\xi_2)\frac{(x-c)^2}{2}\end{aligned}$$

for some ξ_i between c and x. (This includes an error term that we will not justify here. It depends on the existence of a continuous second derivative in our setting.) See Fig. 7.14 on page 219 for an illustration.

Using this we can write

$$\frac{f(x)}{g(x)} = \frac{f'(c)\,(x-c) + f''(\xi_1)\frac{(x-c)^2}{2}}{g'(c)\,(x-c) + g''(\xi_2)\frac{(x-c)^2}{2}}.$$

Taking the limit as $x \to c$ we get

$$\begin{aligned}\lim_{x\to c} \frac{f(x)}{g(x)} &= \lim_{x\to c} \frac{f'(c) + f''(\xi_1)\frac{(x-c)}{2}}{g'(c) + g''(\xi_2)\frac{(x-c)}{2}} \\ &= \frac{f'(c)}{g'(c)}.\end{aligned}$$

We can apply L'Hôpital's rule to other indeterminate forms and we can apply L'Hôpital's rule when we get an indeterminate form as x goes to $\pm\infty$. Each of these cases is slightly different, but all of them can be justified by rewriting the functions under consideration.

First we consider the case of the indeterminate form $\pm\infty/\pm\infty$ for a function $f(x)/g(x)$ as $x \to a$. In this case we must have $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$. This means that the function, which is a simple rewrite of the original function,

$$\frac{\frac{1}{g(x)}}{\frac{1}{f(x)}}$$

has the form $0/0$. By L'Hôpital's rule the limit of this function exists if the ratio of the derivatives has a limit and then the limits are the same.

If we assume that $\lim_{x\to a} f(x)/g(x) = L_1$ and $\lim_{x\to a} f'(x)/g'(x) = L_2$ both exist, both are non zero, and both are finite then,

$$\begin{aligned}
\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{\frac{1}{g(x)}}{\frac{1}{f(x)}} \\
&\overset{(L)}{=} \lim_{x\to a} \frac{\frac{g'(x)}{g^2(x)}}{\frac{f'(x)}{f^2(x)}} \\
&= \lim_{x\to a} \frac{g'(x)}{f'(x)} \frac{f^2(x)}{g^2(x)}
\end{aligned}$$

This implies that $L_1 = L_1^2/L_2$ or $L_1 = L_2$. This type of argument can be made rigorous and it can made to work if the limit is 0 or $\pm\infty$. This gives our second version of L'Hôpital's rule.

Theorem 68 (L'Hôpital's Rule II). *Assume that $f(x)$ and $g(x)$ are defined on an open interval around a, possibly excluding a. Also assume that* $\lim_{a\to x} f(x) = \pm\infty$ *and that* $\lim_{a\to x} g(x) = \pm\infty$. *If*

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

then

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

Remark 21. As with the first version of L'Hôpital's rule, Theorem 67, this also applies to one-sided limits.

This has some interesting uses.

Example 255. Consider the limit $\lim_{x\to 0^+} x\ln(x)$. This can be rewritten as

$$\begin{aligned}
\lim_{x\to 0^+} x\ln(x) &= \lim_{x\to 0^+} \frac{\ln(x)}{\frac{1}{x}}, \qquad \left(\text{Form}\frac{\infty}{\infty}\right) \\
&\overset{(L)}{=} \lim_{x\to 0^+} \frac{\frac{1}{x}}{\frac{-1}{x^2}} \\
&= \lim_{x\to 0^+} -x \\
&= 0.
\end{aligned}$$

Remark 22. Notice that in the preceding example the limit was not set equal to the form ∞/∞. This was not done since the form ∞/∞ has no fixed value and therefore does not exist. However, it was noted as a check that the form of the limit is indeterminate by explicitly stating the form of the limit. This can be very useful to make certain that L'Hôpital's rule applies to the limit we are evaluating.

L'Hôpital's rule also applies when we consider limits with $x \to \pm\infty$.

Theorem 69 (L'Hôpital Rule III). *Assume that $f(x)$ and $g(x)$ are defined and differentiable on an interval (a,∞) or $(-\infty,b)$. Also assume that $f(x)/g(x)$ has one of the indeterminate forms $0/0$ or $\pm\infty/\pm\infty$ as $x \to \pm\infty$. If*

$$\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = L,$$

then

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L,$$

or, if

$$\lim_{x\to-\infty} \frac{f'(x)}{g'(x)} = L,$$

then

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

This form of L'Hôpital's rule can be used to compare growth rates of functions as $x \to \infty$.

Example 256. The function $f(x) = \exp(x)$ grows faster than any polynomial as $x \to \infty$. This is true if $\lim_{x\to\infty} \exp(x)/x^n = \infty$ for any integer $n \geq 0$. We show this using induction. First we consider the case of a polynomial of degree 0, $p(x) = 1$, a constant function. Here

$$\lim_{x\to\infty} \frac{e^x}{1} = \lim_{x\to\infty} e^x = \infty.$$

Now we assume that $\lim_{x\to\infty} e^x/x^n = \infty$ and show that $\lim_{x\to\infty} e^x/x^{n+1} = \infty$. In this case, noting that $\lim_{x\to\infty} x^n = \infty$ and $\lim_{x\to\infty} e^x = \infty$,

$$\begin{aligned}
\lim_{x\to\infty} \frac{e^x}{x^{n+1}} &\overset{(L)}{=} \lim_{x\to\infty} \frac{e^x}{(n+1)\,x^n} \\
&= \frac{1}{n+1} \lim_{x\to\infty} \frac{e^x}{x^n} \\
&= \infty.
\end{aligned}$$

By the induction hypothesis, $\lim_{x\to\infty} e^x/x^n = \infty$ for all nonnegative integers n and $\exp(x)$ grows faster than x^n for any n. It is left to the reader to justify that this implies that $\exp(x)$ grows faster than every polynomial.

Some other indeterminate forms can be changed to a form where L'Hôpital's rule can be used. The basic method for dealing with the form $0 \cdot \infty$ was illustrated in Example 255 on page 221. Here, if $\lim_{x\to a} f(x) = \pm\infty$ we have $\lim_{x\to a} 1/f(x) = 0$ and if $\lim_{x\to a} f(x) = 0$ we *may have* $\lim_{x\to a} 1/f(x) = \infty$. Using these we can translate the form to $0/0$ or ∞/∞.

Example 257. If we rewrite the limit $\lim_{x\to\infty} x \sin(1/x)$, the form $0 \cdot \infty$, as

$$\lim_{x\to\infty} \frac{\sin\left(\frac{1}{x}\right)}{\frac{1}{x}}$$

we get the form $0/0$ and can apply L'Hôpital's rule .

$$\begin{aligned}
\lim_{x\to\infty} x\sin\left(\frac{1}{x}\right) &= \lim_{x\to\infty} \frac{\sin\left(\frac{1}{x}\right)}{\frac{1}{x}} \\
&\overset{(L)}{=} \lim_{x\to\infty} \frac{\cos\left(\frac{1}{x}\right)\frac{-1}{x^2}}{\frac{-1}{x^2}} \\
&= \lim_{x\to\infty} \cos\left(\frac{1}{x}\right) \\
&= 1.
\end{aligned}$$

We could also have used the substitution $y = 1/x$ and the fact that as $x \to \infty$, $1/x$ goes to 0 from the positive side. This translation gives us

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

This limit was considered before Theorem 67 on page 218.

Somewhat more difficult is applying L'Hôpital's rule to the forms 0^0, ∞^0, and 1^∞. In these cases the form arises from a limit of the form $\lim_{x\to a} f(x)^{g(x)}$. Using the fact that $\exp(y)$ is continuous for all y, we can apply Theorem 20 on page 46 to the limit

$$\lim_{x\to a} \exp\left(g(x)\,\ln(f(x))\right).$$

This requires the application of L'Hôpital's rule to

$$\lim_{x\to a} \left(g(x)\,\ln(f(x))\right). \tag{7.4}$$

A few examples illustrate how this is done.

Example 258. The limit $\lim_{x\to-\infty}(1+1/x)^x$ has the form $1^{-\infty}$. Taking the natural log of this function and taking the limit gives

$$\begin{aligned}
\lim_{x\to-\infty} \ln\left(\left(1+\frac{1}{x}\right)^x\right) &= \lim_{x\to-\infty} x\ln\left(1+\frac{1}{x}\right) \\
&= \lim_{x\to-\infty} \frac{\ln\left(1+\frac{1}{x}\right)}{\frac{1}{x}} \\
&\overset{(L)}{=} \lim_{x\to-\infty} \frac{\frac{-1}{x^2}\frac{1}{1+\frac{1}{x}}}{\frac{-1}{x^2}} \\
&= \lim_{x\to-\infty} \frac{1}{1+\frac{1}{x}} \\
&= 1.
\end{aligned}$$

This means that

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

Example 259. A limit that has the form 0^0 is

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

Rewriting this in the form of Eq. 7.4 we obtain the limit

$$\lim_{x\to 0^+} \ln(x)\,\sin(x).$$

This is of the form $(-\infty)\,0$, which we change into

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

Now we have the form $-\infty/\infty$.

Applying L'Hôpital's rule to this last limit gives

$$\begin{aligned}
\lim_{x\to 0^+} \ln(x)\,\sin(x) &= \lim_{x\to 0^+} \frac{\ln(x)}{\frac{1}{\sin(x)}}\\
&\overset{(L)}{=} \lim_{x\to 0^+} \frac{\frac{1}{x}}{\frac{-\cos(x)}{\sin^2(x)}}\\
&= \lim_{x\to 0^+} \left(\sin(x)\,\frac{-1}{\cos(x)}\,\frac{\sin(x)}{x}\right)\\
&= 0.
\end{aligned}$$

The last equality is true since $\lim_{x\to 0^+}\sin(x)=0$, $\lim_{x\to 0^+} -1/\cos(x) = -1$, and $\lim_{x\to 0^+}\sin(x)/x = 1$. We now have

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

Example 260. The limit

$$\lim_{x\to\infty} x^{\left(\frac{1}{x}\right)}.$$

has the form ∞^0. Again, we take the limit in Eq. 7.4,

$$\begin{aligned}
\lim_{x\to\infty} \frac{\ln(x)}{x} &\overset{(L)}{=} \lim_{x\to\infty} \frac{\frac{1}{x}}{1} \qquad \left(\text{Form } \frac{\infty}{\infty}\right)\\
&= 0.
\end{aligned}$$

This means that

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

This will be used for the root test for convergence of series, Sect. 10.5.

The form $\infty - \infty$ is also indeterminate. For example, if we take $f(x) = x + c$ and $g(x) = x$ for any number c, then $\lim_{x\to\infty}\big(f(x)-g(x)\big)$ has the form $\infty-\infty$. In this case $\lim_{x\to\infty}\big(f(x)$

$-g(x)\big)=c$, an arbitrary real number. On the other hand, if we take $h(y) = y^2$ and $s(y) = y$, we again have the form $\infty - \infty$. In this case we get $\lim_{y\to\infty}\big(h(y) - s(y)\big) = \infty$ and $\lim_{y\to\infty}\big(s(y) - h(y)\big) = -\infty$. Clearly we can get any value from the form $\infty - \infty$.

In order to evaluate this indeterminate form using L'Hôpital's rule we need to change this form into either the form $0/0$ or the form ∞/∞. This can be difficult and is often done using methods that are specific to the limit under consideration.

Example 261. Consider the function $h(y) = y - \ln(y)$ as y goes to infinity. We can factor y out of this expression to get $h(y) = y\,(1 - \ln(y)/y)$. Since $\lim_{y\to\infty} \ln(y)/y = 0$, this can be done using L'Hôpital's rule, we have $h(y) > 1/2\,y$ if $y > M$ for some positive M. This means $\lim_{y\to\infty}(y - \ln(y)) = \infty$.

When the form $\infty - \infty$ comes from the difference of functions $f(x)$ and $g(x)$ we can rewrite the difference as

$$\begin{aligned} f(x) - g(x) &= \frac{(f(x) - g(x))\,(f(x) + g(x))}{f(x) + g(x)} \\ &= \frac{f^2(x) - g^2(x)}{f(x) + g(x)}. \end{aligned}$$

The denominator of this expression goes to ∞. If the numerator is bounded, the limit is zero. Otherwise we can attempt to use L'Hôpital's rule to find this limit.

Example 262. Consider the function $h(x) = \sqrt{x^4 - 2x^2 + 2} - x^2$ and $\lim_{x\to\infty} h(x)$. We can rewrite $h(x)$ as

$$\begin{aligned} h(x) &= \frac{\left(\sqrt{x^4 - 2x^2 + 2} - x^2\right)\left(\sqrt{x^4 - 2x^2 + 2} + x^2\right)}{\sqrt{x^4 - 2x^2 + 2} + x^2} \\ &= \frac{-2x^2 + 2}{\sqrt{x^4 - 2x^2 + 2} + x^2} \\ &= -\frac{-2 + \frac{2}{x^2}}{\sqrt{1 - \frac{2}{x^2} + \frac{2}{x^4}} + 1}. \end{aligned}$$

Taking the limit of this expression as $x \to \infty$ gives us

$$\begin{aligned} \lim_{x\to\infty} h(x) &= -\lim_{x\to\infty} \frac{-2 + \frac{2}{x^2}}{\sqrt{1 - \frac{2}{x^2} + \frac{2}{x^4}} + 1} \\ &= -1. \end{aligned}$$

7.4.1 Slant Asymptotes

A horizontal asymptote for a function $f(x)$ is a line $y = c$ such that $\lim_{x\to\infty} f(x) = c$ or $\lim_{x\to-\infty} f(x) = c$. These two limits can be rewritten as $\lim_{x\to\infty}(f(x) - c) = 0$ or $\lim_{x\to-\infty}(f(x) - c) = 0$. We use the same basic idea to define a slant asymptote.

Definition 29 (Slant asymptote). Let f be a function that is defined on (c,∞) or on $(-\infty, c)$ for some real number c. A line $y = ax + b$, with $a \neq 0$, is a *slant asymptote* for $f(x)$ if either $\lim_{x\to\infty}\big(f(x) - (ax + b)\big) = 0$ or $\lim_{x\to-\infty}\big(f(x) - (ax + b)\big) = 0$, see Fig. 7.15 on page 226.

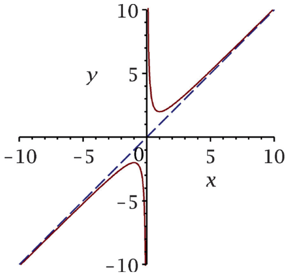


Fig. 7.15

A simple example illustrates how this works.

Example 263. The function $f(x) = (x^2+2x-1)/x$ has slant asymptote $y = x+2$ as x goes to either plus infinity or minus infinity, see Fig. 7.16 on page 226. Since we can write

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

the difference between $f(x)$ and $x+2$ is simply $1/x$ if $x \neq 0$. Because $1/x$ goes to 0 as x goes to plus or minus infinity, the line $y = x+2$ is a slant asymptote for $f(x)$.

If a function $f(x)$ has a slant asymptote $y = ax+b$ as x goes to infinity then, since $\lim_{x\to\infty} f(x) - ax - b = 0$ and $\lim_{x\to\infty} b/x = 0$,

$$\begin{aligned} 0 &= \lim_{x\to\infty} \frac{f(x)-ax-b}{x} \\ &= \lim_{x\to\infty} \frac{f(x)}{x} - a. \end{aligned} \tag{7.4.1}$$

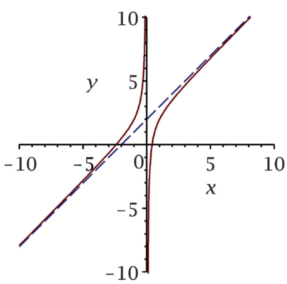


Fig. 7.16 $f(x) = (x^2+2x-1)/x$

This means that, if a slant asymptote exists, the slope of the slant asymptote is

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

In the case when we have an a from Eq. 7.4.1 the constant b can be calculated as the limit

$$b = \lim_{x\to\infty} f(x) - ax.$$

Even if an a exists, the function may not have a slant asymptote. Only if both the a and b exist, does the function have a slant asymptote.

Example 264. The function $h(y) = \sqrt{y^2 - 4y + 8}$ has two slant asymptotes, one as $y \to -\infty$ and one as $y \to \infty$. The a values are

$$\begin{aligned} a_1 &= \lim_{y\to\infty} \frac{h(y)}{y} \\ &= \lim_{y\to\infty} \frac{\sqrt{y^2 - 4y + 8}}{y} \\ &= \lim_{y\to\infty} \sqrt{1 - \frac{4}{y} + \frac{8}{y^2}} \\ &= 1 \ . \end{aligned}$$

and

$$\begin{aligned} a_2 &= \lim_{y\to-\infty} \frac{h(y)}{y} \\ &= \lim_{y\to-\infty} \frac{\sqrt{y^2 - 4y + 8}}{y} \\ &= \lim_{y\to\infty} -\sqrt{1 - \frac{4}{y} + \frac{8}{y^2}} \\ &= -1 \ . \end{aligned}$$

The first b value is

$$\begin{aligned} b_1 &= \lim_{y\to\infty} h(y) - y \\ &= \lim_{y\to\infty} \sqrt{y^2 - 4y + 8} - y \\ &= \lim_{y\to\infty} \frac{\left(\sqrt{y^2 - 4y + 8} - y\right)\left(\sqrt{y^2 - 4y + 8} + y\right)}{\sqrt{y^2 - 4y + 8} + y} \\ &= \lim_{y\to\infty} \frac{-4y + 8}{y\left(\sqrt{1 - \frac{4}{y} + \frac{8}{y^2}} + 1\right)} \\ &= \lim_{y\to\infty} \frac{-4 + \frac{8}{y}}{\left(\sqrt{1 - \frac{4}{y} + \frac{8}{y^2}} + 1\right)} \\ &= -2. \end{aligned}$$

A similar computation shows that $b_2 = 2$.

This gives us two slant asymptotes, $z = y - 2$ and $z = -y + 2$. Figure 7.17 on page 228 illustrates the function and the slant asymptotes.

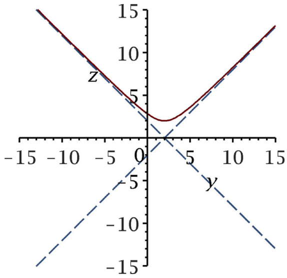


Fig. 7.17 $h(y) = \sqrt{y^2 - 4y + 8}$

Example 265. Consider the function $g(z) = \sqrt{z^2+1} + \cos(z)$. For this function we have

$$\begin{aligned}
\lim_{z\to-\infty} \frac{g(z)}{z} &= \lim_{z\to-\infty} \left(\frac{\sqrt{z^2+1}}{z} + \frac{\cos(z)}{z} \right) \\
&= \lim_{z\to-\infty} \left(-\sqrt{1+\frac{1}{z^2}} + \frac{\cos(z)}{z} \right) \\
&= -1 + 0 \\
&= -1 \ .
\end{aligned}$$

This gives an a of -1.

However,

$$\begin{aligned}
\lim_{z\to-\infty} (g(z) - (-z)) &= \lim_{z\to-\infty} \left(\sqrt{z^2+1} + \cos(z) + z \right) \\
&= \lim_{z\to-\infty} \left(z \left(1 - \sqrt{1+\frac{1}{z^2}} \right) + \cos(z) \right) .
\end{aligned}$$

Using L'Hôpital's rule we can show that

$$\lim_{z\to-\infty} z \left(1 - \sqrt{1+1/z^2} \right) = 0 \ .$$

Therefore

$$\lim_{z\to-\infty} (g(z) - (-z))$$

does not exist. This function does not have a slant asymptote. Figure 7.18 on page 229 is a graph of this function.

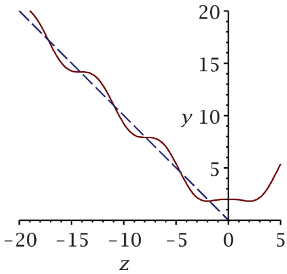


Fig. 7.18 $g(z) = \sqrt{z^2+1} + \cos(z)$

Exercises

1. Evaluate the following limits using L'Hôpital's rule when possible.

(a) $\lim_{\theta \to \frac{\pi}{2}} \dfrac{\cos(\theta)}{\theta - \frac{\pi}{2}}$

(b) $\lim_{\theta \to 0} \theta \cot(\theta)$

(c) $\lim_{x \to 1} \dfrac{\sin(\pi x)}{x^2 - 1}$

(d) $\lim_{x \to 0} \dfrac{1 - e^x}{x^2 - 1}$

(e) $\lim_{y \to 0} \dfrac{2^{4y}}{e^{10y}}$

(f) $\lim_{z \to 0} \dfrac{e^{4z^2}}{3^{10z}}$

(g) $\lim_{x \to 0} \dfrac{5^{4x}}{e^{x^2}}$

2. Evaluate the following limits using L'Hôpital's rule when possible.

(a) $\lim_{\theta \to \infty} \dfrac{\cos(\theta)}{\theta - \frac{\pi}{2}}$

(b) $\lim_{x \to \infty} \dfrac{x^2 + 6x}{2^x}$

(c) $\lim_{x \to \infty} x \sin\left(\dfrac{1}{x}\right)$

(d) $\lim_{z \to \infty} \dfrac{\ln(z)}{z}$

(e) $\lim_{y \to \infty} y^2 \left(1 - \cos\left(\dfrac{2}{y}\right)\right)$

(f) $\lim_{w \to \infty} \left(3w^2 + 6w + 1\right) \exp(-w)$

(g) $\lim_{x \to \infty} \dfrac{x}{\ln(x)}$

(h) $\lim_{x \to \infty} x^2 e^{-x}$

3. Evaluate the following limits using L'Hôpital's rule when possible.

(a) $\lim_{x \to 0^+} x^x$

(b) $\lim_{z \to \infty} \left(1 + \dfrac{4}{z}\right)^z$

(c) $\lim_{x \to \infty} x^{\frac{1}{\ln(x)}}$

(d) $\lim_{x \to \infty} x^{\frac{\ln(x)}{x}}$

(e) $\lim_{y \to 0} \left(e^{7y} + y\right)^{\frac{1}{y}}$

(f) $\lim_{\theta \to 0^+} \left(\tan(\theta)\right)^{\sin(\theta)}$

(g) $\lim_{\theta \to \frac{\pi}{2}^-} (\tan(\theta) - \sec(\theta))$

(h) $\lim_{x \to \infty} \left(\sqrt{4x^2 - 9} - 2x\right)$

4. Do the following functions have slant asymptotes? If they do, what are the equations of the slant asymptotes?

(a) $f(x) = \dfrac{3x^2 - 2x - 3}{x - 3}$

(b) $g(y) = \sqrt{4y^2 - 2y + 3}$

(c) $f(x) = \dfrac{x^3 - 7x^2 + 14x - 5}{x^2 - 4x + 5}$

(d) $h(y) = \sqrt[3]{27y^3 - 74y^2 + 10}$

(e) $r(z) = \sqrt[3]{27z^3 \cos(z) - 74z^2 + 10}$

5. Show that $f(x) = \ln(x)$ does not have a slant asymptote.

Chapter 8
More Techniques of Integration

8.1 A Review of Substitution

Given that integration by substitution is in Sect. 6.5, a short review of substitution for integration is probably appropriate. Some of the substitutions here will be less obvious than those considered in the first semester of calculus. In addition, it is shown how the notation used for separable differential equations follows from integration by substitution.

Recall that integration by substitution is equivalent to reversing the chain rule,

$$\frac{d}{dx}\,\mathbf{f}\circ g(x) = \mathbf{f}'(g(x))\; g'(x).$$

Rewriting this in terms of integrals we have

$$\begin{aligned}\int \mathbf{f}'(g(x))\; g'(x)\,dx &= \int \left(\frac{d}{dx}\;(\mathbf{f}\circ g(x))\right) dx \\ &= \mathbf{f}\circ g(x) + C.\end{aligned}$$

This means that we are trying to find the functions $\mathbf{f}'(y)$, $g'(x)$, and $g(x)$ in the form $\mathbf{f}'(g(x))\; g'(x)$. If we have $\mathbf{f}'(ax+b)$, we do not need the $g'(x)$ since it is a constant that can be obtained using multiplication by a/a. Otherwise, we need to have $g'(x)$ up to a constant multiple.

Example 266. This shows how we can work with the form $\mathbf{f}'(g(x))\; g'(x)$. Consider

$$\int \big(\cos(4x), \sin(4x)\big)\,dx.$$

Here we take $\mathbf{f}'(y) = \big(\cos(y), \sin(y)\big)$ and $g(x) = 4x$. This gives $g'(x) = 4$ and the integral becomes

$$\begin{aligned}\int \big(\cos(4x), \sin(4x)\big)\,dx &= \frac{1}{4}\int \big(\cos(4x), \sin(4x)\big)\,4\,dx \\ &= \frac{1}{4}\int \mathbf{f}'(g(x))\,g'(x)\,dx \\ &= \frac{1}{4}\,\mathbf{f}(g(x)) + C \\ &= \frac{1}{4}\big(\sin(4x), -\cos(4x)\big) + C.\end{aligned}$$

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_8

Frequently people use u to denote $g(x)$, combined with differential notation as in Eq. 2.4 on page 64, in order to simplify the notation.

Example 267. Consider the first component of the Example 266,

$$\int \cos(4x)\,dx.$$

Here we can take $f'(y) = \cos(y)$ and $u = 4x$. This gives $du = 4\,dx$ or $dx = du/4$. The integral now becomes

$$\begin{aligned}\int \cos(4x)\,dx &= \int \cos(u)\,\frac{du}{4}\\ &= \frac{\sin(u)}{4} + C\\ &= \frac{\sin(4x)}{4} + C.\end{aligned}$$

The above example shows how substituting u for $g(x)$ simplifies the notation. This can work for more complicated integrals.

Example 268. Consider

$$\int \frac{\sin\left(\ln(6x+1)\right)}{6x+1}\,dx.$$

The substitution $v = \ln(6x+1)$ works in this case since

$$\frac{dv}{dx} = \frac{6}{6x+1} = 6\frac{1}{6x+1}.$$

Then

$$\frac{1}{6x+1}\,dx = \frac{dv}{6}$$

and the integral becomes

$$\begin{aligned}\int \frac{\sin\left(\ln(6x+1)\right)}{6x+1}\,dx &= \int \sin(v)\,\frac{dv}{6}\\ &= -\frac{\cos(v)}{6} + C\\ &= -\frac{\cos\left(\ln(6x+1)\right)}{6} + C.\end{aligned}$$

Example 269. Consider the integral

$$\int w e^{-w^2} \tan\left(e^{-w^2}\right) \sec^2\left(e^{-w^2}\right) dw.$$

An appropriate substitution here is $u = \tan\left(e^{-w^2}\right)$ since

$$\frac{d}{dw}\tan\left(e^{-w^2}\right) = -2we^{-w^2}\sec^2\left(e^{-w^2}\right).$$

(There is another appropriate substitution. What is it?) This last equation can be rewritten as

$$\frac{du}{-2} = we^{-w^2}\sec^2\left(e^{-w^2}\right) dw.$$

Using this substitution the original integral can be evaluated:

$$\begin{aligned}\int w e^{-w^2} \tan\left(e^{-w^2}\right) \sec^2\left(e^{-w^2}\right) dw &= \int -\frac{u}{2}\, du \\ &= -\frac{u^2}{4} + C \\ &= -\frac{\tan^2\left(e^{-w^2}\right)}{4} + C.\end{aligned}$$

We can use similar substitutions for vector valued functions.

Example 270. Consider the function $\mathbf{r}(t) = \left(t^2 \cos\left(t^3+1\right), t^2 \sin\left(t^3+1\right)\right)$. Substituting $u = t^3 + 1$ and $t^2\, dt = du/3$ we have

$$\begin{aligned}\int \mathbf{r}(t)\, dt &= \int \left(\cos\left(t^3+1\right), \sin\left(t^3+1\right)\right) t^2\, dt \\ &= \frac{1}{3} \int (\cos(u), \sin(u))\, du \\ &= \frac{1}{3} (\sin(u), -\cos(u)) + \mathbf{C} \\ &= \frac{1}{3} \left(\sin\left(t^3+1\right), -\cos\left(t^3+1\right)\right) + \mathbf{C}.\end{aligned}$$

A problem that many people have is using correct limits for definite integrals when doing integration using a substitution. It is required to always have equal quantities. The least confusing way of doing this is to find an antiderivative of the integrand and then do a definite integral with the antiderivative. The other notational ways of dealing with substitutions are to change the limits of integration to match the substitution and to always note that the limits are in terms of the original variable. The following two examples show we can work with the limits using the last two options.

Example 271. Consider

$$\int_0^{\sqrt{\pi/2}} z \cos\left(z^2\right)\, dz.$$

We can use the substitution $u = z^2$ and $z dz = du/2$. In this case, if $z = 0$ then $u = 0$ and if $z = \sqrt{\pi/2}$ then $u = \pi/2$. This means that

$$\begin{aligned}\int_0^{\sqrt{\pi/2}} z \cos\left(z^2\right)\, dz &= \int_0^{\pi/2} \frac{\cos(u)}{2}\, du \\ &= \left.\frac{\sin(u)}{2}\right|_0^{\pi/2} \\ &= \frac{\sin\left(\frac{\pi}{2}\right)}{2} - \frac{\sin(0)}{2} \\ &= \frac{1}{2}.\end{aligned}$$

Example 272. Consider

$$\int_0^{\sqrt{\pi/2}} z \cos\left(z^2\right)\, dz.$$

We can use the substitution $u = z^2$ and $z\,dz = du/2$. This means that

$$\begin{aligned}
\int_0^{\sqrt{\pi/2}} z\cos\left(z^2\right)\,dz &= \int_{z=0}^{z=\sqrt{\pi/2}} \frac{\cos(u)}{2}\,du \\
&= \left.\frac{\sin(u)}{2}\right|_{z=0}^{z=\sqrt{\pi/2}} \\
&= \left.\frac{\sin\left(z^2\right)}{2}\right|_0^{\sqrt{\pi/2}} \\
&= \frac{\sin\left(\frac{\pi}{2}\right)}{2} - \frac{\sin(0)}{2} \\
&= \frac{1}{2}.
\end{aligned}$$

The important thing to notice is that if we do not pay attention to the limits we often get quantities that are not equal. For example, if we did not change the limits in the two preceding examples we would write that the quantities

$$\int_0^{\sqrt{\pi/2}} z\cos\left(z^2\right)\,dz$$

and

$$\int_0^{\sqrt{\pi/2}} \frac{\cos(u)}{2}\,du$$

are equal. However, the quantities are not equal. The first is equal to $1/2$ and the second expression equals $\sin\left(\sqrt{\pi/2}\right)/2 \approx 0.4750$. In order to maintain the value of the integral, we must pay attention to the limits of integration.

8.1.1 Substitution and Separable Differential Equations

Recall that differential equations were discussed in Sect. 6.1. The definition of a differential equation, Definition 22, and the definition of a solution to a differential equation are in that section.

There is an often encountered situation when studying differential equations where a substitution is used without mention. It is in the solution of separable differential equations. A differential equation is *separable* if it can be written in the form

$$\frac{dy}{dx} = f(x)\,g(y).$$

Assuming that $g(y)$ is not zero, we can rewrite this as

$$\frac{1}{g(y)}\frac{dy}{dx} = f(x)$$

or, in integral form when integrating against x,

$$\int \frac{1}{g(y)}\frac{dy}{dx}\,dx = \int f(x)\,dx.$$

Using the substitution $w = y(x)$ and $dw = \frac{dy}{dx}\,dx$ we have

$$\int \frac{1}{g(w)}\,dw = \int f(x)\,dx.$$

This is usually written, using the original symbol for the dependent variable, as

$$\int \frac{1}{g(y)}\,dy = \int f(x)\,dx.$$

Example 273. Consider the separable differential equation

$$\frac{d\,P(t)}{dt} = \frac{P(t)}{100}$$

where $P(0) = 1{,}000$. (This is an *initial condition*.) Using the substitution trick above we get

$$\int \frac{1}{P}\,dP = \int \frac{1}{100}\,dt.$$

Integrating both sides yields

$$\ln(|P|) = \frac{t}{100} + C.$$

Solving for $|P|$ gives

$$|P| = e^{\frac{t}{100}+C} = C_1\,e^{\frac{t}{100}}$$

where C is an unknown constant and C_1 is an unknown positive constant.

Since $|P|$ is $\pm P$, using C to represent an unknown nonzero constant, the last equation becomes

$$P = C\,e^{\frac{t}{100}}.$$

Because $P(t) = 0$ is a solution to the original solution differential equation, all functions of the form

$$P = C\,e^{\frac{t}{100}}.$$

should be solutions to the differential equation. (Check that they are solutions.)

In this example it is assumed that $P(0) = 1,000$. Using the solutions from the last equation we have

$$\begin{aligned} 1{,}000 &= Ce^{\frac{0}{100}} \\ &= C. \end{aligned}$$

The solution to the differential equation with its initial condition is

$$P(t) = 1{,}000\,e^{\frac{t}{100}}.$$

The solution of the differential equation in Example 273 is called exponential growth, i.e. the function is an exponential function that grows as the independent variable goes to infinity if the initial value of the function is positive. Any differential equation of the form

$$\frac{dx}{dt} = \alpha x \tag{8.1}$$

has the same behavior if $\alpha > 0$.

Example 274. Consider the differential equation

$$\frac{dx}{dt} = 0.25x \quad \text{with} \quad x(0) = 1{,}000.$$

To solve the differential equation we separate the variable as

$$\int \frac{dx}{x} = \int 0.25\,dt$$

or, by integration,

$$\ln|x| = 0.25t + C.$$

Solving for x and including the solution $x(t) = 0$ yields

$$x(t) = Ce^{0.25t}.$$

At $t = 0$, we have

$$1{,}000 = Ce^0$$

or $C = 1{,}000$. This means the solution to the differential equation is

$$x(t) = 1{,}000\,e^{0.25t}.$$

See Fig. 8.1 on page 236.

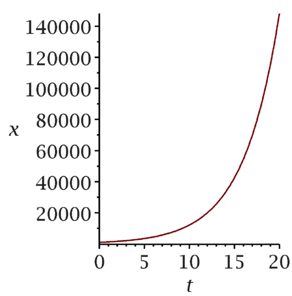


Fig. 8.1 The solution to the differential equation in Example 274 on page 236

On the other hand, if the constant in Eq. 8.1 is negative the solution always goes to 0 as $t \to \infty$. This is used to model, for example, the decay of radioactive isotopes. The following example examines this situation.

Example 275. The Carbon isotope ${}^{14}C$ has radioactive with a half-life of approximately $t_{1/2} = 5{,}730$ years. (The error is ± 40 years.) The half-life is the time it takes for half of the atoms of the isotope to decay. For the decay of radioactive isotopes we assume that the quantity of the isotope present follows the differential equation

$$\frac{dA}{dt} = -kA$$

where A is the amount of the isotope and k is a positive constant. This is simply saying that a certain proportion of the atoms decay over a given time period, no matter when the period starts or how many atoms are present in the sample.

Rewriting the differential equation as

$$\int \frac{dA}{A} = \int -k\,dt$$

and integrating we get

$$\ln|A| = -kt + C.$$

Solving for A and assuming that $A(0) = A_0$ gives us

$$A(t) = A_0 e^{-kt}.$$

To find k we use the fact that half-life is 5,730 years,

$$\frac{A_0}{2} = A_0 e^{-k5{,}730}$$

or

$$\frac{1}{2} = e^{-k5{,}730}.$$

Solving for k gives

$$k = \frac{\ln(2)}{5{,}730}$$

and

$$A(t) = A_0 \exp\left(-\frac{\ln(2)\,t}{5{,}730}\right).$$

(Note that the constant k is always $\ln(2)/t_{1/2}$.)

Finally, if we start with 1 µg, the amount of ^{14}C after 500 years is

$$\begin{aligned} A(500) &\approx 1 \exp\left(-\frac{500\ln(2)}{5{,}730}\right) \\ &\approx 0.941\ \mu g \end{aligned}$$

See Fig. 8.2 on page 237.

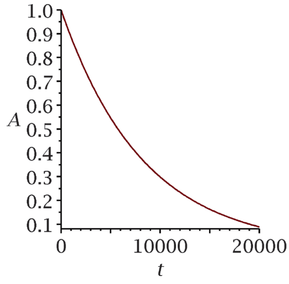


Fig. 8.2 The amount of ^{14}C as a function of time for example 275 on page 236

This method is used, with corrections, for dating in archeology. Corrections are required since the amount of Carbon 14 produced on earth is not constant. This means that using tree ring dating and other methods can give better dating. This other information can then be used to establish more accurate carbon ratios for dating when only ^{14}C is available.

Exercises

1. Find antiderivatives of the following functions.

(a) $f(x) = (2x+1)\sqrt{x^2+x+5}$

(b) $g(y) = \dfrac{\sin(2y)}{1+\cos^2(2y)}$

(c) $h(w) = w\sin\left(w^2\right)\exp\left(\cos\left(w^2\right)\right)$

(d) $f(x) = \dfrac{2x}{\sqrt{1-x^4}}$

(e) $g(y) = \dfrac{\exp(2y)}{\sqrt{\exp(2y)-4}}$

(f) $h(w) = \dfrac{\exp(w)}{1+10\exp(w)}$

(g) $f(x) = \dfrac{x\cos\left(x^2\right)\sin\left(x^2\right)}{\cos^2\left(x^2\right)+4}$

(h) $g(y) = \dfrac{y\exp(y^2+2)}{\sqrt{\exp\left(y^2+2\right)-4}}$

(i) $h(w) = \left(3w^2+2w\right)\tan\left(w^3+w^2\right)\sec^3\left(w^3+w^2\right)$

(j) $f(x) = 2x\tan\left(x^2\right)\sec\left(x^2\right)$

2. Evaluate the following definite integrals. (Give exact answers for all problems.)

(a) $\displaystyle\int_0^2 (2x+1)\sqrt{x^2+x+5}\,dx$

(b) $\displaystyle\int_{-1}^1 \sin(2x)\sqrt{1-\sin^2(2x)}\,dx$

(c) $\displaystyle\int_0^{\frac{\pi}{4}} \tan\left(\frac{t}{2}\right)\sec^2\left(\frac{t}{2}\right)\,dt$

(d) $\displaystyle\int_{-2}^2 \frac{w^2}{1+w^6}\,dw$

(e) $\displaystyle\int_{-1}^{-\frac{1}{2}} \frac{\sin(\pi x)}{\sqrt{4-\cos^2(\pi x)}}\,dx$

(f) $\displaystyle\int_0^{\frac{\pi}{4}} \tan\left(\frac{t}{2}\right)\sec^2\left(\frac{t}{2}\right)\,dt$

(g) $\displaystyle\int_{-1}^3 w^2\,e^{\left(w^3+2\right)}\,dw$

(h) $\displaystyle\int_{\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^2\left(\frac{t}{2}\right)\sec^2\left(\frac{t}{2}\right)\,dt$

(i) $\displaystyle\int_3^{-3} w^3\,e^{\left(w^4+2\right)}\,dw$

(j) $\displaystyle\int_4^5 (1+\cos(x))\sqrt{4+x+\sin(x)}\,dx$

(k) $\displaystyle\int_0^{\frac{\pi}{4}} \cot^3\left(\frac{t}{2}\right)\csc^2\left(\frac{t}{2}\right)\,dt$

(l) $\displaystyle\int_{-2}^0 \frac{\sin(w)\cos(w)}{1+\sin^2(w)}\,dw$

3. Evaluate the following definite integrals.

(a) $\displaystyle\int_0^1 e^t\left(\cos\left(1+e^t\right), \sin\left(1+e^t\right), \sqrt{1+e^t}\right)\,dt$

(b) $\displaystyle\int_{-2}^{0} te^{t^2+1}\left(\cos\left(e^{t^2+1}\right), \sin\left(e^{t^2+1}\right), \sqrt{1+e^{t^2+1}}\right) dt$

(c) $\displaystyle\int_{0}^{\pi} \sin(t)\left(\cos(1+\cos(t)), \exp(1+\cos(t)), \sqrt{1+\cos(t)}\right) dt$

(d) $\displaystyle\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \cos(\theta)\left(\sin^2(\theta), \exp(\sin(\theta)), \sqrt[3]{\sin(\theta)}\right) d\theta$

4. Solve the following differential equations.

(a) $\dfrac{dy}{dx} = x^2 y.$

(b) $\dfrac{dy}{dx} = xe^{-y}.$

(c) $\dfrac{dx}{dt} = \sec(x)te^{-1-t^2}.$

(d) $\dfrac{dx}{dt} = \dfrac{x^2+1}{t}.$

5. Solve the following differential equations with the given initial conditions.

(a) $\dfrac{dy}{dx} = x^2 y$ and $y(1) = 1$.

(b) $\dfrac{dy}{dx} = xe^{-y}$ and $y(0) = 0$.

(c) $\dfrac{dx}{dt} = \dfrac{1+t}{x}$ and $x(0) = 2$.

(d) $\dfrac{dx}{dt} = x^2 t$ and $x(2) = 1$.

6. Newton's law of cooling states that *under ideal conditions* the rate of change of an object in a constant temperature medium is proportional to the difference between the temperature of the object $T(t)$ and the temperature T_A of the medium, the ambient temperature,

$$\frac{dT(t)}{dt} = -k(T(t) - T_A). \tag{8.2}$$

(a) Solve the differential equation (8.2) using the substitution $U(t) = T(t) - T_A$.

(b) Find the solution to the differential equation if $T(0) = 100$, $T_A = 22$, and $T(3) = 35$.

(c) Find the solution to the differential equation if $T(0) = 100$, $T_A = 22$, and $T'(0) = -40$.

7. The half-life of Helium 6, ^6He, is 806.7 ms. If you start with a sample of 7 µg, how much ^6He will you have after 1 min?

8. The half-life of Tin 123, ^{123}Sn, is 129.2 days. If you have 2 ng after 1 year, how much did you have initially?

9. You start with a mixture of two isotopes of Californium, Californium 241 with a half life of 3.78 min and Californium 244 with a half-life of 19.4 min. If the ratio of Californium 241 atoms to Californium 244 atoms is initially 10 to 1, when will the ratio of Californium 241 atoms to Californium 244 atoms be 1 to 1?

10. Show that $\displaystyle\int_0^1 x(1-x)\,dx = -\int_1^0 u(1-u)\,du$ using the substitution $u = 1-x$.

8.2 Integration by Parts

The second technique of integration presented is integration by parts. It can be viewed as the inverse operation of the product rule. If we take the product rule

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

and integrate both sides with respect to x we get

$$\begin{aligned}(f\cdot g)(x) &= \int (f\cdot g)'(x)\,dx\\ &= \int f'(x)\cdot g(x)\,dx + \int f(x)\cdot g'(x)\,dx.\end{aligned}$$

Rearranging gives the integration by parts formula:

$$\int \left(f\cdot g'\right)(x)\,dx = (f\cdot g)\,(x) - \int \left(f'\cdot g\right)(x)\,dx.$$

This is sometimes rewritten by replacing f by u and g by v, and using differentials to get the form

$$\int u\,dv = uv - \int v\,du.$$

It is unclear how integration by parts can be used from the formula. There are two very important cases when it can be used. The first is the case where $f(x)$ is a polynomial of degree n and $g(x)$ is a transcendental function that is easily integrated at least $n+1$ times. In this case, applying integration by parts n times will yield a formula for the antiderivative.

Example 276. Consider the integral

$$\int xe^x\,dx.$$

Here $f(x)=x$ is a polynomial of degree 1 and $g'(x)=e^x$ is infinitely integrable. In this case we can use $g(x)=e^x$ and $f'(x)=1$ to get

$$\begin{aligned}\int xe^x\,dx &= xe^x - \int 1e^x\,dx\\ &= xe^x - e^x + C.\end{aligned}$$

It is simple to check this formula using differentiation.

Remark 23. In the above example the single function $g(x)=e^x$ was used for the antiderivative. No matter which antiderivative of $g(x)$ we use, we always get the same answer. This is relatively easy to show and is left as an exercise.

This technique also works with higher degree polynomials using the integration by parts formula multiple times.

Example 277. Consider the integral

$$\int z^3\cos(3z)\,dz.$$

Here we can take $u=z^3$ and $dv=\cos(3z)\,dz$. Then we have $du=3z^2\,dz$ and $v=\sin(3z)/3$.

The integral now takes the form

$$\int z^3\cos(3z)\,dz = \frac{z^3\sin(3z)}{3} - \int z^2\sin(3z)\,dz.$$

The last term on the right can be integrated using parts with $u_1=z^2$, $dv_1=\sin(3z)\,dz$, $du_1=2z\,dz$, and $v_1=-\cos(3z)/3$ to get

$$\int z^3\cos(3z)\,dz = \frac{z^3\sin(3z)}{3} - \left(-\frac{z^2\cos(3z)}{3} + \int \frac{2}{3}z\cos(3z)\,dz\right)$$
$$= \frac{z^3\sin(3z)}{3} + \frac{z^2\cos(3z)}{3} - \frac{2}{3}\int z\cos(3z)\,dz.$$

Applying integration by parts to the last term on the right with $u_2 = z$, $dv_2 = \cos(3z)\,dz$, $du_2 = 1$, and $v_2 = \sin(3z)/3$ gives

$$\int z^3\cos(3z)\,dz = \frac{z^3\sin(3z)}{3} + \frac{z^2\cos(3z)}{3} - \frac{2}{3}\int z\cos(3z)\,dz$$
$$= \frac{z^3\sin(3z)}{3} + \frac{z^2\cos(3z)}{3} - \frac{2}{3}\left(\frac{z\sin(3z)}{3} - \int \frac{\sin(3z)}{3}\,dz\right)$$
$$= \frac{z^3\sin(3z)}{3} + \frac{z^2\cos(3z)}{3} - \frac{2z\sin(3z)}{9} - \frac{2\cos(3z)}{27} + C.$$

The other main situation where integration by parts is used is when $f(x)$ and $g'(x)$ have the nth derivative of $f(x)$ as a multiple of $f(x)$ and the nth antiderivative of $g'(x)$ is a multiple of $g'(x)$. Applying integration by parts n times to $\int f(x)\cdot g'(x)\,dx$ will yield a term of the form $k\cdot\int f(x)\cdot g'(x)\,dx$ on the right side of the integrals where $k \neq 1$. This term can be taken to the other side to get an equation that can be solved for $\int f(x)\cdot g'(x)\,dx$. The next example illustrates this technique.

Example 278. The integral

$$\int \cos(w)\,e^w\,dw$$

can be evaluated using integration by parts. Let $u = e^w$ and let $dv = \cos(w)\,dw$. Then $du = e^w\,dw$ and $v = \sin(w)$. This gives

$$\int \cos(w)\,e^w\,dw = \sin(w)\,e^w - \int \sin(w)\,e^w\,dw.$$

Integrating by parts again with $u_1 = e^w$ and $dv_1 = \sin(w)\,dw$ gives $du_1 = e^w\,dw$, $v_1 = -\cos(w)$, and

$$\int \cos(w)\,e^w\,dw = \sin(w)\,e^w - \left(-\cos(w)\,e^w + \int \cos(w)\,e^w\,dw\right)$$
$$= \sin(w)\,e^w + \cos(w)\,e^w - \int \cos(w)\,e^w\,dw.$$

Adding $\int \cos(w)\,e^w\,dw$ to both sides of the equation, dividing by 2, and adding the constant of integration gives

$$\int \cos(w)\,e^w\,dw = \frac{\sin(w)\,e^w + \cos(w)\,e^w}{2} + C.$$

Remark 24. An important thing to remember about integration by parts is that if we integrate a function $f(x)$ in the first step, we need to keep integrating the integrals of $f(x)$. We should never need to differentiate an integral of $f(x)$. Remember when we differentiate the antiderivative of a function, we get the original function back.

If we take the integral from the last example with $u = e^w$, $dv = \cos(w)\,dw$, $du = e^w\,dw$ and $v = \sin(w)$ we get

$$\int \cos(w)\,e^w\,dw = \sin(w)\,e^w - \int \sin(w)\,e^w\,dw.$$

Taking $u_1 = \sin(w)$, $dv_1 = e^w\,dw$, $du_1 = \cos(w)\,dw$, and $v_1 = e^w$ gives

$$\begin{aligned}\int \cos(w)\,e^w\,dw &= \sin(w)\,e^w - \left(\sin(w)\,e^w - \int \cos(w)\,e^w\,dw\right)\\ &= \int \cos(w)\,e^w\,dw.\end{aligned}$$

This has not changed the original integral at all and is not helpful.

There is one other common situation where integration by parts is useful, when we integrate a polynomial in x multiplied by the natural logarithm of x, $\ln(x)$. In this case differentiating the natural logarithm of x turns the integral into the integral of a sum of powers of x.

Example 279. Consider the integral

$$\int \ln(x)\,dx.$$

We can take $f(x) = \ln(x)$ and $g'(x) = 1$. Differentiating and integrating gives $f'(x) = 1/x$ and $g(x) = x$. The integral now becomes

$$\begin{aligned}\int \ln(x)\,dx &= x\ln(x) - \int \frac{x}{x}\,dx\\ &= x\ln(x) - \int 1\,dx\\ &= x\ln(x) - x + C.\end{aligned}$$

This section ends with a final example that shows how a substitution can change an integral that looks very difficult into a simple integration by parts.

Example 280. The integral

$$\int_0^2 x^3 e^{x^2}\,dx$$

can be changed into a simple integration by parts using the substitution $w = x^2$, $dw = 2x\,dx$, $w(0) = 0$, and $w(2) = 4$. The integral is transformed to an integral in terms of w:

$$\begin{aligned}\int_0^2 x^3 e^{x^2}\,dx &= \int_0^2 x^2 e^{x^2}\,x\,dx\\ &= \int_0^4 w\,e^w\,\frac{dw}{2}\end{aligned}$$

The integral in terms of w can be integrated by parts using $u = w$, $dv = e^w\,dw$, $du = dw$, and $v = e^w$.

$$\begin{aligned}\int_0^4 w\,e^w\,\frac{dw}{2} &= w\,e^w\Big|_0^4 - \int_0^4 e^w\,dw\\ &= 4e^4 - \left(e^w\Big|_0^4\right)\\ &= 3e^4 + 1.\end{aligned}$$

The final result is

$$\int_0^2 x^3 e^{x^2}\,dx = 3e^4 + 1.$$

Exercises

1. Find antiderivatives of the following functions.

(a) $f(x) = x\sin(5x)$
(b) $f(x) = x^3\cos(5x^2)$
(c) $g(y) = y\exp(5y-2)$
(d) $h(w) = w^2\ln(w)$
(e) $f(x) = x^2\exp\left(x^3\right)$
(f) $g(y) = \cos(2y)\sin(y)$
(g) $h(w) = e^{(2w)}\sin(w)$
(h) $f(x) = \cos(3x)\sin(5x)$
(i) $g(y) = \cos\left(\theta + \frac{\pi}{6}\right)\sin(\theta)$
(j) $h(w) = \arcsin(w)$
(k) $f(x) = \arctan(x)$

2. Evaluate the following definite integrals. (Give exact answers for all problems.)

(a) $\int_0^2 xe^x\,dx$
(b) $\int_{-\pi}^0 2x\sin\left(x^2\right)\,dx$
(c) $\int_0^{\sqrt{\frac{\pi}{4}}} t\sin\left(t^2\right)\cos\left(t^2\right)\,dt$
(d) $\int_{-2}^2 e^w\sin\left(e^w\right)\,dw$
(e) $\int_{-1}^{\frac{1}{2}} \left(x^2+x\right)e^{2x}\,dx$
(f) $\int_1^3 \sin(\pi t)\,e^{2t}\,dt$
(g) $\int_2^6 e^{-2w}\cos(2w)\,dw$
(h) $\int_1^5 x^2\ln(x)\,dx$
(i) $\int_0^T t^2\cos\left(\frac{t}{2}\right)\,dt$
(j) $\int_\alpha^0 (w+1)\exp(-w)\,dw$

3. A 1 kg mass starts at rest at the origin and is constrained to move along the x-axis. The position of the mass given by $x(t)$. If the force acting on the mass is the exponentially decaying function

$$F(t) = 10e^{\left(-\frac{t}{2}\right)}\cos(t)N,$$

find the position of the mass as a function of time.

4. The one dimensional motion of a mass on a spring can be modeled by the equation

$$x'(t) = Ae^{-kt}\sin(\omega t + \phi).$$

If $A=1$, $k=\frac{1}{4}$, $\omega=\frac{\pi}{4}$ and $\phi=\frac{\pi}{12}$, find the distance traveled by the mass from $t=0$ to $t=10$

5. Explain why if $\mathbf{f}(x)$ is a vector valued function and $g(x)$ is a real valued function, the integration by parts formula

$$\int \left(g\cdot\mathbf{f}'\right)(x)\,dx = (g\cdot\mathbf{f})(x) - \int \left(g'\cdot\mathbf{f}\right)(x)\,dx.$$

is valid.

6. Use the equation from Exercise 5 to evaluate the following.

(a) $\int x\,(e^x, \sin(x))\,dx.$

(b) $\int \theta\,(\cos(\theta), \sin(\theta))\,d\theta.$

(c) $\int e^{-\theta}\,(\cos(2\theta), \sin(\theta))\,d\theta.$

(d) $\int e^{-\theta}\left(\cos(\theta), \sin(2\theta), \frac{1}{\pi}\right)d\theta.$

7. Solve the following differential equations.

(a) $\frac{dw}{dt} = (1+w^2)\ln(t), \quad w(1) = 0$

(b) $\frac{dy}{dx} = \frac{x\exp(x)}{y+1}, \quad y(0) = 1$

(c) $\frac{dw}{dt} = (1+w)t^2\ln(t), \quad w(1) = 0$

(d) $\frac{dy}{dx} = \frac{\cos(x)\exp(x)}{y+1}, \quad y(0) = 1$

8.3 Integration by Partial Fractions

Partial fractions is a method used to integrate rational functions. It relies on two important facts. The first allows us to write a general rational function as a sum of simple rational functions. This is called a *partial fractions decomposition* . Some examples of partial fractions decompositions will illustrate the idea.

Example 281. Each of the following equalities shows how we can rewrite a rational function as a sum of simpler rational function. Later in this section you should learn how to do this. For now the equalities can be verified by multiplying both sides of each equality by the denominator of the left side of the equation.

$$\frac{6x+4}{(x-2)^2} = \frac{6}{x-2} + \frac{16}{(x-2)^2}$$
$$\frac{3x+5}{(x+2)(x-1)} = \frac{1}{3(x+2)} + \frac{8}{3(x-1)}$$
$$\frac{x^2-29x-39}{(x^2+4x+8)(x-5)} = \frac{4x+3}{x^2+4x+8} - \frac{3}{x-5}$$
$$\text{and} \qquad \frac{x^3+8x^2+4x-3}{(x-1)^2(x^2+2x+2)} = \frac{3}{x-1} + \frac{2}{(x-1)^2} - \frac{2x+1}{x^2+2x+2}.$$

Finding the partial fractions decomposition is the hard part of this technique. As you will see, it is theoretically possible to find a partial fractions decomposition of any rational function. The biggest problem is completely factoring the denominator of the rational function.

The second fact is that once we have a partial fractions decomposition, integrating the simpler rational functions is always possible. Each of the terms on the right sides of the equations in the example above can be integrated with the information you already have. Terms of the types

$$\frac{6}{x-2} \quad \text{and} \quad \frac{16}{(x-2)^2}$$

have antiderivatives

$$6\ln(x-2) \quad \text{and} \quad -\frac{16}{x-2}.$$

In a similar way we can integrate any rational function of the form $A/(x-b)^n$ for a positive integer n.

The terms similar to $(4x+3)/(x^2+4x+8)$ are not hard to integrate if we can complete the square. In this case we can rewrite x^2+4x+8 as $(x+2)^2+2^2$ by completing the square. Using this we can rewrite this rational function as

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

The first term on the right of this equation can be integrated using the substitution $u = x^2+4x+8$ and $du = 2(x+2)\,dx$ to get

$$\begin{aligned}\int \frac{4(x+2)}{x^2+4x+8}\,dx &= 2\int \frac{1}{u}\,du \\ &= 2\ln|u| + C \\ &= 2\ln\left|x^2+4x+8\right| + C\end{aligned}$$

The second term on the right can be integrated using the substitution $v = (x+2)/2$ and $dv = dx/2$:

$$\begin{aligned}\int \frac{5}{(x+2)^2+2^2}\,dx &= \frac{5}{2}\int \frac{1}{\left(\frac{x+2}{2}\right)^2+1}\,\frac{dx}{2} \\ &= \frac{5}{2}\int \frac{1}{v^2+1}\,dv \\ &= \frac{5}{2}\tan^{-1}(v) + C \\ &= \frac{5}{2}\tan^{-1}\left(\frac{x+2}{2}\right) + C.\end{aligned}$$

Combining all of the above we have

$$\begin{aligned}\int \frac{x^2-29x-39}{(x^2+4x+8)(x-5)}\,dx &= \int \frac{4x+3}{x^2+4x+8}\,dx - \int \frac{3}{x-5}\,dx \\ &= \frac{4(x+2)}{(x+2)^2+2^2} - \frac{5}{(x+2)^2+2^2} - 3\int \frac{1}{x-5}\,dx \\ &= 2\ln\left|x^2+4x+8\right| - \frac{5}{2}\tan^{-1}\left(\frac{x+2}{2}\right) - 3\ln|x-5| + C.\end{aligned}$$

There are three results that allow us to rewrite a rational function as a partial fractions decomposition. The first result is a consequence to the division algorithm for polynomials.

Lemma 4. *Let $r(x) = p(x)/q(x)$ be a rational function. Then $r(x)$ can be written as*

$$r(x) = f(x) + \frac{h(x)}{q(x)},$$

where $f(x)$ and $h(x)$ are polynomials with the degree of $h(x)$ less than the degree of $q(x)$.

An example makes this easier to understand.

Example 282. Let

$$r(x) = \frac{x^3+4x^2-3x+5}{x^2+x-2}$$

Then

$$r(x) = \frac{x^3+4x^2-3x+5}{x^2+x-2} = x+3+\frac{-4x+11}{x^2+x-2}.$$

This can be done by long division of polynomials

$$\begin{array}{r|l} & \quad\quad\quad\quad\quad x \;\; +3. \\ \hline x^2+x-2) & x^3+4x^2-3x \;\; +5 \\ & -x^3 \;\; -x^2+2x \\ \hline & \quad\quad 3x^2 \;\; -x \;\; +5 \\ & \quad -3x^2-3x \;\; +6 \\ \hline & \quad\quad\quad -4x+11 \end{array}$$

The second fact allows us to factor the denominator of the rational function. The proof is beyond this class and is omitted.

Theorem 70. *If $q(x)$ is a polynomial in x with real coefficients, then $q(x)$ can be factored as*

$$\begin{aligned} q(x) = h\,(x+a_1)\cdot(x+a_2)\cdots(x+a_n)\cdot \\ (x^2+b_1x+c_1)\cdot(x^2+b_2x+c_2)\cdots(x^2+b_mx+c_m) \end{aligned}$$

where h, the a_i's, b_j's, and c_k's are real numbers and each of the quadratics $x^2+b_jx+c_j$ has no real roots, each is irreducible.

Note that the $x-a_i$'s and the $x^2+b_jx+c_j$'s may repeat. This means that we may write

$$q(x) = d\,(x+a_1)^{k_1}\cdot(x+a_2)^{k_2}\cdots(x+a_n)^{k_n} \tag{8.3}$$
$$\cdot\,(x^2+b_1x+c_1)^{\ell_1}\cdot(x^2+b_2x+c_2)^{\ell_2}\cdots(x^2+b_mx+c_m)^{\ell_m} \tag{8.4}$$

where the a_i' are all distinct and all of the (b_j,c_j) pairs are distinct.

Assuming that $q(x)$ can be factored as in Eq. 8.3, we can theoretically write

$$\begin{aligned} \frac{p(x)}{q(x)} = f(x) + \sum_{i=1}^{k_1}\frac{A_{1,i}}{(x+a_1)^i} + \sum_{i=1}^{k_2}\frac{A_{2,i}}{(x+a_2)^i} + \cdots + \sum_{i=1}^{k_n}\frac{A_{n,i}}{(x+a_n)^i} \\ + \sum_{i=1}^{\ell_1}\frac{B_{1,i}+C_{1,i}x}{(x^2+b_1x+c_1)^i} + \sum_{i=1}^{\ell_2}\frac{B_{2,i}+C_{2,i}x}{(x^2+b_2x+c_2)^i} + \cdots + \sum_{i=1}^{\ell_m}\frac{B_{m,i}+C_{m,i}x}{(x^2+b_mx+c_m)^i}. \end{aligned}$$

Here $f(x)$ is a polynomial. The terms of the forms $A/(x+a)^i$, $(B+Cx)(x^2+bx+c)$, and $(C\,(2x+b))(x^2+bx+c)^\ell$ can be integrated as above.

The problem facing us now is how to find the A's, B's, and C's. At this point it is assumed that the degree of $p(x)$ is less than the degree of $q(x)$. A procedure that "always" works is to set the rational function equal to the desired partial fractions decomposition and multiply both sides of the equation by the denominator of the rational function. This gives two polynomials that must be equal. We then need to find A's, B's, and C's that make the coefficients on both sides of the equation equal.

For example, consider

$$\frac{x-3}{(x+1)\,(x+2)} = \frac{A}{x+1} + \frac{B}{x+2}.$$

Multiplying both sides of the equation by $(x+1)\,(x+2)$ gives

$$x-3 = A\,(x+2) + B\,(x+1) = (A+B)\,x + (2A+B).$$

Since the polynomials are equal, their coefficients for each power of x must be equal,

$$\begin{aligned} x &= (A+B)\,x, \quad \text{and} \\ -3 &= 2A+B. \end{aligned}$$

For the example this gives two linear equations in two unknowns

$$\begin{aligned} 1 &= A+B, \quad \text{and} \\ -3 &= 2A+B. \end{aligned}$$

which can be solved giving $A=-4$ and $B=5$.

Using the values of A and B we have

$$\frac{x-3}{(x+1)\,(x+2)} = -\frac{4}{x+1} + \frac{5}{x+2}.$$

The terms on the right can easily be integrated.

In the general case, by setting the coefficients for each power of x equal, we get a system of n linear equations in n unknowns that can be solved. Here n is the degree of the denominator of the rational function. Although this technique always works, in many cases it can be simplified by using the roots of the denominator of the rational function.

Again consider

$$\frac{x-3}{(x+1)\,(x+2)} = \frac{A}{x+1} + \frac{B}{x+2}.$$

Multiplying both sides of the equation by $(x+1)\,(x+2)$ gives

$$x-3 = A\,(x+2) + B\,(x+1).$$

Since this equation holds at all points besides $x=-1$ and $x=-2$ and both sides are continuous functions, the equation must also hold at $x=-1$ and $x=-2$. Plugging $x=-1$ and $x=-2$ into the equation gives

$$\begin{aligned} -4 &= A \quad \text{and} \\ -5 &= -B, \end{aligned}$$

or $A=-4$ and $B=5$.

This leaves us at the point of having seen all of the pieces necessary to do integration by partial fractions. What is left is to see some examples of how this works. The next three examples are restricted to cubic denominators, but the techniques are usable for much more general denominators.

First, the example of

$$\int \frac{x-3}{(x+1)\,(x+2)}\,dx$$

is completed.

Example 283. As was shown above

$$\frac{x-3}{(x+1)\,(x+2)} = -\frac{4}{x+1} + \frac{5}{x+2}.$$

This means that

$$\int \frac{x-3}{(x+1)(x+2)}\,dx = -\int \frac{4}{x+1}\,dx + \int \frac{5}{x+2}\,dx$$
$$= -4\ln|x+1| + 5\ln|x+2| + C$$
$$= \ln\left|\frac{(x+2)^5}{(x+1)^4}\right| + C.$$

The second example shows that there are cases when we cannot use only the roots of factors of the denominator. Here the root is a double root.

Example 284. Consider the integral

$$\int \frac{3x-1}{x^2-6x+9}\,dx.$$

The first thing to do is factor the denominator, $x^2-6x+9=(x-3)^2$. The partial fractions decomposition of the integrand has the form

$$\frac{3x-1}{x^2-6x+9} = \frac{A}{x-3} + \frac{B}{(x-3)^2}.$$

Multiplying both sides of this equation by $(x-3)^2$ gives the equality

$$3x-1 = A(x-3)+B. \tag{8.5}$$

We can match the x coefficients to get $A=3$ and we can substitute $x=3$ into the equation to get $B=8$.

With this information we can integrate the rational function,

$$\int \frac{3x-1}{x^2-6x+9}\,dx = \int \frac{3}{x-3}\,dx + \int \frac{8}{(x-3)^2}\,dx$$
$$= 3\ln|x-3| - \frac{8}{x-3} + C.$$

To avoid confusion we note that substituting $x=3$ into Eq. 8.5 only gives a value for B, not a value for A. Only knowing B does not give enough information to integrate the rational function.

When we deal with a denominator that has an irreducible quadratic factor and another factor, it is always necessary to solve a system of equations for unknown coefficients. The next example shows how this can be done in the simplest case.

Example 285. The integral

$$\int \frac{5z^2-z+4}{z^3-z^2-7z+15}\,dz$$

can be integrated using partial fractions. To do this we need to factor the denominator, $z^3-z^2-7z+15$ as $(z+3)(z^2-4z+5)$ and complete the square for z^2-4z+5 to get $z^2-4z+5=(z-2)^2+1$. This gives a partial fractions decomposition form

$$\frac{5z^2-z+4}{z^3-z^2-7z+15} = \frac{A}{z+3} + \frac{B(z-2)+C}{(z-2)^2+1}.$$

(Why can the writing the numerator of the second term as $B(z-2)+C$ be helpful?)

Multiplying both sides of the equation by $(z+3)(z^2-4z+5)$ and expanding the right side of the equation gives

$$\begin{aligned}5z^2-z+4 &= A((z-2)^2+1)+(z+3)(B(z-2)+C)\\ &= (A+B)z^2+(-4A+B+C)z+(5A-6B+3C).\end{aligned}$$

Setting the coefficients on both sides of the equation equal produces three equations in three unknowns.

$$\begin{aligned}5 &= A+B\\ -1 &= -4A+B+C\\ 4 &= 5A-6B+3C.\end{aligned}$$

Adding four times the first equation to the second equation and subtracting five times the first equation from the third equation eliminates A from the equations.

$$\begin{aligned}19 &= 5B+C\\ -21 &= -11B+3C\end{aligned}$$

Then C can be eliminated by subtracting three times $19=5B+C$ from $-21=-11B+3C$ to get

$$-78=-26B$$

or $B=3$.

The value of C is obtained by setting $B=3$ in $19=5B+C$. We find that $C=4$. Substituting $B=3$ into $5=A+B$ gives $A=2$. These values of A, B, and C give a partial fractions decomposition

$$\frac{5z^2-z+4}{z^3-z^2-7z+15}=\frac{2}{z+3}+\frac{3(z-2)+4}{(z-2)^2+1}.$$

The partial fractions decomposition can be integrated using the substitutions $w=z^2-4z+5$, $dw/2=(z-2)\,dz$, $u=z-2$, and $du=dz$ as follows.

$$\begin{aligned}\int\frac{5z^2-z+4}{z^3-z^2-7z+15}\,dz &= \int\frac{2}{z+3}+\frac{3(z-2)}{z^2-4z+5}+\frac{4}{(z-2)^2+1}\,dz\\ &= \int\frac{2}{z+3}\,dz+\frac{3}{2}\int\frac{1}{w}\,dw+4\int\frac{1}{u^2+1}\,du\\ &= 2\ln|z+3|+\frac{3}{2}\ln(w)+4\tan^{-1}(u)+C\\ &= 2\ln|z+3|+\frac{3}{2}\ln(z^2-4z+5)+4\tan^{-1}(z+2)+C\end{aligned}$$

The basic method used here can be extended to any rational function, even though this can be quite difficult for many rational functions. As will be demonstrated in the next section, partial fractions can also be used to evaluate integrals involving trigonometric functions.

8.3.1 The Logistic Equation

One application of partial fractions integration is solving the logistic differential equation. This differential equation is used for modeling population growth when there are limits on the size of a population. As the next example shows, exponential growth without limits is not a good model of reality.

Example 286. A well know type of bacteria is E. coli, Escherichia coli, that grows in mammals. It can cause disease and is used in experiments. Some varieties reproduce very quickly. Assume that a variety of E. coli has a mass per cell of approximately 1×10^{-15} kg and has a doubling time of 20 h. Assuming its reproduction follows the exponential growth model, see Sect. 8.1.1, the mass of bacteria at a given time follows the results of Example 273 on 235.

The solution of the differential equation with a doubling time of 20 h is

$$P(t) = P(0)\exp\left(\frac{\ln(2)\,t}{20}\right),$$

where t is time measured in hours. If we measure the mass in kg and assumes we start with a single bacterium, the total mass of the bacteria as a function of time is

$$M(t) = 10^{-15}\exp\left(\frac{\ln(2)\,t}{20}\right).$$

If growth is unlimited, after 1 week there would be approximately 3.38×10^{-13} kg of E. coli, after 1 month there would be approximately 6.87×10^{-5} kg of E. coli, after 3 months there would be approximately 7.45×10^{17} kg of E. coli, and after 6 months there would be approximately 9.31×10^{50} kg of E. coli. The mass of the earth is approximately 5.97×10^{24} kg, and an informed estimate for the amount of mass in the visible universe is 1.54×10^{53} kg. From this it should be clear that the exponential growth model has limits on its range of usefulness.

The logistic equation is a variant of the exponential growth differential equation that includes an extra term on the right that limits growth. This extra term can be justified in a couple different ways. A simple way to justify the extra term is to assume that interactions between members of the population under consideration have a negative effect on reproduction. This could take the form of the death of an individual or increasing the stress on an individual, thereby lowering the reproductive rate for that individual. If we assume that the number of interactions between members of the population is proportional to the square of the population, this can be justified using probability, we get the differential equation

$$\frac{dP}{dt} = \alpha P - \beta P^2 \tag{8.6}$$

$$= k_1 P\,(k_2 - P). \tag{8.7}$$

In the second form of the equation, which is used here to find the solution, $k_1 = \beta$ and $k_2 = \alpha/\beta$. It is assumed that α, β, k_1 and k_2 are all positive.

Some of the properties of $P(t)$ are easy to discern immediately from the differential equation. Since $P(t)$ is supposed to be a population, $P(t) \geq 0$ for all t. If the population starts at 0 or k_2, the population will stay at that value since $dP/dt = 0$ for all time. (These are called equilibrium populations.) If $P \in (0, k_2)$, $dP/dt > 0$ and the population is increasing. On the other hand, if $P > k_2$, then $dP/dt < 0$ and the population is decreasing. This indicates that a population modeled with the logistic growth differential equation will not display unbounded growth.

Noting that this is a separable differential equation, see Sect. 8.1.1, we can solve it using partial fractions integration. First we rewrite the differential equation as

$$\int \frac{dP}{P\,(k_2 - P)} = \int k_1\,dt.$$

The left side integration is as follows:

$$\begin{aligned}\int \frac{dP}{P(k_2-P)} &= \frac{1}{k_2}\int\left(\frac{1}{P}+\frac{1}{k_2-P}\right)dP\\ &= \frac{1}{k_2}\left(\ln(P)-\ln(k_2-P)\right)+C\\ &= \frac{1}{k_2}\ln\left(\frac{P}{k_2-P}\right)+C.\end{aligned}$$

Since the integral of the right side of the separated equation is

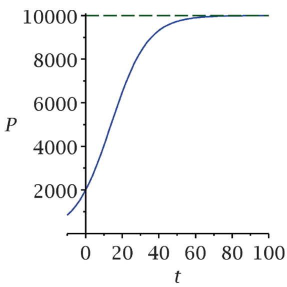


Fig. 8.3

$$\int k_1\,dt = k_1\,t + C,$$

we get the equation

$$\ln\left(\frac{P}{k_2-P}\right) = k_1k_2\,t + C$$

to solve for P.

Taking the natural exponential of both sides of the equation and multiplying by k_2-P gives

$$P = C\exp(k_1k_2\,t)\,(k_2-P).$$

Adding $C\exp(k_1k_2\,t)\,P$ to both sides, dividing by $1+C\exp(k_1k_2\,t)$ and setting $\alpha = k_1k_2$ gives the desired expression

$$\begin{aligned}P &= \frac{k_2\,C\exp(k_1k_2\,t)}{1+C\exp(k_1k_2\,t)}\\ &= \frac{k_2}{C_1\exp(-k_1k_2\,t)+1}\\ &= \frac{k_2}{C_1\exp(-\alpha\,t)+1}.\end{aligned}$$

Example 287. Let $k_1 = 10^{-5}$ and $k_2 = 10^4$ in Eq. 8.7 and let t be in hours. If the initial population is 2,000 individuals, the solution of the logistic equation is

$$P(t) = \frac{10^4}{1+4\exp\left(\frac{-t}{10}\right)}.$$

It should be fairly easy to see that as time goes toward infinity the population tends toward the equilibrium population of $P_E = 10^4$. This is shown in Fig. 8.3 on page 251.

For illustrative purposes, assume that instead of the $k_1 = 2\times 10^{-5}$ in Example 287, we take $k_2 = 10^4$. The solution of the differential equation is then

$$P(t) = \frac{10{,}000}{1 + C\mathrm{e}^{-\frac{t}{5}}}.$$

The solution curves for different initial conditions are shown in Fig. 8.4 on page 252. From the figure it appears that if we start with any population greater than zero, the solution to the logistic equation tends to the positive equilibrium population. This is in fact true, but showing it is true involves the existence and uniqueness theorem for solutions of differential equations. Since that is beyond this course, it is left for further reading or for later classes.

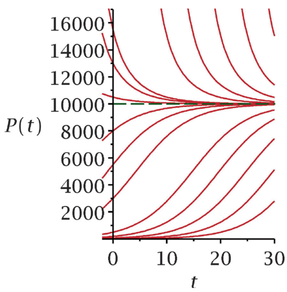


Fig. 8.4

Exercises

1. Evaluate the following integrals.

(a) $\displaystyle\int \frac{5x-18}{x^2-36}\,dx$

(b) $\displaystyle\int \frac{7y+2}{(y-4)(y+2)}\,dy$

(c) $\displaystyle\int 2\,\frac{3z+5}{(z-2)(z+9)}\,dz$

(d) $\displaystyle\int \frac{8t+45}{(t+5)(t+6)}\,dt$

(e) $\displaystyle\int \frac{6t-53}{(t-9)^2}\,dt$

(f) $\displaystyle\int \frac{x+12}{(x+6)^2}\,dx$

(g) $\displaystyle\int \frac{9y+28}{(y+3)^2}\,dy$

(h) $\displaystyle\int \frac{z+7}{(z+5)^2}\,dz$

(i) $\displaystyle\int \frac{19t^2+33t-42}{(t+3)(t^2-4)}\,dt$

(j) $\displaystyle\int \frac{16y^2+3y-124}{(y-1)(y^2-16)}\,dy$

(k) $\displaystyle\int \frac{22z^2+33z-124}{(z+1)(z^2-16)}\,dz$

(l) $\displaystyle\int \frac{19x^2-27x+2}{(x-2)(x^2-1)}\,dx$

(m) $\displaystyle\int \frac{5y^2-32y-51}{(y-7)(y+2)(y-9)}\,dy$

(n) $\displaystyle\int \frac{21z^2-10z-359}{(z+5)(z-7)(z+3)}\,dz$

(o) $\displaystyle\int \frac{16w^2+57w-289}{(w-5)(w+6)(w+7)}\,dw$

(p) $\displaystyle\int \frac{2t^2-9t-19}{(t-5)(t+9)(t-7)}\,dt$

(q) $\displaystyle\int \frac{2x+9}{x^2+6x+13}\,dx$

(r) $\displaystyle\int \frac{y-2}{y^2-8y+17}\,dy$

(s) $\displaystyle\int \frac{z+2}{z^2+2z+10}\,dz$

(t) $\displaystyle\int \frac{2w+3}{w^2+2w+5}\,dw$

(u) $\displaystyle\int \frac{2t^2-10t+15}{(t^2-6t+10)(t-2)}\,dt$

(v) $\displaystyle\int \frac{4x^2-25x+49}{(x^2-8x+25)(x-3)}\,dx$

(w) $\displaystyle\int \frac{4y^2-19y+30}{(y^2-6y+10)(y-1)}\,dy$

(x) $\displaystyle\int \frac{6z^2+35z+50}{(z^2+6z+10)(z+2)}\,dz$

2. Evaluate the following integrals.

(a) $\displaystyle\int \frac{13x+56}{x^2+9x+14}\,dx$

(b) $\displaystyle\int \frac{14y^2-52y+16}{y^3-16y-3y^2+48}\,dy$

(c) $\displaystyle\int \frac{6x^2-11x+13}{x^3-3x^2+7x-5}\,dx$

(d) $\displaystyle\int \frac{z^2+2z-5}{z^3+3z^2-15z-5}\,dz$

(e) $\displaystyle\int \frac{4z+31}{z^2+12z+36}\,dz$

(f) $\displaystyle\int \frac{10t-21}{t^2-5t+6}\,dt$

(g) $\displaystyle\int \frac{10x^2-185-x}{x^3-25x-2x^2+50}\,dx$

(h) $\displaystyle\int \frac{12w-24}{w^2-3w}\,dw$

(i) $\displaystyle\int \frac{4t^2-25t+75}{t^3-9t^2+33t-25}\,dt$

(j) $\displaystyle\int \frac{19w^2+37w-36}{w^3-w+4w^2-4}\,dw$

(k) $\displaystyle\int \frac{20z^2+3z-231}{z^3-37z-84}\,dz$

(l) $\displaystyle\int \frac{8y-4}{y^2-2y+1}\,dy$

(m) $\displaystyle\int \frac{15x^2-28x-64}{x^3-16x-2x^2+32}\,dx$

(n) $\displaystyle\int \frac{w^3+w}{w^4+2w^2+1}\,dw$

(o) $\displaystyle\int \frac{13y^3-98y}{y^4-15y^2+56}\,dy$

(p) $\displaystyle\int \frac{6x^3-16x}{x^4-8x^2+20}\,dx$

(q) $\displaystyle\int \frac{5t+3}{t^2-2t+1}\,dt$

(r) $\displaystyle\int \frac{5z^2-23z+36}{z^3-8z^2+30z-36}\,dz$

3. Evaluate the following definite integrals.

(a) $\displaystyle\int_0^2 \frac{x-9}{x^2-3x-4}\,dx$

(b) $\displaystyle\int_{-1}^0 \frac{2y+21}{y^2+y-6}\,dy$

(c) $\displaystyle\int_2^4 \frac{3z+4}{z^2+6z+9}\,dz$

(d) $\displaystyle\int_{-5}^{-2} \frac{3w+7}{w^2+2w+1}\,dw$

(e) $\displaystyle\int_0^4 \frac{3x^2+10x+10}{x^3+3x^2+4x+2}\,dx$

(f) $\displaystyle\int_2^5 \frac{7y^2-18x-30y+89}{(y-5)(y^2-6x+13)}\,dy$

4. Solve the following differential equations.

(a) $\displaystyle\frac{dy}{dx}=\frac{100x-x^2}{10}.$

(b) $\displaystyle\frac{dy}{dx}=\frac{10x-30x^2}{10}.$

(c) $\displaystyle\frac{dy}{dx}=2x-\frac{x^2}{37}.$

(d) $\displaystyle\frac{dP}{dt}=0.05P-0.00001P^2.$

(e) $\displaystyle\frac{dP}{dt}=0.2P-0.005P^2.$

5. A population follows the logistic growth model, Eqs. 8.6 and 8.7. Given the following information, find the population P as a function of time.

(a) $\alpha = 0.01$, $\beta = 3 \times 10^{-4}$ and $P(0) = 200$.
(b) $\alpha = 0.05$, $\beta = 2 \times 10^{-5}$ and $P(0) = 150$.
(c) $\alpha = 0.05$, $\beta = 2 \times 10^{-5}$ and $P(0) = 3{,}500$.
(d) $k_1 = 10^4$, $k_2 = 3 \times 10^3$ and $P(0) = 200$.
(e) $k_1 = 2 \times 10^4$, $k_2 = 4 \times 10^3$ and $P(0) = 250$.
(f) $k_1 = 2 \times 10^4$, $k_2 = 4 \times 10^3$ and $P(0) = 4{,}000$.

6. What happens to the solutions when the differential equation (8.6) is changed to

$$\frac{dP}{dt} = \alpha P + \beta P^2 \ ?$$

Recall that α and β are both positive. Is this a more realistic model than exponential growth?

7. A population $P(t)$ might following either the exponential growth model $P'(t) = kP(t)$ or the logistic model $P'(t) = kP(t)\left(1 - \frac{P(t)}{P_\infty}\right)$. Here k is the same constant. If $P(0) = 20$ and $P'(0) = 10$, what is the difference in the predicted populations when $t = 10$?

8.4 Trigonometric Integrals

Many models for physical phenomena involve periodic or almost periodic functions. A large number of the functions in these models can be approximated by sums of trigonometric functions or functions built with trigonometric functions. For this reason it is important to be able to evaluate some integrals involving trigonometric functions exactly. In this section it is shown how to do some of these integrals.

In the section on integration by parts some integrals involving trigonometric functions were done. In this section the list of types of integrals involving trigonometric integrals that can be integrated is increased. The basic idea is to rewrite the integrals through trigonometric identities and substitutions into a form that has already been considered. Although all rational functions of trigonometric functions can be reduced to integrals involving rational functions, the ideas considered here are somewhat limited.

The easiest forms to work with are those with an integrand of the form $f(x) = \cos^m(x)\sin^\alpha(x)$ or $g(x) = \cos^\alpha(x)\sin^m(x)$ where $m = 2n+1$ is an odd positive integer and α is a real number. For the form $f(x) = \cos^m(x)\sin^\alpha(x)$ we can write

$$\begin{aligned}\cos^m(x)\sin^\alpha(x) &= \cos(x)\,(\cos^2(x))^n \sin^\alpha(x)\\ &= \cos(x)\,(1-\sin^2(x))^n \sin^\alpha(x).\end{aligned}$$

Since $\cos(x) = \frac{d}{dx}\sin(x)$, using the substitution $u = \sin(x)$, this gives

$$\begin{aligned}\int \cos^m(x)\sin^\alpha(x)\,dx &= \int (1-\sin^2(x))^n \sin^\alpha(x)\cos(x)\,dx\\ &= \int (1-u^2)^n u^\alpha\,du.\end{aligned}$$

The last integral is simply a sum of constants times powers of u and can be integrated. Using the same method where one exchanges the roles of $\sin(x)$ and $\cos(x)$ can be used to integrate functions of form $g(x) = \cos^{\alpha}(x)\sin^{m}(x)$. A couple examples illustrate this adequately.

Example 288. Consider the integral

$$\int \cos^3(x)\, \sin^2(x)\, dx.$$

Rewriting $\cos^3(x)$ and using the substitution $u = \sin(x)$ allows us to evaluate the integral.

$$\begin{aligned}
\int \cos^3(x)\, \sin^2(x)\, dx &= \int \left(1-\sin^2(x)\right) \sin^2(x) \cos(x)\, dx \\
&= \int \left(1-u^2\right) u^2\, du \\
&= \int u^2 - u^4\, du \\
&= \frac{u^3}{3} - \frac{u^5}{5} + C \\
&= \frac{\sin^3(x)}{3} - \frac{\sin^5(x)}{5} + C.
\end{aligned}$$

The next example shows how this works when α is not an integer and when there are more terms.

Example 289. The integral

$$\int \left(\frac{1}{\sqrt{\cos(z)}} + \sqrt{\cos(z)}\right) \sin^3(z)\, dz$$

can be evaluated by rewriting the $\sin^3(z)$ term and using the substitution $w = \cos(z)$.

$$\begin{aligned}
&\int \left(\frac{1}{\sqrt{\cos(z)}} + \sqrt{\cos(z)}\right) \sin^3(z)\, dz \\
&\qquad = \int \left(\frac{1}{\sqrt{\cos(z)}} + \sqrt{\cos(z)}\right) \left(1-\cos^2(z)\right) \sin(z)\, dz \\
&\qquad = \int \left(\frac{1}{\sqrt{w}} + \sqrt{w}\right) \left(1-w^2\right) (-dw) \\
&\qquad = \int \left(-\frac{1}{\sqrt{w}} - \sqrt{w} + w^{\frac{3}{2}} + w^{\frac{5}{2}}\right) dw \\
&\qquad = \frac{2}{7} w^{\frac{7}{2}} + \frac{2}{5} w^{\frac{5}{2}} - \frac{2}{3} w^{\frac{3}{2}} - 2 w^{\frac{1}{2}} + C \\
&\qquad = \frac{2}{7} \sin^{\frac{7}{2}}(z) + \frac{2}{5} \sin^{\frac{5}{2}}(z) - \frac{2}{3} \sin^{\frac{3}{2}}(z) - 2 \sin^{\frac{1}{2}}(z) + C.
\end{aligned}$$

This idea can also be used to deal with integrals of the forms $f(x) = \cos^m(x)\sin^n(x)$ and $g(x) = \cos^n(x)\sin^m(x)$ where m is an odd negative integer and n is an integer. We simply multiply the numerator or denominator by $\cos(x)$ or $\sin(x)$ and use the Pythagorean Theorem to get a form that can be changed into a rational function. The next example show this works.

Example 290. The integral

$$\int \sec(\theta)\,d\theta = \int \frac{1}{\cos(\theta)}\,d\theta$$

has a odd power of $\cos(\theta)$ in the denominator. Multiplying the numerator and denominator by $\cos(\theta)$, applying the Pythagorean Theorem, and using the substitution $u = \sin(\theta)$ we get

$$\begin{aligned}\int \frac{1}{\cos(\theta)}\,d\theta &= \int \frac{\cos(\theta)}{\cos^2(\theta)}\,d\theta \\ &= \int \frac{\cos(\theta)}{1-\sin^2(\theta)}\,d\theta \\ &= \int \frac{1}{1-u^2}\,du.\end{aligned}$$

The last integral can be done using partial fractions with

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

(Check that this is correct.) The integral now becomes

$$\begin{aligned}\int \frac{1}{\cos(\theta)}\,d\theta &= \frac{1}{2}\int \frac{1}{(1+u)} + \frac{1}{(1-u)}\,du \\ &= \frac{1}{2}\left(\ln|1+u| - \ln|1-u|\right) + C \\ &= \frac{1}{2}\ln\left|\frac{1+u}{1-u}\right| + C \\ &= \frac{1}{2}\ln\left|\frac{1+\sin(\theta)}{1-\sin(\theta)}\right| + C.\end{aligned}$$

The common form for this integral is

$$\int \sec(\theta)\,d\theta = \ln|\sec(\theta)+\tan(\theta)| + C.$$

Everyone should show that the two forms are equivalent.

The last examples show how we can deal with an odd power of $\sin(x)$ or $\cos(x)$. Dealing with an expression that only has positive even powers of both $\sin(x)$ and $\cos(x)$ is more complicated, but doable. We reduce the even powers of $\sin(x)$ and $\cos(x)$ using the half angle formulas

$$\begin{aligned}\cos^2(x) &= \frac{1+\cos(2x)}{2} \quad \text{and} \\ \sin^2(x) &= \frac{1-\cos(2x)}{2}.\end{aligned}$$

These formulas can be used until we only have odd powers of $\cos(2nx)$ for various n's. The only case considered here is positive powers of $\sin(x)$ and $\cos(x)$. A fairly simple example illustrates this technique.

Example 291. The integral

$$\int \sin^2(\theta)\cos^2(\theta)\,d\theta$$

can be rewritten using the double angle formulas.

$$\begin{aligned}
\int \sin^2(\theta)\cos^2(\theta)\,d\theta &= \int \left(\frac{1-\cos(2\theta)}{2}\right)\left(\frac{1+\cos(2\theta)}{2}\right)d\theta \\
&= \int \frac{1}{4}-\frac{\cos^2(2\theta)}{4}\,d\theta \\
&= \int \frac{1}{4}-\frac{1+\cos(4\theta)}{8}\,d\theta \\
&= \frac{\theta}{8}-\frac{\sin(4\theta)}{32}+C.
\end{aligned}$$

There are also forms involving $\tan(\theta)$, $\sec(\theta)$, $\cot(\theta)$, and $\csc(\theta)$ that can be simplified using the Pythagorean theorem. The two identities used here are

$$\begin{aligned}
&\sec^2(\theta) = 1+\tan^2(\theta) \\
\text{and}\qquad &\csc^2(\theta) = 1+\cot^2(\theta).
\end{aligned}$$

Using these identities combined with the facts that $\sec'(\theta) = \tan(\theta)\sec(\theta)$, $\tan'(\theta) = \sec^2(\theta)$, $\csc'(\theta) = -\cot(\theta)\csc(\theta)$, and $\cot'(\theta) = -\csc^2(\theta)$ allows us to convert some integrals to integrals of polynomials and rational functions.

Example 292. Consider the integral

$$\int \tan^4(\theta)\,d\theta.$$

Here we change the $\tan^2(\theta)$ factors to $\sec^2(\theta)-1$ factors one at a time.

$$\begin{aligned}
\int \tan^4(\theta)\,d\theta &= \int \tan^2(\theta)\left(\sec^2(\theta)-1\right)d\theta \\
&= \int \left(\tan^2(\theta)\sec^2(\theta)-\tan^2(\theta)\right)d\theta \\
&= \int \left(\tan^2(\theta)\sec^2(\theta)-\left(\sec^2(\theta)-1\right)\right)d\theta \\
&= \int \left(\tan^2(\theta)\sec^2(\theta)-\sec^2(\theta)\right)d\theta+\int 1\,d\theta.
\end{aligned}$$

Substituting $u = \tan(\theta)$ and $du = \sec^2(\theta)\,d\theta$ gives

$$\begin{aligned}
\int \tan^4(\theta)\,d\theta &= \int \left(u^2-1\right)du+(\theta+C) \\
&= \frac{u^3}{3}-u+(\theta+C) \\
&= \frac{\tan^3(\theta)}{3}-\tan(\theta)+\theta+C.
\end{aligned}$$

We can also reduce other problems to rational integrals.

Example 293. Consider the integral

$$\int \frac{\tan(x)}{3-\tan^2(x)}\,dx.$$

The idea is to rewrite the rational function in terms of $\tan(x)$ to one in terms of $u = \sec(x)$ where the numerator contains $du = \sec(x)\tan(x)\,dx$.

$$\begin{aligned}\int \frac{\tan(x)}{3-\tan^2(x)}\,dx &= \int \frac{\sec(x)\tan(x)}{\sec(x)\,(4-\sec^2(x))}\,dx \\ &= \int \frac{1}{u\,(4-u^2)}\,du \\ &= \int \left(-\frac{1}{8\,(u+2)} + \frac{1}{4u} - \frac{1}{8\,(u-2)}\right) du \\ &= -\frac{1}{8}\ln|u+2| + \frac{1}{4}\ln|u| - \frac{1}{8}\ln|u-2| + C \\ &= \frac{1}{8}\ln\left|\frac{u^2}{u^2-4}\right| + C \\ &= \frac{1}{8}\ln\left|\frac{\sec^2(x)}{\sec^2(x)-4}\right| + C\end{aligned}$$

If we want to do this integral in a different way we can multiply the numerator and denominator of

$$\frac{\tan(x)}{3-\tan^2(x)}$$

by $\cos^2(x)$ and use the substitution $v = \cos(x)$ with $dv = -\sin(x)\,dx$. We still get a partial fractions integral to evaluate.

Exercises

1. Evaluate the following integrals.

(a) $\displaystyle\int \cos^3(x)\,dx$

(b) $\displaystyle\int \sin^3(y)\cos^2(y)\,dy$

(c) $\displaystyle\int \cos(2z)\sin^2(z)\,dz$

(d) $\displaystyle\int \sin^4(t)\,dt$

(e) $\displaystyle\int \csc^4(t)\,dt$

(f) $\displaystyle\int \cos(2x)\sin^3(x)\,dx$

(g) $\displaystyle\int \frac{\cos(y)}{1+\cos^2(y)}\,dy$

(h) $\displaystyle\int \sec^4(z)\tan^3(z)\,dz$

(i) $\displaystyle\int \cos^2(3t)\,dt$

(j) $\displaystyle\int \frac{1}{\sin^2(5y)}\,dy$

(k) $\displaystyle\int \frac{1}{\cos(z)\,\sin(z)}\,dz$

(l) $\displaystyle\int \frac{1}{\cos^3(z)\,\sin(z)}\,dz$

(m) $\displaystyle\int \sqrt{\sin(x)}\,\cos(x)\,dx$

(n) $\displaystyle\int \sqrt{\sin(y)}\,\cos^3(y)\,dy$

(o) $\displaystyle\int \cos(2z)\tan(2z)\,dz$

(p) $\displaystyle\int \tan^3(3w)\,dw$

(q) $\displaystyle\int e^t\cos^2(t)\,dt$

(r) $\displaystyle\int \frac{1}{\cos^3(2\theta)}\,d\theta$

(s) $\displaystyle\int e^t\cos^2(t)\,dt$

(t) $\displaystyle\int \frac{10}{\sin^3(y)}\,dy$

2. Evaluate the following definite integrals.

(a) $\displaystyle\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \sin^3(x)\cos^2(x)\,dx$

(b) $\displaystyle\int_{0}^{\frac{\pi}{6}} \sin^2(z)\cos^2(z)\,dz$

(c) $\displaystyle\int_{\frac{\pi}{4}}^{\frac{2\pi}{3}} \csc^4(\theta)\cot(\theta)\,d\theta$

3. Solve the following differential equations.

(a) $\dfrac{dy}{dx} = x\cos^2(2y)$. (b) $\dfrac{d\theta}{dt} = \tan^2(t)\sin^2(\theta)$.

8.5 Integration Using Trigonometric Substitution

The substitution technique presented in this section is a little different. Most of the substitutions in the previous sections have involved replacing a function of x with a variable u. In this section the variable x is replaced with a function of u.

The substitution is accomplished by setting x equal to a trigonometric function of, for example, θ. A simple example is setting $x = \sin(\theta)$ in the expression $\sqrt{1-x^2}$ to get

$$\begin{aligned}\sqrt{1-x^2} &= \sqrt{1-\sin^2(\theta)}\\ &= \cos(\theta),\end{aligned}$$

if $\theta \in [-\pi/2, \pi/2]$. This type of substitution is used to get rational trigonometric functions that can be integrated with the techniques from the previous sections. A simple example will illustrate how this works.

Example 294. Consider the integral

$$\int \sqrt{1-x^2}\,dx$$

Applying the substitution $x = \sin(\theta)$ and $dx = \cos(\theta)\,d\theta$ and using $\cos(\theta) = \sqrt{1-\sin^2(\theta)}$ gives

$$\int \sqrt{1-x^2}\,dx = \int \cos^2(\theta)\,d\theta.$$

Using the trigonometric identities $\cos^2(\theta) = (1+\cos(2\theta))/2$ and $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ yields

$$\begin{aligned}\int \sqrt{1-x^2}\,dx &= \int \cos^2(\theta)\,d\theta\\ &= \int \frac{1+\cos(2\theta)}{2}\,d\theta\\ &= \frac{\theta}{2} + \frac{\sin(2\theta)}{4} + C\end{aligned}$$

$$
\begin{aligned}
&= \frac{\theta}{2} + \frac{\sin(\theta)\cos(\theta)}{4} + C \\
&= \frac{\sin^{-1}(x)}{2} + \frac{x\sqrt{1-x^2}}{4} + C.
\end{aligned}
$$

By the definition $x = \sin(\theta)$, we have $\theta = \sin^{-1}(x)$.

The substitutions in this section are restricted to those involving three forms of the Pythagorean Theorem:

$$
\begin{aligned}
\cos^2(\theta) &= 1 - \sin^2(\theta), \\
\sec^2(\theta) &= 1 + \tan^2(\theta), \text{ and} \\
\tan^2(\theta) &= \sec^2(\theta) - 1.
\end{aligned}
$$

These can be used to remove a square root around the corresponding expressions: $\sqrt{1-x^2}$, $\sqrt{1+x^2}$, and $\sqrt{x^2-1}$. This has already been demonstrated for the form $\sqrt{1-x^2}$. The next two examples show how the other two substitutions can be used. One of these integrals does not include a square root.

Example 295. Consider the integral

$$
\int \frac{1}{1+z^2}\,dz.
$$

(You should know what this is from Calculus I, Example 212.) This has the form $1+z^2$ which corresponds to the identity $\sec^2(\theta) = 1+\tan^2(\theta)$

Setting $z = \tan(\theta)$ and $dz = \sec^2(\theta)\,d\theta$ makes the integral simple.

$$
\begin{aligned}
\int \frac{1}{1+z^2}\,dz &= \int \frac{1}{1+\tan^2(\theta)}\,\sec^2(\theta)\,d\theta \\
&= \int \frac{1}{\sec^2(\theta)}\,\sec^2(\theta)\,d\theta \\
&= \int d\theta \\
&= \theta + C \\
&= \tan^{-1}(z) + C.
\end{aligned}
$$

(Since $z = \tan(\theta)$, $\theta = \tan^{-1}(z)$.)

Example 296. The integral

$$
\int \frac{1}{\sqrt{w^2-1}}\,dw
$$

contains a term that corresponds to $\tan^2(\mu) = \sec^2(\mu) - 1$ with $w = \sec(\mu)$. Noting that $dw = \sec(\mu)\tan(\mu)\,d\mu$ allows us to integrate the expression.

$$
\begin{aligned}
\int \frac{1}{\sqrt{w^2-1}}\,dw &= \int \frac{1}{\tan(\mu)}\,\sec(\mu)\tan(\mu)\,d\mu \\
&= \int \frac{1}{\cos(\mu)}\,d\mu \\
&= \int \frac{\cos(\mu)}{1-\sin^2(\mu)}\,d\mu.
\end{aligned}
$$

Replacing $\sin(\mu)$ with u gives

$$\begin{aligned}\int \frac{1}{\sqrt{w^2-1}}\,dw &= \int \frac{1}{1-u^2}\,du\\ &= \frac{1}{2}\int \frac{1}{1+u}+\frac{1}{1-u}\,du\\ &= \frac{1}{2}\left(\ln|1+u|-\ln|1-u|\right)+C\\ &= \frac{1}{2}\ln\left|\frac{1+\sin(\mu)}{1-\sin(\mu)}\right|+C.\end{aligned}$$

In order to return to a function of w we must translate $\sin(\mu)$ into a function of w. Recalling that $w=\sec(\mu)$ and that in right triangle geometry the $\sec(\mu)$ is the hypotenuse over the adjacent, we can set $\sec(\mu)=w/1$ and use Fig. 8.5 on page 261.

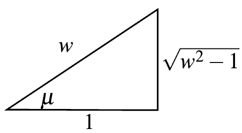


Fig. 8.5

From the figure we can see that $\sin(\mu)=(\sqrt{w^2-1})/w$ and hence

$$\begin{aligned}\frac{1+\sin(\mu)}{1-\sin(\mu)} &= \frac{1+\frac{\sqrt{w^2-1}}{w}}{1-\frac{\sqrt{w^2-1}}{w}}\\ &= \frac{w+\sqrt{w^2-1}}{w-\sqrt{w^2-1}}\\ &= \left(w+\sqrt{w^2-1}\right)^2.\end{aligned}$$

The final formula we get for the integral is

$$\int \frac{1}{\sqrt{w^2-1}}\,dw = \ln\left|w+\sqrt{w^2-1}\right|+C.$$

Using methods similar to the $\sec(\mu)$ substitution above we can integrate some expressions involving $1+w^2$ and $1-w^2$ using the substitutions $w=\tan(\theta)$ and $w=\sin(\theta)$. The next two examples show these substitutions can be applied.

Example 297. The integral

$$\int_{-1}^{1} \frac{1}{\sqrt{1+w^2}}\,dw$$

can be evaluated using the substitution $w=\tan(\theta)$. This gives $1+w^2=1+\tan^2(\theta)=\sec^2(\theta)$ and $dw=\sec^2(\theta)\,d\theta$. The integral is now

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{1+w^2}}\,dv &= \int_0^{\frac{\theta}{4}} \frac{1}{\sqrt{\sec^2(\theta)}}\sec^2(\theta)\,d\theta\\ &= \int_0^{\frac{\theta}{4}} \sec(\theta)\,d\theta\end{aligned}$$

$$
\begin{aligned}
&= \ln|\sec(\theta)+\tan(\theta)|\,\bigg|_0^{\frac{\theta}{4}}\\
&= \ln\left(\sqrt{2}+1\right)-\ln(1)\\
&= \ln\left(\sqrt{2}+1\right).
\end{aligned}
$$

If we use the substitution $w=\tan(\theta)$ and we get an expression involving trigonometric functions, the triangle in Fig. 8.6 on page 262 can be used to change the result into functions of w.

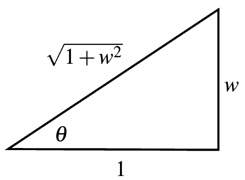


Fig. 8.6

In this figure $\tan(\theta)=w$, $\cos(\theta)=1/\sqrt{1+w^2}$, and $\sin(\theta)=w/\sqrt{1+w^2}$. This gives all of the required trigonometric functions for a $w=\tan(\theta)$ substitution.

All that is left for this section is to demonstrate a $w=\sin(\theta)$ substitution for the form $1-w^2$.

Example 298. To evaluate the integral

$$\int \frac{1}{w\sqrt{1-w^2}}\,dw$$

we can use the substitution $w=\sin(\theta)$. This gives, assuming appropriate domains and ranges, $\sqrt{1-w^2}=\sqrt{\cos^2(\theta)}=\cos(\theta)$ and $dw=\cos(\theta)\,d\theta$.

The integral now becomes

$$
\begin{aligned}
\int \frac{1}{w\sqrt{1-w^2}}\,dw &= \int \frac{\cos(\theta)}{\sin(\theta)\cos(\theta)}\,d\theta\\
&= \int \csc(\theta)\,d\theta\\
&= \int \frac{\csc^2(\theta)+\csc(\theta)\cot(\theta)}{\csc(\theta)+\cot(\theta)}\,d\theta.
\end{aligned}
$$

In the last integral the numerator is the negative of the derivative of the denominator. This means that

$$\int \frac{1}{w\sqrt{1-w^2}}\,dw = -\ln|\csc(\theta)+\cot(\theta)|+C.$$

Using Fig. 8.7 on page 263 we can change the result into a function of w.
From the figure we get that $\csc(\theta)=1/w$ and $\cot(\theta)=\sqrt{1-w^2}/w$. Replacing $\csc(\theta)$ and $\cot(\theta)$ gives a final expression for the integral,

$$\int \frac{1}{w\sqrt{1-w^2}}\,dw = -\ln\left|\frac{1+\sqrt{1-w^2}}{w}\right|+C.$$

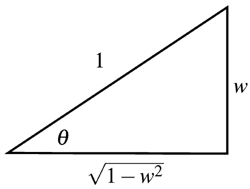


Fig. 8.7

Exercises

1. Evaluate the following integrals.

(a) $\int \sqrt{4-x^2}\,dx$

(b) $\int \sqrt{4+y^2}\,dy$

(c) $\int \frac{x}{\sqrt{1-x^2}}\,dx$

(d) $\int \frac{x^2}{\sqrt{-1+x^2}}\,dx$

(e) $\int \frac{x^2}{-1+x^2}\,dx$

(f) $\int \frac{w}{\sqrt{9+4w^2}}\,dw$

(g) $\int \frac{x}{\sqrt{-1+x^2}}\,dx$

(h) $\int \frac{(z-3)^2}{\sqrt{-5-z^2+6z}}\,dz$

(i) $\int \frac{(w-2)^2}{\sqrt{10+w^2+2w}}\,dw$

(j) $\int \sqrt{x^2-4}\,dx$

(k) $\int \frac{(w-2)^2}{10+w^2+2w}\,dw$

(l) $\int \frac{\sqrt{1-x^2}}{x^2}\,dx$

(m) $\int \frac{\sqrt{y^2-1}}{y^2}\,dy$

(n) $\int \frac{\sqrt{1+w^2}}{w^2}\,dw$

(o) $\int \sqrt{y^2-4y+8}\,dy$

(p) $\int z^3\sqrt{z^2-4}\,dy$

2. Evaluate the following definite integrals.

(a) $\int_{-1}^{1} \sqrt{4-x^2}\,dx$

(b) $\int_{0}^{4} \frac{z^2}{\sqrt{z^2+9}}\,dz$

(c) $\int_{3}^{5} \frac{1}{\sqrt{w^2+4w+3}}\,dw$

(d) $\int_{-4}^{-2} \sqrt{-y^2-4y}\,dy$

(e) $\int_{-3}^{-1} \frac{\sqrt{x^2+25}}{x^2}\,dx$

(f) $\int_{3}^{5} \frac{\sqrt{w^2-4w+5}}{(w-2)^2}\,dw$

3. Solve the following differential equations.

(a) $\frac{dz}{dt} = \frac{z^2+1}{t^2+4}$

(b) $\frac{dy}{dx} = \sin(x)\sqrt{y^2-4}$

(c) $\frac{dw}{ds} = \mathrm{e}^s\left(w^2+4w+5\right)$

(d) $\frac{dy}{dt} = \left(t^2+\mathrm{e}^t\right)\sqrt{1-4y^2}$

8.6 Integration Using Tables and Computer Algebra Systems

There are cases when we do not recognize the form of an integral or when we simply do not know how to find an antiderivative. In these cases we can use a table of integrals or a computer algebra system to help find an antiderivative. The main skills required for using a table of integrals are the ability to recognize which form is similar to the integrand we have and the ability to use substitutions to transform the integrand into a form in a table.

This is a skill that is no longer required in many cases. However, it can still be useful when we need to reduce an integral to a "nicer" form for numerical integration. We have already used this skill for integration using trigonometric substitutions. As an example we consider an integral that can be done using a trigonometric substitution.

Example 299. Consider the integral

$$\int x^2\sqrt{4x^2+1}\,dx. \tag{8.8}$$

This integral can be evaluated using the substitution $2x=\tan(\theta)$. That is a long process.

However, in a table of integrals we can usually find the form

$$\int u^2\sqrt{u^2+a^2}\,du = \frac{u}{4}\left(u^2+a^2\right)^{3/2} - \frac{a^2}{8}u\sqrt{u^2+a^2} - \frac{a^4}{8}\ln\left(u+\sqrt{u^2+a^2}\right). \tag{8.9}$$

Setting $u=2x$, $dx=du/2$ and $a=1$, then substituting into Eq. 8.8, we get

$$\int x^2\sqrt{4x^2+1}\,dx = \frac{1}{8}\int u^2\sqrt{u^2+1}\,du.$$

Applying the form in Eq. 8.9 we have

$$\begin{aligned}\int x^2\sqrt{4x^2+1}\,dx &= \frac{1}{8}\left(\frac{u}{4}\left(u^2+1\right)^{3/2} - \frac{1}{8}u\sqrt{u^2+1} - \frac{a^4}{8}\ln\left(u+\sqrt{u^2+1}\right)\right)+C\\ &= \frac{1}{8}\left(\frac{x}{2}\left((2x)^2+1\right)^{3/2} - \frac{x}{4}\sqrt{(2x)^2+1}\right.\\ &\qquad \left.-\frac{1}{8}\ln\left(2x+\sqrt{(2x)^2+1}\right)\right)+C\end{aligned}$$

Note that the forms in a table of integrals normally do not contain a constant of integration.

The biggest problem in using a table of integrals is recognizing if there is a form matching the integral you are attempting. Tables of integrals can be very large, containing hundreds of forms. It may be difficult to find an extensive table of integrals on the internet. Table 8.1 is a small portion of one table that we will use in the examples and exercises. These are not standard integrals, but some of the less common integrals.

To use this table we must translate an integral into one of these forms.

Example 300. We consider the integral

$$\int y\sin\left(\ln\left(\frac{y^2+4}{5}\right)\right)dy.$$

1	$\int \frac{1}{(cx^2+bx+a)^2}\,dx = \frac{2cx+b}{(cx^2+bx+a)\,(4ac-b^2)} + \frac{2c}{4ac-b^2}\int \frac{1}{cx^2+bx+a}\,dx$
2	$\int \frac{1}{c^3+x^3}\,dx = \frac{1}{6c^2}\ln\left(\frac{(c+x)^3}{c^3+x^3}\right) + \frac{1}{c\sqrt{3}}\tan^{-1}\left(\frac{2x-c}{c\sqrt{3}}\right)$
3	$\int \frac{1}{x\sqrt{x^2+a^2}}\,dx = -\frac{1}{a}\ln\left(\frac{a+\sqrt{x^2+a^2}}{x}\right)$
4	$\int \frac{1}{x\left(\sqrt{a^2-x^2}\right)^3}\,dx = \frac{1}{a^2\sqrt{a^2-x^2}} + \frac{1}{a^3}\ln\left(\frac{a+\sqrt{a^2-x^2}}{x}\right)$
5	$\int \frac{1}{\sqrt{2ax-x^2}}\,dx = \arcsin\left(\frac{x-a}{\lvert a\rvert}\right)$
6	$\int \frac{1}{(2ax-x^2)^{\frac{3}{2}}}\,dx = \frac{x-a}{a^2\sqrt{2ax-x^2}}$
7	$\int (\arccos(ax))^2\,dx = x\,(\arccos(ax))^2 - 2x - \frac{2\sqrt{1-a^2x^2}}{a}\arccos(ax)$
8	$\int \sin(\ln(x))\,dx = \frac{x}{2}\left(\sin(\ln(x)) - \cos(\ln(x))\right)$
9	$\int x^2\sqrt{ax+b}\,dx = \frac{2\left(8a^2-12abx+15b^2x^2\right)\sqrt{(a+bx)^3}}{105b^3}$
10	$\int \frac{x^2}{\sqrt{ax+b}}\,dx = \frac{2\left(8a^2-4abx+3b^2\right)}{15b^3}\sqrt{a+bx}$

Table 8.1 A small table of integrals

This is similar to the eighth integral in the table, except for the fact that there is $(y^2+4)/5$ inside the natural logarithm function. Using the substitution $u=(y^2+4)/5$, and $y\,dy = 5\,du/2$ we get the integral

$$\begin{aligned}\int y\sin\left(\ln\left(\frac{y^2+4}{5}\right)\right)dy &= \frac{5}{2}\int \sin(\ln(u))\,du\\ &= \frac{5}{2}\left(\frac{u}{2}\left(\sin(\ln(u)) - \cos(\ln(u))\right)\right) + C\\ &= \frac{y^2+4}{4}\left(\sin\left(\ln\left(\frac{y^2+4}{5}\right)\right) - \cos\left(\ln\left(\frac{y^2+4}{5}\right)\right)\right) + C\end{aligned}$$

One of the main reasons that tables of integrals are not used very much is that calculators and computers with computer algebra systems can do most integrals. These programs are usually better and more accurate than humans. For illustration purposes we will see how a TI-Nspire CAS calculator and the computer algebra system Maple can be used to evaluate integrals. There are other choices, but two examples should be enough since the patterns of use are usually similar.

Example 301. If we want to find the integral of

$$\int \frac{w^3-5w+4}{\sqrt{w^2+2w+2}}\,dw$$

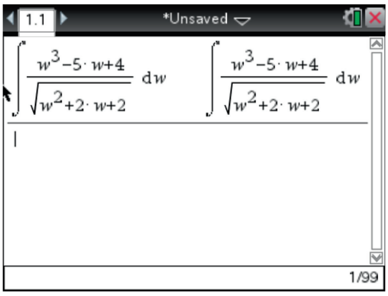


Fig. 8.8

using Maple we enter the command *int((w^3-5*w+4)/sqrt(w^2+2*w+2),w);* . The output is

$$\frac{1}{3}w^2\sqrt{w^2+2w+2}-\frac{5}{6}w\sqrt{w^2+2w+2}-\frac{23}{6}\sqrt{w^2+2w+2}+\frac{19}{2}\operatorname{arcsinh}(w+1).$$

When we try this integral with a TI-Nspire CAS we get no result. The calculator cannot do this integral see Fig. 8.8 on page 266. It is important to realize that there is nothing wrong with the calculator. Typically calculators do not contain the full algorithms used for integration. This means that there are more integrals that they cannot do than a full CAS, e.g. Mathematica or Maple, cannot do.

Example 302. If we try to evaluate the integral

$$\int \frac{1}{\cos(y)}\,dy$$

with a TI-Nspire CAS we get the result in Fig. 8.9 on page 266.

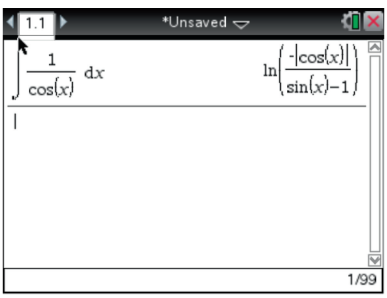


Fig. 8.9

We get

$$\ln(\sec(y)+\tan(y))$$

from Maple. This is almost the common form that we see in most calculus classes. (What is different?)

Exercises

1. Translate each of the following into one of the forms in the integral Table 8.1 on page 265. Then find the integral.

(a) $\displaystyle\int \frac{3}{x^3 - 3x^2 + 3x + 3}\,dx$

(b) $\displaystyle\int \frac{7}{\sqrt{8z - 4 - z^2}}\,dz$

(c) $\displaystyle\int \sin(4y)\left(\arccos\left(2\cos(4y) + 2\right)\right)^2 dy$

(d) $\displaystyle\int \frac{w}{(4w^4 + 10w^2 + 3)^2}\,dw$

(e) $\displaystyle\int s^2 \sin\left(\ln\left(4s^3 - 2\right)\right) ds$

(f) $\displaystyle\int \frac{\sin^2(x)\cos(x)}{\sqrt{4\sin(x) + 1}}\,dx$

(g) $\displaystyle\int \frac{2}{x(81 - x^4)^{3/2}}\,dx$

(h) $\displaystyle\int \frac{3}{(y+1)\sqrt{y^2 + 2y + 4}}\,dy$

(i) $\displaystyle\int \left(t^2 + 4t + 4\right)\sqrt{2t + 10}\,dt$

(j) $\displaystyle\int \frac{w^5}{\sqrt{4w^2 + 7}}\,dw$

2. Use a computer algebra to evaluate the each of the following integrals and then compare that result with the result for the integral from Exercise 1.

(a) $\displaystyle\int \frac{3}{x^3 - 3x^2 + 3x + 3}\,dx$

(b) $\displaystyle\int \frac{7}{\sqrt{8z - 4 - z^2}}\,dz$

(c) $\displaystyle\int \sin(4y)\left(\arccos\left(2\cos(4y) + 2\right)\right)^2 dy$

(d) $\displaystyle\int \frac{w}{(4w^4 + 10w^2 + 3)^2}\,dw$

(e) $\displaystyle\int s^2 \sin\left(\ln\left(4s^3 - 2\right)\right) ds$

(f) $\displaystyle\int \frac{\sin^2(x)\cos(x)}{\sqrt{4\sin(x) + 1}}\,dx$

(g) $\displaystyle\int \frac{2}{x(81 - x^4)^{3/2}}\,dx$

(h) $\displaystyle\int \frac{3}{(y+1)\sqrt{y^2 + 2y + 4}}\,dy$

(i) $\displaystyle\int \left(t^2+4t+4\right)\sqrt{2t+10}\,dt$

(j) $\displaystyle\int \frac{w^5}{\sqrt{4w^2+7}}\,dw$

3. In this exercise you will see how one can get a TI CAS calculator to integrate

$$\int \frac{w^3-5w+4}{\sqrt{w^2+2w+2}}\,dw.$$

See Example 301 on page 265 for more information.

(a) Evaluate

$$\int \frac{w^n}{\sqrt{w^2+2w+2}}\,dw$$

on your calculator for $n=0,1,2$ and 3.

(b) Take an appropriate sum of the integrals from part (a) to evaluate the integral and compare the answer to what one gets from Maple, Mathematica or another CAS on a computer.

8.7 Improper Integrals

In some situations we may want to find the integral of a function when our definition of an integral using Riemann sums, Sect. 6.2, does not work. This can occur when the interval of integration is infinite. Two examples of this are escape velocity, the speed an object must have to escape the gravity well of a planet, and the probability that a light bulb will last longer than 5 years. Another situation that we consider in this section is when a function is unbounded. This happens when a function has an asymptote.

Recall that in the definition of the integral of $f(x)$ from a to b using Riemann sums it is required that the function $f(x)$ is bounded and that a and b are both finite. Improper integrals are used when we wish to weaken these conditions on the integral. Examples of improper integrals are the area between the x axis and $y=\sin(x)$ on the interval $[0,\infty)$, the area under $g(x)=1/(1+x^2)$ from $-\infty$ to ∞, the area under $f(x)=1/x$ on $[-1,1]$, and the area under $h(x)=1/x^{2/3}$ from 0 to $1/2$. See Figs. 8.10 and 8.11 on pages 268 and 269.

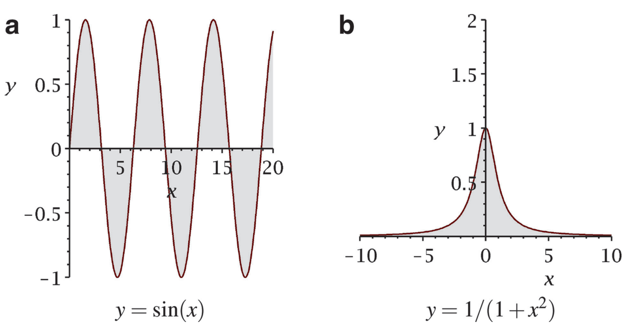


Fig. 8.10

There are some basic problems when trying to use Riemann sums with these functions on the given intervals. The first problem is that we cannot divide an infinite interval into a finite

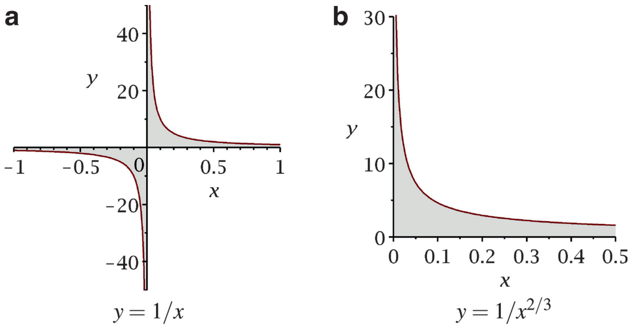


Fig. 8.11

number of intervals with finite length. We cannot even write a Riemann sum for an integral over an unbounded interval.

The second problem is with unbounded function values near an asymptote. In this case, assume an interval $[x_{i-1}, x_i]$ used for a Riemann sum contains an asymptote at a where $\lim_{x\to a} f(x) = \infty$. The term $f(\xi_i)(x_i - x_{i-1})$ in the Riemann sum can be as large as we want by simply taking the value of $f(\xi_i)$ to be very large. See Fig. 8.12 on page 269. For these reasons, we cannot simply use Riemann sums to find the "area" abutting an asymptote. As is demonstrated later in this section, some of the areas under the graphs of functions around asymptotes are finite.

There are several cases that must be considered. In all of the cases the idea is to define the integral as the limit of integrals that can be found through the original definition of the Riemann integral. The first case considered is when the interval of integration is unbounded.

Definition 30. Assume that $a \in \mathbb{R}$ and that for all $b \geq a$

$$I_b = \int_a^b f(x)\,dx$$

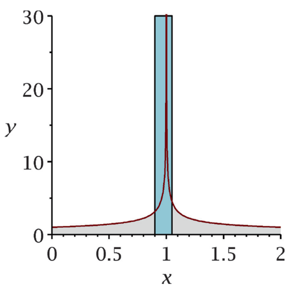


Fig. 8.12

exists. If $\lim_{b\to\infty} I_b = L$, then the integral of $f(x)$ on $[a, \infty)$ exists and equals L. In this case we write

$$\int_a^\infty f(x)\,dx = L.$$

Similarly, assume that $b \in \mathbb{R}$. If $\int_a^b f(x)\,dx$ exists for all $a \le b$ and the limit as $a \to -\infty$ of these integrals exist, we define

$$\int_{-\infty}^{b} f(x)\,dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx.$$

If $\int_a^b f(x)\,dx$ exists for all a and b with $b \ge a$ and for some c in $\mathbb{R}$ we have $\int_{-\infty}^{c} f(x)\,dx = L$ and $\int_c^\infty f(x)\,dx = M$, then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{c} f(x)\,dx + \int_c^\infty f(x)\,dx \\ &= \lim_{a\to-\infty}\int_a^c f(x)\,dx + \lim_{b\to\infty}\int_c^b f(x)\,dx \\ &= L + M.\end{aligned}$$

In any of the above cases, if the limit(s) does not exit, we say the integral does not exist or *does not converge*. The value of the integral can be $\pm\infty$ in any of these cases if we need to add ∞ and $-\infty$.

Remark 25. Since the functions in these notes are only defined on the real numbers, the expressions $F(-\infty)$ and $F(\infty)$ are meaningless. The improper integrals in this section are *only* done through limits.

The first two integrals presented at the beginning of the section can now be evaluated.

Example 303. Consider

$$\int_0^\infty \sin(x)\,dx.$$

Rewriting this as a limit gives

$$\begin{aligned}\int_0^\infty \sin(x)\,dx &= \lim_{b\to\infty}\int_0^b \sin(x)\,dx \\ &= \lim_{b\to\infty} -\cos(x)\Big|_0^b \\ &= \lim_{b\to\infty}\big(1-\cos(b)\big).\end{aligned}$$

Since the limit as b goes to ∞ of $\cos(b)$ does not exist, the integral does not converge.

Example 304. Unlike the first example,

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx$$

exists. Since $f(x) = 1/(1+x^2)$ is a positive even function the limits

$$\lim_{a\to-\infty}\int_a^0 \frac{1}{1+x^2}\,dx \quad \text{and} \quad \lim_{b\to\infty}\int_0^b \frac{1}{1+x^2}\,dx$$

have the same value if one of them exists.

Looking on the interval $[0,\infty)$ we have

$$\begin{aligned}\lim_{b\to\infty}\int_0^{\infty}\frac{1}{1+x^2}\,dx &= \lim_{b\to\infty}\int_0^{b}\frac{1}{1+x^2}\,dx\\ &= \lim_{b\to\infty}\arctan(x)\Big|_{x=0}^{b}\\ &= \lim_{b\to\infty}\big(\arctan(b)-\arctan(0)\big)\\ &= \frac{\pi}{2}.\end{aligned}$$

This means that

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{1+x^2}\,dx &= \int_{-\infty}^{0}\frac{1}{1+x^2}\,dx+\int_0^{\infty}\frac{1}{1+x^2}\,dx\\ &= \frac{\pi}{2}+\frac{\pi}{2}\\ &= \pi.\end{aligned}$$

The next example shows that the above definition removes some ambiguity. Here the problem is that infinite values can be approached in many different ways.

Example 305. When we try to integrate $h(z)=2z/(1+z^2)$, see Fig. 8.13 on page 272, from negative infinity to positive infinity without following the above definition carefully, we can get different answers. If we follow the definition using the substitution $u=1+z^2$ we get

$$\begin{aligned}\int_{-\infty}^{\infty}h(z)\,dz &= \lim_{a\to-\infty}\int_a^{0}\frac{2z}{1+z^2}\,dz+\lim_{b\to\infty}\int_0^{b}\frac{2z}{1+z^2}\,dz\\ &= \lim_{r\to\infty}\int_r^{1}\frac{1}{u}\,du+\lim_{s\to\infty}\int_1^{s}\frac{1}{u}\,du\\ &= \lim_{r\to\infty}\ln|u|\Big|_{u=r}^{1}+\lim_{s\to\infty}\ln|u|\Big|_{u=1}^{s}\\ &= \lim_{r\to\infty}-\ln(r)+\lim_{s\to\infty}\ln(s).\end{aligned}$$

Since the first term goes to negative infinity and the second term goes to positive infinity, this integral does not exist.

On the other hand, if we take the limit of the integral

$$\int_{-R}^{R}\frac{2z}{1+z^2}\,dz,$$

see Fig. 8.14 on page 272, as R goes to infinity we get

$$\begin{aligned}\lim_{R\to\infty}\int_{-R}^{R}\frac{2z}{1+z^2}\,dz &= \lim_{R\to\infty}\ln(1+z^2)\Big|_{z=-R}^{R}\\ &= \lim_{R\to\infty}\Big(\ln\big(1+R^2\big)-\ln\big(1+R^2\big)\Big)\\ &= 0.\end{aligned}$$

This does not match what was obtained using the definition.

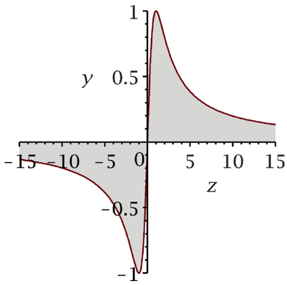


Fig. 8.13 $y = 2z/(1+z^2)$

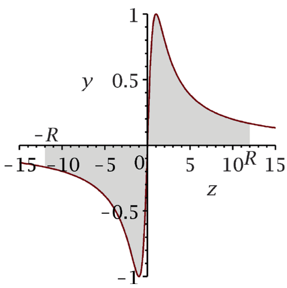


Fig. 8.14 The area under $y = 2z/(1+z^2)$ from $-R$ to R

In fact, by choosing limits appropriately, we can get any value for $\int_a^b 2z/(1+z^2)\,dz$ as both a and b go to infinity. Only the case of positive values is considered. See Fig. 8.15 on page 273. Let $\alpha > 0$ and assume $R > 0$. Since

$$\lim_{b\to\infty} \int_R^b \frac{2z}{1+z^2}\,dz = \infty,$$

there is a $b(R) > R$ such that

$$\int_R^{b(R)} \frac{2z}{1+z^2}\,dz = \alpha.$$

Then, since $h(z)$ is an odd function,

$$\begin{aligned}\lim_{R\to\infty} \int_{-R}^{b(R)} \frac{2z}{1+z^2}\,dz &= \lim_{R\to\infty} \left(\int_{-R}^{R} \frac{2z}{1+z^2}\,dz + \int_{R}^{b(R)} \frac{2z}{1+z^2}\,dz \right) \\ &= \lim_{R\to\infty} \int_{R}^{b(R)} \frac{2z}{1+z^2}\,dz \\ &= \alpha.\end{aligned}$$

Since there is not a unique value that we get out of these different ways of evaluating the limit, it is reasonable to say the integral does not exist.

The second case is when the values of the integrand are unbounded on a bounded interval. For simplicity it is assumed that there is a single point inside of the interval of integration or at an endpoint of the interval of integration where $f(x)$ is not defined.

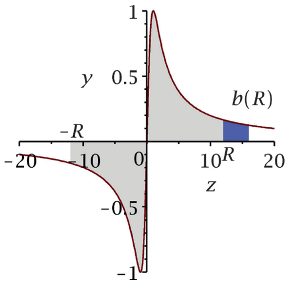


Fig. 8.15 The area under $y = 2z/(1+z^2)$ from $-R$ to $b(R)$

Definition 31. Assume that f is defined on an interval $[a,b]$ except at one point $c \in [a,b]$. In addition assume that for any r and s in $[a,b]$ such that $c \notin [r,s]$ the integral

$$I_{r,s} = \int_r^s f(x)\,dx$$

exists.

If $c = a$ and $\lim_{r \to a^+} \int_r^b f(x)\,dx = L$, then the integral of $f(x)$ on $[a,b]$ exists and equals L. In this case we write

$$\int_a^b f(x)\,dx = L.$$

Similarly, if $c = b$ and $\lim_{r \to b^-} \int_a^r f(x)\,dx = M$, then the integral of $f(x)$ on $[a,b]$ exists and equals L. In this case we write

$$\int_a^b f(x)\,dx = M.$$

If $c \in (a,b)$ and both $\lim_{r \to c^+} \int_r^b f(x)\,dx = L$ and $\lim_{s \to c^-} \int_a^s f(x)\,dx = M$ exist with finite values, then the integral $\int_a^b f(x)\,dx$ exists and

$$\begin{aligned}\int_a^b f(x)\,dx &= \lim_{s \to c^-} \int_a^s f(x)\,dx + \lim_{r \to c^+} \int_r^b f(x)\,dx \\ &= L + M.\end{aligned}$$

In any of the above cases, if a limit(s) does not exit, we say the integral does not exist. The value of the integral can be $\pm\infty$ in any of these cases if we do not need to add ∞ and $-\infty$.

With this definition we can evaluate the third and fourth integrals at the beginning of this section.

Example 306. Let $h(x) = x^{-2/3}$ and consider the integral

$$\int_0^{\frac{1}{2}} x^{-2/3}\,dx.$$

The function $h(x)$ has a asymptote at $x = 0$ and is continuous on $[a, 1/2]$ if $a \in (0, 1/2)$. This means that

$$\int_a^{\frac{1}{2}} x^{-2/3}\,dx$$

exists for all $a \in (0, 1/2)$.

By the definition,

$$\begin{aligned}\int_0^{\frac{1}{2}} x^{-2/3}\,dx &= \lim_{a\to 0^+}\int_a^{\frac{1}{2}} x^{-2/3}\,dx\\ &= \lim_{a\to 0^+} 3x^{1/3}\Big|_a^{\frac{1}{2}}\\ &= \lim_{a\to 0^+}\left(3\left(\frac{1}{2}\right)^{1/3} - 3a^{1/3}\right)\\ &= 3\left(\frac{1}{2}\right)^{1/3}.\end{aligned}$$

The next example is a situation where the integral does not exist.

Example 307. Consider the integral

$$\int_{-1}^{1}\frac{1}{x}\,dx.$$

Since $1/x$ has a single asymptote at $x=0$ on the interval $[-1,1]$, the definition says that we should evaluate the two limits

$$\lim_{a\to 0^-}\int_{-1}^{a}\frac{1}{x}\,dx \qquad \text{and} \qquad \lim_{b\to 0^+}\int_{b}^{1}\frac{1}{x}\,dx$$

and add the results if we get a unique quantity.

The first limit evaluates to

$$\begin{aligned}\lim_{a\to 0^-}\int_{-1}^{a}\frac{1}{x}\,dx &= \lim_{a\to 0^-}\ln|x|\Big|_{-1}^{a}\\ &= \lim_{a\to 0^-}\left(\ln|a| - \ln|-1|\right)\\ &= -\infty.\end{aligned}$$

The second integral yields

$$\begin{aligned}\lim_{b\to 0^b}\int_{b}^{1}\frac{1}{x}\,dx &= \lim_{b\to 0^+}\ln|x|\Big|_{b}^{1}\\ &= \lim_{b\to 0^+}\left(\ln|1| - \ln|b|\right)\\ &= \infty.\end{aligned}$$

The fact that the sum of these to quantities is not well defined means the integral does not exist.

The integrals

$$\int_1^{\infty}\frac{1}{x^{\alpha}}\,dx$$

will be important when discussing series in Chap. 10. Even if you do not work Exercises 4 and 5 in this section, you should know the results stated in these exercises.

Remark 26. Many students forget to look for asymptotes when evaluating integrals over finite intervals. A few seconds checking for asymptotes can save many points on an exam.

Exercises

1. Evaluate the following integrals with infinite limits.

(a) $\int_1^\infty \frac{1}{x^2}\,dx$

(b) $\int_0^\infty \exp(y)\cos(y)\,dy$

(c) $\int_0^\infty \exp(-z)\cos(y)\,dz$

(d) $\int_{-2}^\infty \frac{15}{x+3}\,dt$

(e) $\int_{-2}^\infty t-2\,dt$

(f) $\int_4^\infty \frac{x}{1+x^2}\,dx$

(g) $\int_4^\infty \frac{4}{1+y^2}\,dy$

(h) $\int_{-\infty}^1 \frac{3}{4+z^2}\,dz$

(i) $\int_{-\infty}^{-2} \frac{2}{t^{\left(\frac{3}{2}\right)}}\,dt$

(j) $\int_{-\infty}^0 y\exp(1-y^2)\,dy$

(k) $\int_{-\infty}^{-3} \ln(-z)\,dz$

(l) $\int_1^\infty \frac{4}{x^2+6x+8}\,dx$

(m) $\int_{-\infty}^{-6} \frac{4}{y^2+6y+8}\,dy$

(n) $\int_2^\infty \frac{5z+1}{z^2-1}\,dz$

(o) $\int_{-\infty}^{-1} \frac{3w-1+2w^2}{w^3-3w^2+4w-12}\,dw$

(p) $\int_{-\infty}^\infty \frac{3t}{1+t^4}\,dt$

(q) $\int_{-\infty}^\infty \frac{3}{x^2+6}\,dx$

(r) $\int_{-\infty}^\infty \frac{2}{y^2-4}\,dy$

(s) $\int_{-\infty}^\infty \frac{3}{z\sqrt{z^2-2}}\,dz$

(t) $\int_{-\infty}^\infty \frac{4}{\left(\sqrt{w^2-2}\right)^3}\,dw$

2. Evaluate the following integrals.

(a) $\int_0^1 \frac{1}{\sqrt[3]{x^2}}\,dx$

(b) $\int_{-1}^1 \frac{1}{\sqrt[3]{x^2}}\,dx$

(c) $\int_{-1}^3 \frac{1}{z}\,dz$

(d) $\int_1^2 \frac{2}{x^2-3x+2}\,dx$

(e) $\int_{-1}^1 \sqrt[5]{\frac{1}{w}}\,dw$

(f) $\int_{-3}^3 \sqrt[5]{\frac{1}{w^7}}\,dw$

(g) $\int_2^5 \frac{1}{y^2-11y+10}\,dy$

(h) $\int_{-4}^0 \sqrt[3]{\frac{1}{z^2-2z+1}}\,dz$

(i) $\int_0^4 \sqrt[3]{\frac{1}{z^2-2z+1}}\,dz$

(j) $\int_{-2}^3 \frac{10x}{x^2+x-2}\,dx$

(k) $\int_2^\infty \frac{3x}{x^2+x-2}\,dx$

(l) $\int_{-3}^{-1} \sec^2(w)\,dw$

3. Find a solution to the differential equation

$$\frac{dy}{dx} = \frac{2}{\sqrt[3]{(xy)^2}} \quad \text{and} \quad y(0)=1.$$

Does this solution differ from solutions to differential equations that you have previously found?

4. If $\alpha \le 1$ show that the integral
$$\int_1^\infty \frac{1}{x^\alpha}\,dx$$
diverges to infinity.

5. If $\alpha > 1$ show that the integral
$$\int_1^\infty \frac{1}{x^\alpha}\,dx$$
is finite.

6. If we have a finite number of asymptotes for $f(x)$, $x = c_1, c_2, \ldots, c_k$, in (a,b) we can simply break the integral of $f(x)$ from $x = a$ to $x = b$ into $2k$ integrals. Taking d_i between c_1 and c_{i+1} we define the integral as a sum of improper integrals,
$$\int_a^b f(x)\,dx = \int_a^{c_1} f(x)\,dx + \int_{c_1}^{d_1} f(x)\,dx + \int_{d_1}^{c_2} f(x)\,dx + \cdots + \int_{d_{k-1}}^{c_k} f(x)\,dx + \int_{c_k}^{b} f(x)\,dx\,.$$
This integral exists if the sum makes sense.
Evaluate the following integrals.

(a) $\displaystyle\int_{-2}^{2} \frac{x}{\sqrt[3]{x^2-1}}\,dx$ (c) $\displaystyle\int_{-5}^{5} \frac{x}{\sqrt{|x^2-4|}}\,dx$

(b) $\displaystyle\int_{-3}^{2} \frac{1}{x^2-1}\,dx$ (d) $\displaystyle\int_{0}^{2\pi} \tan(z)\,dz$

8.8 Practice for Integration

When we need to find a closed form integral of a function without a computer algebra system, the obvious question is "What technique should we use?" This is not a simple question to answer. No matter what rules someone gives you, there are exceptions and cases that do not follow any of the patterns you have learned. Given those restrictions, here is a suggested order of techniques to try. It is certainly not the only reasonable order that you can use.

(I) Does the function have a simple antiderivative? This includes such things as $\int e^x\,dx$, $\int 1/(1+z^2)\,dz$, $\int 24z^2+z-10\,dz$, $\int 1/\left(t\sqrt{t^2-1}\right)dt$, and $\int 1/x^\alpha\,dx$. In this case, simply use the antiderivative.

(II) Is there a substitution that turns the integrand into a function with a simple antiderivative, Sect. 6.5? Integrals that fall into this category include such things as $\int \cos(20\theta+1)\,d\theta$, $\int x e^{x^2}\,dx$ and $\int 2/(x^2+4x+8)\,dx$, through completing the square.

(III) Can we do the integral using integration by parts, Sect. 8.2? This can be used for integrals where we have a polynomials times a transcendental, $\int z^n \sin(z)\,dz$, the product of two transcendentals, $\int e^x \sin(x)\,dx$, or a polynomial times $\ln(x)$, $\int \left(4x^3+x\right)\ln(x)\,dx$.

(IV) Is the function a rational function, e.g. $f(x) = (x^4+3x^2-10)(x^3+3x-36)$? In these cases, try using partial fractions, Sect. 8.3.

(V) Does the integrand only involve powers of $\sin(x)$ and $\cos(x)$ or can it be turned into a function that only contains powers of $\sin(x)$ and $\cos(x)$? If so, we have a trigonometric integral, Sect. 8.4.

(VI) Does the integrand involve square roots of one of the forms x^2-1, $1+x^2$ or $1-x^2$? If so, we can try using a trigonometric substitution, Sect. 8.5.

The rest of this section is a collection of integrals where all of the techniques to get closed form integrals that you have seen in this chapter are used. It is suggested that you do a variety of these problems. This is important, if you want to be proficient at integrating functions, since integration is about pattern recognition.

1. $\displaystyle\int \frac{x^2+8x+33}{(x^2+6x+18)(x+3)}\,dx$
2. $\displaystyle\int \sin(10x)\sin(5x)\,dx$
3. $\displaystyle\int \frac{w^2}{\sqrt{9+w^2}}\,dw$
4. $\displaystyle\int \frac{-1}{10-5z}\,dz$
5. $\displaystyle\int \frac{1}{x\sqrt{x^2-16}}\,dx$
6. $\displaystyle\int \frac{y^2}{\sqrt{16-y^2}}\,dy.$
7. $\displaystyle\int \frac{7}{3w^2-w}\,dw$
8. $\displaystyle\int \frac{x^2}{(x^2-25)^{3/2}}\,dx$
9. $\displaystyle\int \sec(2\theta-2)\tan(2\theta-2)\,d\theta$
10. $\displaystyle\int \frac{14x+14}{x^2+2x-24}\,dx$
11. $\displaystyle\int \tan^3(\omega)\,d\omega$
12. $\displaystyle\int \frac{1}{100+64x^2}\,dx$
13. $\displaystyle\int x^2\sqrt{81-16x^2}\,dx$
14. $\displaystyle\int s^2\cos(4s)\,ds$
15. $\displaystyle\int \frac{1}{x^2-12x+37}\,dx.$
16. $\displaystyle\int \sec^2(4\theta)\,d\theta$
17. $\displaystyle\int \frac{\cos^3(2x)}{\sin^{16}(2x)}\,dx$
18. $\displaystyle\int \frac{w^2+6w+56}{(w^2+16)(w-4)}\,dw$
19. $\displaystyle\int \frac{\cos^2(x)}{\sin(x)}\,dx$
20. $\displaystyle\int \frac{x^2-9x+11}{(x-4)(x-1)^2}\,dx$
21. $\displaystyle\int \frac{1}{y^3\sqrt{y^2-1}}\,dy$
22. $\displaystyle\int \frac{3z-1}{z^2+4}\,dz$
23. $\displaystyle\int \sin(6x)\sin(7x)\,dx$
24. $\displaystyle\int \sqrt{100-9t^2}\,dt$
25. $\displaystyle\int \cos^3(x)\,dx$
26. $\displaystyle\int e^{2x}\sin(x)\,dx$
27. $\displaystyle\int \frac{1+4y}{y^2+9}\,dy$
28. $\displaystyle\int \cos^2(2x)\sin^3(2x)\,dx$
29. $\displaystyle\int e^x\cos(5x)\,dx$
30. $\displaystyle\int (-3z-7)^n\,dz$ if $n \neq -1$
31. $\displaystyle\int \frac{-10x^2+7x+8}{x^3-x^2-2x}\,dx$
32. $\displaystyle\int \frac{7w+9}{(w+3)(w-1)}\,dw$

33. $\int e^t \sec^2\left(e^t\right)\,dt$

34. $\int \sec^2(x)\tan^5(x)\,dx$

35. $\int x\ln(z)\,dz$

36. $\int \frac{1}{x^3\sqrt{x^2-1}}\,dx$

37. $\int \frac{2x-3}{x^2+9}\,dx$

38. $\int \frac{z^2+1}{(z+1)^3}\,dz$

39. $\int 2\sqrt{9-4x^2}\,dx$

40. $\int e^w e^{-e^w+7}\,dw$

41. $\int \frac{-4}{\left(x^2+1\right)(x+1)}\,dx$

42. $\int \cos(y)\cos(\sin(y))\,dy$

43. $\int \frac{1+x}{x^2-6x+13}\,dx$

44. $\int_{-9}^{-2} \frac{1}{\sqrt[3]{x+4}}\,dx$

45. $\int \frac{1}{9+4s^2}\,ds$

46. $\int \csc^2(x)\,dx$

47. $\int_{-7}^{-1} x^7\cos\left(2x^8\right)\,dx$

48. $\int \sin(w)\,e^{5\cos(w)+1}\,dw$

49. $\int \frac{5x+14}{x^2+2x-8}\,dx$

50. $\int \cos(\theta)\,d\theta$.

51. $\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \sec^2(\phi)\,d\phi$

52. $\int \cos^3(x)\,dx$

53. $\int_{-3}^{-1} x^5 e^{-4x^6}\,dx$

54. $\int_{\frac{\pi}{4}}^{\frac{5\pi}{6}} \csc^2(\theta)\,d\theta$

55. $\int_{-2}^{4} (x-3)^{\frac{-9}{5}}\,dx$

56. $\int \frac{1}{3z-2}\,dz$

57. $\int \sin(x)\sec^2(4\cos(x)-2)\,dx$

58. $\int \sec(z)\tan(z)\,dz$

59. $\int \sec^2(3x)\,dx$

60. $\int (az)^n\,dz$ if $n\neq -1$

61. $\int_0^1 \frac{1}{\sqrt[4]{y}}\,dy.$

62. $\int_0^1 \frac{1}{x}\,dx$

63. $\int_{\frac{2\pi}{3}}^{\pi} \cos(2x)\,dx$

64. $\int_5^9 z\,dz$

65. $\int (14x-5)\sin^3\left(7x^2-5x\right)\,dx$

66. $\int w^2 e^{8w}\,dw.$

67. $\int \frac{x-1}{(x-2)^2}\,dx$

68. $\int \frac{6z^2+2z+32}{z^3+2z^2+9z+18}\,dz$

69. $\int \frac{-3x^2+3x-2}{x^3-2x^2+x}\,dx$

70. $\int \sin^3(w)\,dw$

71. $\int x^2\cos(x)\,dx.$

72. $\int \frac{3x^2-2x-1}{(x+1)x^2}\,dx$

73. $\int -4s^4\sin\left(5s^5\right)\,ds$

74. $\int_{-4}^{0} x^{\frac{-11}{6}}\,dx$

75. $\int \csc(z)\cot(z)\,dz$

76. $\int \csc^2(x-1)\,dx.$

77. $\displaystyle\int \frac{1}{49+x^2}\,dx$

78. $\displaystyle\int -\cos(6w)\sin(7w)\,dw$

79. $\displaystyle\int x\sin(4x)\,dx$

80. $\displaystyle\int_{\frac{2\pi}{3}}^{\frac{7\pi}{6}} \cos(z)\,dz$

81. $\displaystyle\int \cos(\theta)\cos(4\sin(\theta)-6)\,d\theta.$

82. $\displaystyle\int \cos(6x)\,\cos(3x)\,dx$

83. $\displaystyle\int_4^{\infty} \frac{1}{x^2}\,dx.$

84. $\displaystyle\int_{\ln(2)}^{\ln(10)} \exp(-3x-8)\,dx$

85. $\displaystyle\int \frac{1}{x^3\sqrt{x^2-64}}\,dx$

86. $\displaystyle\int x\sin(x)\,dx$

87. $\displaystyle\int z^{15}\,dz$

88. $\displaystyle\int \sin(6w)\,dw.$

89. $\displaystyle\int_0^{\frac{\pi}{16}} \sec^2(4\theta)\,d\theta$

90. $\displaystyle\int x^2\sqrt{81-64x^2}\,dx$

91. $\displaystyle\int \frac{10x-15}{x^2-3x-54}\,dx$

92. $\displaystyle\int \cot^3(w)\,dw$

93. $\displaystyle\int e^x\sec^2\left(3e^x+1\right)\,dx.$

94. $\displaystyle\int \cos(6x)\,dx$

95. $\displaystyle\int \sec^2(2x)\tan^2(2x)\,dx$

96. $\displaystyle\int_3^4 \frac{1}{u^4}\,du$

97. $\displaystyle\int (14z-1)\cos^3\left(7z^2-z\right)\,dz$

98. $\displaystyle\int_0^1 \arccos^2(y)\,dy$

8.9 Numerical Approximations of Integrals

There are many cases when we cannot get an exact value for an integral. The reasons for this include the facts than many functions do not have simple antiderivatives and we may only have data, function values, at a limited number of points. Because of this, we want to be able to get accurate approximations for definite integrals without using the Fundamental Theorem of Calculus, Theorem 65. It is important to realize that the techniques in this section are among the simplest techniques. There are many techniques that are more sophisticated and more accurate. We should look, if possible, to those techniques when we are actually doing numerical integration.

In Sect. 6.2 the idea of approximating the area under a curve with a sum representing areas was introduced. The three types of Riemann sums commonly used for approximations are the left and right endpoint rules and the midpoint rule. These rules are easy to understand. The left and right endpoint rules have problems with accuracy. The results do not improve quickly as the number of intervals increases. The midpoint rule is somewhat more accurate but has the problem that we usually cannot use previously calculated function values when increasing the number of intervals to improve accuracy.

Example 308. The left, right and midpoint sums for $f(x)=\cosh(x)$ on $[2,8]$ with 10 equal subintervals are

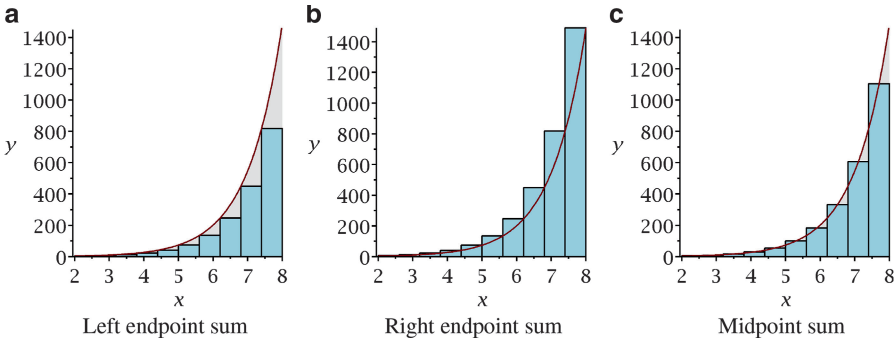


Fig. 8.16

$$\begin{aligned} L_{10} &= \frac{6}{10}\sum_{n=0}^{9} f(2+0.6n) \\ &\approx 1{,}085.177 \\ R_{10} &= \frac{6}{10}\sum_{n=1}^{10} f(2+0.6n) \\ &\approx 1{,}977.207 \\ M_{10} &= \frac{6}{10}\sum_{n=0}^{9} f(2.3+0.6n) \\ &\approx 1{,}464.781. \end{aligned}$$

Figure 8.16 on page 280 shows the sums with the right endpoints in Fig. 8.16a, the left endpoints in Fig. 8.16b and the midpoints in Fig. 8.16c. The exact value is

$$\int_2^8 \cosh(x)\,dx = \sinh(8) - \sinh(2) \approx 1{,}486.852.$$

The theory of the errors for these approximations predicts that the midpoint rule should give a better estimate than either endpoint estimate. This is exactly what happens here.

The expressions used to give upper estimates on the errors for numerical integration techniques are usually not easy to derive. The derivation of the error formula for the left endpoint rule does, however, give the flavor of these derivations and is fairly simple. Assume that the integral of a function $f(x)$ with a continuous first derivative over an interval from a to $a+h$ is to be approximated with $\int_a^{a+h} f(x)\,dx = f(a)\,h$. The absolute value of the error, using the Mean Value Theorem (Sect. 5.5), is

$$\begin{aligned} \left|\left(\int_0^h f(a+x)\,dx\right) - f(a)\,h\right| &= \left|\int_0^h \big(f(a) + f'(\mu(x))x - f(a)\big)\,dx\right| \\ &= \left|\int_0^h f'(\mu(x))\,x\,dx\right| \\ &\le \left|M\int_0^h x\,dx\right| \\ &\le \frac{Mh^2}{2}. \end{aligned}$$

Here M is any number greater than $|f'(x)|$ on $[a, a+h]$.

If the distance from a to b is divided into n equal intervals of length $h = (b-a)/n$ and M is any number greater than $|f'(x)|$ on $[a,b]$, the expression for the error becomes

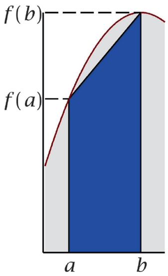


Fig. 8.17 A single trapezoid area approximation

$$\left|\left(\int_a^b f(x)\,dx\right) - h\sum_{i=0}^{n-1} f(a+ih)\right| \le n\,\frac{Mh^2}{2} = \frac{M\,(b-a)\,h}{2}.$$

In a similar manner we can show that the formula of the error estimate for the right endpoint rule is the same as for the left endpoint rule. The most important feature of these estimates is that if you divide the length of the intervals by 2, double the number of equal intervals, the error estimate is halved.

Using the first degree Taylor polynomial as an estimate for the midpoint rule we can show that the error E_M for the midpoint rule using equal intervals of length h satisfies

$$|E_M| \le \frac{M_2 h^2\,(b-a)}{24}.$$

Here M_2 is any upper bound on the absolute value of the second derivative of $f(x)$ on $[a,b]$. In this case if we double the number of intervals, the error estimate is divided by 4.

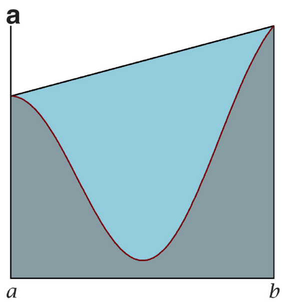
The two numerical integration techniques introduced in this section are the trapezoid method and Simpson's method. If we consider a Riemann sum as integrating a piecewise constant function that approximates the integrand $f(x)$, the trapezoid method can be considered as integrating a piecewise linear approximation to $f(x)$ and Simpson's method can be view as integrating a piecewise quadratic approximation for $f(x)$. We integrate these more accurate polynomial approximations for $f(x)$ to get more accurate approximations for $\int_a^b f(x)\,dx$.

The trapezoid method uses a secant line approximation for $f(x)$. For an interval from a to b we use the linear approximation $f(a+h) \approx g(h) = f(a) + h\,(f(b) - f(a))(b-a)$. We have $g(0) = f(a)$ and $g(b-a) = f(b)$. See Fig. 8.17 on page 281.
The area under the linear function is a trapezoid with area $\frac{f(a)+f(b)}{2}\,(b-a)$. You can find the area either by the geometry of a trapezoid or by integration.

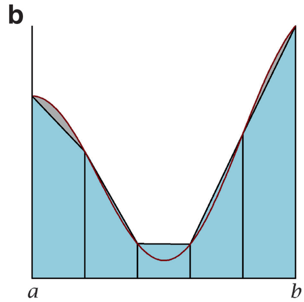
As with Riemann sums, we usually get a better approximation by using more intervals, each with a trapezoid approximating the area under the curve on the subinterval. See Fig. 8.18 on page 282 where there is a comparison of one trapezoid and five trapezoids for the same function.

Using the formula for the area of a trapezoid and dividing the interval $[a,b]$ with a partition of n intervals and corresponding points $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ we have

$$\begin{aligned}\int_a^b f(x)\,dx &\approx \frac{f(x_0)+f(x_1)}{2}\,(x_1 - x_0) + \frac{f(x_1)+f(x_2)}{2}\,(x_2 - x_1)\\ &\quad + \cdots + \frac{f(x_{n-1})+f(x_n)}{2}\,(x_n - x_{n-1})\\ &\approx \frac{1}{2}\Big(f(x_0)\,h_1 + f(x_1)\,(h_1+h_2) + f(x_2)\,(h_2+h_3)\end{aligned}$$



Area approximation with one trapezoid



Area approximation with five trapezoids

Fig. 8.18

$$+\cdots+f(x_{n-1})\,(h_{n-1}+h_n)+f(x_n)\,h_n\Big).$$

To simplify the expression, h_k is used to denote $x_k - x_{k-1}$. If the subintervals all have the same length, the partition points are equally spaced, the trapezoid rule becomes

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\Big(f(a)+2f(a+h)+2f(a+2h)+\cdots+2f(b-h)+f(b)\Big).$$

or

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\left(f(a)+f(b)+2\sum_{i=1}^{n-1} f(a+ih)\right).$$

The error for the trapezoid method with subintervals of equal length is

$$|E_T| \le \frac{M_2\,h^2\,(b-a)}{12}.$$

Here M_2 is a number at least as large as the maximum of $|f''(x)|$ on $[a,b]$.

Example 309. Let $f(x) = x\sin\left(x^2\right)$ and consider

$$\int_0^4 x\sin\left(x^2\right)\,dx.$$

See Fig. 8.19 on page 282.

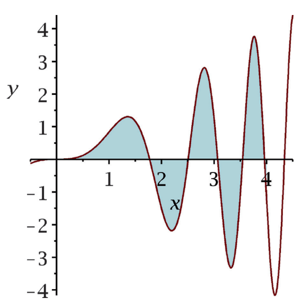


Fig. 8.19 The function $f(x) = x\ \sin(x)$

If we divide the interval into n equal subintervals, then the trapezoid method has the formula

$$\int_0^4 x\sin\left(x^2\right)\,dx \approx \frac{4-0}{2n}\left(f(0)+f(4)+2\sum_{k=1}^{n-1} f\left(k\frac{4}{n}\right)\right).$$

Taking $n = 2,4,8,16,32$, and 64 we get the following approximations with errors in the Table 8.2 on page 283. In the table T_n is the trapezoid approximation and $E_n = Value - T_n$

n	T_n	E_n
2	–4.1788	5.1576
4	–0.0115	.9903
8	0.05369	.92506
16	.80547	.17344
32	.93790	0.04086
64	.96850	0.01012

Table 8.2

is the error.

Using the fact that $h = b - a/n$, the theoretical error term $M_2\,h^2\,(b-a)/12$ can be rewritten in terms of the number of equal subintervals n as

$$\frac{M_2\,(b-a)^3}{12n^2}\ .$$

If this is true, doubling the number of intervals should approximately quarter the error of the approximation. The last four estimates in the Table 8.2 on page 283 follow this pattern.

A nice feature of the trapezoid method is that when we have unequally spaced points, the formula is the same as the original formula,

$$\begin{aligned}\int_a^b f(x)\,dx &\approx \frac{f(x_0)+f(x_1)}{2}\,(x_1-x_0)+\frac{f(x_1)+f(x_2)}{2}\,(x_1-x_1)\\ &\quad+\cdots+\frac{f(x_{n-1})+f(x_n)}{2}\,(x_n-x_{n-1})\\ &\approx \frac{1}{2}\Big(f(x_0)\,h_1+f(x_1)\,(x_2-x_0)+f(x_2)\,(x_3-x_1)\\ &\quad+\cdots+f(x_{n-1})\,(x_n-x_{n-2})+f(x_n)\,(x_n-x_{n-1})\Big).\end{aligned}$$

Example 310. Assume that we want to integrate a function from $x = 0$ to $x = 3$ when we only have the information in Table 8.3.

x	0	0.2	0.5	0.7	0.9	1.3	1.4	1.7	1.9	2
$f(x)$	1.	–.392	–.942	–.065	.562	–.022	–.102	.098	–.078	–.303

Table 8.3

Applying the formula for the trapezoid rule gives

$$\int_a^b f(x)\,dx \approx \frac{1}{2}\big(1\,(0.2) - 0.392\,(0.5) - 0.924\,(0.5) + \cdots - 0.078\,(0.3) - 0.303\,(.1)\big)$$
$$\approx 0.158.$$

For Simpson's method we use a quadratic to approximate the function of interest using three points $(x_{-1}, f(x_{-1}))$, $(x_0, f(x_0))$, and $(x_1, f(x_1))$. This can be done in a number of ways, but the calculations are omitted here in favor of illustrative graphs. In Fig. 8.20 on page 284 are the graph of a function and then the graph with the quadratic function that matches the function at $x = 1$, $x = 1.5$, and $x = 2$.

If we have three points $(x-h, f_{-1})$, (x, f_0), and $(x+h, f_1)$, there is a unique quadratic $q(y)$ through the three points. When we integrate this quadratic from $x-h$ to $x+h$, we get

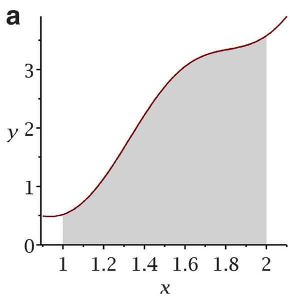
$$\int_{x-h}^{x+h} q(y)\,dy = \frac{h}{3}\,(f_{-1} + 4f_0 + f_1)\,. \tag{8.10}$$

Since each quadratic covers two intervals, it is standard to assume that the interval $[a,b]$ is divided into an even number of subintervals. Dividing $[a,b]$ into $2n$ equal intervals with partition points $a = x_0 < x_1 < \cdots < x_{2n-1} < x_{2n} = b$ we can use formula (8.10) on pairs of intervals to get

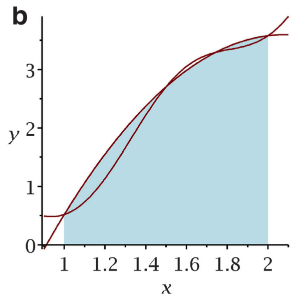
$$\begin{aligned}
\int_a^b f(x)\,dx &= \frac{h}{3}\Big(f(x_0) + 4f(x_1) + f(x_2)\Big) + \frac{h}{3}\Big(f(x_2) + 4f(x_3) + f(x_4)\Big) \\
&\qquad + \cdots + \frac{h}{3}\Big(f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})\Big) \\
&= \frac{h}{3}\Big(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) \\
&\qquad + \cdots + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})\Big).
\end{aligned}$$

Assuming that the integrand $f(x)$ has four continuous derivatives on $[a,b]$, that M_4 is an upper bound for $|f^{(4)}(x)|$ on $[a,b]$, and that $[a,b]$ is divided into $2n$ equal intervals of length h, a bound on the error E_S for Simpson's method is

$$|E_S| \le \frac{M_4\,(b-a)^5}{180\,(2n)^4} = \frac{M_4(b-a)}{180}\,h^4.$$

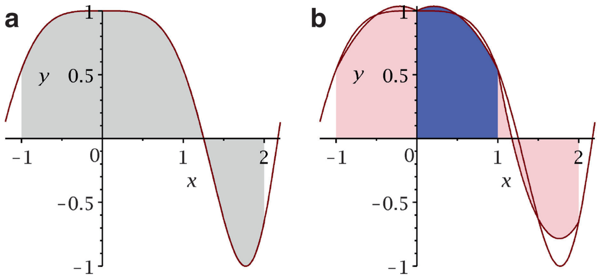


The area under a function



The area under a quadratic approximation to the function

Fig. 8.20



a The area under $f(x) = \cos\left(x^2\right)$

b The Simpson's approximation to the area under $f(x) = \cos\left(x^2\right)$

Fig. 8.21

Example 311. Consider the integral of $f(x) = \cos\left(x^2\right)$ from -1 to 2. This integral is to be approximated using Simpson's method using six equal intervals. Figure 8.21 on page 285 shows the function and the three quadratic approximations with areas.

The approximation to the integral is

$$\int_{-1}^{2} \cos\left(x^2\right)\,dx \approx \frac{1}{6}\left(\cos(-1) + 4\cos\left(-\frac{1}{4}\right) + 2\cos(0) + 4\cos\left(\frac{1}{4}\right)\right.$$
$$\left. + 2\cos(1) + 4\cos\left(\frac{9}{4}\right) + \cos(4)\right)$$
$$\approx 1.3676447\,.$$

To estimate the error we need a bound on $|f^{(4)}|$. Since the fourth derivative of $\cos\left(x^2\right)$ is $16x^4\cos(x^2) + 48x^2\sin(x^2) - 12\cos(x^2)$, we can find bounds on the absolute value of each term and add those bounds to get a bound for the fourth derivative. The absolute value of the first term is less than $16 \cdot 2^4 \cdot 1 = 256$ since $x^4 < 2^4$ and $\left|\cos\left(x^2\right)\right| < 1$ on the interval $[-1,2]$. Similarly, $|48x^2\sin(x^2)| < 192$ and $|12\cos(x^2)| < 12$ on $[-1,2]$. This means that $M_4 = 256 + 192 + 12 = 460$ is an upper bound for $|f^{(4)}(x)|$ on $[-1,2]$ and the error must satisfy

$$|E_S| < \frac{460\left(\frac{1}{2}\right)^4}{180} = \frac{23}{144} = 0.15972.$$

An approximation to the integral exact to 10 digits is 1.365985700. The absolute value of the difference between the Simpson's approximation and this value is less than 0.00166. As is often the case, the actual error is much smaller than the calculated error bound. In this case the smallest M_4 value is around 305. A better M_4 value would not greatly improve the error bound from the formula.

Example 312. When we calculate the errors for Simpson's method approximation to

$$\int_{-2}^{3} \frac{\cosh(x)}{1+x^2}\,dx$$

with $2, 4, 8, 16, 32, 64, 128, 256, 512,$ and $1{,}024$ intervals we get Table 8.4 on page 286. Figure 8.22a on page 286 is a plot of the function.

The last column of the table is the ratio of the nth error to the $n-1$st error. The ratios are approaching $0.0625 = 1/16$. The formula for the error bound on Simpson's method is of the form

$$|E_n| < Kh^4.$$

The h's used here are shrinking by a factor of 2. Plugging this into the error estimate we have

n	S_n	E_n	$ratio_n$
2.	5.4449670452599	$-0.50141\ 10^{-1}$	
4.	5.4468108129080	$-0.51984\ 10^{-1}$	1.0368
8.	5.4037071669079	$-0.88807\ 10^{-2}$	.17083
16.	5.3950003230246	$-0.17386\ 10^{-3}$	$0.19578\ 10^{-1}$
32.	5.3948292382800	$-0.27782\ 10^{-5}$	$0.15979\ 10^{-1}$
64.	5.3948266320541	$-1.7193\ 10^{-7}$	$0.61885\ 10^{-1}$
128.	5.3948264708767	$-1.0748\ 10^{-8}$	$0.62516\ 10^{-1}$
256.	5.3948264608005	$-6.7179\ 10^{-10}$	$0.62504\ 10^{-1}$
512.	5.3948264601707	$-4.1988\ 10^{-11}$	$0.62501\ 10^{-1}$
1024.	5.3948264601313	$-2.6242\ 10^{-12}$	$0.62500\ 10^{-1}$

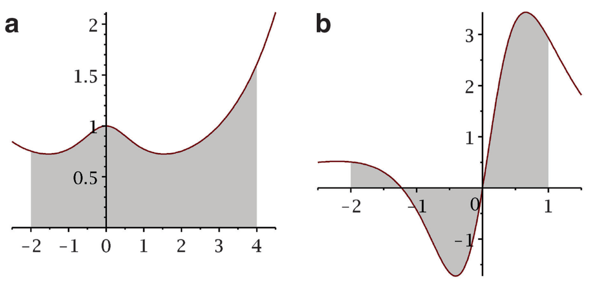
Table 8.4

$$\left|\frac{E_n}{E_{n-1}}\right| \approx \frac{\left(\frac{h}{2}\right)^4}{h^4} = \frac{1}{16}.$$

The calculated values match the theoretical rate of decrease in the error.

In order to see the difference between the trapezoid method and Simpson's method we should calculate the values of both for a single function with the same number of intervals. This is equivalent to using the same number of function evaluations.

Example 313. Let $f(x) = \left(5w^2 + 8\sin(w)\right)/\left(w^4 + 2w^2 + 1\right)$ and approximate $\int_{-2}^{1} f(x)\,dx$. The value of this integral is approximately 1.646323100, see Fig. 8.22b on page 286. The table in Table 8.5 on page 286 contains the approximations and errors for the trapezoid rule and Simpson's method using 10, 20, 30, and 40 intervals.



The graph of $f(x) = \cosh(x)/(1+x^2)$

The graph of $f(x) = \left(5w^2 + 8\sin(w)\right)/\left(w^4 + 2w^2 + 1\right)$

Fig. 8.22

The errors follow the expected pattern in that the errors for Simpson's method are much smaller than those for the trapezoid rule.

n	T_n	$E_{T,n}$	S_n	$E_{S,n}$
10	1.6298407311	$0.1648236\,10^{-1}$	1.6489280774	$-0.2604977\,10^{-2}$
20	1.6422460174	$0.4077082\,10^{-2}$	1.6463811128	$-0.5801259\,10^{-4}$
30	1.6445147110	$0.1808389\,10^{-2}$	1.6463347360	$-0.1163585\,10^{-4}$
40	1.6453066014	$0.1016498\,10^{-2}$	1.6463267961	$-0.3695938\,10^{-5}$

Table 8.5

Exercises

1. Use the trapezoid method to approximate the following integrals using n equal intervals.

(a) $\displaystyle\int_0^2 \cos^3(x)\,dx,\ n=5$

(b) $\displaystyle\int_{-1}^0 \sin^3(y)\cos^2(y)\,dy,\ n=7$

(c) $\displaystyle\int_2^4 exp\left(-z^2\right)dz,\ n=4$

(d) $\displaystyle\int_{-2}^2 e^t - t^2\,dt,\ n=6$

(e) $\displaystyle\int_{-4}^{-1} \operatorname{arccot}^4\left(\sqrt{-t}\right)dt,\ n=8$

(f) $\displaystyle\int_0^4 \cos\left(\frac{x^2}{3}\right)dx,\ n=20$

(g) $\displaystyle\int_{-2}^2 \frac{\sin(y)}{1+\cos^2(y)}\,dy,\ n=21$

(h) $\displaystyle\int_{-2}^3 z^2\sec\left(\frac{z}{4}\right)dz,\ n=16$

(i) $\displaystyle\int_4^{10} \cos^2\left(3\sqrt{t^2-1}\right)dt,\ n=13$

2. In each of the following a table of data is given for a function over a given interval. Use the trapezoid method to approximate the following integral of the function over $[a,b]$.

(a) $[a,b]=[0,1]$, using Table 8.6.

x	0	1/2	3/4	1
$f(x)$	0.5500	1.209	0.5036	−0.2068

Table 8.6 The values for Exercise 2(a)

(b) $[a,b]=[-1,2]$, using Table 8.7.

x	−1.	−0.4545	−0.1818	0.6364	1.455	2.
$f(x)$	1.756	1.669	1.564	1.838	1.802	1.093

Table 8.7 The values for Exercise 2(b)

(c) $[a,b]=[0.5,4.1]$, using Table 8.8.

x	0.5000	1.100	1.700	2.300	2.600	2.900	3.200	4.100
$f(x)$	0.4188	0.8199	1.244	1.244	1.041	0.8199	0.6161	−0.1223

Table 8.8 The values for Exercise 2(c)

(d) $[a,b]=[-1.1,3.3]$, using Table 8.9.

x	−1.10	−0.4714	−0.05238	0.3667	0.5762	1.205	1.624	1.833	2.462	3.09	3.30
$f(x)$	2.062	0.876	1.654	3.107	3.042	2.317	2.90	3.468	4.210	3.06	2.668

Table 8.9 The values for Exercise 2(d)

3. Use Simpson's method to approximate the following integrals using n equal intervals.

(a) $\displaystyle\int_0^2 \cos^3(x)\,dx,\ n=4$

(b) $\displaystyle\int_{-1}^0 \sin^3(y)\cos^2(y)\,dy,\ n=6$

(c) $\displaystyle\int_2^4 exp\left(-z^2\right)dz,\ n=8$

(d) $\displaystyle\int_{-2}^2 e^t - t^2\,dt,\ n=6$

(e) $\displaystyle\int_{-4}^{-1} \operatorname{arccot}^4\left(\sqrt{-t}\right)dt,\ n=10$

(f) $\displaystyle\int_0^4 \cos\left(\frac{x^2}{3}\right)dx,\ n=20$

(g) $\displaystyle\int_{-2}^2 \frac{\sin(y)}{1+\cos^2(y)}\,dy,\ n=24$

(h) $\displaystyle\int_{-2}^3 z^2\sec\left(\frac{z}{4}\right)dz,\ n=16$

(i) $\displaystyle\int_4^{10} \cos^2\left(3\sqrt{t^2-1}\right)dt,\ n=28$

4. Use the trapezoid method to approximate each of the following integrals with $4,8,16,32$, and 64 equal intervals. Do the errors follow the expected pattern?

(a) $\displaystyle\int_{-1}^2 \left(x^2+e^{-x}\right)dx$

(b) $\displaystyle\int_0^\pi \sin^2(z)\cos^2(z)\,dz$

(c) $\displaystyle\int_{-2}^0 \frac{x}{1+x^2}\,dx$

(d) $\displaystyle\int_1^5 \frac{1}{x^4}\,dx$

5. Use Simpson's method to approximate each of the following integrals with $4,8,16,32$, and 64 equal intervals. Do the errors follow the expected pattern?

(a) $\displaystyle\int_{-1}^2 \left(x^2+e^{-x}\right)dx$

(b) $\displaystyle\int_0^\pi \sin^2(z)\cos^2(z)\,dz$

(c) $\displaystyle\int_{-2}^0 \frac{x}{1+x^2}\,dx$

(d) $\displaystyle\int_1^5 \frac{1}{x^4}\,dx$

6. In this problem one derives the approximation used in Simpson's method. The function

$$q(x) = a\,\frac{x\,(x-h)}{2h^2} - b\,\frac{(x+h)\,(x-h)}{h^2} + c\,\frac{x\,(x+h)}{2h^2}$$

is a quadratic function. It is the Lagrange form of the quadratic function whose graph goes through $(-h,a)$, $(0,b)$ and (h,c).

(a) Show that the quadratic function $q(x)$goes through the three points $(-h,a)$, $(0,b)$ and (h,c).

(b) Find the integral of $q(x)$ from $x=-h$ to $x=h$.

7. Use the error term for Simpson's method to show that it gives exact answers for any cubic function over any interval when using equally spaced points.

Chapter 9
Applications of Integration

9.1 Work

Work can be defined as the change of energy in a system. In an elementary physics class it is often calculated as force times distance for the movement of a mass along a straight line under the application of a constant force. For example, if a 2 kg mass is moved up 2 m from the surface of the earth we can approximate the force on the mass as $F = 9.8 \cdot 2 = 19.6\,\mathrm{N}$. Since the mass is moved 2 m, the energy involved in moving the mass is $E = 19.6 \cdot 2 = 39.2\,\mathrm{J}$. This section begins with a review of the ideas in Sect. 1.3 on page 19.

There are two points of view on work: it is the amount of energy put into a system or it is the amount of energy taken out of a system. Each of these points of view is used in some fields of study. In these notes work will be the amount of energy taken out of the system. In the case of the mass being moved up 2 m, the work will be negative since the increase of the potential energy of the mass is 39.2 J. The work is $-39.2\,\mathrm{J}$. (See Fig. 9.1 on page 290.)

If we consider the mass moving up from the surface of the earth in terms of vectors, the change in position can be taken as $\mathbf{r} = (0,2)\,\mathrm{m}$ and the force can be written as $\mathbf{F} = 2\,(0,-9.8)\,\mathrm{N}$. Since the displacement vector $\mathbf{r}$ and the force vector $\mathbf{F}$ are in opposite directions, their dot product is $-39.2\,\mathrm{J}$, the work. In fact, if we move a mass along a straight line from $\mathbf{a}$ to $\mathbf{b}$ with a constant force $\mathbf{F}$, the work is $(\mathbf{b}-\mathbf{a}) \cdot \mathbf{F}$. (See Sect. 1.3 and Fig. 9.2 on page 290.)

To understand this, we can consider a car constrained to a straight track that goes in a direction $\mathbf{r}$ with $\|\mathbf{r}\| = 1$. If a force $\mathbf{F}$ is applied to the car and there is no friction, the force that moves the car down the track is the projection of $\mathbf{F}$ onto $\mathbf{r}$, $\operatorname{proj}_{\mathbf{r}} \mathbf{F}$, see Eq. 1.15. This means that the work moving the car a displacement of $\mathbf{w} = s\mathbf{r}$ with a constant force $\mathbf{F}$ is given by

$$\begin{aligned} W &= s \operatorname{comp}_{\mathbf{r}} \mathbf{F} \\ &= s\,\frac{\mathbf{F} \cdot \mathbf{r}}{\|\mathbf{r}\|} \\ &= \mathbf{F} \cdot (s\mathbf{r}) \\ &= \mathbf{F} \cdot \mathbf{w}. \end{aligned}$$

(See Fig. 9.3 on page 290.)

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-09438-0_9) contains supplementary material, which is available to authorized users.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_9

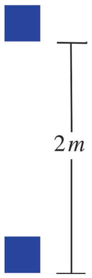


Fig. 9.1 A mass raised 2 m

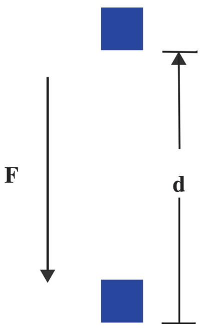


Fig. 9.2 The force **F** and displacement **d** of the mass

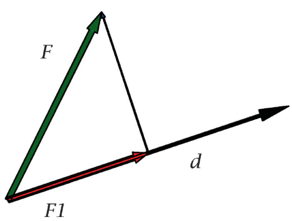


Fig. 9.3 Projection of a force onto the direction of travel

Example 314. A mass is moved along the line segment from $(1,0,-3)$ to $(2,2,5)$ with a constant force of $(-3,2,-1)\,\mathrm{N}$ acting on the mass. The work done in moving the mass is

$$\begin{aligned} W &= \big((2,2,5)-(1,0,-3)\big)\cdot(-3,2,-1) \\ &= (1,2,8)\cdot(-3,2,-1) \\ &= -7\ \mathrm{J}. \end{aligned}$$

Example 315. Assume that a 5 kg mass is moved 15 m west and 2 m up. If we assume that the acceleration due to gravity is constant and the surface of the earth is flat for 15 m, the work done is

$$\begin{aligned} W &= 5\,(0,-9.8)\cdot(-15,2) \\ &= -98.0\ \mathrm{J}. \end{aligned}$$

We can consider the generalizations of this definition of work to the cases of linear motion in a non-constant force field and to the case of nonlinear motion in a non-constant force field. In the case of spacecraft and planetary systems, the forces are not constant and the motion of

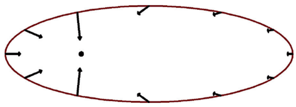


Fig. 9.4 Force of gravity from a sun on a planet always points toward the sun

objects is clearly nonlinear. This is easy to see since the force from a star acting on a planet orbiting the star always points toward the star. (See Fig. 9.4 on page 291.)

Keeping things simple, the case of linear motion in a non-constant force field is considered first. Assume that a mass is moving along a straight line segment from **a** to **b** parametrized by $\ell(s) = \mathbf{a} + s\mathbf{w}$ with $\mathbf{w} = \mathbf{b} - \mathbf{a}$ and $s \in [0,1]$. Also assume that the force field is continuous along the line segment, $\mathbf{F}(\ell(s))$ is continuous. This implies that the dot product of $\mathbf{F}$ with $\mathbf{w}$ is continuous on the given line segment.

If $0 = s_0 < s_1 < \cdots < s_n = 1$, we can approximate the work from $\ell(s_{i-1})$ to $\ell(s_i)$ by

$$\begin{aligned} W_i &\approx \mathbf{F}(\ell(\xi_i)) \cdot (\ell(s_i) - \ell(s_{i-1})) \\ &\approx \mathbf{F}(\ell(\xi_i)) \cdot \big((s_i - s_{i-1})\,\mathbf{w}\big) \\ &\approx (s_i - s_{i-1})\,\mathbf{F}(\ell(\xi_i)) \cdot \mathbf{w}. \end{aligned}$$

Here ξ_i is any number between s_{i-1} and s_i. Adding these approximate values for work over each of the intervals $[s_{i-1}, s_i]$ gives

$$\begin{aligned} W &= \sum_{i=1}^{n} W_i \\ &\approx \sum_{i=1}^{n} \mathbf{F}(\ell(\xi_i)) \cdot \mathbf{w}\,(s_i - s_{i-1}). \end{aligned}$$

This expression is a Riemann sum for the integral of $\mathbf{F}(\ell(s)) \cdot \mathbf{w}$ for $s \in [0,1]$. Taking a "limit" as the maximum of the $s_i - s_{i-1}$'s goes to zero yields the expression

$$W = \int_0^1 \mathbf{F}(\ell(s)) \cdot \mathbf{w}\,ds. \tag{9.1}$$

Remark 27. It is important to note that this formula holds for any parametrization $\mathbf{r}(t)$ of the line segment from **a** to **b** where $\mathbf{r}'(t) = \alpha(t)\mathbf{w}$ with $\alpha(t) \geq 0$. With a substitution, change of variable, we can rewrite Eq. 9.1 as

$$W = \int_{t_0}^{t_1} \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\,dt.$$

Here $\mathbf{r}(t_0) = \mathbf{a}$ and $\mathbf{r}(t_1) = \mathbf{b}$.

Example 316. Assume that a 1,000 kg mass is moved from the surface of the earth to a height of 350 km above the earth along a straight line perpendicular to the surface of the earth. What is the work required?

First note that the gravitational constant times the mass of the earth is $\mu = GM_E = 3.98600 \times 10^5\,\mathrm{km}^3/\mathrm{s}^2$ and the earth's radius is approximately $R_E = 6{,}371$ km. Then the force of gravity acting on a mass m at a point $\mathbf{x}$ when the origin is the center of the earth is

$$\mathbf{F}_G := -\frac{GM_E\,m}{\|\mathbf{x}\|^3}\mathbf{x} = -\frac{GM_E\,m}{\|\mathbf{x}\|^2}\,\frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

This is the usual inverse square law in vector form.

Without loss of generality we can take the initial point at the surface of the earth as $(R_E, 0, 0)$ km and the final point as $(R_E + 350, 0, 0)$ km. The line segment can be parametrized as $\ell(r) = (R_E + r, 0, 0)$ km for $r \in [0, 350]$. In terms of r we have

$$\mathbf{F}_G := -\frac{GM_E\, m}{(R_E + r)^3}\,(R_E + r, 0, 0).$$

Since $\ell\,'(r) = (1, 0, 0)$, the work done is

$$\begin{aligned} W &= \int_0^{350} -\frac{GM_E\, m}{(R_E + r)^3}\,(R_E + r, 0, 0)\cdot(1, 0, 0)\, dr \\ &= \int_0^{350} -\frac{GM_E\, m}{(R_E + r)^2}\, dr \\ &= \left.\frac{GM_E\, m}{(R_E + r)}\right|_0^{350} \\ &= GM_E\, m\left(\frac{1}{R_E + 350} - \frac{1}{R_E}\right) \\ &\approx -3.0662 \times 10^{41}\,\mathrm{J}. \end{aligned}$$

The last formula for work in this section is meant for finding the work when an object is moved through a force field along a nonlinear path. The justification for this formula is similar to the formula for linear motion and is omitted. If the path can be parametrized as a vector valued function $\mathbf{r}(t)$ on an interval $[a, b]$ and the force at a point $\mathbf{x}$ is $\mathbf{F}(\mathbf{x})$, the formula for the work done is

$$W = \int_a^b \mathbf{F}(\mathbf{r}(t))\cdot\mathbf{r}'(t)\, dt. \tag{9.2}$$

This reduces to the formula for movement along a straight line since if $\mathbf{r}(t) = \mathbf{w}t + \mathbf{c}$, $\mathbf{r}'(t) = \mathbf{w}$ for all $t \in [a, b]$. This can be fairly easy to use.

Example 317. Find the work done when a mass is moved along the curve $\mathbf{r}(t) = (t, 2t, t^2)$ m from $t = 0$ to $t = 2$ through a force field $\mathbf{F}(x, y, z) = (x - y, y - z, z - x)$ N. Here $\mathbf{F}(\mathbf{r}(t)) = (-t, 2t - t^2, t^2 - t)$ N and $\mathbf{r}'(t) = (1, 2, 2t)$ m/s. The work is

$$\begin{aligned} W &= \int_0^2 (-t, 2t - t^2, t^2 - t)\cdot(1, 2, 2t)\, dt \\ &= \int_0^2 3t - 4t^2 + 2t^3\, dt \\ &= \left.\frac{3t^2}{2} - \frac{4t^3}{3} + \frac{t^4}{2}\right|_0^2 \\ &= 6 - \frac{32}{3} + 8 \\ &= \frac{10}{3}\ \mathrm{J}. \end{aligned}$$

Exercises

1. Find the work done in moving a mass along a straight line from point A to point B under the influence of a constant force $\mathbf{F}$.

   (a) $A = (-2,3)$, $B = (3,-2)$, and $F = (1,-2)$
   (b) $A = (1,-2,3)$, $B = (-1,3,-2)$, and $F = (-2,1,-2)$
   (c) $A = (0,0,0)$, $B = (5,1,5)$, and $F = (1,-2,3)$
   (d) $A = (-1,2,-4)$, $B = (0,1,-1)$, and $F = (4,-4,6)$
   (e) $A = (1,1)$, $B = (3,3)$, and $F = (3,-1)$
   (f) $A = (4,1)$, $B = (-4,-1)$, and $F = (-1,-2)$

2. Find the work done in moving a mass along a straight line from point A to point B under the influence of a nonconstant force $\mathbf{F}(\mathbf{x})$.

   (a) $A = (-2,3)$, $B = (3,-2)$, and $F(x,y) = (x,-y)$
   (b) $A = (1,1)$, $B = (4,-2)$, and $F(x,y) = \left(x^2,y^2\right)$
   (c) $A = (1,1,1)$, $B = (4,-2,2)$, and $F(x,y,z) = (x,y,z)$
   (d) $A = (2,7,-6)$, $B = (5,1,-3)$, and $F(x,y,z) = \left(z^2,x,-y\right)$
   (e) $A = (0,3,7)$, $B = (7,-2,1)$, and $F(x,y,z) = (e^x,y,e^{x-z})$
   (f) $A = (0,1,-1)$, $B = (0,-2,2)$, and $F(x,y,z) = \left(\cos\left(x^2+y^2\right),x+y,z-y\right)$
   (g) $A = (7,-1,2)$, $B = (3,-5,-6)$, and $F(x,y,z) = \left(\cos(\pi x),\sin(4\pi y),1-z^2\right)$

3. Find the work done in moving a mass along the given curve $\mathbf{r}(t)$ from point A to point B, or from t_1 to t_2, under the influence of a nonconstant force $\mathbf{F}(\mathbf{x})$.

   (a) $A = (1,0,0)$, $B = (1,0,2\pi)$, $\mathbf{r}(t) = (\cos(t),\sin(t),t)$, and $\mathbf{F}(\mathbf{x}) = -(x,y,z)$.
   (b) $A = (-1,1,0)$, $B = (2,4,3)$, $\mathbf{r}(t) = (t,t^2,t^2-1)$, and $\mathbf{F}(\mathbf{x}) = (z,x+y,z-x)$.
   (c) $A = (1,e,1)$, $B = (\sqrt{10},e^{10},10)$, $\mathbf{r}(t) = (\sqrt{(}t),\exp(t),t)$, and $\mathbf{F}(\mathbf{x}) = (x^2,-y,-z^2)$.
   (d) $t_1 = -\pi$, $t_2 = \pi$, $\mathbf{r}(t) = (\sin(t),4,\cos(t))$, and $\mathbf{F}(\mathbf{x}) = (z-1,x,y-1)$.
   (e) $A = (0,0)$, $B = (1,1)$, $\mathbf{r}(t) = (t,t^2)$, and $\mathbf{F}(\mathbf{x}) = (x^2,y^3)$.

4. Let x be the position of a mass on a spring relative to the equilibrium position of the mass. Assume that the force that the spring exerts on a mass is equal to $F(x) = -kx$. This means that the spring pulls or pushes a mass towards its equilibrium position. Assuming that $k = 2$, what is the work done in moving the mass from $x = -1$ to $x = 2$? See Fig. 9.5 on page 293.

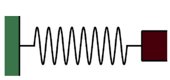


Fig. 9.5 A mass on a spring

5. Let x be the position of a mass on a spring relative to the equilibrium position of the mass. Assume that the force that the spring exerts on a 1 kg mass is equal to $F(x) = -kx$. This means that the spring pulls or pushes a mass towards its equilibrium position. Assuming that $k = 2$ and the mass is released at the point $x = 3$, how fast is the mass moving when is passes the point $x = 0$?

6. Find the work done when moving a 2 m long chain that is flat on a table onto the arc of a circle of radius 2 m that is above and tangent to the table. One end of the chain will still be on the table. The density of the chain is 1 kg/m. See Fig. 9.6 on page 294.

9.2 Work 2

In the previous section the work required to move a point object through a force field was considered. In this section the work required to move a mass that is not a point object is considered. Examples of this include the work required to take the water out of a tank or the amount of energy required to take a chain on the ground and hang it from a hook above the ground. In each of these cases, we can approximate the work done by considering the work done when moving thin layers of the water or small pieces of the chain from their original positions to their final positions.

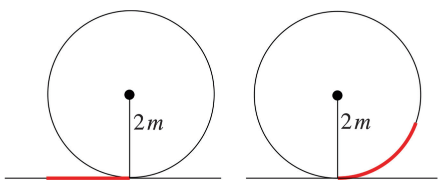


Fig. 9.6 Moving a chain onto a circle

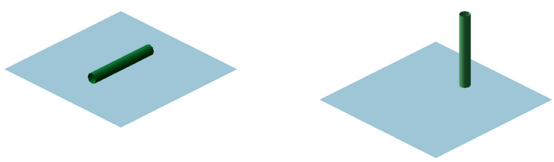


Fig. 9.7 Positions of a cable on a table and hung above the table

Consider the problem of taking a 1 m cable that has mass density of $\rho = 1/4$ kg/m that is lying on the ground and is moved so that one end of the cable is 1 m above the ground and the other end is touching the ground. See Fig. 9.7 on page 294.

Let x be the height above the ground where a point on the cable is after it has been moved. A short length of the cable, from x to $x+\Delta x$, has mass $\rho \Delta x$ and has been moved approximately $x\,m$ off the ground. This means, assuming the acceleration due to gravity is constant, that the work done moving the piece of cable is approximately $-g\rho x \Delta x$. If the cable is divided into n pieces with lengths Δx_i the total work done is approximately

$$W \approx \sum_{i=1}^{n} -g\rho \xi_i \Delta x_i,$$

where ξ_i is a point in the i the interval. The right hand side of this equation is a Riemann sum for the integral

$$\int_0^1 -g\rho x dx.$$

This means that the work is given by

$$\begin{aligned} W &= \int_0^1 -9.8 \cdot \frac{1}{4} \cdot x\,dx \\ &= -\frac{9.8}{4}\frac{x^2}{2}\bigg|_0^1 \\ &= -\frac{9.8}{4}\text{ J} \end{aligned}$$

We can also calculate work done in other situations. Two of these are done here.

Consider the amount of work it takes to empty a right circular cylinder of water by pumping the water over the top edge of the cylinder. See Fig. 9.8 on page 295. Let ρ represent the density of water in the system of measurement used. If the radius of the cylinder is r, the work required to lift a section of water with thickness Δy_i a height h_i is

$$\Delta W_i \approx -g\rho\pi r^2 \Delta y_i h_i.$$

If the height of the tank is H and the tank is initially filled to a height of $H_0 \le H$, where y is the

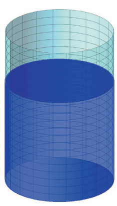


Fig. 9.8 A cylinder partially full of a liquid

height above the base, we have $h(y) = H - y$ where y goes from 0 to H_0. Writing out a Riemann sum for the work we get

$$\begin{aligned} W &= \sum_{i=1}^{n} \Delta W_i \\ &\approx \sum_{i=1}^{n} -g\rho\pi r^2 (H - y_i)\Delta y_i. \end{aligned}$$

Taking the limit as Δy_i goes to zero gives an integral for the total work.

$$\begin{aligned} W &= \int_0^{H_0} -g\rho\pi r^2 (H-y)\,dy \\ &= g\rho\pi r^2 \frac{(H-y)^2}{2}\bigg|_0^{H_0} \\ &= g\rho\pi r^2 \frac{\left((H-H_0)^2 - H^2\right)}{2}. \end{aligned} \tag{9.3}$$

Example 318. Consider a cylindrical tank with height 2 m and diameter 1/2 m. If the tank is half full of water, how much work does it take to empty the tank by pumping the water over the top rim of the tank?

The density of water is $\rho = 1\,\mathrm{g/cm^3} = 1{,}000\,\mathrm{kg/m^3}$. Plugging into Eq. 9.3 we have

$$\begin{aligned} W &= 9.8 \cdot 1{,}000 \cdot \pi \cdot \frac{1}{4} \cdot \frac{\left(\frac{1}{4}-1\right)}{2} \\ &= -918.75\,\pi. \end{aligned}$$

Now consider the problem of a cable that was resting partially on a table with a length ℓ_1 m hanging over the edge and a length ℓ_2 m on the table. Assume that there is no friction between the cable and the table top and the cable has density ρ kg/m. See Fig. 9.9 on page 296. A question in this setting is, what is the speed of the cable when it leaves the table?

This can be done by calculating the kinetic energy of the cable when it leaves the table. We can consider the cable in two pieces, the portion initially on the table and the portion initially hanging over the edge. First consider the portion of the cable that is initially hanging over the edge of the table. Each piece of length $\Delta\ell$ of this portion of the cable has moved down a distance of ℓ_2 when the cable leaves the table. This means that the work done for this piece of the cable is

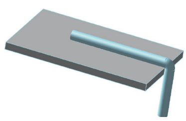


Fig. 9.9 A cable hanging over the edge of a table

$$\int_0^{\ell_1} 9.8\rho\ell_2\,d\ell = 9.8\rho\,\ell_1\,\ell_2\,\mathrm{J}.$$

A section of the cable on the table that is x away from the edge ends up $\ell_2 - x$ below the table top. This means that the work done getting the length of cable on the table off the table is

$$\int_0^{\ell_2} 9.8\rho\,(\ell_2 - x)\,dx = 4.9\rho\,\ell_2^2\,,\ \mathrm{J}.$$

Combining the two energies and using $E = mv^2/2$ gives

$$\frac{\rho\,(\ell_1+\ell_2)\,v^2}{2} = 9.8\rho\,\ell_1\,\ell_2 + 4.9\rho\,\ell_2^2\,\mathrm{J}.$$

Solving for v gives

$$v = \sqrt{\frac{19.6\ell_1\,\ell_2 + 9.8\,\ell_2^2}{\ell_1+\ell_2}}\,,\ \mathrm{m/s}.$$

Exercises

1. A 6 m cable with constant density is taken from a flat surface and hung from a point 3 m above the surface. Explain why the work done is the same as taking a 3 m cable with the same constant density and hanging it from a point 3 m above the surface.

2. A 3 m cable with constant density is taken from a flat surface and hung from a point 3 m above the surface. If the density is 0.2 kg/m and the acceleration due to gravity is $9.8\,\mathrm{m}/\mathrm{s}^2$, how much work is done moving the cable?

3. A 3 m cable has density $\rho(x) = \left(3 - \frac{x}{3}\right)$ kg/m, where x is the distance from end A of the cable. The cable starts on a table at the surface of the earth. It is then moved so that the cable is hanging from end A of the cable with end A of the cable 4 m above the table. What is the work that has been done?

4. A conical tank is to be emptied by taking water over the top of the tank. The tank is 10 m tall, has its point at the bottom and the top is a circle with radius 5 m. The density of water is 1 g per milliliter. If the tank is half full, how much work must be done to empty the tank? See Fig. 9.10.

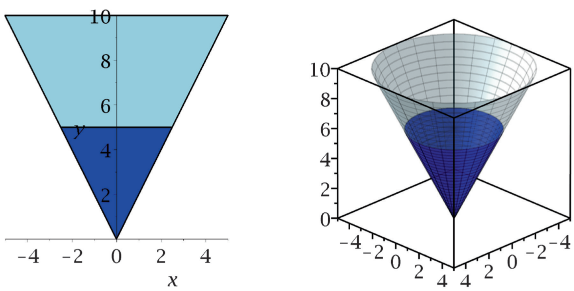


Fig. 9.10 An upside down right circular cone partially filled with fluid

5. A conical tank is to be emptied by taking water out through the bottom of the tank. The tank is 10 m tall, has its point at the bottom and the top is a circle with radius 5 m. The density of water is 1 g per milliliter. If the tank is half full, how much work must be done to empty the tank? See Fig. 9.10.

6. A hemispherical tank is emptied by draining a liquid with density $1.8\,\mathrm{g}/\mathrm{cm}^3$ out through a hole in the bottom of the tank. The tank has a radius of 0.4 m and is oriented with its flat circular side as the top of the tank. How much work is done in emptying the tank? Assume that the tank is on the surface of the Earth.

7. A 3 m length of cable with uniform mass density is partially hanging off of a platform. Assume that the surface of the platform has no friction. If 1 m of the cable is hanging over the edge of the platform when the cable is not moving, how fast will the cable be moving when the cable leaves platform. (Use $g = 9.8\,\mathrm{m}/\mathrm{s}^2$.)

8. A 3 m length of cable with uniform mass density 0.2 kg/m and is partially hanging off of a platform. There is a 2 kg mass on the end of the cable off the platform. Assume that the surface of the platform has no friction. If 1 m of the cable is hanging over the edge of the platform when the cable is not moving, how fast will the cable be moving when the cable leaves platform. (Use $g = 9.8\,\mathrm{m}/\mathrm{s}^2$.)

9. A 3 m length of cable with uniform mass density 0.2 kg/m is partially hanging off of a platform. There is a 2 kg mass on the end of the cable on the platform. Assume that the surface of the platform has no friction. If 1 m of the cable is hanging over the edge of the platform when the cable is not moving, how fast will the cable be moving when the cable leaves platform. (Use $g = 9.8\,\mathrm{m}/\mathrm{s}^2$.)

10. A 2 m cable is partially on incline and partially hanging over the top of the incline, see Fig. 9.11. Assume there is no friction between the cable and the incline. If the incline is 30° from horizontal and 1 m of the cable is hanging off the top of the incline, how much work will have been done when the end of the cable leaves the incline?

11. A 2 m cable is partially on incline and partially hanging over the top of the incline, see Fig. 9.11. Assume there is no friction between the cable and the incline. If the incline is 30° from horizontal and 1 m of the cable is hanging off the top of the incline, how fast will the cable be traveling when the cable leaves the incline?

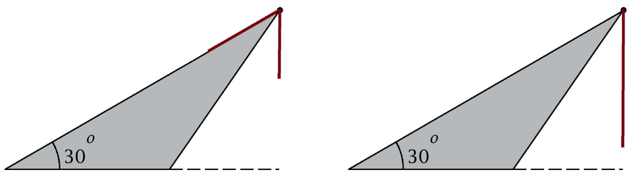


Fig. 9.11 A cable hanging over the top of an incline

12. A company is planning on building a space elevator from a point on the equator of the surface of the earth to a station in geosynchronous orbit 42,164 km from the center of the earth. Assume the radius of the earth is 6,378 km. If the density of the "cable" used is 1,000 kg/m for the full length of the "cable," how much work is done in moving the components of the "cable" from the surface of the earth to their positions on the space elevator?

13. A company is planning on building a space elevator from a point on the equator of the surface of the earth to a station in geosynchronous orbit 42,164 km from the center of the earth. Assume the radius of the earth is 6,378 km. If the density of the "cable" used is 1,000 kg/m at the bottom of the cable and decreases linearly to a density of 250 kg/m at the top of the "cable," how much work is done in moving the components of the "cable" from the surface of the earth to their positions on the space elevator?

9.3 Volumes by Slices

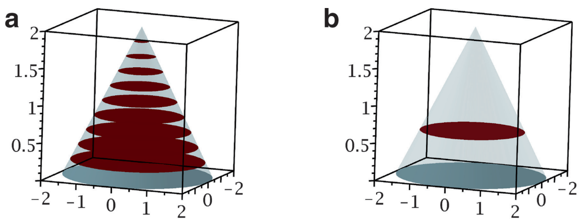
Integration can be used to find the volumes of many objects. In this text two methods are presented, volumes by slices and volumes by shells. The technique in this section, volumes by slices, assumes that we can describe a three dimensional object as the union of figures in the planes $x = t$ (or $y = t$ or $z = t$) as t varies between a and b. (See Fig. 9.12a on page 299.) We will assume that each function $A(t)$ is integrable over the interval $[a, b]$.

In the case of the volume between the cone $z = 2 - \sqrt{x^2 + y^2}$ and the xy-plane, the intersection of the figure with a plane $z = t$ is a circle when $t \in [0, 2]$. See Fig. 9.12b on page 299. We can approximate the volume between the planes $z = t$ and $z = t\Delta t$ as a tablet with constant area $A(t) = \pi r(t)^2 = \pi(2 - t)^2$ parallel to the xy-plane and height Δt. (See Fig. 9.13a on page 299.)

Stacking a series of these tablets together gives an approximation for the volume. (See Fig. 9.13b on page 299.) The sum of the volume of these tablets is

$$V \approx \sum_{i=1}^{n} \pi(2-t)^2 \Delta t_i.$$

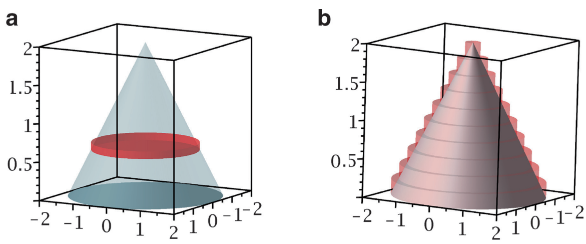
This is a Riemann sum for the area function $A(t) = \pi(2-t)^2$ as t goes from 0 to 2. Taking the limit as the mesh of the partition goes to 0 gives an integral for the volume



Multiple horizontal planes intersecting a cone

A single horizontal planes intersecting a cone

Fig. 9.12



A tablet approximating part of the volume of the cone

A stack of tablets approximating the volume of the whole cone

Fig. 9.13

$$\begin{aligned} V &= \int_0^2 \pi(2-t)^2\,dt \\ &= -\pi \left.\frac{(2-t)^3}{3}\right|_0^2 \\ &= \pi\left(-\frac{0}{3} + \frac{2^3}{3}\right) \\ &= \frac{8\pi}{3}. \end{aligned}$$

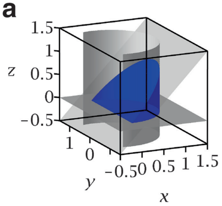
We can use the same basic formula,

$$V = \int_a^b A(t)\,dt,$$

to find the volume of other figures. As an example consider the volume in 3 dimensions that is inside the cylinder $x^2 + y^2 \leq 1$, above the xy-plane, and below the plane $z = x$. See Fig. 9.14a on page 300.

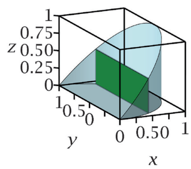
If we take slices in the planes perpendicular to the x-axis, $x = c$, the areas are rectangles. See Fig. 9.14b on page 300. These rectangles have height $z = x$ and length going from $y = -\sqrt{1-x^2}$ to $y = \sqrt{1-x^2}$. Putting these facts together leads to the volume,

a



The volume inside $x^2 + y^2 = 1$, above the xy-plane and below $z = x$

b



A slice of the figure parallel to the yz-plane

Fig. 9.14

$$\begin{aligned} V &= \int_0^1 2x\sqrt{1-x^2}\,dx \\ &= -\frac{2}{3}\left(1-x^2\right)^{(3/2)}\Big|_0^1 \\ &= \frac{2}{3}. \end{aligned}$$

For many people the biggest problem with this technique is visualizing the object and the finding appropriate slices with which to build the volume.

9.3.1 Volumes of Cones

A special case is when the volume is a cone. A cone is generated by connecting every point in a plane figure, the base, with a single point, the vertex, outside the plane. The idea here is build a cone by stacking areas similar to the base together and thereby generate the cone. See Fig. 9.15 on page 301. Let the area of the base of the cone be A_{Base} and let h the distance from the plane to the vertex. If we intersect the cone with a plane parallel to the plane with the figure that is a distance x from the plane, $x \in [0,h]$, the area of the intersection is $A(x) = A_{Base}(h-x)^2/h^2$. (This formula must include the $(h-x)^2/h^2$ since we are dealing with areas, not the length of segments.) Integrating these areas as x goes from 0 to h gives the volume,

$$\text{Vol} = \int_0^1 A(x)\,dx$$

$$
\begin{aligned}
&= \frac{A_{Base}}{h^2} \int_0^1 (h-x)^2\,dx \\
&= \frac{A_{Base}}{h^2} \left.\frac{-(h-x)^3}{3}\right|_{x=0}^{h} \\
&= \frac{1}{3} A_{Base}\, h.
\end{aligned}
$$

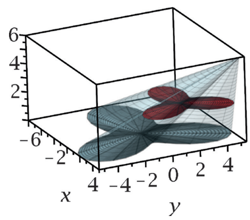
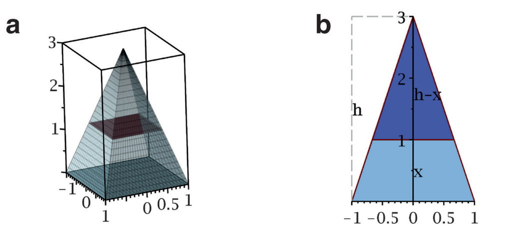


Fig. 9.15 An example of a slice of a cone parallel to the base



A pyramid with a square base and a slice of the pyramid parallel to the base

A side view of the pyramid

Fig. 9.16

This is rather abstract, but the idea can be illustrated using the case of a pyramid that has a square base where the lengths of the sides of the base are 2 m and height is 3 m. See Fig. 9.16a on page 301. From a side view this looks like Fig. 9.16b on page 301.

By similar triangles, the length of a side of the square at a distance x above the xy-plane is $\ell(x) = 2(h-x)/h = 2(3-x)/3$. This means that the area of a slice of the pyramid that is x units above the xy-plane is $A(x) = 4(3-x)^2/9$. By the method of slices, the volume is

$$
\begin{aligned}
\text{Vol} &= \int_0^3 \frac{4}{9}(3-x)^2\,dx \\
&= -\frac{4}{27}(3-x)^3 \Big|_{x=0}^{3} \\
&= 4.
\end{aligned}
$$

Since the area of the base is 4 and the height is 3, the formula for the volume of a cone

$$
\text{Vol} = \frac{1}{3} A_{Base}\, h
$$

gives a volume of 4. This confirms that the formula is correct in this case.

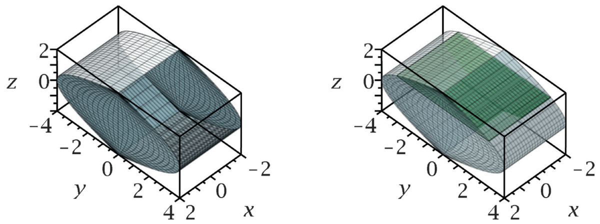


Fig. 9.17 The volume for Exercise 1

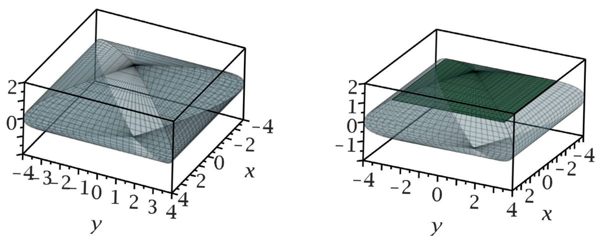


Fig. 9.18 The volume for Exercise 2

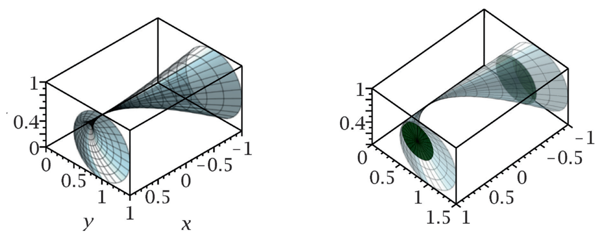


Fig. 9.19 The volume for Exercise 3

Exercises

1. Use slices to find the volume of the region in three dimensions bounded by $y = 4 - z^2$, $y = z^2 - 4$, $x = -2$, and $x = 2$. See Fig. 9.17 on page 302.

2. Use slices to find the volume of the region in three dimensions bounded by $y = 4 - z^2$, $y = z^2 - 4$, $x = z^2 - 4$, and $x = 4 - z^2$. See Fig. 9.18 on page 302.

3. A volume touches the xy-plane along the curve $y = x^2$ such that each slice by a plane $x = c$ is a circle of radius $r = \frac{c^2}{2}$. Find the volume of this figure for $x \in [-1, 1]$. See Fig. 9.19 on page 302.

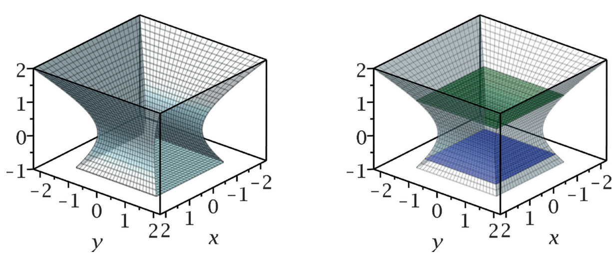


Fig. 9.20 The volume for Exercise 4

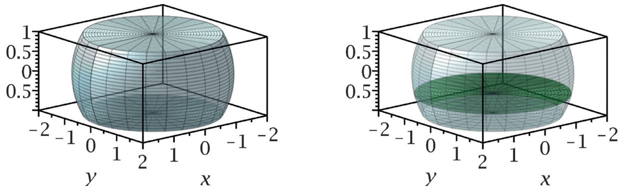


Fig. 9.21 The volume for Exercise 5

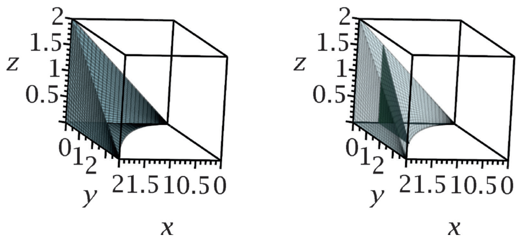
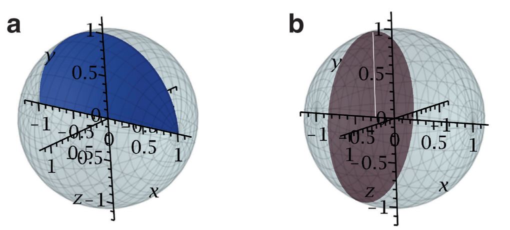


Fig. 9.22 The volume for Exercise 6

4. Find the volume of the region bounded by the hyperbolas $y^2 - z^2 = 1$ and $x^2 - z^2 = 1$ for $z \in [-1, 2]$. See Fig. 9.20 on page 302.
5. Find the volume of the section of the ball of radius 2 centered at the origin between the planes $z = -1$ and $z = 1$. See Fig. 9.21 on page 303.
6. A volume is made of triangles with one vertex at $(x, 0, 0)$, a second vertex at $(x, 0, x)$ and the third vertex at $(x, x^2, 0)$. Find the volume of this figure if $x \in [0, 2]$. See Fig. 9.22 on page 303.
7. Find the volume of the region inside $4 = x^2 + y^2 + \frac{z^2}{4}$ using slices parallel to the xy-plane.



The area rotated to fill a sphere

A disk formed by rotating a line segment around the x-axis

Fig. 9.23

8. Find the volume in the first octant with $z \geq 9 - x^2 - y^2$, $z \leq 10$, $x \leq 3$ and $y \leq 3$
9. Find the volume in the first octant with $z \leq 9 - x^2 - y^2$, $x \leq 3$ and $y \leq 3$

9.4 Volumes of Revolution: Washers

The idea here is to build volumes by stacking areas, slices as in Sect. 9.3, generated by rotating line segments perpendicular to an axis of rotation around that axis. If these line segments depend nicely on a parameter along the axis of rotation, the integral of the areas will give a volume.

Consider the case of the volume of a sphere. We can generate a sphere by rotating a half disk, say $x \in [-r, r]$ and $y \in [0, \sqrt{r^2 - x^2}]$, around an axis, the x-axis. See Fig. 9.23a on page 303.

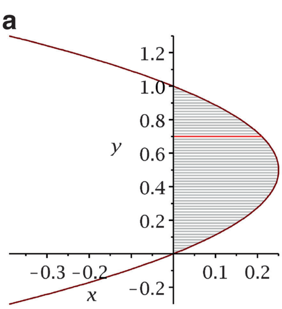
Since the half disk is made up of the line segments from $(x, 0)$ to $(x, \sqrt{r^2 - x^2})$, the sphere is made up of the disks generated by rotating these line segments around the x-axis. See Fig. 9.23b on page 303. The area of each of these disks as a function of x is $A(x) = \pi(r^2 - x^2)$. The volume is then

$$\begin{aligned}
\text{Vol} &= \int_{-r}^{r} A(x)\,dx \\
&= \int_{-r}^{r} \pi(r^2 - x^2)\,dr \\
&= \pi\left(x r^2 - \frac{x^3}{3}\right)\Bigg|_{x=-r}^{r} \\
&= \frac{4\pi}{3} r^3.
\end{aligned}$$

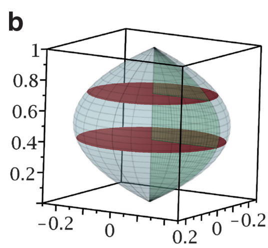
This is the correct formula for the volume of a sphere.

This idea is divided into two parts. The first is rotating line segments perpendicular to the axis of rotation starting at the axis of rotation, say the x-axis, and extending out to a distance $r(x)$ from the x-axis around the x-axis. This is called the method of disks. Since each of the areas, a disk with radius $r(x)$, has area $\pi r^2(x)$, the volume formula is

$$\text{Volume} = \int_a^b \pi r(x)^2\,dx.$$



The area to be rotated around the y-axis



Disks formed by rotating a line segments

Fig. 9.24

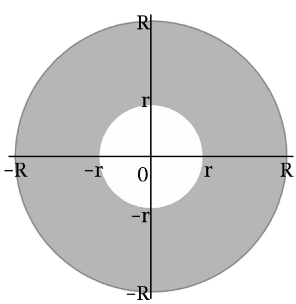


Fig. 9.25 A washer formed by removing a disk from inside another disk

Example 319. What is the volume generated by rotating the region in the first quadrant bounded by the y-axis and $x = y^2 - y$ around the y-axis? See Fig. 9.24a on page 304.

Here y goes from 0 to 1 and each line segments parallel to the x-axis goes from a point $(0, y)$ to $(y - y^2, y)$. This means $r(y) = y - y^2$ and the areas of the slices are $A(y) = \pi \left(y - y^2\right)^2$. See Fig. 9.24b on page 304.

The volume is

$$\begin{aligned}
\text{Vol} &= \int_a^b \pi r(x)^2 \, dx \\
&= \pi \int_0^1 (y - y^2)^2 \, dy \\
&= \pi \left(\frac{y^3}{3} - \frac{y^4}{2} + \frac{y^5}{5} \right) \Bigg|_{y=0}^{1} \\
&= \frac{\pi}{30}.
\end{aligned}$$

The second part of this general method is rotating line segments perpendicular to the axis of rotation that start at a distance r, the inner radius, away from the axis and extend out to a distance R, the outer radius, from the axis of rotation. The areas we get are washers and the method is called the method of washers. This is illustrated in Fig. 9.25 on page 304.

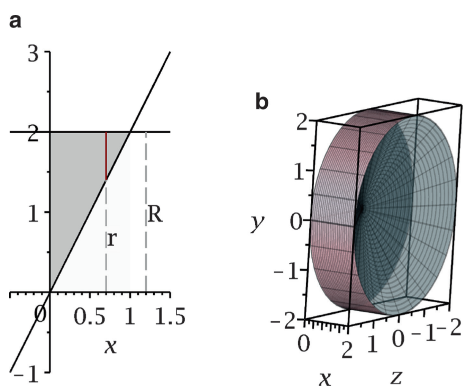


Fig. 9.26 The area and volume for Example 320 on page 305

The area of this washer is the area of a disk of radius R minus the area of a disk of radius r, $A = \pi \left(R^2 - r^2\right)$. If the axis of rotation is the x-axis, the inner and outer radii are functions of x, and if x varies from a to b we get a volume formula

$$\text{Vol} = \int_a^b \pi \left(R^2(x) - r^2(x)\right) dx.$$

This formula may be difficult to remember, but geometrically it is very simple. This integral can be view as taking the volume obtained by rotating $R(w)$ around an axis, say the w-axis, and the removing the volume obtained by rotating $r(w)$ around the w-axis,

$$\text{Vol} = \int_a^b \pi R^2(w) \, dw - \int_a^b \pi r^2(w) \, dw$$

Example 320. What is the volume of the region in 3 dimensions obtained by rotating the region bounded by the y-axis, $y=2$ and $y=2x$ in the xy-plane around the x-axis? The lines $y=2$ and $y=2x$ meet at $(1,2)$. This gives the region in Fig. 9.26a on page 305.

The outer radius is $R=2$ and the inner radius is $r=2x$. This means the volume is

$$\begin{aligned}\text{Vol} &= \int_0^1 \pi\left(2^2-(2x)^2\right)\,dx\\ &= \pi\left(4x-\frac{4x^3}{3}\right)\Bigg|_0^1\\ &= \pi\left(4-\frac{4}{3}\right)\\ &= \frac{8\pi}{3}.\end{aligned}$$

If we look at the region, see Fig. 9.26b on page 305, we can see that it is a cylinder of radius 2 and height 1 with a right circular cone of radius 2 and height 1 removed. Since the volume of the cylinder is $\text{Vol}_1=\pi r^2 h=4\pi$ and the volume of the cone is $\text{Vol}_2=\pi r^2 h/3=4\pi/3$, the volume of the solid of revolution is $8\pi/3$, the answer obtained through the integral.

In the previous examples in this section the axis of rotation was always one of the coordinate axes. This is not a requirement. If the axis of rotation is parallel to one of the coordinate axes, finding the volume of a volume of revolution follows the same procedure as above.

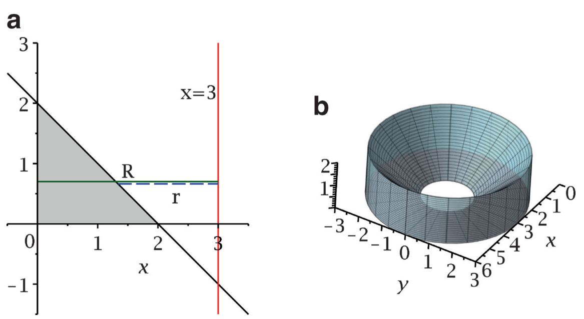


Fig. 9.27

Example 321. What is the volume of the solid generated by rotating the region bounded by the y-axis, $y=2-x$, and $y=0$ around the line $x=3$. The region in the plane is illustrated in Fig. 9.27a on page 306. The volume of revolution is shown in Fig. 9.27b on page 306.

If we use y as the variable of integration, the inner radius is $r(y)=1+y$ and the outer radius is always $R(y)=3$. (How do we get the inner radius?) We also have the limits on y at $y=0$ and $y=2$. The volume is then calculated as

$$\begin{aligned}\text{Vol} &= \int_0^2 \pi\left(3^3-(1+y)^2\right)\,dy\\ &= \pi\int_0^2\left(8-2y-y^2\right)\,dy\\ &= \pi\left(8y-y^2-\frac{y^3}{3}\right)\Bigg|_{y=0}^2\end{aligned}$$

$$= \pi \left(16 - 4 - \frac{8}{3}\right)$$
$$= \frac{28\pi}{3}.$$

Exercises

1. Find the volume of the region generated by rotating the area in the first quadrant between the x-axis and the curve $y = 3 + 2x - x^2$ around the x-axis. See Fig. 9.28 on page 307.

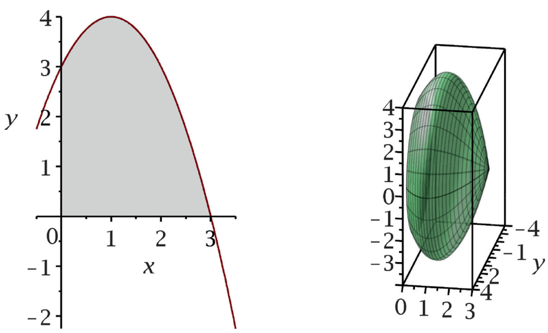


Fig. 9.28 The area and volume for Exercise 1

2. Find the volume of the region generated by rotating the area in the second quadrant between the x-axis and the curve $y = 3 + 2x - x^2$ around the x-axis. See Fig. 9.29 on page 307.

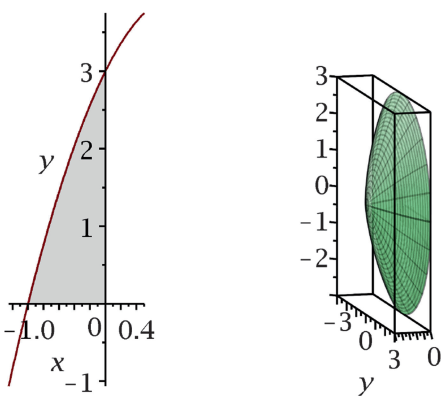


Fig. 9.29 The area and volume for Exercise 2

3. Find the volume of the region generated by rotating the area in the first quadrant between the x-axis and the curve $y = \sin(x)$ with $x \in [0, \pi]$ around the x-axis. See Fig. 9.30 on page 308.
4. Find the volume of the region generated by rotating the area in the first quadrant between the line $y = x$ and the curve $y = 4 + 2x - x^2$ with $x \in [0, \pi]$ around the x-axis. See Fig. 9.31 on page 308.

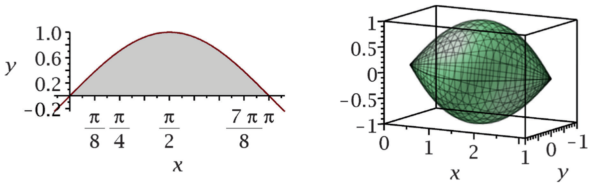


Fig. 9.30 The area and volume for Exercise 3

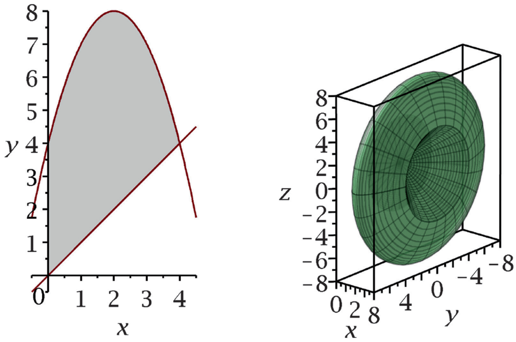


Fig. 9.31 The area and volume for Exercise 4

5. Find the volume of the region generated by rotating the area in the first quadrant between the line $y = x$ and the curve $y = 4 + 2x - x^2$ with $x \in [0, \pi]$ around the line $y = -2$. See Fig. 9.32 on page 308.

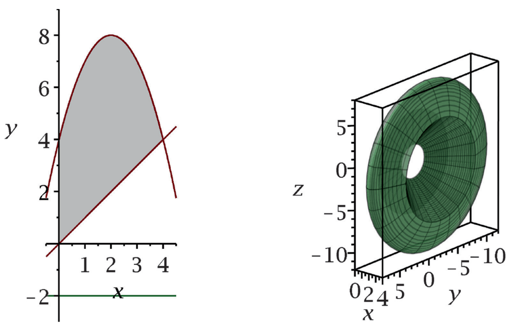


Fig. 9.32 The area and volume for Exercise 5

6. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the x-axis. Find the resulting volume.
7. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the y-axis. Find the resulting volume.
8. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the line $y = -1$. Find the resulting volume.
9. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the line $x = 2$. Find the resulting volume.

9.5 Volumes of Revolution: Shells

The idea here is to build volumes by stacking areas generated by rotating line segments parallel to an axis or rotation around that axis. The objects generated are called shells. If these line segments depend nicely on a parameter perpendicular to the axis of rotation, the integral of the areas will give a volume.

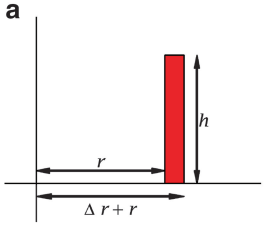
To write the volume as the limit of Riemann sums we need the volume obtained from rotating a thin strip parallel to the axis of rotation. See Figs. 9.33a on page 309 and 9.33b on page 309. This is the volume between two shells, both with height h, the inner with radius r_i and the outside with radius $r_i + \Delta r_i$. Since this is the difference in the volumes of two cylinders, the volume is

$$\begin{aligned}\Delta V_i &= \pi (r_i + \Delta r_i)^2 h - \pi r_i^2 h \\ &= \pi 2 r_i h \Delta r_i + \pi (\Delta r_i)^2 h.\end{aligned}$$

Consider using the shells to approximate the volume obtained by rotating an area bounded by $x = a$, $x = b$ ($b > a > 0$), $\ell(x)$ and $u(x)$ with $u(x) \geq \ell(x)$ around the y-axis. The interval $[a,b]$ can be partitioned into $a = r_0 < r_1 < r_2 < \cdots < r_n = b$ and, with $\Delta r_i - r_{i-1}$, the total volume can be approximated by

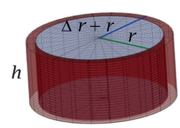
$$\begin{aligned}\text{Vol} &\approx \sum_{i=1}^{n} \Big(\big(\pi 2 r_i \Delta r_i + \pi (\Delta r_i)^2\big) \big(u(\xi_i) - \ell(\xi_i)\big) \Big). \\ &\approx \sum_{i=1}^{n} \Big(\pi 2 r_i \Delta r_i \big(u(\xi_i) - \ell(\xi_i)\big) \Big) + \sum_{i=1}^{n} \Big(\pi \, (\Delta r_i)^2 \big(u(\xi_i) - \ell(\xi_i)\big) \Big).\end{aligned}$$

a



A rectangle to rotate around an axis

b



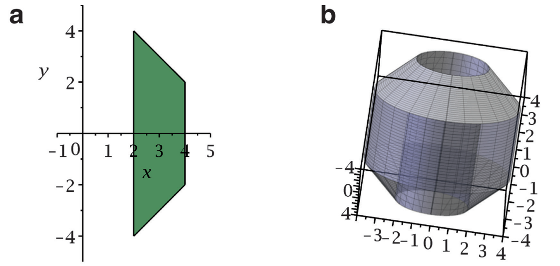
The volume generated by rotating a rectangle around an axis parallel to a side of the rectangle

Fig. 9.33

When we assume $u(x) - \ell(x) < M$ for all $x \in [a,b]$ and $\Delta r_i < K$ for all i the second term is always between 0 and $\pi K M (b-a)$. Noting that we can take $K = \max_i \Delta r_i \to 0$, the second sum goes to zero as we take $\max \Delta r_i$ going to zero. Since the second term goes to zero as $\max_i \Delta r_i \to 0$, and the first term is a Riemann sum, in the limit we should have

$$\text{Vol} = \int_a^b 2\pi x \big(u(x) - \ell(x)\big)\, dx.$$

In fact, this is the correct formula.



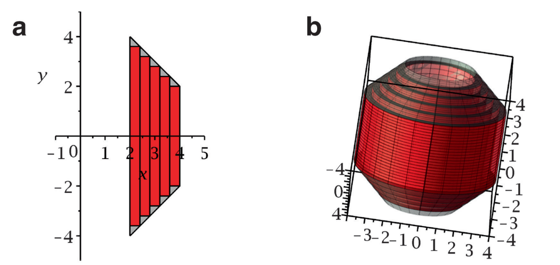
A region to rotate around an axis

The volume generated by rotating the region around an axis

Fig. 9.34

Example 322. Let A be the region in the xy-plane bounded by $x = 2$, $x = 4$, $y = 6 - x$, and $y = x - 6$, see Fig. 9.34a on page 310. The problem is to find the volume of the region generated by rotating this plane figure around the y-axis, Fig. 9.34b on page 310.

The volume can be approximated by approximating the plane region with strips, see Fig. 9.35a on page 310. Those strips can be rotated to get an approximate volume for the 3-dimensional figure, see Fig. 9.35b on page 310.



The region approximately filled with rectangles

The approximation to the volume generated by rotating the rectangles around the axis

Fig. 9.35

Finding the volumes of these approximations and taking the limit as the widths go to zero will give the volume,

$$\begin{aligned}
\text{Vol} &= \int_2^4 2\pi x\big((6-x)-(x-6)\big)\,dx \\
&= 2\pi \int_2^4 12x - 2x^2\,dx \\
&= 2\pi \left(6x^2 - \frac{2x^3}{3}\right)\Bigg|_{x=2}^{4} \\
&= \frac{208}{3}\pi.
\end{aligned}$$

This volume can be described as a cylinder of radius 4 and height 4 that has had a cylinder of radius 2 and height 4 removed form the center. Then two additional volumes obtained by rotating the triangles bounded by the lines $y = 4$, $x = 4$, and $y = 6 - x$, and by $y = -4$, $x = 4$, and $y = x - 6$ around the y-axis. Since these two volumes of revolution have the same volume

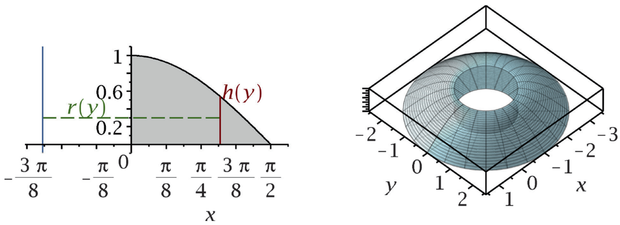


Fig. 9.36 The area and volume for Example 323 on page 311

and that volume can be calculated using washers, the volume can easily be calculated without using shells. (Draw a picture to see the triangles.)

Example 323. Consider the volume generated by rotating the plane region in the first quadrant bounded by $y = \cos(x)$, the y-axis and the x-axis around the line $x = -1$, see Fig. 9.36 on page 311.

If we try to use washers to find the volume of this region we need to integrate against y with an outer radius of $R(y) = 1 + \arccos(y)$ and an inner radius of $r(y) = 1$. When simplified, the integral for the volume using washers is the rather unpleasant

$$\text{Volume} = \pi \int_0^1 2\arccos(y) + \arccos^2(y)\,dy.$$

If we use cylinders, the radius is $r(x) = 1 + x$ and the height is $h(x) = \cos(x)$ with x in $[0, \pi/2]$. This gives the much simpler integral

$$\text{Volume} = 2\pi \int_0^{\frac{\pi}{2}} \cos(x) + x\cos(x)\,dx.$$

We can use integration by parts to get

$$\begin{aligned}
\text{Volume} &= 2\pi \int_0^{\frac{\pi}{2}} \cos(x) + x\cos(x)\,dx \\
&= 2\pi\left((\sin(x) + x\sin(x))\Big|_{x=0}^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin(x)\,dx \right) \\
&= 2\pi\left(1 + \frac{\pi}{2} + \left(\cos(x)\Big|_{x=0}^{\frac{\pi}{2}} \right)\right) \\
&= \pi^2.
\end{aligned}$$

This is a illustration of the fact that finding the volume of a volume of revolution is often much easier when using one of the two techniques given here rather than the other. If one of the methods is too difficult, try the other.

9.5.1 Moment of Inertia

Moment of inertia plays the same role in rotational mechanics that mass does in the motion of a point mass in mechanics. For example, the angular momentum of an object rotating around an axis with moment of inertia I and angular velocity (in radians/s) ω is $I\omega$. This corresponds to the momentum formula mv. The kinetic energy of the rotation is given by $1I\omega^{2/2}$, a direct correspondence to the energy formula $E = mv^{2/2}$.

If a point mass with mass m is a distance r from the axis of rotation of a mechanical system, the moment of inertia of that mass is mr^2. See Fig. 9.37 on page 312.

When we consider a thin shell around an axis or rotation the total mass of the shell is at a fixed radius from the axis of rotation. This means that the moment of inertia of this shell of mass m should be $I = mr^2$. If it has thickness Δr, height h and density ρ, the moment of inertia can be approximated by $\Delta I \approx \rho\,\pi\,r^3\,h\,\Delta r$. See Fig. 9.38 on page 312.

If we has a solid with constant density ρ that is obtained by rotating a plane region around an axis in the plane, we can use the approximation for the moment of inertia for thin shells to get the approximation

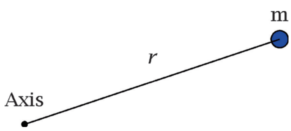


Fig. 9.37 $I = mr^2$

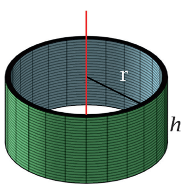


Fig. 9.38 A shell used to calculate a moment of inertial

$$I \approx \sum_{i=1}^{n} \rho\,2\pi\,r_i^3\,h(r_i)\,\Delta r_i.$$

The left hand side of this equation is almost the Riemann sum for the volume using shells. Taking the limit of this sum as the maximum of the Δr_i's goes to 0 gives the following equation.

$$I = \int_a^b 2\pi\,\rho\,r^3\,h(r)\,dr.$$

The errors in this derivation are constants times the forms $\Delta\rho\,\Delta r$ and $(\Delta r)^2$. These terms will disappear when we take the limit.

Example 324. Consider a solid cylinder with constant density and mass m that is rotating around it's axis of symmetry. Assume the cylinder has radius r and height h. The density of the cylinder is $\rho = m/\left(\pi r^2 h\right)$, the mass divided by the volume. We can calculate the moment of inertia:

$$\begin{aligned} I &= \int_0^r 2\pi\,\frac{m}{\pi r^2 h}\,x^3\,h\,dx \\ &= \frac{2m}{r^2}\int_0^r x^3\,dx \end{aligned}$$

$$= \frac{2m}{r^2}\frac{x^4}{4}\bigg|_0^r$$
$$= \frac{mr^2}{2}.$$

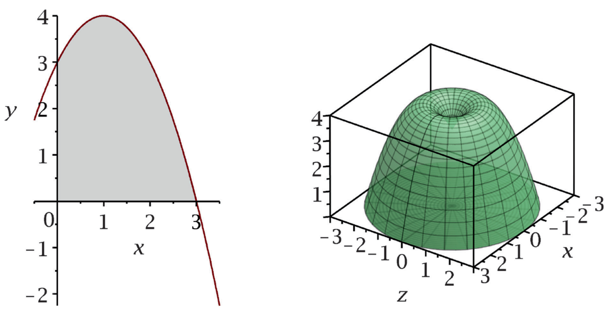


Fig. 9.39 The area and volume for Exercise 1

Exercises

1. Find the volume of the region generated by rotating the area in the first quadrant between the x-axis and the curve $y = 3 + 2x - x^2$ around the y-axis. See Fig. 9.39 on page 313.

2. Find the volume of the region generated by rotating the area in the second quadrant between the x-axis and the curve $y = 3 + 2x - x^2$ around the y-axis. See Fig. 9.40 on page 314.

3. Find the volume of the region generated by rotating the area in the first quadrant between the x-axis and the curve $y = \sin(x)$ with $x \in [0, \pi]$ around the y-axis. See Fig. 9.41 on page 314.

4. Find the volume of the region generated by rotating the bounded area in the first quadrant between the line $y = x$ and the curve $y = 4 + 2x - x^2$ around the y-axis.

5. Find the volume of the region generated by rotating the area in the first quadrant between the line $y = x$ and the curve $y = 4 + 2x - x^2$ around the line $x = -2$.

6. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the x-axis. Find the resulting volume.

7. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the y-axis. Find the resulting volume.

8. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the line $y = -1$. Find the resulting volume.

9. The region between the curves $y = x^3$ and $y = x$ in the xy-plane is rotated around the line $x = 2$. Find the resulting volume.

10. Find the moment of inertia of a ball of radius r with uniform density that has mass m around a line running through the center of the ball.

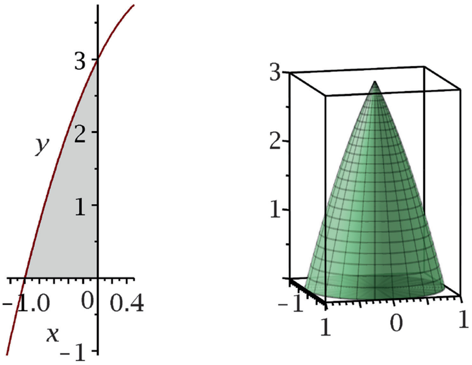


Fig. 9.40 The area and volume for Exercise 2

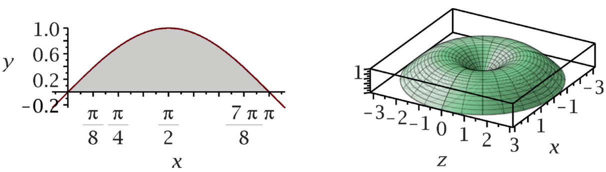


Fig. 9.41 The area and volume for Exercise 3

11. A right circular cone has height 2 m, base radius $\frac{1}{2}$ m and mass 9 kg. Find the moment of inertia of the cone around the line running through the tip of the cone and the center of its base.

12. A three dimensional object is generated by rotating a figure in the xy-plane around the y-axis. The region in the xy-plane is bounded by the x-axis, the line $x = b > 0$ and the curve $y = h(x)$ with $h(x) > 0$. Assuming that the mass density in the object is a function ρ of the distance from the y-axis, explain why the moment of inertia of the object is given by

$$I = \int_0^b 2\pi\,\rho(x)\,h(x)\,x^3\,dx.$$

9.6 Arc Length and Areas of Revolution

An important question in applications is given a position $\mathbf{r}(t)$ of an object as a function of time, how far has the object traveled from $t = a$ to $t = b$? Since we have already seen that the speed of the of the object is given by $\|\mathbf{r}'(t)\|$ and since the distance traveled is the integral of speed, the distance traveled is

$$L = \int_a^b \|\mathbf{r}'(t)\|\,dt.$$

This can be illustrated with a simple example.

Example 325. Let the position of an object be given by $\mathbf{r}(t) = \big(3\cos(t), 3\sin(t), t\big)$. The velocity of the object is $\mathbf{r}'(t) = \big(-3\sin(t), 3\cos(t), 1\big)$ and the speed is

$$\begin{aligned} s(t) &= \big\|\big(-3\sin(t), 3\cos(t), 1\big)\big\| \\ &= \sqrt{9\left(\cos^2(t) + \sin^2(t)\right) + 1} \\ &= \sqrt{10}. \end{aligned}$$

This means that the length of the helix from $t = 0$ to $t = 2\pi$ is

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{10}\, dt \\ &= 2\pi\sqrt{10}. \end{aligned}$$

Note the contrast with the distance from the initial point to the final point,

$$\begin{aligned} d &= \big\|\mathbf{r}(2\pi) - \mathbf{r}(0)\big\| \\ &= \big\|(1,0,2\pi) - (1,0,0)\big\| \\ &= 2\pi. \end{aligned}$$

The object here is to define the length for curves, arc length, when the function may or may not represent motion. In this case, the formula is exactly the same. If $\mathbf{x}(t)$ for $t \in [a,b]$ defines a reasonable curve as its image, then the length of the curve is given by

$$L = \int_a^b \big\|\mathbf{x}'(t)\big\|\, dt.$$

The idea of a reasonable curve will be discussed in later courses. At this point it is sufficient to assume that $\mathbf{x}(t)$ is continuous and is continuously differentiable on a finite number of segments $[t_{i-1}, t_i]$ for $i = 1, 2, 3, \ldots, n$ with $t_0 = a$ and $t_n = b$.

If we go back and look at the length of a curve through Riemann sums we should get the same formula. The approximation we use for a time interval from t_i to $t_i + \Delta t$ for the length is

$$\Delta L_i \approx \|\mathbf{r}(t_i + \Delta t) - \mathbf{r}(t_i)\|,$$

see Fig. 9.42a on page 316. Adding these approximations for a partition of $[a,b]$, see Fig. 9.42b on page 316, we get an approximation for the whole length of a curve

$$\begin{aligned} L &\approx \sum_{i=1}^{n} \|\mathbf{r}(t_i + \Delta t) - \mathbf{r}(t_i)\| \\ &\approx \sum_{i=1}^{n} \left\|\frac{\mathbf{r}(t_i + \Delta t_i) - \mathbf{r}(t_i)}{\Delta t_i}\right\| \Delta t_i. \end{aligned}$$

If the approximations are good enough, taking the limit as the largest size of Δt_i goes to zero and $\lim_{\Delta t_i \to 0} (\mathbf{r}(t_i + \Delta t_i) - \mathbf{r}(t_i))/\Delta t_i = \mathbf{r}'(t_i)$ gives

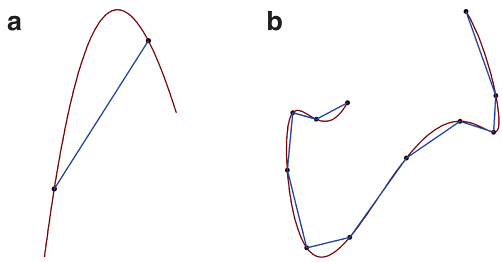
$$\begin{aligned} L &= \lim_{max|\Delta t_i| \to 0} \sum_{i=1}^{n} \left\|\frac{\mathbf{r}(t_i + \Delta t_i) - \mathbf{r}(t_i)}{\Delta t_i}\right\| \Delta t_i \\ &= \int_a^b \big\|\mathbf{r}'(t)\big\|\, dt. \end{aligned}$$

The reasons why the approximations work are left for later courses.

Example 326. The length of the graph of the function $f(x) = x^3 + x - 1$ from $x = -2$ to $x = 3$ is found by finding the length of the curve given by $\mathbf{r}(t) = (t, f(t))$. We get the general form of the length integral when $y = f(x)$ (or $x = f(y)$),

$$L = \int_a^b \sqrt{1 + (f'(t))^2}\, dt.$$

In this case the integral is



A secant line from $\mathbf{r}(t_i)$ to $\mathbf{r}(t_{i+1})$

Secant lines from $\mathbf{r}(t_i)$ to $\mathbf{r}(t_{i+1})$

Fig. 9.42

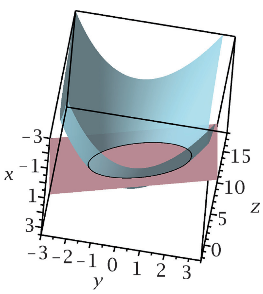


Fig. 9.43 The intersection of $z = x^2 + y^2$ and $z = x + y + 7/2$

$$\begin{aligned} L &= \int_{-2}^{3} \sqrt{1 + (3x^2 - 1)^2}\, dx \\ &\approx 32.85678. \end{aligned}$$

As is often the case, the indefinite integral does not have a simple closed form.

This form works for all nice graphs of functions.

Example 327. What is the length of the intersection of the paraboloid $z = x^2 + y^2$ and the plane $z = x + y + 7/2$? This is illustrated in Fig. 9.43 on page 316. In this case the points in the intersection of the two surfaces satisfy $x^2 + y^2 = x + y + 7/2$ or $x^2 - x + y^2 - y = 7/2$. Completing the square yields

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

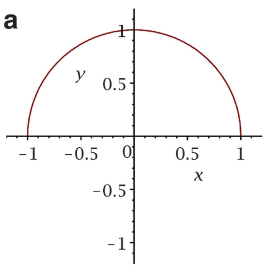
This is a circle of radius 2 centered at $(1/2, 1/2)$ in the xy-plane that can be parametrized as $\mathbf{r}(t) = (1/2 + 2\cos(t),\, 1/2 + 2\sin(t))$.

The z coordinate as a function of t is $z(t) = 9/2 + 2\big(\cos(t) + \sin(t)\big)$ and a parametrization of the curve is $\mathbf{s}(t) = \big(1/2 + 2\cos(t),\, 1/2 + 2\sin(t), 9/2 + 2\big(\cos(t) + \sin(t)\big)\big)$. Taking the derivative of $\mathbf{s}(t)$ gives

$$\mathbf{s}'(t) = \big(-2\sin(t), 2\cos(t), 2(\cos(t) - \sin(t)\big).$$

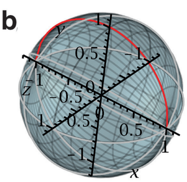
The integral for the length of the curve is

a



A half circle to rotate around the x-axis

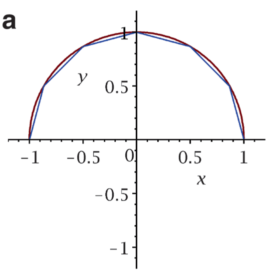
b



The sphere obtained from rotating a half circle around the x-axis

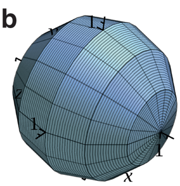
Fig. 9.44

a



A piecewise linear approximation to a half circle

b



The approximation to a sphere obtained from rotating a the piecewise linear approximation to the half circle around the x-axis

Fig. 9.45

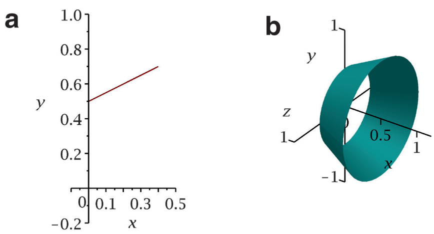
$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{\big|\big(-2\sin(t), 2\cos(t), 2(\cos(t) - \sin(t)\big)\big|}\, dt \\ &= \int_0^{2\pi} \sqrt{8 - 4\sin(2t)}\, dt \\ &\approx 17.4755. \end{aligned}$$

Again, this integral does not evaluate to a simple form.

We can use the ideas from the first part of this section to find the area of a surface of revolution. A surface of revolution is obtained by rotating a plane curve around a line in the same plane in three dimensions. A simple example is that a sphere of radius 1 can be generated by rotating the graph of $f(x) = \sqrt{1-x^2}$ in the xy-plain around the x-axis. See Fig. 9.44 on page 317.

If we approximate the curve by a sequence of secant lines and then rotate those secant lines around the x-axis we get a surface whose area approximates the area of the sphere, see Fig. 9.45 on page 317 for an example.

The thing we are missing at this point is the area of the surface obtained by rotating a line segment in the same plane as the axis of rotation. To motivate the formula for this area consider the area obtained by rotating a line segment from (x_0, y_0) to $(x_0 + \Delta x, y_0 + \Delta y)$ around the x-axis. See Fig. 9.46 on page 318 for an illustration.



A line segment to rotate around the x-axis

The area obtained by rotating a line segment around the x-axis

Fig. 9.46

At each point $(x_0 + t\Delta x, y_0 + t\Delta y)$ along the line from (x_0, y_0) to $(x_0 + \Delta x, y_0 + \Delta y)$ a circle of circumference $2\pi(y_0 + t\Delta y)$ is generated by the rotation. Since the length of the line segment being rotated is $s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ and since the circumference of the circle depends linearly on the distance from an endpoint of the segment being rotated, the area can be put on a plane as a trapezoid with height s, with top length $2\pi y_0$ and with bottom length $2\pi(y_0 + \Delta y)$. (See Fig. 9.47 on page 318.)

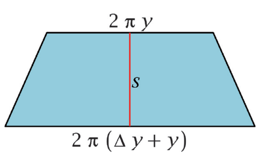


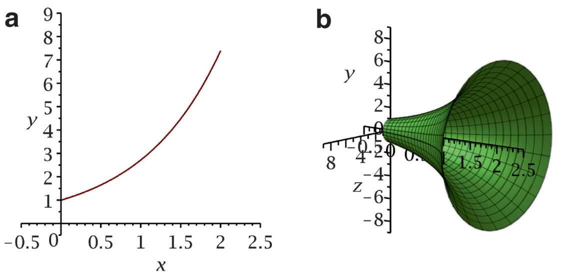
Fig. 9.47 The trapezoid representing area of the figure in Fig. 9.46b on page 318

The area of the trapezoid is

$$A = s\,\frac{(2\pi y_0) + 2\pi(y_0 + \Delta y)}{2} = s\big(2\pi y_0 + \pi\Delta y\big).$$

This is also the area of the revolution generated by rotating the line segment around the x-axis.

Now we return to the area of a surface of revolution. Assuming that $y = y(x)$ is a differentiable function of x, the area obtained by rotating one curve segment for $x \in [x_i, x_i + \Delta x_i]$ around the x axis is approximated using a secant line as



A portion of the graph of $y = e^x$ used in Example 328

The area for Example 328

Fig. 9.48

$$\begin{aligned}\Delta A_i &\approx \left(2\pi y(x_i) + \pi y'(\mu_i)\Delta x_i\right)\sqrt{1+\left(\frac{\Delta y_i}{\Delta x_i}\right)^2}\,\Delta x_i \\ &\approx \left(2\pi y(x_i) + \pi y'(\mu_i)\Delta x_i\right)\sqrt{1+(y'(\xi_i))^2}\,\Delta x_i.\end{aligned}$$

Here μ_i and ξ_i are points guaranteed to exist by the Mean Value Theorem. The Riemann sum for this is

$$\begin{aligned}A &= \sum_{i=1}^{n}\Delta A_i \\ &\approx \sum_{i=1}^{n}\left(2\pi y(x_i)\sqrt{1+(y'(\xi_i))^2}\,\Delta x_i + \pi y'(\mu_i)\sqrt{1+(y'(\xi_i))^2}\,(\Delta x_i)^2\right) \\ &\approx \sum_{i=1}^{n}\left(2\pi y(x_i)\sqrt{1+(y'(\xi_i))^2}\,\Delta x_i\right) + \sum_{i=1}^{n}\left(\pi y'(\mu_i)\sqrt{1+(y'(\xi_i))^2}\,(\Delta x_i)^2\right).\end{aligned}$$

Consider the second term in last line. It can be bounded in absolute value as

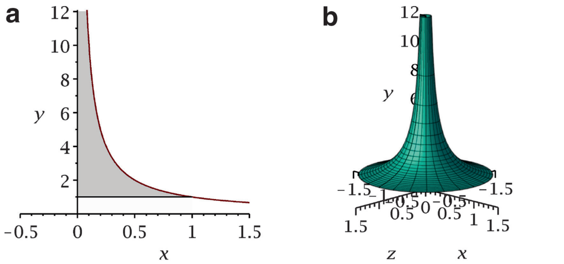
$$\begin{aligned}\left|\sum_{i=1}^{n}\left(\pi y'(\mu_i)\sqrt{1+(y'(\xi_i))^2}\,(\Delta x_i)^2\right)\right| &\le \pi\left|\max_i(\Delta x_i)\right|\left(\sum_{i=1}^{n} M\sqrt{1+M^2}\Delta x_i\right) \\ &\le \pi\left|\max_i(\Delta x_i)\right| M\sqrt{1+M^2}\,(b-a)\,,\end{aligned}$$

where M is the maximum of the absolute value of y' and $[a,b]$ is the interval on the axis of rotation associated with the figure. This term goes to zero as $\max(\Delta x_i)$ goes to zero. This means that only the first term is important and the limit of that Riemann gives the area of the surface of rotation. From this we conclude that formula for the area of the surface of revolution generated by rotating the curve $y = y(x)$ for $x \in [a,b]$ around the x-axis is

$$A = \int_a^b 2\pi y(x)\sqrt{1+(y'(x))^2}\,dx.$$

Example 328. The objective is to find the area of the surface generated by rotating the curve $y = \exp(x)$ from $x = 0$ to $x = 2$ around the x-axis. See Fig. 9.48 on page 319 for an illustration.

We can compute the surface area using the formula, the substitutions $\exp(x) = \tan(\theta)$ and $u = \sin(\theta)$, and partial fractions:



The area for Example 329

The surface and volume for Example 329

Fig. 9.49

$$\begin{aligned}
A &= 2\pi \int_0^2 \exp(x)\sqrt{1+\exp(2x)}\,dx \\
&= 2\pi \int_{\pi/4}^{\tan^{-1}(\exp(2))} \frac{1}{\cos^3(\theta)}\,d\theta \\
&= 2\pi \int_{\sqrt{2}/2}^{\frac{\exp(2)}{\sqrt{1+\exp(4)}}} \frac{1}{(1-u^2)^2}\,du \\
&= 2\pi \left(\frac{1}{4}\ln\left(\frac{1+u}{1-u}\right) - \frac{1}{2}\frac{u}{u^2-1}\right)\Bigg|_{u=\frac{\sqrt{2}}{2}}^{\frac{\exp(2)}{\sqrt{1+\exp(4)}}} \\
&= -\frac{\ln\left(-\exp(2)+\sqrt{\exp(4)+1}\right)}{4} + \frac{\ln\left(2-\sqrt{2}\right)}{4} \\
&\quad + \frac{\ln\left(\exp(2)+\sqrt{\exp(4)+1}\right)}{4} - \frac{\ln\left(2+\sqrt{2}\right)}{4} \\
&\quad + \frac{\exp(2)\sqrt{\exp(4)+1}}{2} - \frac{\sqrt{2}}{2} \\
&\approx 27.7500
\end{aligned}$$

This formula can be combined with improper integrals to get an interesting example concerning volumes and areas.

Example 329. Find the surface area and volume of the region generated by rotating the region in the first quadrant of the xy-plane that is bounded by $y = 1/x$, $y = 1$ and $x = 1$. See Fig. 9.49 on page 320 for an illustration.

We first calculate the volume by disks with integration against y. The radius of the disk of intersection of the volume with the plane $y = y_0$ is $r = 1/y_0$. The integral for the volume is

$$\begin{aligned}
\text{Volume} &= \int_1^\infty \frac{\pi}{y^2}\,dy \\
&= \lim_{b\to\infty} \int_1^b \frac{\pi}{y^2}\,dy
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{b\to\infty} \left(-\frac{\pi}{y}\right)\Bigg|_1^b \\
&= \lim_{b\to\infty} \pi \left(-\frac{1}{b}+1\right) \\
&= \pi.
\end{aligned}
$$

The region in three dimensions has a finite volume.

We now calculate the surface area of the region. The base of the volume is a disk of radius 1 and has area π. The area of the other surface of the volume is given by

$$\text{Area} = \int_1^\infty 2\pi \frac{1}{y} \sqrt{1+\left(\frac{-1}{y^2}\right)^2}\, dy.$$

This is integrable in closed form, however, it is easier to show the integral is infinite by noting that

$$\frac{1}{y}\sqrt{1+\left(\frac{-1}{y^2}\right)^2} \geq \frac{1}{y}$$

when $y \neq 0$.

This means that we can compute

$$
\begin{aligned}
\text{Area} &\geq \int_1^\infty 2\pi \frac{1}{y}\, dy \\
&\geq 2\pi \lim_{b\to\infty} \int_1^b \frac{1}{y}\, dy \\
&\geq 2\pi \lim_{b\to\infty} (\ln(b) - \ln(1)) \\
&\geq \infty.
\end{aligned}
$$

The surface area of the region is infinite.

This is an example of a region with a finite volume and an infinite surface area.

Exercises

1. Find the length of the curve $\mathbf{r}(t) = (t, t^2)$ for $t \in [-1, 2]$
2. Find the length of the curve $\mathbf{r}(t) = (t^3, t^2)$ for $t \in [-2, 3]$
3. Find the length of the curve $y = \cosh(x)$ for $x \in [-\ln(2), \ln(3)]$.
4. Find the length of the curve $\mathbf{r}(t) = \left(\frac{(t^2+2)^{3/2}}{3}, t\right)$ for $t \in [-1, 0]$.
5. Find the length of the curve $\mathbf{r}(t) = \left(t^4, t^2\right)$ for $t \in [1, 4]$.
6. Find the length of the curve $\mathbf{r}(t) = \left(\ln(2t), t^2\right)$ for $t \in [1, 4]$.
7. Find the length of the curve $\mathbf{r}(t) = \left(t^4, \frac{t^5}{5}\right)$ for $t \in [-1, 3]$.
8. Use Simpson's method with 16 equal intervals to approximate the length of the curve $y = x^3$ from $x = -2$ to $x = 1$.

9. Use Simpson's method with 30 equal intervals to approximate the length of the curve $\mathbf{s}(w) = (w^3, w)$ from $w = -1$ to $w = 2$.

10. Use Simpson's method with 26 equal intervals to approximate the length of the curve $\mathbf{s}(t) = (\sin(t), 4\cos(t))$ from $t = 0$ to $t = \frac{\pi}{3}$.

11. Find the area of the surface obtained by rotating the curve $y = 2x+1$ from $x \in [0,2]$ around the x-axis.

12. Find the area of the surface obtained by rotating the curve $y = x^2+1$ from $x \in [0,2]$ around the x-axis.

13. Find the area of the surface obtained by rotating the curve $y = x^3$ from $x \in [-4,-1]$ around the x-axis.

14. Find the area of the surface obtained by rotating the hyperbola $y^2 = x^2+1$ with $y > 0$ and $x \in [0,2]$ around the x-axis.

15. Find the area of the surface generated by rotating the curve $y = \sqrt{x}$ for $x \in [2,4]$ around the x-axis.

16. Find the area of the surface generated by rotating the curve $y = \sqrt{x}$ for $x \in [0,4]$ around the x-axis.

17. Find the area of the surface generated by rotating the curve $y = \cosh(x)$ for $x \in [-1,1]$ around the x-axis.

18. Find the area of the surface generated by rotating the curve $y = e^{-x}$ for $x \in [1,\infty]$ around the x-axis.

19. Use Simpson's method with 32 equal intervals to approximate the area obtained by rotating the curve $y = \sinh(x)$ with $x \in [0,5]$ around the x-axis.

9.7 Differential Equations

We have already seen differential equations a number of times in this text. Differential equations were introduced in Sect. 6.1, solving separable differential equations was discussed in Sect. 8.1, and the logistic equation was introduced in Sect. 8.3.1. In this section we look at the geometry of solutions of differential equations and we look at a simplistic numerical method for approximating solutions to differential equations.

Recall that in our context a differential equation is an equation that relates a function of a real variable, the real variable and the first n derivatives of the function,

$$\mathbf{F}\big(t, \mathbf{x}(t), \mathbf{x}'(t), \mathbf{x}''(t), \ldots, \mathbf{x}^{(n)}(t)\big) = 0.$$

For example, the equation relating force to acceleration is

$$\mathbf{x}'' = \frac{1}{M}\mathbf{F}\big(t, \mathbf{x}(t), \mathbf{x}'(t)\big)$$

where $\mathbf{F}$ is the force. The case where $\mathbf{F}\big(t, \mathbf{x}(t), \mathbf{x}'(t)\big)$ is a constant was discussed in Sect. 6.1.

If we consider a differential equation of the form

$$x'(t) = f(t, x(t)),$$

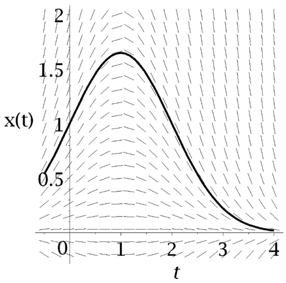


Fig. 9.50 The slope field and solution for Example 330

the graph of a solution can be parametrized as $t \to (t, x(t))$. This means that the vector $(1, x'(t_0))$ is tangent to the graph of $x(t)$ at $(t_0, x(t_0))$. We start with the following separable differential equation.

Example 330. Consider the differential equation

$$x' = x(1-t)$$

with initial condition $x(0) = 1$. Using separation of variables we can write

$$\int \frac{dx}{x} = \int (1-t)\,dt$$

or

$$\ln|x| = \frac{(1-t)^2}{2} + C.$$

Solving for x gives

$$x(t) = C\exp\left(\frac{(1-t)^2}{2}\right)$$

Since $x(0) = 1$, $C = \exp(-1/2)$ and

$$x(t) = \exp\left(\frac{t^2 - 2t}{2}\right).$$

Plotting this curve along with the tangent vectors to solutions of the differential equation at points in the plane gives Fig. 9.50 on page 323.

A plot of short line segments centered at (t, x) whose slopes are the slopes of the solutions to the differential equation, i.e. the slope is $f(t, x)$ is called a *slope field*. This is what is plotted in Fig. 9.50 on page 323.

We can get a crude approximation to the solution of a differential equation of the form $x' = f(t, x)$ in the tx-plane by plotting a slope field on a grid and then sketching a curve or curves that are tangent to the segments in the slope field plot.

Example 331. We plot the slope field and approximate solution curves satisfying $y(0) = 1$, $y(1) = 1/2$, and $y(0) = -1$ for the differential equation

$$\frac{dy}{dx} = \frac{x}{y^2 + 1}.$$

We need the slopes of some line segments for the slope field. For example at $(1,1)$ the slope of the line segment is $1/\left(1^1+1\right)=1/2$ and the slope at $(0,-1)$ is $-1/(0+1)=-1$. The plots in Fig. 9.51 on page 324 show the slope field and the slope field with the solution curves added.

Slope fields can also give information about how solutions behave over long periods of time.

Example 332. The logistic equation, see page 8.3.1, is an *autonomous* differential equation, one where the $f(t,P)$ does not depend on t. Because of this, looking at the slope field over a finite time interval can give information about solutions to the differential equation over long periods of time. The slope field in Fig. 9.52 on page 324 for the logistic equation $P'(t)=P(1-P)$ gives some very useful information even if we do not know the exact solutions.

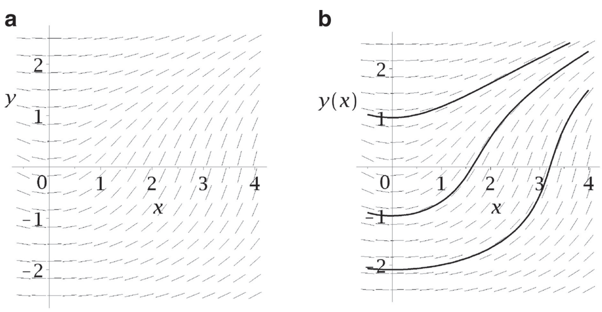
The slope field shows three basic things: if $P>1$ then $P'(t)<0$, if $P<0$ then $P'(t)<0$, and if $P\in(0,1)$ then $P'(t)>0$. This means that if $P(t)\in(0,1)$ for some t, then $P(t)$ is increase. In this case, $P(t)$ tends to $P=1$ as $t\to\infty$. Also, if $P(t)>1$, then $P(t)$ is decreasing. In fact, $P(t)$ tends to $P=1$ as $t\to\infty$. This means that any solution to the differential equation that starts near $P=1$ stays near $P=1$ as $t\to\infty$. In this case the steady state solution $P(t)=1$ to the differential equation is a *stable equilibrium* for the differential equation.

On the other hand, if a solution to the differential equation $P_1(t)$ is close but not equal to the solution $P(t)=0$, then $P_1(t)$ moves away from the solution $P(t)=0$. This means that $P(t)=0$ is an *unstable equilibrium* for the differential equation.

We next consider a simple method, Euler's method, for numerically approximating the solution to a differential equation with a given an initial condition. This method is close to calculating a Riemann sum using left endpoints.

Consider the differential equation

$$\frac{dx}{dt}=f(t,x) \tag{9.4}$$



The slope field for Example 331

The slope field and some solutions for Example 331

Fig. 9.51

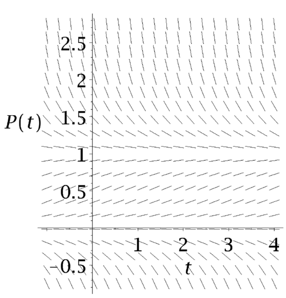


Fig. 9.52 The slope field for the logistic equation, Example 332

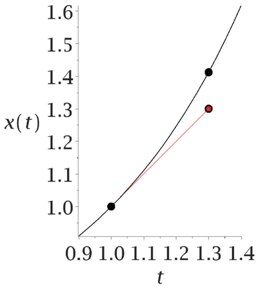


Fig. 9.53 One tangent line approximation for Euler's method

with initial condition $x(a) = x_0$. Fix a step size $h > 0$. Using the linear approximation

$$\begin{aligned} x(a+h) &\approx x(a) + x'(a)\,h \\ &\approx x_0 + f(a, x_0)\,h \end{aligned}$$

we get an approximation for $x(a+h)$. This is illustrated in Fig. 9.53 on page 325.

To find $x(a+k\,h)$ we simply repeat this tangent line approximation k times using the iteration

$$\begin{aligned} x_m &\approx x_{m-1} + x'(a+(m-1)h)\,h \\ &\approx x_{m-1} + f\left(a+(m-1)h, x_{m-1}\right)\,h. \end{aligned}$$

Note that there are two different approximations being put into the tangent line approximation, the approximation for $x(a+(m-1)\,h)$ and the approximation for $x'(a+(m-1)h)\,h$. Because of this, the error estimates for Euler's method in this form are not easy to derive and are left for later courses.

If we have a differential equation of the form $x'(t) = f(t)$, at a given t value we are not approximating $x'(t)$. We have the exact value since f does not depend on x. In this case the formula for the mth x_k is

$$x_m = x_0 + \sum_{i=0}^{m-1} f(t+i\,h)\,h.$$

The sum $\sum_{i=0}^{m-1} f(t+i\,h)\,h$ is the left endpoint Riemann sum for $f(t)$ on the interval from a to $a+m\,h$.

Example 333. Consider the differential equation $x'(t) = f(t) = \cos\left(\sqrt{t}\right)$ with $x(1) = -1$. We take 10 steps of Euler's method to approximate $x(4)$. This means that $h = (4-1)/10 = 0.3$. In Table 9.1 on page 325 are the x_k's from Eq. 9.7 for $i = 0, 1, 2, \ldots, 10$. The x_k's are rounded to three decimal places.

t	1	1.3	1.6	1.9	2.2	2.5	2.8	3.1	3.4	3.7	4.0
x_k	−1	−0.838	−0.713	−0.622	−0.565	−0.539	−0.542	−0.573	−0.629	−0.710	−0.814

Table 9.1 Euler's approximations for $x'(t) = f(t) = \cos\left(\sqrt{t}\right)$ with $x(1) = -1$

Figure 9.54 on page 326 shows how these values approximate the solution of the differential equation. In the figure the dots are the approximations and the solid line is the actual solution.

A basic format for the algorithm for Euler's method is contained in the following sequence of steps. Assume that we have a differential equation as in Eq. 9.4, an initial point t_0, an initial value $x_0 = x(t_0)$, a final point t_1 where we want to approximate $x(t_1)$, and a number of equal subintervals n to use.

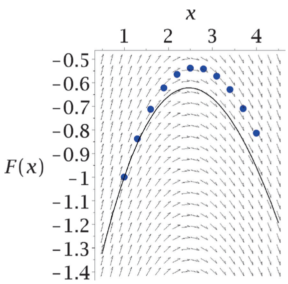


Fig. 9.54 The Euler's approximation and solution for Example 333

1. Set $h = (t_1 - t_0)/n$ and set (t_0, x_0) to be the initial point for the sequence to be constructed. Also set $i = 1$.
2. Repeat the following until $i = n + 1$.
   (a) Set
   $$x_i = x_{i-1} + hf(t_{i-1}, x_{i-1}).$$
   (b) Set
   $$t_i = t_{i-1} + h$$
   and record the point (t_i, x_i)
   (c) Set $i = i + 1$.
3. Return (t_n, x_n).

A couple examples will give a feel for how this works.

Example 334. This example shows how the Euler's method approximation can slowly drift away from an exact solution. Consider the differential equation

$$\frac{dy}{dx} = \frac{y}{\sqrt{x}} \qquad \text{and} \qquad y(1) = -2.$$

This is a separable differential equation with solution $y(x) = -2\exp\left(2\left(\sqrt{x} - 1\right)\right)$.

If we approximate $y(3)$ using 10 steps of Euler's method, we use $h = 1/5$ and get the approximations and the exact values in Table 9.2 on page 326. This is plotted in Fig. 9.55 on page 327.

x	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3
y_n	−2	−2.40	−2.84	−3.32	−3.84	−4.42	−5.04	−5.72	−6.46	−7.26	−8.13
$y(x)$	−2	−2.42	−2.89	−3.40	−3.96	−4.58	−5.26	−6.00	−6.81	−7.69	−8.65

Table 9.2 Euler's approximations for $dy/dx = y/\sqrt{x}$ with $y(1) = -2$

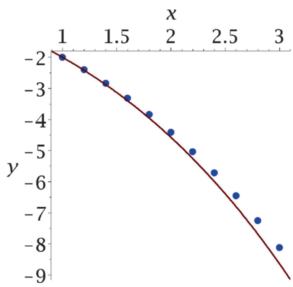


Fig. 9.55 The solution and Euler's approximation for Example 334

n	2	4	8	16	32
x_n	4.	5.063	5.960	6.583	6.959
E_n	3.389	2.327	1.429	0.8058	0.4304

Table 9.3 Euler approximations for $x(2)$ with n equal steps

Under certain conditions, the approximations from Euler's method will converge to the exact values of the solution of a differential equation as the number of equal intervals goes to infinity (reference?). The next example shows how the values from Euler's method can get better as n gets larger.

Example 335. We use the simple differential equation

$$\frac{dx}{dt} = x \quad \text{and} \quad x(0) = 1$$

as our example. We look at the approximations obtained for $x(2) = e^2$ for $n = 2, 4, 8, 16,$ and 32 intervals. The graph in Fig. 9.56 on page 328 shows these Euler approximations. The points are the values and the connecting lines are to emphasize which points are for each n. The the values from Euler's method at $t = 2$, x_n and the errors at $t = 2$, $E_n = x(2) - x_n$, are in Table 9.3 on page 327. The errors are decreasing and the ratio between the errors is getting close to $1/2$. This is the expected rate of decrease for the errors.

While it is true that Euler's method does converge to the solution, under fairly mild conditions over a region where the solution to the differential equation exists, it is also true that Euler's method has trouble in certain situations. We illustrate what can happen with Euler's method when a solution has an asymptote.

Example 336. To illustrate what happens when Euler's method is applied to a differential equation whose solutions have asymptotes we use the differential equation and initial condition

$$\frac{dy}{dx} = xy^2 \quad \text{and} \quad y(1) = 1.$$

This is a separable differential equation with solution

$$y(x) = \frac{2}{3 - x^2}.$$

(Check that this is the correct solution.) This solution has an asymptote at $x = \sqrt{3}$.

When we apply Euler's method to the differential equation on $[1, 1.875]$ with 7 equal steps of size $h = 1/8$, we get the table of values in Table 9.4 on page 328 and the graph in Fig. 9.57 on page 328.

As usually happens, Euler's method goes through the asymptote without any problem and gives unacceptable values after the asymptote.

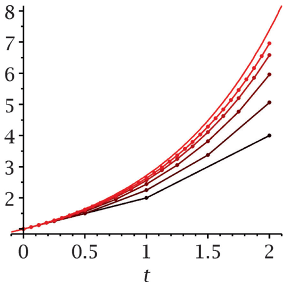


Fig. 9.56 Graphs of Euler's approximation for different numbers of points with the exact solution, Example 335

n	0	1	2	3	4	5	6	7
x_n	1	1.125	1.25	1.375	1.5	1.625	1.75	1.875
y_n	1.	1.125	1.303	1.568	1.991	2.734	4.253	8.209

Table 9.4 The values of the Euler approximation for Example 336

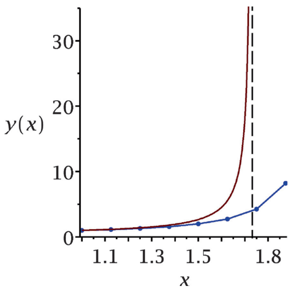


Fig. 9.57 The graph of the exact solution and the Euler approximation for Example 336. The Euler approximation goes through the asymptote

n	0	1	2	3	4	5	6	7	8	9	10
$x_{n,1}$	1	1.3	1.64	2.022	2.4476	2.9179	3.4333	3.9934	4.5968	5.2411	5.9223
$x_{n,2}$	2	2.1	2.18	2.234	2.2552	2.236	2.1678	2.0412	1.864	1.5709	1.2039

Table 9.5 Euler approximation values for $\mathbf{x}'(t) = \big(x_1(t)+x_2(t),\ x_2(t)-x_1(t)\big)$

An important thing to remember is that Euler's method does not converge quickly to the solution of a differential equation. For most situations we should not use Euler's method to approximate the solution to a differential equation, we should use other methods. Most numerical computation programs use good routines for approximating solutions to differential equations. Use one of these rather than Euler's method.

Euler's method and its analysis are the same for vector valued functions. This is not the case for most numerical methods for solving differential equations. Most algorithms for numerically solving vector valued differential equations must be analyzed carefully. The following is an example of Euler's method applied to a simple vector valued differential equation.

Example 337. Consider the differential equation $\mathbf{x}'(t) = \big(x_1(t)+x_2(t),\ x_2(t)-x_1(t)\big)$ with the initial condition $\mathbf{x}(0) = (1,2)$. This differential equation has solution

$$\mathbf{x}(t) = \Big(e^t\big(2\sin(t)+\cos(t)\big),\ -e^t\big(\sin(t)-2\cos(t)\big)\Big)\,.$$

(You should check that it really is the solution.)

The recursive formula for Euler's method is

$$\begin{aligned}\mathbf{x}_{i+1} &= \mathbf{x}_i + h\left(x_{i,1} + x_{i,2},\ x_{i,2} - x_{i,1}\right) \\ &= \left((1+h)x_{i,1} + hx_{i,2},\ -hx_{i,1} + (1+h)x_{i,2}\right) .\end{aligned}$$

Using a step size of $h = 0.1$ and ten steps gives us the approximations in Table 9.5 on page 328.

Exercises

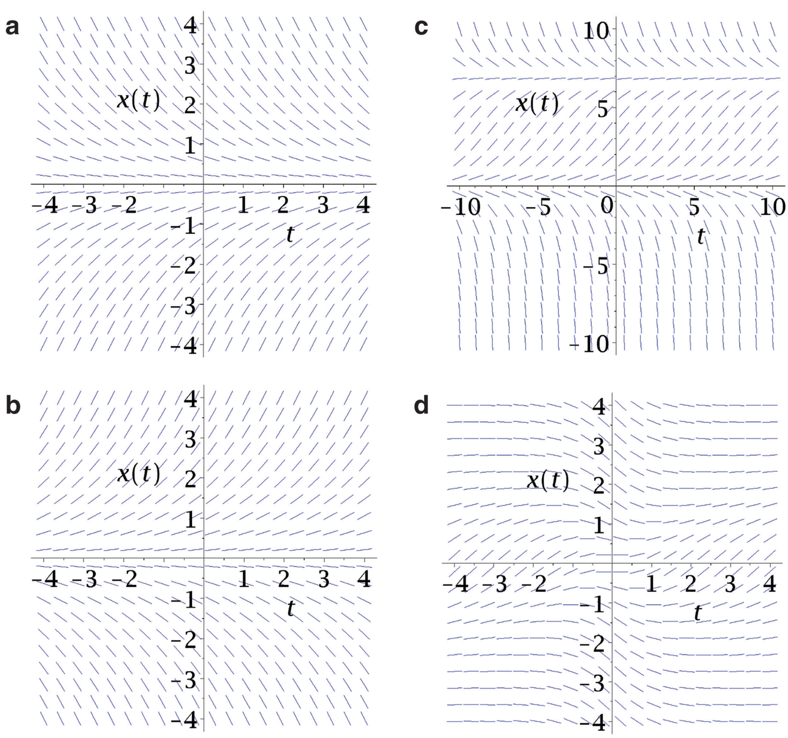
1. Show that each of the following functions is a solution to the associated differential equation.

(a) $x(t) = 4\cos(2t) - 3\sin(2t), \quad \dfrac{d^2x(t)}{dt^2} + 4x(t) = 0$

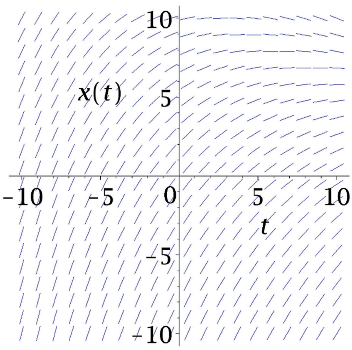
(b) $y(x) = 2xe^x + 4e^x + 2, \quad y'' - 2y' + y = 2$

(c) $z(t) = t + \dfrac{4}{t}, \quad z'(t) + \dfrac{z(t)}{t} = 2$

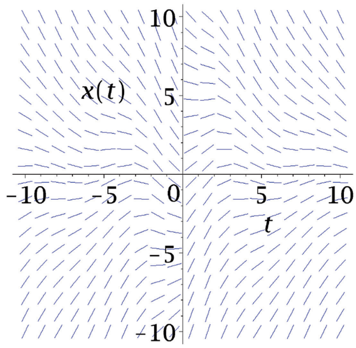
2. Sketch several solutions to the differential equation that has the slope field in each of the following graphs.



e



f



3. Use Euler's method with n equal steps on each of the following initial value problems to approximate $x(b)$. Evaluate the accuracy of your answer using a calculator or computer algebra system.

(a) $x' = x+t$, $x(0) = 1$, $b = 1$, $n = 5$
(b) $x' = x^2$, $x(1) = -1$, $b = 3$, $n = 7$
(c) $x' = x^2$, $x(1) = 2$, $b = 3$, $n = 7$
(d) $x' = \dfrac{x-t^2}{x^2+1}$, $x(-1) = 1$, $b = 2$, $n = 20$
(e) $x' = tx - t^2x^2$, $x(1) = -2$, $b = 5$, $n = 30$
(f) $x' = x(1{,}000 - x)$, $x(0) = 10$, $b = 20$, $n = 60$

4. Use Euler's method with n equal steps on each of the following initial value problems to approximate the function value at b. Evaluate the accuracy of your answer using a calculator or computer algebra system.

(a) $\mathbf{x}' = \big(x_1x_2,\ x_1+x_2\big)$, $\mathbf{x}(0) = (1,0)$, $b = 1$, $n = 5$
(b) $\mathbf{y}' = \big(-y_2,\ y_1+\cos(t)\big)$, $y(1) = (-1,1)$, $b = 3$, $n = 7$
(c) $\mathbf{x}' = \big(\sin(x_2)+\frac{1}{t},\ 2\sin(x_1)-\frac{1}{t}\big)$, $x(1) = (0,2)$, $b = 3$, $n = 12$
(d) $\mathbf{y}' = \left(\dfrac{y_1}{y_2},\ y_1+y_2-1\right)$, $x(-1) = 1$, $b = 2$, $n = 15$

5. Consider the differential equation $x'(t) = x(t)$ with initial condition $x(0) = 1$. Show that the approximation to $x(1)$ obtained through Euler's method with n equal steps going from $t = 0$ to $t = 1$ is

$$x_n = \left(1+\frac{1}{n}\right)^n.$$

In Sect. 7.4 you learned how to show that this value converges to e.

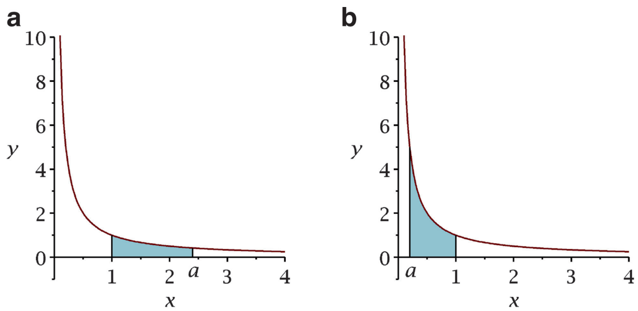
9.8 Ln(x) and Exp(x)

One of the things that the fundamental theorem of Calculus allows us to do is to formally define the natural logarithm and natural exponential functions. This is of theoretical interest and shows an interesting application of calculus.

We start with the natural logarithm. To define the natural logarithm we use the integral

$$\ln(x) = \int_1^x \frac{1}{t}\,dt \tag{9.5}$$

and then show that this function has all of the properties of a logarithm function. Since $1/t$ is continuous on $(0,\infty)$, the function $\ln(x)$ is a well defined differentiable function on $(0,\infty)$. There are a number of properties that follow from this definition. When considering these properties the graphs in Fig. 9.58 on page 331 may help. Figure 9.58a shows the area for $\ln(a)$ when $a > 1$ and Fig. 9.58b shows the area for $\ln(a)$ when $a \in (0,1)$. The area in Fig. 9.58b is negative since we are integrating from right to left.



The positive area under $y = 1/x$ for $a > 1$

The negative area under $y = 1/x$ for $a < 1$

Fig. 9.58

Some of the properties of $\ln(x)$ are stated in the following result.

Theorem 71 (Properties of the natural logarithm function). *Let* $\ln(x)$ *be defined as in Eq. 9.5 on page 331. Then the following properties are true.*

(i) *The domain of* $\ln(x)$ *is* $(0,\infty)$.

(ii) $\ln(1) = 0$

(iii) $\ln(x)$ *is an increasing function that is concave down.*

(iv) *The image of* $\ln(x)$ *is* $(-\infty,\infty)$.

To prove the last result in this theorem we will use the following proposition.

Lemma 5. *Let* $\ln(x)$ *be defined as above. Then, for any* $b > 1$,

$$\ln(b) = -\ln\left(\frac{1}{b}\right).$$

Proof. We use the change of variables $u = 1/x$ and $du = -1/x^2\,dx$ where $x = 1$ implies $u = 1$ and $x = b$ implies $u = 1/b$ to change the integral.

$$\begin{aligned}
\ln\left(\frac{1}{b}\right) &= \int_1^{\frac{1}{b}} \frac{1}{u}\,du \\
&= \int_1^{b} \frac{1}{\frac{1}{x}}\,\frac{-1}{x^2}\,dx
\end{aligned}$$

$$= -\int_1^b \frac{1}{x}\,dx$$
$$= -\ln(b).$$

Proof. (Of Theorem 71.)

(i) The domain of $\ln(x)$ is $(0,\infty)$.
Since $1/t$ is continuous on both $[1,b]$ and $[1/b,0]$ for any $b \geq 1$, the integral in Eq. 9.5 is well defined. Thus $\ln(x)$ is defined for all $x > 0$.

(ii) $\ln(1) = 0$
In this case $\int_1^1 1/t\,dt$ is defined as 0.

(iii) $\ln(x)$ is an increasing function that is concave down.
The derivative of $\ln(x)$ is, by the Fundamental Theorem of Calculus 65, $1/x > 0$ for all $x > 0$. This means that $\ln(x)$ is increasing. Taking the second derivative we get $\ln''(x) = -1/x^2$, a function that is negative for $x > 0$. This means that $\ln(x)$ is concave down on $(0,\infty)$.

(iv) The image of $\ln(x)$ is $(-\infty,\infty)$.
We show that $\lim_{b\to\infty}\ln(b) = \infty$. To show this we simply note that, for every $n > 0$ and every $t \in \left[2^{n-1}, 2^n\right]$, $1/t \geq 2^{-n}$. This means that

$$\int_{2^{n-1}}^{2^n} \frac{1}{t}\,dt \geq \int_{2^{n-1}}^{2^n} \frac{1}{2^n}\,dt$$
$$= \frac{1}{2}.$$

From this we find that $\lim_{n\to\infty}\ln\left(2^n\right) = \lim_{n\to\infty}\sum_{k=1}^n 1/2 = \lim_{n\to\infty} n/n = \infty$. Since $\ln(x)$ is an increasing continuous function, $\lim_{x\to\infty}\ln(x) = \infty$.
Since $\ln(1/b) = -\ln(b)$, we also have $\lim_{x\to 0}\ln(x) = -\infty$. The continuity of $\ln(x)$ implies that its image is $(-\infty,\infty)$.

We now proceed to show that the basic properties used for computing logarithms hold for the function $\ln(x)$.

Theorem 72 (Rules of logarithms). *Let* $\ln(x)$ *be defined by Eq. 9.5, then*

(i) $\ln(xy) = \ln(x) + \ln(y)$ *and*
(ii) $\ln(x^y) = y\ln(x)$.

Proof. We only prove the first statement. Consider the integral for $\ln(xy)$,

$$\ln(xy) = \int_1^{xy} \frac{1}{t}\,dt$$
$$= \int_1^x \frac{1}{t}\,dt + \int_x^{xy} \frac{1}{t}\,dt.$$

Using the substitution $t = xu$ and $dt = x\,du$ for the second integral gives us

$$\ln(xy) = \int_1^x \frac{1}{t}\,dt + \int_x^{xy} \frac{1}{t}\,dt$$
$$= \int_1^x \frac{1}{t}\,dt + \int_1^y \frac{1}{u}\,du$$
$$= \ln(x) + \ln(y).$$

To prove the second statement we can use the substitution $t = u^y$ and $dt = yu^{y-1}\,du$ in a similar manner.

We now define the function $\operatorname{Exp}(x)$ as the inverse of the function $\ln(x)$. Since the $\ln(x)$ function is strictly increasing with image $(-\infty,\infty)$, it has an inverse function that has domain $(-\infty,\infty)$. The derivative of the function $\operatorname{Exp}(x)$, using the derivatives of inverse functions from Sect. 4.4.1, is calculated as

$$\begin{aligned}\frac{d}{dx}\operatorname{Exp}(x) &= \frac{1}{\ln'(\operatorname{Exp}(x))}\\ &= \frac{1}{\frac{1}{\operatorname{Exp}(x)}}\\ &= \operatorname{Exp}(x).\end{aligned}$$

The basic properties of the natural exponential function follow directly from the properties of the natural exponential function.

Theorem 73. *The function* $\operatorname{Exp}(x)$ *is a continuous increasing function that is convex up on its domain.*

Proof. Since the domain of $\ln(x)$ is $(0,\infty)$, the range of $\operatorname{Exp}(x)$ is $(0,\infty)$. Given that the first and second derivatives of $\operatorname{Exp}(x)$ are $\operatorname{Exp}(x) > 0$, we see that the function is increasing and convex up.

We only need to prove the next theorem to have all of the basic properties of the natural exponential function.

Theorem 74 (Rules of exponents). *Let x and y be any two real numbers. Then*

(i) $\operatorname{Exp}(x+y) = \operatorname{Exp}(x)\operatorname{Exp}(y)$ *and*
(ii) $\operatorname{Exp}(x)^y = \operatorname{Exp}(xy)$.

Proof. Let a and b be numbers such that $x = \ln(a)$ and $y = \ln(b)$. Then we have

$$\begin{aligned}\operatorname{Exp}(x+y) &= \operatorname{Exp}(\ln(a)+\ln(b))\\ &= \operatorname{Exp}(\ln(ab))\\ &= ab\\ &= \operatorname{Exp}(x)\operatorname{Exp}(y).\end{aligned}$$

This proves the first statement.

To prove the second statement simply note that

$$\begin{aligned}\ln(\operatorname{Exp}(x)^y) &= y\ln(\operatorname{Exp}(x))\\ &= yx.\end{aligned}$$

Since $\ln(x)$ is one-to-one, this shows that $\operatorname{Exp}(x)^y = \operatorname{Exp}(xy)$.

If we let $e = \operatorname{Exp}(1)$ then $\operatorname{Exp}(x) = \operatorname{Exp}(1\cdot x) = \operatorname{Exp}(1)^x = e^x$. This is why we can use the notation $\exp(x) = e^x$ where e is Euler's number, $\operatorname{Exp}(1)$.

Exercises

1. Use Lemma 5 on page 331 and Theorem 72 on page 332, to show that, if $a, b > 0$,

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b).$$

2. Show that a function of the form $\ln(|x|) + C$ is not an antiderivative of $\frac{1}{|x|}$.
3. Show that

$$f(x) = \begin{cases} \ln(-x) + 5 & \text{if } x < 0 \\ \ln(x) - 2 & \text{if } x > 0 \end{cases}$$

is an antiderivative of $\frac{1}{x}$. Use this example to help explain why the functions $h(x) = \ln(|x|) + C$ are not the only antiderivatives of $\frac{1}{x}$.

4. Use the material in this section to show that, if $x, y > 0$, then

$$\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)\ .$$

5. Use the material in this section to show that, if $x, y > 0$, then

$$\exp(x - y) = \frac{\exp(x)}{\exp(y)}\ .$$

Chapter 10
Series

10.1 Review of Sequences

An important part of working with infinite series is understanding sequences and understanding the difference between series and sequences. Because of this, we start our study of series with a review of sequences. The material in this section follows the ideas in Sect. 2.1.

The definition is the same.

Definition 32. An *infinite sequence* is a function a from the natural numbers to a set. The values of this function are often denoted by $\{a_n\}_{n=1}^{\infty}$ instead of writing $a(n)$.

There are many examples of sequences in Sect. 2.1. We present a few example as a refresher.

Example 338. Let $a_n = \exp(-n) + \frac{n+2}{n}$ for $n = 1, 2, 3, \ldots$. The first five terms of this sequence are $a_1 = \frac{1}{e} + 3$, $a_2 = \frac{1}{e^2} + 2$, $a_3 = \frac{1}{e^3} + \frac{5}{3}$, $a_4 = \frac{1}{e^4} + \frac{3}{2}$, and $a_5 = \frac{1}{e^5} + \frac{7}{5}$.

Example 339. This sequence is defined by a recursive formula. Let $\mathbf{a}_1 = (2, -1)$ and $\mathbf{a}_2 = (-1, -3)$. Then set

$$\mathbf{a}_n = \mathbf{a}_{n-1} - \mathbf{a}_{n-2}.$$

The first five terms of this sequence are $\mathbf{a}_1 = (2, -1)$, $\mathbf{a}_2 = (-1, -3)$, $\mathbf{a}_3 = (-3, -2)$, $\mathbf{a}_4 = (-2, 1)$, and $\mathbf{a}_5 = (1, 3)$.

The definition of an infinite sequence converging is the same as in Sect. 2.1.

Definition 33. A sequence $\{\mathbf{a}_n\}_{n=1}^{\infty}$ *converges* to $\mathbf{L}$ if, given any distance $r > 0$, there is an N, such that when $n \geq N$, $\mathbf{a}_n$ is within the distance r of $\mathbf{L}$. If this condition is not satisfied, the sequence *diverges*.

As a reminder, here are two examples.

Example 340. The sequence $b_n = (n^2 + 1)/(n^2 + 4n + 55)$ converges to $L = 1$. Dividing the numerator and denomination of each element in the sequence by n^2, this sequence can be rewritten as

$$b_n = \frac{1 + \frac{1}{n^2}}{1 + \frac{4}{n} + \frac{55}{n^2}}. \tag{10.1}$$

Since c/n and c/n^2 go to zero as n goes to ∞ for any constant c, the numerator and denominator in Eq. 10.1 both go to 1 as n goes to ∞. The limit is now easily seen to be $L = 1$.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0_10

Example 341. The sequence $c_n = \cos(n\pi)$ does not converge. We can use Theorem 12 on page 38 since $|c_n - c_{n+1}| = \left|\cos(n\pi) - \cos\left((n+1)\pi\right)\right| = 2 > 0$ for all $n \geq 1$. The theorem immediately tells us that the sequence diverges.

It is important to remember that there are sequences where the differences between the consecutive terms go to 0, but the sequence diverges.

Example 342. Consider the sequence a_n defined by $a_{n+1} = a_n + 1/2^{k(n)}$ with $a_1 = 1$. Here $k(n) = m$ when $n \in [2^m, 2^{m+1})$. All n's must satisfy this condition for some m and if $n_1 > 2^m$ then $1/2^{k(n)} \leq 1/2^m$. Since $1/2^m$ goes to 0 as m goes to ∞, $|a_{n+1} - a_n| \leq 1/2^{k(n)}$ goes to 0 as n goes to ∞.

However, there are 2^m values of n in $[2^m, 2^{m+1})$, so

$$\begin{aligned} a_{2^{m+1}-1} - a_{2^m} &= \sum_{i=1}^{2^m} \frac{1}{2^m} \\ &= 1. \end{aligned}$$

This means that, using Theorem 12 on page 38, the sequence diverges.

We will see similar examples when we talk about series.

The basic rules saying that term wise sums and scalar multiples of sequences converge to the desired quantities when the limits exist. This is Theorem 9 on page 36.

Theorem 75. *(Theorem 9) Let $\{\mathbf{a}_n\}$, $\{\mathbf{b}_n\}$, and $\{c_n\}$ be sequences that converge to $\mathbf{K}$, $\mathbf{L}$, and M respectively. Also, let s be a number and assume that $M \neq 0$. Then the following hold:*

(i) $\lim_{n\to\infty} (\mathbf{a}_n + \mathbf{b}_n) = \mathbf{K} + \mathbf{L}$

(ii) $\lim_{n\to\infty} s\mathbf{a}_n = s\mathbf{K}$

(iii) $\lim_{n\to\infty} \mathbf{a}_n \cdot \mathbf{b}_n = \mathbf{K} \cdot \mathbf{L}$

(iv) $\lim_{n\to\infty} \dfrac{\mathbf{a}_n}{c_n} = \dfrac{\mathbf{K}}{M}$

Another important result states that a sequence of vectors converges if and only if each of the component sequences converges. This is Theorem 13 on page 38. The following is one example of this.

Example 343. Consider the sequence

$$\mathbf{a}_n = \left(\frac{1}{n}, \exp(-n), \sin\left(\frac{\pi n}{2}\right)\right).$$

Since $\sin(\pi(4k+1)/2) = 1$ for all positive integers k and $\sin(\pi(4m-1)/2) = -1$ for all positive integers m, the sequence with terms $\sin(\pi n/2)$ does not converge. Therefore the sequence with terms $\mathbf{a}_n$ does not converge.

There is a method for finding the limits of sequences that relies on limits of functions. In this case we rely on the limit of a function $f(x)$ as x goes to ∞ to find the limit of a sequence. The following theorem gives us the setting.

Theorem 76. *Let $\{\mathbf{a}_n\}_{n=1}^{\infty}$ be a sequence and let $\mathbf{f}(x)$ be a function defined on some interval (m, ∞) such that $\mathbf{f}(n) = a_n$ after some N. If the limit $\lim_{x\to\infty} f(x) = \mathbf{L}$, then the limit $\lim_{n\to\infty} a_n$ exists and equals $\mathbf{L}$.*

Proof. By definition, since $\lim_{x\to\infty} f(x) = \mathbf{L}$, for all sequences $b_n \to \infty$ we have $\lim_{n\to\infty} \mathbf{f}(b_n) = \mathbf{L}$. Setting $b_n = n$ gives us that

$$\lim_{n\to\infty} \mathbf{f}(n) = \lim_{n\to\infty} a_n = \mathbf{L}\ .$$

Example 344. Consider the sequence $b_n = \arctan(n)$. Since $\lim_{x\to\infty} \arctan(x) = \pi/2$, we have $\lim_{n\to\infty} b_n = \pi/2$.

Using functions as in Theorem 76, we can apply L'Hôpital's rule to find the limits of sequences.

Example 345. Consider the sequence $a_n = n^2 \exp(-n)$. This sequence is the values of the function $f(x) = x^e \exp(x)$ at the integers. To take the limit of the function we use L'Hôpital's rule.

$$\begin{aligned}\lim_{x\to\infty} \frac{x^2}{\exp(x)} &\overset{(L)}{=} \lim_{x\to\infty} \frac{2x}{\exp(x)} \\ &\overset{(L)}{=} \lim_{x\to\infty} \frac{2}{\exp(x)} \\ &= 0\ .\end{aligned}$$

This means that the sequence converges to 0.

It is not true that if a sequence with $a_n = f(n)$ converges that the limit of $f(x)$ as x goes to infinity exists. Consider the function $f(x) = \cos(2\pi x)$. Here $f(n) = 1$ for all integers n, but $\lim_{x\to\infty} \cos(2\pi x)$ does not exist.

Exercises

1. Do the following sequences converge? If a sequence converges, what is the limit? Justify your answers.

(a) $a_n = \dfrac{n+1}{n-2}$

(b) $b_m = \dfrac{m^2+1}{m-2m^2}$

(c) $c_k = \dfrac{k-k^3}{4k^3+100}$

(d) $a_n = \dfrac{(-1)^n (n+1)}{5n+5}$

(e) $b_m = \dfrac{3^m}{4^m}$

(f) $c_n = \dfrac{4^n}{n^3+3^n}$

(g) $r_n = \ln(\ln(n))$ for $n = 8, 9, 10, \ldots$

(h) $s_m = \dfrac{m!}{4^m}$

(i) $w_k = \dfrac{\sqrt[8]{k^2+1}}{\sqrt[4]{k+2}}$

(j) $a_n = \dfrac{e^n - 3^n}{2^n + \sqrt{n!}}$

2. Do the following sequences converge? If they converge, find the limit. If they do not converge, explain why they do not converge.

(a) $\mathbf{a}_n = \left(\dfrac{1}{n}, \dfrac{2n}{n^2+2}\right)$

(b) $\mathbf{b}_m = \left(2^{-m}, \cos(\pi m)\right)$

(c) $\mathbf{c}_k = \left(\sin\left(\dfrac{\pi}{k}\right), \cos\left(\dfrac{\pi}{k^2}\right)\right)$

(d) $\mathbf{r}_t = \left(\frac{1}{t}, \frac{t^2+t}{t-t^2}\right)$

(e) $\mathbf{a}_n = \left(\frac{e^n}{n^3}, \frac{n^3}{e^n}\right)$

(f) $\mathbf{b}_m = \big(\ln(m), \exp(-m), \cos(2\pi m)\big)$

(g) $\mathbf{a}_n = \left(\sin(\pi n), \left(1+\frac{1}{n}\right)^2, \exp\left(\frac{n+1}{n+2}\right)\right)$

(h) $\mathbf{c}_k = \left(\tan\left(\frac{1}{k}\right), \arctan(k), \frac{1}{1+k^2}\right)$

(i) $\mathbf{r}_t = \left(\frac{t^2}{t-t^2}, \frac{t^2+2^t}{t^3+2^{t/2}}, \frac{e^t+1}{1+5^t}\right)$

(j) $\mathbf{m}_s = \left(\frac{s}{e^s}, \frac{s^3}{10+(-s)^s}, \cos\left(\pi s + \frac{1}{s}\right)\right)$

3. Use the squeeze theorem, Theorem 11, to show that the following sequences converge.

(a) $a_n = 3 + \frac{(-1)^n}{n^2}$

(b) $b_m = \frac{2+(-1)^m}{2m}$

(c) $c_k = \frac{\lfloor k \rfloor}{k^2+k}$

(d) $c_k = \frac{\cos(k)}{\sqrt{k}}$

(e) $b_m = \frac{m+1}{m^2+4}$

(f) $a_n = \frac{1}{n} + \frac{(-1)^n}{n^2}$

4. Evaluate the following limits of sequences.

(a) $a_n = \frac{n}{2^n}$

(b) $b_m = \left(1+\frac{1}{m}\right)^m$

(c) $c_k = \left(1+\frac{1}{3k}\right)^k$

(d) $c_k = \ln(k)\exp(-k)$

(e) $b_m = \frac{\exp(m)}{\sqrt{m}}$

(f) $a_n = \frac{\cos(x)}{x}$

10.2 Definitions of Series and Convergence

An *infinite series*, is the "sum of an infinite sequence" taken in the order of the indices. For example, if we add the terms of the sequence defined by $a_n = 1/n$ we write

$$\sum_{n=1}^{\infty} \frac{1}{n}.$$

It is not clear what is meant by this infinite sum. The following definition is what is meant by an infinite series.

Definition 34 (Infinite series). The sum of an infinite sequence, *an infinite series*, is defined as

$$\sum_{n=1}^{\infty} \mathbf{a}_n = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{a}_n.$$

This definition holds even if the limit on the right does not exist. The sum of the sequence can start at any value to get an infinite series, for example,

$$\sum_{n=53}^{\infty} \mathbf{a}_n \ .$$

For a concrete example we can return to Zeno's paradox from Sect. 2.1. In that example a person was walking across a room where the distance traveled in the nth time interval was $a_n = 1/2^n$. Since we had an infinite number of time intervals, the total distance traveled is, as calculated in Sect. 2.1,

$$\begin{aligned} D &= \sum_{n=1}^{\infty} \frac{1}{2^n} \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{2^n} \\ &= \lim_{N\to\infty} \left(1 - \frac{1}{2^N}\right) \\ &= 1. \end{aligned}$$

We now state the formal definition of convergence of a series.

Definition 35 (Series convergence). Let $\{\mathbf{a}_n\}_{n=m}^{\infty}$ be a sequence of vectors. The *infinite series*

$$\sum_{n=m}^{\infty} \mathbf{a}_n$$

is defined by the limit

$$\sum_{n=m}^{\infty} \mathbf{a}_n = \lim_{N\to\infty} \sum_{n=m}^{N} \mathbf{a}_n.$$

If the limit exists, the series is said to *converge*. When the limit does not exist, the series *diverges*.

Instead of working directly with the definition we will often work with partial sums as sequences. The Nth *partial sum* of the series $\sum_{n=1}^{\infty} \mathbf{a}_n$ is the sum of the first N terms of the series,

$$\mathbf{S}_N = \sum_{i=1}^{N} \mathbf{a_i}.$$

Example 346. Consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n\,(n+1)}.$$

By partial fractions decomposition we have

$$\frac{1}{n\,(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

This means that the partial sums are

$$S_N = \sum_{n=1}^{N} \frac{1}{n\,(n+1)}$$

$$
\begin{aligned}
&= \left(\frac{1}{1}-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{N-1}-\frac{1}{N}\right)+\left(\frac{1}{N}-\frac{1}{N+1}\right)\\
&= \frac{1}{1}+\left(-\frac{1}{2}+\frac{1}{2}\right)+\left(-\frac{1}{3}+\frac{1}{3}\right)+\cdots+\left(-\frac{1}{N}+\frac{1}{N}\right)-\frac{1}{N+1}\\
&= 1-\frac{1}{N+1}.
\end{aligned}
$$

Calculating the limit of the $S_N{}'$s gives

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)} = \lim_{N\to\infty}\left(1-\frac{1}{N+1}\right) = 1.$$

The equivalence between the converge of the series and the convergence of the partial sums is important and is stated as a theorem.

Theorem 77 (Convergence and partial sums). *A series $\sum_{i=m}^{\infty}\mathbf{a}_i$ converges if and only if the sequence of partial sums $\mathbf{S}_N = \sum_{i=m}^{m+N}\mathbf{a}_i$ converges.*

The next example is the case of geometric series. This is a case where we can always get an exact value for the sum of a series, if the series converges.

Example 347 (Geometric series). We consider a series of the form

$$\sum_{i=0}^{\infty} a^i$$

where a is a fixed real number and work with the Nth partial sum

$$S_N = \sum_{i=1}^{N} a^i = 1+a+a^2+\cdots+a^{N-1}+a^N.$$

Multiplying both sides of this equation by $1-a$ yields

$$
\begin{aligned}
(1-a)\,S_N &= 1+a+a^2+\cdots+a^{N-1}+a^N\\
&\qquad -a-a^2-\cdots-a^{N-1}-a^N-a^{N+1}\\
&= 1-a^{N+1}.
\end{aligned}
$$

Thus $S_N = \left(1-a^{N-1}\right)/(1-a)$. There are four cases to consider.

1. If $|a|<1$, then $\lim_{N\to\infty} a^N = 0$ and taking the limit as N goes to ∞ of S_N gives

$$
\begin{aligned}
\sum_{i=1}^{\infty} a^n &= \lim_{N\to\infty}\frac{1-a^{N+1}}{1-a}\\
&= \frac{1}{1-a}.
\end{aligned}
$$

2. When we assume that $|a|>1$, the difference between the Nth and $N+1$st terms is

$$|S_{N+1}-S_N| = |a^{N+1}-a^N| = |a^N|\,|1-a| > |1-a|.$$

By Theorem 12 on page 38 the sequence of partial sums diverges and therefore the series diverges.

3. When $a_n = 1$, then $S_N = N$ for all N and the series diverges.

4. When $a_n = -1$, then $S_N = 1$ for even N and $S_N = 0$ for all odd N. The series diverges since $|S_{N+1} - S_N| = 1$ for all N.

Using the example of a general geometric series we can discuss the convergence of the series $\sum_{i=0}^{\infty} 1/3^i$ and of the series $\sum_{i=0}^{\infty} (-2)^i$.

Example 348. The series $\sum_{i=0}^{\infty} 1/3^i$ can be rewritten as the geometric series $\sum_{i=0}^{\infty} (1/3)^i$. Since the a here is $1/3 \in (-1,1)$, the series converges. Calculating the value we get

$$\sum_{i=0}^{\infty} \left(\frac{1}{3}\right)^i = \frac{1}{1-\frac{1}{3}} = \frac{3}{2}.$$

Example 349. The series $\sum_{i=0}^{\infty} (-2)^i$ is a geometric series with $a = -2$. Since $|a| > 1$, the series diverges. The first few partial sums for this series are $S_0 = 1$, $S_1 = -1$, $S_2 = 3$, $S_3 = -5$, $S_4 = 11$, and $S_5 = -22$.

We can restate Theorem 12 on page 38 when applied to the partial sums of series by simply noting that $|S_{N+1} - S_N| = |a_{N+1}|$ and get the following result. It is important to note that this result does not say that $|a_n| \to 0$ implies that the series converges. In the next section it will become clear that the converse of this theorem is not true.

Theorem 78. *Let $\sum_{i=m}^{\infty} \mathbf{a}_i$ be an infinite series such that $\lim_{i\to\infty} \mathbf{a}_i$ does not exist or does not equal $\mathbf{0}$. Then the series does not converge.*

Example 350. Consider the series $\sum_{i=m}^{\infty} \sin(n\pi/2)$. Since

$$\sin((4k+1)\pi/2) = \sin(\pi/2) = 1$$

for any integer k, the difference between consecutive partial sums does not go to zero and the series does not converge.

As with sequences of vectors, series of vectors can be considered coordinate by coordinate. This is simply because a partial sum of vectors is a vector of partial sums. If $\mathbf{a}_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,n})$ and $S_{i,n} = \sum_{k=m}^{m+N} a_{i,N}$ then

$$\begin{aligned} \mathbf{S}_N &= \sum_{m}^{m+N} \mathbf{a}_i \\ &= \left(\sum_{m}^{m+N} a_{i,1}, \sum_{m}^{m+N} a_{i,2}, \ldots, \sum_{m}^{m+N} a_{i,n}\right) \\ &= (S_{1,N}, S_{2,N}, \ldots, S_{n,N}). \end{aligned}$$

We can now rewrite Theorem 13 on page 38 for series.

Theorem 79 (Convergence of vector valued series). *Let $\sum_{i=m}^{\infty} \mathbf{a}_i$ be a series with $\mathbf{a}_i$ having n components. Then $\sum_{i=m}^{\infty} \mathbf{a}_n = \mathbf{L}$ if and only if each of the component series of $\sum_{i=m}^{\infty} \mathbf{a}_i$ converges to the corresponding component of $\mathbf{L}$. Writing this out gives $\lim_{n\to\infty} \sum_{i=m}^{n} a_{i,j} = \mathbf{L}_j$.*

Example 351. Let $\mathbf{a}_i = \left(1/2^i, 1/(-2)^i\right)$. Each component of this series is a geometric series. Since the a values both have absolute values less than 1, each component series converges. Therefore the series converges.

Example 352. Let $\mathbf{a}_i = \left((2/3)^i, (3/2)^i\right)$ for $i = 0, 1, 2, \ldots$. In this case the second component series is $\sum_{i=0}^{\infty} (3/2)^i$, a geometric series with $|a| = 3/2 > 1$. This means the second component of the series diverges and the whole series diverges.

The next result concerns computations with series. This is the equivalent of Theorem 75 on page 336. The proof of this result involves rewriting the series in terms of limits of partial sums. For example, if $\sum_{i=m}^{\infty} a_i$ and $\sum_{i=m}^{\infty} b_i$ are two series then a partial sum of $\sum_{i=m}^{\infty}(a_i + b_i)$ can be written as follows.

$$\begin{aligned} S_N &= \sum_{i=m}^{m+N} (a_i + b_i) \\ &= \sum_{i=m}^{m+N} a_i + \sum_{i=m}^{m+N} b_i. \end{aligned}$$

This is the sum of partial sums for the series $\sum_{i=m}^{\infty} a_i$ and $\sum_{i=m}^{\infty} b_i$. From this we can conclude that, if the partial sums for the series $\sum_{i=m}^{\infty} a_i$ and $\sum_{i=m}^{\infty} b_i$ converge, then so do the partial sums for $\sum_{i=m}^{\infty}(a_i + b_i)$.

Theorem 80 (Computations with series). *Let $\sum_{i=m}^{\infty} \mathbf{a}_i$ and $\sum_{i=m}^{\infty} \mathbf{b}_i$ be series that converge to $\mathbf{K}$ and $\mathbf{L}$. Also, let s be a number. Then the following hold:*

(i) $\displaystyle\sum_{i=m}^{\infty} (\mathbf{a}_i + \mathbf{b}_i) = \mathbf{K} + \mathbf{L}$

(ii) $\displaystyle\sum_{i=m}^{\infty} (s\,\mathbf{a}_i) = s\mathbf{K}$

Example 353. Consider the series $\sum_{i=m}^{\infty} a^i$ for a fixed m. If $|a| < 1$ then the series $\sum_{i=0}^{\infty} a^i$ converges and thus, since $a^{m+i} = a^m a^i$,

$$\begin{aligned} \sum_{i=m}^{\infty} a^i &= \sum_{k=0}^{\infty} a^m a^k \\ &= a^m \sum_{k=0}^{\infty} a^k \\ &= \frac{a^m}{1-a}. \end{aligned}$$

In the case when $a = -1/4$ this means that

$$\begin{aligned} \sum_{i=1}^{\infty} \left(-\frac{1}{4}\right)^i &= -\frac{1}{4} \sum_{k=0}^{\infty} \left(-\frac{1}{4}\right)^k \\ &= \frac{-\frac{1}{4}}{1+\frac{1}{4}} \\ &= -\frac{1}{5}. \end{aligned}$$

Example 354. The series

$$\sum_{n=1}^{\infty} \left(\frac{2}{n\,(n+1)} + (0.67)^n\right)$$

converges since the series $\sum_{n=1}^{\infty} 1/(n\,(n+1))$ converges, see Example 346 on page 339, and the series $\sum_{n=1}^{\infty} (0.67)^n$ is a convergent geometric series.

An important idea when considering series is that of the sum of the "tail" of a series $\sum_{n=0}^{\infty}\mathbf{a}_n$. This is the sum $\sum_{n=N}^{\infty}\mathbf{a}_n$ for an arbitrarily large N. It is equivalent in some ways to the tail part of an integral of the form

$$\int_a^{\infty} f(x)\,dx.$$

This relationship is considered in the next section. For the moment we only want the following result.

Theorem 81 (Sums of tails of series). *Let $\sum_{n=0}^{\infty}\mathbf{a}_n$ be an infinite series. This series converges if and only if the sums of the tails of the series converge to* $\mathbf{0}$, *i.e.*

$$\lim_{N\to\infty}\sum_{n=N}^{\infty}\mathbf{a}_n = \mathbf{0}.$$

Example 355. Consider the series $\sum_{n=1}^{\infty} 2^n/5^n$. The tails for this series sum to

$$\begin{aligned}\sum_{n=m}^{\infty}\frac{2^n}{5^n} &= \left(\frac{2}{5}\right)^m \sum_{k=0}^{\infty}\left(\frac{2}{5}\right)^k\\ &= \left(\frac{2}{5}\right)^m \frac{5}{3}.\end{aligned}$$

Since the sums of these tails go to zero, the series converges.

Exercises

1. Do the following series converge? If a series converges, what is the limit? Justify your answers.

(a) $\displaystyle\sum_{n=1}^{\infty}\frac{4}{3^n}$

(b) $\displaystyle\sum_{n=1}^{\infty}\frac{4^n}{3^n}$

(c) $\displaystyle\sum_{n=0}^{\infty}\frac{3^n+5}{3^n}$

(d) $\displaystyle\sum_{m=1}^{\infty}\frac{3^m}{4^m}$

(e) $\displaystyle\sum_{k=1}^{\infty}\left(\frac{-1}{3}\right)^k$

(f) $\displaystyle\sum_{p=0}^{\infty}\sin\left(\frac{\pi p}{2}\right)$

(g) $\displaystyle\sum_{n=0}^{\infty}\frac{5}{n(n+1)}$

(h) $\displaystyle\sum_{m=1}^{\infty}\frac{1}{m(m+1)(m+2)}$

(i) $\displaystyle\sum_{k=1}^{\infty}\left(\left(\frac{-1}{5}\right)^k+\left(\frac{4}{3}\right)^k\right)$

(j) $\displaystyle\sum_{p=0}^{\infty}\left(\left(\frac{1}{2}\right)^p-\left(\frac{-4}{5}\right)^p\right)$

2. Do the following series converge? If a series converges, what is the limit? Justify your answers.

(a) $\displaystyle\sum_{n=1}^{\infty}\left(\frac{4}{3^n},-\frac{10}{2^n}\right)$

(b) $\displaystyle\sum_{n=1}^{\infty}\left(\frac{4^n}{3^n},\frac{2^n}{3^n}\right)$

(c) $\displaystyle\sum_{n=0}^{\infty}\left(\frac{3^n+5}{3^n},\frac{1}{n(n+1)}\right)$

(d) $\displaystyle\sum_{m=1}^{\infty}\left(\frac{3^m}{4^m},\frac{1}{m(m+1)}\right)$

(e) $\sum_{k=1}^{\infty}\left((-1)^k, \left(\frac{-2}{3}\right)^k\right)$

(f) $\sum_{p=0}^{\infty}\left(\sin(\pi p), \frac{(-1)^p}{3^{\frac{p}{2}}}\right)$

(g) $\sum_{p=0}^{\infty}\left(\left(\frac{1}{2}\right)^p, \left(\frac{-4}{5}\right)^p\right)$

(h) $\sum_{k=1}^{\infty}\left(\left(\frac{-1}{5}\right)^k, \left(\frac{4}{3}\right)^k\right)$

3. A decimal expansion of a number between 0 and 1, $z = 0.a_1a_2a_3\ldots$, can be viewed as the infinite sum

$$z = \sum_{i=1}^{\infty} \frac{a_i}{10^i}.$$

Use this representation to show that $z = 0.99\overline{9}$ equals 1.

4. We can use Geometric series to convert repeating decimals into rational numbers. For example, the repeating decimal $x = 1.81818\overline{18}$ can be rewritten as

$$\begin{aligned} 1.81818\overline{18} &= \frac{1}{10}\sum_{n=0}^{\infty} \frac{18}{100^n} \\ &= \frac{9}{5}\sum_{n=0}^{\infty} \frac{1}{100^n} \\ &= \frac{9}{5}\frac{1}{1-\frac{1}{100}} \\ &= \frac{180}{99}. \end{aligned}$$

Convert the following repeating decimals to fractions.

(a) $0.1212\overline{12}$

(b) $123.123\overline{123}$

(c) $0.2424\overline{24}$

(d) $5.9898\overline{98}$

(e) $1.6551\overline{51}$

(f) $57.1573\overline{73}$

5. Assume that an infinite series $\sum_{n=0}^{\infty} b_n$ converges. The Nth remainder for the series is the sum of the *tail* of the series after the first N terms

$$R_N = \sum_{n=N+1}^{\infty} b_n.$$

Show that $\lim_{N\to\infty} R_N = 0$. (Hint: What is $S_N + R_N$?)

10.3 The Integral Test

In the last section infinite series were defined, their convergence was defined and some simple results were given. In this section we move on to tests that tell us if a series converges or diverges. There are more cases where we cannot find the sum of a convergent series than cases where we can find the sum of a convergent series. For a large number of situations, approximating the sum of a series is good enough. When working in that situation, knowing that a series converges is very important.

It is interesting to realize that almost all of the tests for convergence of series, and power series, are comparisons between the series of interest and a series that we already know converges or diverges. Most of the tests will involve a comparison with geometric series, Example 347. In this section the integral test is presented. Here the sum of an infinite series is compared with the integral of a function. This uses what we already know about improper integrals, Sect. 8.7. This test works for some cases when the rest of our tests do not work.

Before starting the work on tests for convergence it is important to have a result about increasing sequences in our collection of tools. This will be used implicitly in many of the results in the next few sections without being explicitly mentioned. Remember that results about sequences are used for series through partial sums.

Theorem 82. *Let $\{a_n\}_{n=m}^{\infty}$ be a sequence of real numbers that is increasing , $a_{n+1} \geq a_n$, and is bounded above, i.e. there is a K such that $a_n < K$ for all $n \geq m$. Then the sequence $\{a_n\}_{n=m}^{\infty}$ converges to some L less than or equal to K,*

$$\lim_{n\to\infty} a_n = L \leq K.$$

Similarly, if $\{b_n\}_{n=0}^{\infty}$ is a sequence of real number that is decreasing ($b_n \geq b_{n+1}$) that is bounded below ($b_n \geq M$ for some real M), then the sequence B_n converges to some L with $L \geq M$.

Example 356. Consider the series with terms $a_n = 1/\left(2n^2\right)$ for $n = 1,2,3,\ldots$. From Example 346 on page 339 we know that the series with terms $b_n = 1/(n(n+1))$ for $n = 1,2,3,\ldots$ converges.

Noting that $n+1 \leq 2n$ and thus $1/\left(2n^2\right) \leq 1/(n(n+1))$ for $n = 1,2,3,\ldots$ we see that the partial sums maintain this order,

$$S_N = \sum_{n=1}^{N} \frac{1}{2n^2} \leq \sum_{n=1}^{N} \frac{1}{n(n+1)} = 1 - \frac{1}{N+1} \leq 1.$$

Since the S_N's are increasing and bounded above by 1, they converge and $\sum_{n=1}^{\infty} 1/\left(2n^2\right)$ converges.

Before starting with the integral test we look at how integrals of functions on intervals of the form $[c,\infty)$ are linked to specific series. We assume that we have a function $f(x)$ defined on $[c,\infty)$ that is integrable on $[c,\infty)$ and that $f(x) \geq 0$ on $[c,\infty)$. Note that $F(y) = \int_c^y f(x)\,dx$ is an increasing function and for any number $b \geq c$, $\int_b^{b+1} f(x)\,dx \geq 0$. If we define a sequence for integers $n \geq 0$ by

$$a_n = \int_{c+n}^{c+n+1} f(x)\,dx,$$

the Nth partial sum is $S_N = \int_c^{c+N} f(x)\,dx$. The series $\sum_{n=1}^{\infty} a_n$ converges to the same value as $\int_c^{\infty} f(x)\,dx$ or both the series and the integral diverge to infinity. This relationship between convergence of the integral and convergence of the series is what we use in the integral test.

Example 357. To generate a series from an integral on $[0,\infty)$ we use the function

$$f(x) = \lfloor x \rfloor \exp(-x).$$

On the interval $[n,n+1]$ this function is $f_n(x) = n\exp(-x)$ and the integral over the interval is

$$\begin{aligned} a_n &= \int_n^{n+1} n\exp(-x)\,dx \\ &= n\left(\exp(-n) - \exp(-n-1)\right) \\ &= n\left(1 - \exp(-1)\right)\exp(-n). \end{aligned}$$

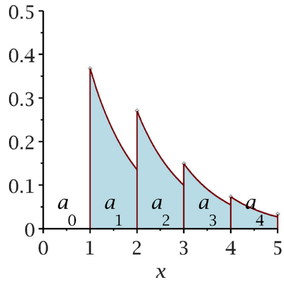
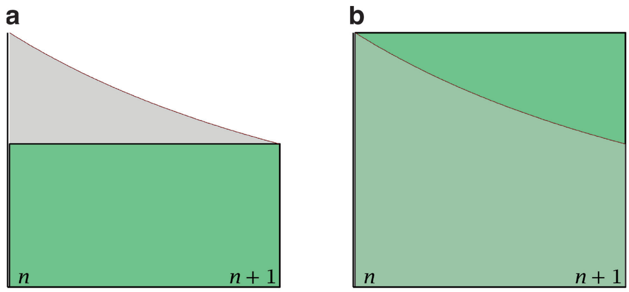


Fig. 10.1 A graph of $f(x) = \lfloor x \rfloor \exp(-x)$ with areas under the curve



A rectangle below a decreasing function from n to $n = 1$

A rectangle above a decreasing function from n to $n = 1$

Fig. 10.2

This is illustrated in Fig. 10.1 on page 346.

We will discuss the convergence of this series in Example 359.

We now proceed to the integral test for convergence of series. The idea is the same as in Example 356. We assume that the summation starts with a_1. Assume that $\sum_{i=1}^{\infty} a_i$ is a series with the a_i's decreasing and greater than or equal to 0. In addition assume that there is a decreasing function $f(x)$ such that $f(n) = a_n$ for all n greater than some N. On the interval from n to $n+1$ we have pictures like those in Fig. 10.2 on page 346.

Since $f(x) \le f(n)$ on the interval from n to $n+1$, we have

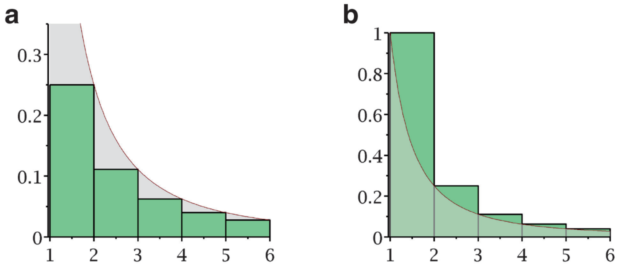
$$b_n = \int_n^{n+1} f(x)\,dx \le \int_n^{n+1} a_n\,dx = a_n$$

and

$$c_n = \int_n^{n+1} f(x)\,dx \ge \int_n^{n+1} a_{n+1}\,dx = a_{n+1}.$$

See Fig. 10.3 on page 347. The first of these inequalities means that

$$\sum_{i=1}^{n} a_i \ge \int_1^{n+1} f(x)\,dx, \tag{10.2}$$



A series of rectangles below a decreasing function from n to $n = 1$

A series of rectangles above a decreasing function from n to $n = 1$

Fig. 10.3

or, in the limit, when we assume that the limits exist,

$$\sum_{i=1}^{\infty} a_i \geq \int_1^{\infty} f(x)\,dx. \tag{10.3}$$

In a similar manner, using the second inequality, we have that

$$\sum_{i=2}^{\infty} a_i \leq \int_1^{\infty} f(x)\,dx.$$

Since $f(x) \geq 0$ for all $x \geq 1$ and $a_i \geq 0$ for all i, all of the limits exist or all of the limits are equal to infinity.

The above yields the first part of the following theorem, the integral test. The second and third parts of the theorem simply extend inequalities (10.2) and (10.3) to the cases when $a_n > f(x)$ on $[n, n+1]$ and when $a_{n+1} < f(x)$ on $[n, n+1]$.

Theorem 83 (Integral Test). *Assume that there is a nonnegative function $f(x)$ on $[m, \infty]$, m an integer, such that $f(x)$ is decreasing after some α and assume that $f(x)$ is integrable on $[m, \beta]$ for any $\beta > m$. Let $a_n = f(n)$ for $n = m, m+1, m+1, \ldots$. Then the series $\sum_{n=m}^{\infty} a_n$ converges if and only if the integral $\int_m^{\infty} f(x)\,dx$ is finite.*

Assume that $f(x)$ satisfies the above conditions and we have $a_n \in [0, f(n)]$. If the integral $\int_m^{\infty} f(x)\,dx$ is finite, then the series $\sum_{n=m}^{\infty} a_n$ converges.

Assume that $f(x)$ satisfies the above conditions and we have $a_n \geq f(n)$. If the integral $\int_m^{\infty} f(x)\,dx$ is infinite, then the series $\sum_{n=m}^{\infty} a_n$ diverges.

The last two statements of the theorem can be viewed as saying that an area contained in a finite area is finite and that an area containing an infinite area is infinite.

Here are some simple illustrations of this method. It is important to note that the functions must be integrable. This heavily restricts the functions we can use for the integral test.

Example 358. Consider the series $a_n = \exp(-n)$ for $n = 0, 1, 2, \ldots$. The function $f(x) = \exp(-x)$ is positive and decreasing on $\mathbb{R}$ with $f(n) = a_n$. We can apply Theorem 83 on page 347 since

$$\begin{aligned}\int_0^{\infty} f(x)\,dx &= \lim_{b \to \infty} \int_0^b \exp(-x)\,dx \\ &= \lim_{b \to \infty} -\exp(-x)\Big|_{x=0}^{b}\end{aligned}$$

$$\begin{aligned}
&= \lim_{b\to\infty} (1 - \exp(-b)) \\
&= 1.
\end{aligned}$$

The series converges because the integral is finite.

Example 359. We now return to Example 357. In that series the terms are $a_n = n\,(1-\exp(-1))\exp(-n)$. Since each a_n is less than $n\exp(-n)$ and $f(x) = x\exp(-x)$ is a decreasing function after $x = 1$ (Show this is true.), we only need to show that $\int_1^\infty x\exp(-x)\,dx$ is finite to show that the series $\sum_{n=1}^\infty a_n$ converges.

The integral is

$$\begin{aligned}
\int_1^\infty x\exp(-x)\,dx &= \lim_{b\to\infty} \int_1^b x\exp(-x)\,dx \\
&= \lim_{b\to\infty} (-x\exp(-x) - \exp(-x))\Big|_{x=1}^{b} \\
&= 2e.
\end{aligned}$$

(Fill in the steps using integration by parts and L'Hôpital's rule.) This shows that the series converges.

Example 360. We consider the series $a_n = 1/n^2$. For the function $f(x) = 1/x^2$ we have $f(n) = a_n$, $f(x) > 0$ if $x > 0$, and $f(x)$ is decreasing if $x > 0$ since $f'(x) = -1/x^3 < 0$. This means the hypotheses of Theorem 83 on page 347 are satisfied. Integrating $f(x)$ from 1 to ∞ gives

$$\begin{aligned}
\int_1^\infty f(x)\,dx &= \lim_{b\to\infty} \int_1^b \frac{1}{x^2}\,dx \\
&= \lim_{b\to\infty} -\frac{1}{x}\Big|_{x=1}^{b} \\
&= \lim_{b\to\infty} \left(1 - \frac{1}{b}\right) \\
&= 1.
\end{aligned}$$

The series converges.

Example 361. Now we look at the series $\sum_{m=1}^\infty 1/m$. The decreasing positive function for this series is $f(x) = 1/x$. It satisfies the hypotheses of Theorem 83 on page 347. Integrating this function from 1 to ∞ gives

$$\begin{aligned}
\int_1^\infty f(x)\,dx &= \lim_{b\to\infty} \int_1^b \frac{1}{x}\,dx \\
&= \lim_{b\to\infty} \ln(x)\Big|_{x=1}^{b} \\
&= \lim_{b\to\infty} (\ln(b) - 0) \\
&= \infty.
\end{aligned}$$

The series diverges.

Example 362. If we consider the series $a_n = \exp(-n)/n$ for $n = 1,2,3,\ldots$. The function $f(x) = \exp(-x)$ is positive and decreasing on $\mathbb{R}$ with $f(n) \geq a_n$ for all $n \geq 1$. From Example 358 on page 347 we know that $\int_1^\infty \exp(-x)\,dx$ is finite. This implies that the series converges.

We now consider a general case that includes Examples 360 on page 348 and 361 on page 348. The situation we consider is the case of *p-series*. A p-series is an infinite series of the form

$$\sum_{i=1}^{\infty} \frac{1}{n^p}.$$

If $p \leq 0$, the term are n^q with $q \geq 0$. This means that $a_n = n^q > 1$ for all n and the series diverges since the terms do not go to 0.

We also know from Example 361 on page 348 that if $p = 1$, the series diverges. This leaves the cases when $p > 1$ and when $p \in (0,1)$. For both of these cases we can apply the integral test with $f(x) = 1/x^p$ since these $f(x)$'s are positive decreasing functions. When $p \in (0,1)$ the integral of $f(x)$ from 1 to ∞ is infinity. (Check this yourself.) In addition, when $p > 1$ the integral of $f(x)$ from 1 to ∞ is finite. (Check this yourself.) Combining all of this information gives the following result.

Theorem 84 (p-series test). *Let p be a real number and consider the p-series*

$$\sum_{n=1}^{\infty} \frac{1}{n^p}.$$

This series converges if $p > 1$ and the series diverges if $p \leq 1$.

Example 363. Let $a_n = \frac{-6}{n^{3/2}}$. By Theorem 80 on page 342, the series $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} \frac{1}{n^{3/2}}$ converges. Since $\sum_{n=1}^{\infty} \frac{1}{n^{3/2}}$ is a p-series with $p > 1$, the series converges.

Example 364. Let $a_n = 4/n^{1/2}$. By Theorem 80 on page 342, the series $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} 1/n^{1/2}$ converges. Since $\sum_{n=1}^{\infty} 1/n^{1/2}$ is a p-series with $p = 1/2 < 1$, the series diverges.

We conclude this section by showing how to estimate the difference between a partial sum of a convergent series and the sum of the convergent series. Here we use the idea of the tail of a series or a sequence. The *tail* of a series or of a sequence is all of the terms (elements) after some N.

Again, we assume that there is a nonnegative function $f(x)$ that is decreasing on $[M,\infty)$ such that $f(n) = a_n \geq 0$ if $n \geq M$. If we fix an integer $N > M$, we can write, assuming the series indices start at 1,

$$\sum_{n=1}^{\infty} a_n = S_N + \sum_{n=N+1}^{\infty} a_n.$$

From the discussion before the statement of the integral test we know that the sum of the tail of the series $R_N = \sum_{n=N+1}^{\infty} a_n$ satisfies

$$\int_{N+1}^{\infty} f(x)\,dx \leq R_N \leq \int_N^{\infty} f(x)\,dx.$$

Adding S_N to all three terms in these inequalities gives

$$S_N + \int_{N+1}^{\infty} f(x)\,dx \leq \sum_{n=1}^{\infty} a_n \leq S_N + \int_N^{\infty} f(x)\,dx.$$

This gives an interval estimate for the sum of the series.

Example 365. We look at how close the 50th partial sum of $\sum_{i=1}^{\infty} 1/n^2$ is to the value of the series. To 10 digits we have

$$\sum_{i=1}^{50} \frac{1}{n^2} \approx 1.625132734.$$

The integral from $N > 0$ to infinity of $f(x) = 1/x^2$ is

$$\begin{aligned}\int_N^{\infty} f(x)\,dx &= \lim_{b\to\infty} \int_N^b \frac{1}{x^2}\,dx \\ &= \lim_{b\to\infty} -\frac{1}{x}\bigg|_{x=N}^{b} \\ &= \lim_{b\to\infty} \left(\frac{1}{N} - \frac{1}{b}\right) \\ &= \frac{1}{N}.\end{aligned}$$

We use this value at $N = 50$ and $N = 51$ to calculate that, to 10 decimal digits,

$$1.644740577 \le \sum_{i=1}^{\infty} \frac{1}{n^2} \le 1.645132734.$$

We can use the midpoint of this interval as an estimate to get

$$\sum_{i=1}^{\infty} \frac{1}{n^2} = 1.644936656 \pm 0.000196079.$$

The actual value of the sum is $\pi^2/6 \approx 1.644934067$.

Exercises

1. Do the following series converge?

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{4n}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{4}{3n^2}$

(c) $\displaystyle\sum_{n=0}^{\infty} \frac{n}{n^2+5}$

(d) $\displaystyle\sum_{m=1}^{\infty} \frac{15}{7\sqrt{m}}$

(e) $\displaystyle\sum_{k=1}^{\infty} k\,e^{-k^2}$

(f) $\displaystyle\sum_{p=2}^{\infty} \frac{1}{p\ln(p)}$

(g) $\displaystyle\sum_{n=2}^{\infty} \frac{n}{n^2+1}$

(h) $\displaystyle\sum_{m=0}^{\infty} \frac{1}{m^2+4m+8}$

(i) $\displaystyle\sum_{n=2}^{\infty} \frac{n}{n+1}$

(j) $\displaystyle\sum_{m=0}^{\infty} \frac{m^3}{m^2+4m+8}$

(k) $\displaystyle\sum_{n=0}^{\infty} \frac{5^n}{10^n}$

(l) $\displaystyle\sum_{m=1}^{\infty} -\frac{7^{m-2}}{2^m}$

(m) $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2}$

(n) $\displaystyle\sum_{n=1}^{\infty} \frac{\arctan(n)}{n^2}$

2. Do the following series converge?

(a) $\sum_{n=1}^{\infty}\left(\frac{4}{3^n}, -\frac{10}{n^{\frac{3}{2}}}\right)$

(b) $\sum_{n=1}^{\infty}\left(2^{-n}, \frac{4^n}{3^n}\right)$

(c) $\sum_{n=0}^{\infty}\left(n^{\frac{-3}{2}}, n^{\frac{3}{2}}\right)$

(d) $\sum_{m=3}^{\infty}\left(\frac{3^m}{4^m}, \frac{1}{m\ln^2(m)}\right)$

(e) $\sum_{k=1}^{\infty}\left(\left(\frac{-1}{3}\right)^k, \frac{-2}{3\sqrt{k}}\right)$

(f) $\sum_{p=0}^{\infty}\left(p^2 e^{-p}, p e^{-p^2}\right)$

3. This is a simple example to show that the requirement that $a_n \geq 0$ in the integral test is important. Consider the series with terms a_n for $n = 1, 2, 3, \ldots$ where

$$a_n = \begin{cases} \frac{1}{n^2} & \text{if } n \text{ is odd} \\ \frac{-1}{n} & \text{if } n \text{ is even.} \end{cases}$$

Let $f(x) = \frac{1}{x^2}$ and show that $a_n \leq f(n)$ for $n = 1, 2, 3, \ldots$ with $\int_1^{\infty} f(x)\,dx < \infty$. Then show, by separating the series into positive and negative parts, that the series does not converge.

4. This is an example to show that infinite series may converge even though the hypotheses of the integral test do not hold. Consider the series with terms a_n for $n = 1, 2, 3, \ldots$ where

$$a_n = \begin{cases} \frac{1}{n} & \text{if } m \text{ is } 2^k \text{ for some } k \\ \frac{1}{n^2} & \text{otherwise.} \end{cases}$$

First conclude that series does not satisfy the hypotheses of the integral test. Rewrite this series as the sum of series $\sum_{m=0}^{\infty} \frac{1}{2^m}$ and a series whose mth term is in $\left[0, \frac{1}{m^2}\right]$. Use this to conclude that the series converges. (If you are ambitious, you can show that for any decreasing integrable $g(x)$ with $g(n) \geq a_n$, $\int_1^{\infty} g(x)\,dx = \infty$.)

5. Use an area argument to justify the following theorem.

Theorem 85. *Assume that there is a nonnegative function $f(x)$ on $[m, \infty]$, m an integer, such that $f(x)$ is decreasing after some α and assume that $f(x)$ is integrable on $[m, \beta]$ for any $\beta > m$. Assume that $a_n \in [0, f(n)]$ for $n = m, m+1, m+1, \ldots$. If the integral $\int_m^{\infty} f(x)\,dx$ is finite, then the series $\sum_{n=m}^{\infty} a_n$ converges.*

6. Use the integral test to show that

$$\sum_{n=0}^{\infty}\left(\frac{7}{8}\right)^n$$

converges.

10.4 The Comparison Tests

The idea used in the previous section was that by comparing the area represented by a series to the area represented by an integral we can get information about the convergence of the series. In this section that idea is extended to comparing series with series. Assume that a series $\sum_{i=1}^{\infty} a_i$

with non-negative terms converges to S and that a series $\sum_{i=1}^{\infty} b_i$ satisfies $a_i \geq b_i \geq 0$. Let A_n and B_n be the partial sums for the series. We have

$$B_n \leq A_n \leq S$$

and B_n is an increasing sequence. This means that the sequence $\{B_n\}$ is a bounded increasing sequence and hence the series $\sum_{i=1}^{\infty} b_i$ converges.

In a similar manner, if $\sum_{i=1}^{\infty} a_i$ diverges and $a_i \leq b_i$, then $\sum_{i=1}^{\infty} b_i$ diverges. This follows from the fact that a region with area that contains a region with infinite area has infinite area. These two statements give us the basic comparison test.

Theorem 86 (Comparison test). *Let $\sum_{i=1}^{\infty} a_i$ be a series with non-negative terms and let $\sum_{i=1}^{\infty} b_i$ be another series with non-negative terms.*

(i) If $b_i \leq a_i$ after some N and $\sum_{i=1}^{\infty} a_i$ converges, then $\sum_{i=1}^{\infty} b_i$ converges.
(ii) If $b_i \geq a_i$ after some N and $\sum_{i=1}^{\infty} a_i$ diverges, then $\sum_{i=1}^{\infty} b_i$ diverges.

Remark 28. It is *very important* to note that the condition that $b_i \leq a_i$ after some N can be used for the integral test or any other test for convergence of a series. This condition can be viewed as saying that the behavior of the tail of the series, $\sum_{i=N+1}^{\infty} a_i$ for arbitrary N, determines the convergence of a series.

In other words, given a positive integer N, the first N terms of any series do not affect the convergence of the series.

We now look at several examples showing how we use the comparison test. The first example demonstrates how to show that a series converges using the comparison test.

Example 366. We know from Example 358 on page 347 that the series $\sum_{n=0}^{\infty} e^{(-n)}$ converges. The series $\sum_{n=0}^{\infty} 1/(e^n+3)$ converges since $e^{(-n)} \geq 1/(e^n+3)$ for all n.

This can be extend to say that any series of the form $\sum_{n=0}^{\infty} C/(e^n+g_n)$ where $C > 0$ and $g_n \geq 0$ converges. This is done by using Theorem 80 on page 342 to show that $\sum_{n=0}^{\infty} C/e^n$ converges and noting that $C/(e^n+g_n) \leq C/e^n$.

The second example shows how we can demonstrate that a series diverges using the comparison test.

Example 367. If we compare the series $\sum_{m=1}^{\infty} 1/(\sqrt{m}+10)$ to the harmonic series $\sum_{n=1}^{\infty} 1/n$, we can show the first series diverges. Since $\sqrt{n}+10 \leq n$ when $n \geq 14$ (check this), we have $1/(\sqrt{n}+10) \geq 1/n$ when $n \geq 14$. Since $\sum_{n=1}^{\infty} 1/n$ diverges, $\sum_{m=1}^{\infty} 1/(\sqrt{m}+10)$ diverges.

The last example gives a case where the comparison test does not apply.

Example 368. Consider the series $\sum_{n=1}^{\infty} 1/n^{3/2}$ and $\sum_{m=1}^{\infty} 1/\sqrt{m}$. These are both p-series with the first converging, $p > 1$, and the second diverging, $p < 1$. If $n \geq 1$ we have $m^{3/2} \geq m \geq \sqrt{m}$, or $1/m^{3/2} \leq 1/\sqrt{m}$. The inequality is in the wrong direction to conclude anything about the convergence of $\sum_{m=1}^{\infty} 1/\sqrt{m}$ given that $\sum_{n=1}^{\infty} 1/n^{3/2}$ converges. It is also in the wrong direction to conclude anything about the convergence of $\sum_{n=1}^{\infty} 1/n^{3/2}$ given that $\sum_{m=1}^{\infty} 1/\sqrt{m}$ diverges.

We now look at a slightly less general comparison test where finding the N in Theorem 86 on page 352 is not required. The idea is that if, given two non-negative series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ and given that

$$r = \lim_{n\to\infty} \frac{a_n}{b_n} > 0$$

exists, we have $a_n < 2rb_n$ after some N and $a_n > rb_n/2$ after some M. Because $\sum_{n=0}^{\infty} 2rb_n$ and $\sum_{n=0}^{\infty} rb_n/2$ converge if and only if $\sum_{n=0}^{\infty} b_n$ converges, these two inequalities mean that if

$\sum_{n=1}^{\infty} b_n$ converges then $\sum_{n=1}^{\infty} a_n$ converges and that if $\sum_{n=1}^{\infty} a_n$ converges then $\sum_{n=1}^{\infty} b_n$ converges. This is the main part of the following theorem.

The same reasoning shows that if $r > 0$ is finite then $\sum_{n=1}^{\infty} b_n$ diverges to infinity if and only if $\sum_{n=1}^{\infty} a_n$ diverges to infinity.

Theorem 87 (Limit comparison test). *Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be two series with non-negative terms. If*

$$r = \lim_{n\to\infty} \frac{a_n}{b_n}$$

exists and is greater than zero, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges. If $r = 0$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.

It is also true that if $r > 0$, then $\sum_{n=1}^{\infty} a_n$ diverges if and only if $\sum_{n=1}^{\infty} b_n$ diverges.

This is used in many situations. A common case is when we have terms that are similar but are changed in a regular manner. First is an example where terms of the form $C/\left(an^2+bn+c\right)$ are compared with $1/n^2$.

Example 369. Take the series with terms $a_n = 3/\left(2n^2+10n+100\right)$. This can be compared to the series with terms $b_n = 1/n^2$, a convergent series, using the limit comparison test. The limit is

$$\begin{aligned} r &= \lim_{n\to\infty} \frac{a_n}{b_n} \\ &= \lim_{n\to\infty} \frac{\frac{3}{2n^2+10n+100}}{\frac{1}{n^2}} \\ &= \frac{3}{2} \lim_{n\to\infty} \frac{1}{1+\frac{10}{n}+\frac{100}{n^2}} \\ &= \frac{3}{2}. \end{aligned}$$

Since $r > 0$, the series $\sum_{n=0}^{\infty} 3/\left(2n^2+10n+100\right)$ converges.

Now we compare two series where the limit of the ratio goes to zero.

Example 370. The series we are working with are $\sum_{n=1}^{\infty} 1/n^3$, which converges and $\sum_{n=1}^{\infty} 1/n!$. If we assume that $n \geq 4$, then

$$\begin{aligned} r &= \lim_{n\to\infty} \frac{\frac{1}{n!}}{\frac{1}{n^3}} \\ &= \lim_{n\to\infty} \frac{n^3}{(n-3)!\,(n-2)\,(n-1)\,n} \\ &= \lim_{n\to\infty} \left(\frac{1}{(n-3)!}\,\frac{n}{n-2}\,\frac{n}{n-1}\,\frac{n}{n}\right) \\ &\leq \lim_{n\to\infty} \frac{2}{n-3} \\ &= 0. \end{aligned}$$

Since $r = 0$, the series $\sum_{n=1}^{\infty} 1/n!$ converges.

Next we look at two series that diverge where the limit comparison test makes it easy to show that one of the series diverges.

Example 371. Since the series $\sum_{i=1}^{\infty} 1/\sqrt[3]{i}$ is a p-series with $p < 1$, the series diverges. We can compare the series $\sum_{i=100}^{\infty} 20/\left(\sqrt[3]{5i}+\sqrt[6]{i}\right)$. Taking the limit of the ratio of corresponding terms gives an $r > 0$.

$$\begin{aligned} r &= \lim_{i\to\infty} \frac{\frac{20}{\sqrt[3]{5i}+\sqrt[6]{i}}}{\frac{1}{\sqrt[3]{i}}} \\ &= 20 \lim_{i\to\infty} \left(\frac{1}{\frac{\sqrt[3]{5i}}{\sqrt[3]{i}} + \frac{\sqrt[6]{i}}{\sqrt[3]{i}}} \right) \\ &= 20 \lim_{i\to\infty} \left(\frac{1}{\sqrt[3]{5} + \frac{1}{\sqrt[6]{i}}} \right) \\ &= \frac{20}{\sqrt[3]{5}}. \end{aligned}$$

Since the series $\sum_{i=1}^{\infty} 1/\sqrt[3]{i}$ diverges, so does the series $\sum_{i=100}^{\infty} 20/\left(\sqrt[3]{5i}+\sqrt[6]{i}\right)$.

The next example shows how limit values of 0 and ∞ may not give any information.

Example 372. Consider the convergent series $\sum_{i=0}^{\infty} a_i = \sum_{i=0}^{\infty} 1/2^i$ and the divergent series $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} 1/n$. In this case we have

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0$$

and

$$\lim_{n\to\infty} \frac{b_n}{a_n} = \infty.$$

The limit values of 0 and ∞ give us no information on the convergence of the series.

In the first limit we are simply noting that the terms of the divergent series are growing faster than the terms of the convergent series. Again, in the second limit we are seeing that the terms of the divergent series grow faster than the terms of the convergent series.

There are many series where the limit comparison test does not help. This is especially true when we have many terms in both series that are zero. When that happens there are terms where the ratio is not defined.

Example 373. Consider the series $\sum_{i=1}^{\infty} a_i$ and $\sum_{i=1}^{\infty} b_i$ with terms

$$a_i = \begin{cases} 0 & \text{if } i \text{ is even} \\ \frac{1}{i^2} & \text{if } i \text{ is odd} \end{cases}$$

and

$$b_i = \begin{cases} 0 & \text{if } i \text{ is odd} \\ \frac{1}{i^2} & \text{if } i \text{ is even} \end{cases}.$$

Both series converge since $1/i^2 \geq a_i \geq 0$ and $1/i^2 \geq b_i \geq 0$ for all $i \geq 1$. However, the ratio $a_i/b_i = 0$ for all even i and the ratio $b_i/a_i = 0$ for all odd i. This means the limits of the ratios do not exist. The ratio test does not apply to this pair of series.

The final result in this section is an extension of the comparison test for series with nonnegative terms to comparing a series whose terms can have any sign with a series whose terms are all nonnegative. In Sect. 10.2 we looked at geometric series and showed that, for example, the series

$$\sum_{n=0}^{\infty}\left(-\frac{1}{2}\right)^n$$

converges. This series has terms that alternate signs between positive and negative. As stated the comparison test does not apply to this type of series.

We use the fact that for any finite sum of terms a_n we have

$$\left|\sum_{n=M}^{N} a_n\right| \leq \sum_{n=M}^{N} |a_n|\,.$$

In terms of partial sums of the series $\sum_{n=0}^{\infty} a_n$ this means that, for $M > N$,

$$|S_M - S_N| \leq \sum_{n=N+1}^{M} |a_n|\ .$$

If the series $\sum_{n=0}^{\infty} |a_n|$ converges, for any $r > 0$ there is a K such that if $k \geq K$,

$$\sum_{n=k}^{\infty} |a_n| < r\ .$$

This means that if $M, N > K$,

$$|S_M - S_N| \leq \sum_{n=K}^{\infty} |a_n| < r\ .$$

At this point, we need to use the concept of a Cauchy sequence and a theorem about convergence of Cauchy sequences, which we do not have. Since those are beyond this class, we simply assume that condition represented in Eq. 10.4 on page 355 implies the convergence of the sequence $\{S_N\}$.

Restating this, a sequence $\{\mathbf{c}_n\}_{n=0}^{\infty}$ is a Cauchy sequence if, for all $r > 0$ there is an N such that if m and n are greater than N, then $\|\mathbf{c}_m - \mathbf{c}_n\| < r$. The result states that any Cauchy sequence in $\mathbb{R}^k$ converges.

In our setting, the S_n's are a Cauchy sequence and must converge. Since the S_n's are our partial sums for $\sum_{n=0}^{\infty} a_n$, the series $\sum_{n=0}^{\infty} a_n$ converges. The statement of this result is the following theorem.

Theorem 88 (Comparison test 2). *Assume that the series $\sum_{n=0}^{\infty} |a_n|$ converges. Then the series $\sum_{n=0}^{\infty} a_n$ converges.*

Example 374. Consider the series

$$\sum_{k=1}^{\infty} \frac{(-1)^{\lfloor\sqrt{k}\rfloor}}{k^2}. \tag{10.4}$$

This series has both positive and negative terms, so our first comparison test and the integral test do not apply. However, we can compare this series with the nonnegative series $\sum_{n=1}^{\infty} 1/n^2$, which converges. Since

$$\left|\frac{(-1)^{\lfloor\sqrt{k}\rfloor}}{k^2}\right| = \frac{1}{k^2},$$

Theorem 88 applies and the series in Eq. 10.4 converges.

A series $\sum_{n=0}^{\infty} a_n$ such that $\sum_{n=0}^{\infty} |a_n|$ converges is called an *absolutely convergent* series.

Exercises

1. Use a comparison test to show that the following series converge or to show that the series diverge.

(a) $\sum_{n=1}^{\infty} \frac{1}{4n+3}$

(b) $\sum_{n=1}^{\infty} \frac{4}{3n^2-1}$

(c) $\sum_{n=3}^{\infty} \frac{n}{n^2-5}$

(d) $\sum_{m=1}^{\infty} \frac{15}{7\sqrt{m+10}}$

(e) $\sum_{k=1}^{\infty} (k+10)\, e^{-k^2}$

(f) $\sum_{p=3}^{\infty} \frac{1}{(p-2)\ln(p)}$

(g) $\sum_{n=2}^{\infty} \frac{n}{n^2+4n+11}$

(h) $\sum_{m=0}^{\infty} \frac{1}{m^2+4m+11}$

(i) $\sum_{n=2}^{\infty} \frac{n}{n+1}$

(j) $\sum_{m=0}^{\infty} \frac{m^3}{m^2+4m+8}$

(k) $\sum_{n=3}^{\infty} \frac{5^n}{10^n-11}$

(l) $\sum_{m=1}^{\infty} -\frac{7^{m-2}}{2^m}$

(m) $\sum_{k=1}^{\infty} \frac{2^k}{k\,3^{k+2}}$

(n) $\sum_{p=10}^{\infty} \frac{5^p}{4^p+3}$

2. Determine if the following series converge or diverge.

(a) $\sum_{n=1}^{\infty} \frac{n+2}{n^2+4n-2}$

(b) $\sum_{n=3}^{\infty} \frac{n^2+3}{4-n^2}$

(c) $\sum_{n=0}^{\infty} \frac{4^n}{n^4+20^{\frac{n}{2}}}$

(d) $\sum_{m=1}^{\infty} \frac{(-1)^m}{2^{m/2}}$

(e) $\sum_{k=1}^{\infty} \frac{(-1)^{\lfloor \sqrt[3]{k} \rfloor}}{k^3}$

(f) $\sum_{p=0}^{\infty} \frac{p^3}{(-1)^p}$

3. Determine if the following series converge or diverge.

(a) $\sum_{n=1}^{\infty} \left(\frac{n+2}{n^3+4n-2}, \frac{n}{2^n}\right)$

(b) $\sum_{n=3}^{\infty} \left(\frac{n+10}{n!}, \frac{n+3}{4-n^2}\right)$

(c) $\sum_{k=2}^{\infty} \left(\left(\frac{4}{k}\right)^k, \frac{k^4}{k^4+20^{\frac{k}{3}}}\right)$

(d) $\sum_{m=1}^{\infty} \left(\frac{(-1)^m}{2^{m/2}}, \frac{100}{m^{5/4}}\right)$

(e) $\sum_{k=1}^{\infty} \left(\frac{k}{10+k^{11/5}}, \frac{(-1)^{\lfloor \sqrt[3]{k} \rfloor}}{k^3}\right)$

(f) $\sum_{p=0}^{\infty} \left(0.97^p, \frac{p+247}{(-1)^p}\right)$

4. Assume

$$a_n = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{2}{n^{\frac{3}{2}}} & \text{if } n \text{ is odd.} \end{cases}$$

Does $\sum_{n=1}^{\infty} a_n$ converge?

5. Assume

$$b_n = \begin{cases} \frac{1}{n} & \text{if } n \text{ is a multiple of } 4 \\ 0 & \text{otherwise.} \end{cases}$$

Does $\sum_{n=1}^{\infty} b_n$ converge? (Hint: Replace n by $4m$ for the nonzero terms.)

6. Consider the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ that has terms with alternating signs.
   (a) By combining consecutive terms, a_{2n-1} and a_{2n}, show that this series converges.
   (b) Explain why this convergent series is not absolutely convergent.

7. We can show that any series that is absolutely convergent is convergent without using Cauchy sequences.

   Given an absolutely convergent series $\sum_{n=1}^{\infty} a_n$, define two series. First set $b_n = a_n$ if $a_n > 0$ and set $b_n = 0$ is $a_n \le 0$. Next, set $c_n = 0$ if $a_n \ge 0$ and $c_n = -a_n$ if $a_n < 0$.
   (a) By comparing $\sum_{n=1}^{\infty} b_n$ and $\sum_{n=1}^{\infty} c_n$ with $\sum_{n=1}^{\infty} |a_n|$ show that $\sum_{n=1}^{\infty} b_n$ and $\sum_{n=1}^{\infty} c_n$ converge.
   (b) By taking the difference of $\sum_{n=1}^{\infty} b_n$ and $\sum_{n=1}^{\infty} c_n$, show that $\sum_{n=1}^{\infty} a_n$ converges.

10.5 The Ratio and Root Tests

In the last section we compared two different series to decide if a series converged. In this section we will see how we can compare consecutive terms of a series to decide if a series converges. The idea behind this test is checking if the series under consideration looks like a geometric series. If the limit of the ratios of consecutive terms is less than one, we can compare the series to a geometric series that converges. In this test we avoid finding the geometric series used for the comparison as is done in Sect. 10.4.

As an example consider the series $\sum_{n=0}^{\infty} n/2^n$. The ratio of two consecutive terms is

$$\frac{a_{n+1}}{a_n} = \frac{\frac{n+1}{2^{n+1}}}{\frac{n}{2^n}} = \frac{n+1}{n}\frac{1}{2}.$$

Using induction we can show that when $m > 10$, $a_m \le (11/20)^{m-n}\, a_{10}$. This means that

$$\begin{aligned}\sum_{n=10}^{\infty} a_n &\le a_{10} \sum_{k=0}^{\infty} \left(\frac{11}{20}\right)^k \\ &\le a_{10} \frac{1}{1-\frac{11}{20}} \\ &\le \frac{20}{9} a_{10}.\end{aligned}$$

This means that, by the comparison test, the series $\sum_{n=0}^{\infty} n/2^n$ converges.

The important thing to note in this example is that after some N, the ratio $r_n = a_{n+1}/a_n$ is less than some $R < 1$. This was used to make a comparison with the geometric series $\sum_{k=0}^{\infty} R^k$ to prove convergence. If, using Theorem 88 on page 355, $\lim_{n\to\infty} |a_{n+1}/a_n| = r < 1$, this always happens. We can use this approach for every situation with $\lim_{n\to\infty} a_{n+1}/a_n = r < 1$. This leads to the statement of the ratio test for the convergence of series.

Theorem 89 (Limit ratio test). *Let $\sum_{n=M}^{\infty} a_n$ be an infinite series and assume that the limit*

$$r = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists. Then

(i) *If $r < 1$, the series converges.*
(ii) *If $r > 1$, the series diverges.*
(iii) *If $r = 1$, the limit gives no information on the convergence of the series.*

Remark 29. The argument before the theorem actually shows that we can use the condition that

$$\left|\frac{a_{n+1}}{a_n}\right| < r < 1$$

for all n greater than some N to guarantee convergence.

The second statement in Theorem 89 is true since it implies that the terms of the series do not go to zero. This is true since if $|a_{n+1}/a_n| > s > 1$, then $|a_{n+1}| > s\,|a_n| > |a_n|$.

Example 375. First we consider the two geometric series $\sum_{i=0}^{\infty}(2/3)^i$ and $\sum_{n=0}^{\infty}(3/2)^n$ to show the ratio test works for geometric series. For the first series we have

$$\begin{aligned} r_1 &= \lim_{i\to\infty}\left|\frac{\left(\frac{2}{3}\right)^{i+1}}{\left(\frac{2}{3}\right)^{i}}\right| \\ &= \lim_{i\to\infty}\left(\frac{2}{3}\right) \\ &= \frac{2}{3}. \end{aligned}$$

As it should, the ratio test shows that this series converges.

For the second series we have

$$\begin{aligned} r_1 &= \lim_{n\to\infty}\left|\frac{\left(\frac{3}{2}\right)^{n+1}}{\left(\frac{3}{2}\right)^{n}}\right| \\ &= \lim_{i\to\infty}\left(\frac{3}{2}\right) \\ &= \frac{3}{2}. \end{aligned}$$

As it should, the ratio test shows that this series diverges.

Now we consider a slightly more complicated series.

Example 376. Consider the series

$$\sum_{i=1}^{\infty}\frac{(i+1)^3}{5^i}.$$

Applying the ratio test gives

$$\begin{aligned} r &= \lim_{i\to\infty}\frac{\frac{(i+1)^3}{5^{i+1}}}{\frac{i^3}{5^i}} \\ &= \lim_{i\to\infty}\left[\frac{(i+1)^3}{i^3}\,\frac{5^i}{5^{i+1}}\right] \\ &= \left[\lim_{i\to\infty}\left(\frac{i+1}{i}\right)\right]^3\left[\lim_{i\to\infty}\frac{1}{5}\right] \\ &= 1^3\,\frac{1}{5} \\ &= \frac{1}{5}. \end{aligned}$$

Since $r < 1$, the series converges.

There are some important cases where the ratio test shows that series converge or diverge.

Example 377. A very important series is $\sum_{n=0}^{\infty} a^n/n!$ for any fixed a. Recalling that $(n+1)! = n!\,(n+1)$, we have

$$\begin{aligned} r &= \lim_{n\to\infty} \left| \frac{\frac{a^{n+1}}{(n+1)!}}{\frac{a^n}{n!}} \right| \\ &= \lim_{n\to\infty} \frac{|a|}{n+1} \\ &= 0. \end{aligned}$$

This series converges for all possible values of a. This series will appear again.

The next example show what can happen when $r = 1$.

Example 378. Consider the series $\sum_{i=1}^{\infty} 1/i^2$ and the series $\sum_{n=1}^{\infty} 1/\sqrt{n}$. These are both p-series. Thus, the first series converges with $p = 2$ and the second series diverges with $p = 1/2$. When we apply the ratio test to the first series we have

$$\begin{aligned} r &= \lim_{i\to\infty} \left| \frac{\frac{1}{(i+1)^2}}{\frac{1}{i^2}} \right| \\ &= \lim_{i\to\infty} \left(\frac{i}{i+1} \right)^2 \\ &= \left(\lim_{i\to\infty} \frac{i}{i+1} \right)^2 \\ &= 1. \end{aligned}$$

In a similar manner, for the second series we have

$$\begin{aligned} r &= \lim_{n\to\infty} \left| \frac{\frac{1}{\sqrt{n+1}}}{\frac{1}{\sqrt{n}}} \right| \\ &= \lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{n+1}} \\ &= \lim_{n\to\infty} \sqrt{1 - \frac{1}{n}} \\ &= 1. \end{aligned}$$

This means that we can have $r = 1$ for both convergent and divergent series.

We can use another technique to compare a series with a geometric series, the root test. This idea is even simpler than the idea of the ratio test. For a moment assume that the terms a_n of a series are nonnegative. If

$$\lim_{n\to\infty} \sqrt[n]{a_n} = r < 1$$

and $R \in (r, 1)$, then $0 \le a_n < R^n$ after some N. Comparing $\sum_{n=0}^{\infty} a_n$ with $\sum_{n=0}^{\infty} R^n$, a convergent series, shows that $\sum_{n=0}^{\infty} a_n$ converges. The same argument used to justify using $|a_n|$ for the ratio test also works here. This leads us to the root test.

Theorem 90 (Root test). *Let $\sum_{n=M}^{\infty} a_n$ be an infinite series and assume that the limit*

$$r = \lim_{n\to\infty} \sqrt[n]{|a_n|}$$

exists. Then

(i) *If $r < 1$, the series converges.*
(ii) *If $r > 1$, the series diverges.*
(iii) *If $r = 1$, the limit gives no information on the convergence of the series.*

Remark 30. The root test for convergence as derived using a comparison with a geometric series can be stated in a more general manner. To apply the comparison test we only needed that $\sqrt[n]{|a_n|} < r < 1$ after some N. Restating this we have that the series $\sum_{n=M}^{\infty} a_n$ converges if for some $r \in (0,1)$, $\sqrt[n]{|a_n|} < r$ after some N.

Most people find the root test harder to use than the ratio test. Using one fact, that can be proven using L'Hôpital's rule, makes it much easier to apply the root test. The result that we use is that for any nonnegative sequence $\{s_n\}_{n=0}^{\infty}$ such that $s_n = p(n)$ for some polynomial, then

$$\lim_{n\to\infty} \sqrt[n]{s_n} = 1.$$

With this, we do several examples. The first example is an application to a geometric series that confirms the result.

Example 379. Consider any series of the form $\sum_{n=0}^{\infty} C\, r^n$ with $C, r \neq 0$. The nth root of the nth term satisfies

$$\begin{aligned} \lim_{n\to\infty} \sqrt[n]{|a_n|} &= |r| \lim_{n\to\infty} \sqrt[n]{|C|} \\ &= |r|. \end{aligned}$$

The root test says that this series converges when $r < 1$. This is true since it is a geometric series. Note that the series diverges when $|r| \geq 1$. This is covered by the last two cases of the theorem.

We now have an example that shows we can use the root test when the ratio test does not apply.

Example 380. Consider a series $\sum_{n=0}^{\infty} a_n$ where $a_n = 2^{-n}$ is n is even and $a_n = 3^{-n}$ if n is odd. The ratio $|a_{n+1}/a_n|$ is $1/2\,(3/2)^n$ if n is odd. This goes to infinity. That means that the ratio test does not apply.

On the other hand $\sqrt[n]{|a_n|}$ is either $1/2$ or $1/3$. The remark after the root test says that this series converges.

Finally we have an example that shows some of the power of the root test.

Example 381. Consider the series $\displaystyle\sum_{n=1}^{\infty} \frac{2^n\,(n^2+5)}{3^{n-2}\,(2n^2+3)}$. Here

$$\begin{aligned} \sqrt[n]{|a_n|} &= \sqrt[n]{\frac{2^n\,(n^2+5)}{3^{n-2}\,(2n^2+3)}} \\ &= \sqrt[n]{\left(\frac{2}{3}\right)^n \frac{9\left(1+\frac{5}{n^2}\right)}{2\left(1+\frac{3}{2n^2}\right)}} \end{aligned}$$

$$= \frac{2}{3}\sqrt[n]{\frac{9}{2}\frac{\left(1+\frac{5}{n^2}\right)}{\left(1+\frac{3}{2n^2}\right)}}$$

Since

$$\frac{\left(1+\frac{5}{n^2}\right)}{\left(1+\frac{3}{2n^2}\right)} < 2$$

if $n > 2$, we have

$$\lim_{n\to\infty}\sqrt[n]{\frac{9}{2}\frac{\left(1+\frac{5}{n^2}\right)}{\left(1+\frac{3}{2n^2}\right)}} = 1.$$

This means that

$$\lim_{n\to\infty}\sqrt[n]{|a_n|} = \frac{2}{3} < 1.$$

By the root test, this series converges.

Exercises

1. Use the ratio or root test to decide which of the following series converge, which diverge, and those for which the ratio and root tests gives no information.

(a) $\sum_{n=1}^{\infty}\frac{5^n}{4^n}$

(b) $\sum_{n=1}^{\infty}\frac{6^n}{11^n}$

(c) $\sum_{n=1}^{\infty}\frac{2^n+10}{4-3^n}$

(d) $\sum_{n=1}^{\infty}\frac{4-3^n}{2^n+10}$

(e) $\sum_{n=1}^{\infty}\frac{n^3}{n^2\left(n^2-5\right)}$

(f) $\sum_{m=1}^{\infty}\frac{5^m+3^m}{m^4\,4^m}$

(g) $\sum_{k=1}^{\infty}(k+10)\,e^{-k^2}$

(h) $\sum_{p=3}^{\infty}\frac{1}{(p-2)\ln(p)}$

(i) $\sum_{n=2}^{\infty}\frac{n^3}{n!}$

(j) $\sum_{m=0}^{\infty}\frac{2^m}{m!}$

(k) $\sum_{n=2}^{\infty}\frac{2^{\frac{n}{3}}}{3^{\frac{n}{5}}}$

(l) $\sum_{m=0}^{\infty}\frac{m!}{4^m}$

(m) $\sum_{n=3}^{\infty}\frac{10^n}{4^n+n!}$

(n) $\sum_{m=1}^{\infty}(-1)^m\frac{7^{m-2}+m}{2^m-1}$

(o) $\sum_{m=1}^{\infty}(-1)^m\frac{7^{m/3}+m^2}{2^m+10}$

(p) $\sum_{k=0}^{\infty}\frac{3^{m-2}+4^{m+1}}{5^m}$

2. Determine if the following series converge or diverge.

(a) $\displaystyle\sum_{n=1}^{\infty}\left(\frac{1}{\ln(n+1)}\right)^n$

(b) $\displaystyle\sum_{n=3}^{\infty}\frac{1}{1+3^n}$

(c) $\displaystyle\sum_{n=0}^{\infty}\frac{\sin(n)}{n^{\frac{3}{2}}}$

(d) $\displaystyle\sum_{m=1}^{\infty}10m^{-m}$

(e) $\displaystyle\sum_{k=1}^{\infty}\left(\frac{k}{k+1}\right)^{k+1}$

(f) $\displaystyle\sum_{p=1}^{\infty}\frac{\ln(p)}{p}$

(g) $\displaystyle\sum_{k=2}^{\infty}\frac{\ln(k)}{k^2}$

(h) $\displaystyle\sum_{p=2}^{\infty}\sin^2\left(\frac{1}{p^2}\right)$

(i) $\displaystyle\sum_{n=0}^{\infty}\frac{n^3}{n!}$

(j) $\displaystyle\sum_{m=1}^{\infty}\frac{m!}{(m+2)!}$

(k) $\displaystyle\sum_{k=1}^{\infty}\frac{k^6+3k-2}{k^7-k^8+1}$

3. Explain why the ratio test does not apply to the series $\sum_{k=1}^{\infty} a_n$ if

$$a_n = \begin{cases} \left(\frac{2}{3}\right)^n & \text{if } n \text{ is even} \\ \left(\frac{1}{3}\right)^n & \text{if } n \text{ is odd.} \end{cases}$$

4. Explain why Remark 30 after the root test allows one to show that the series $\sum_{k=1}^{\infty} a_n$ with

$$a_n = \begin{cases} \left(\frac{2}{3}\right)^n & \text{if } n \text{ is even} \\ \left(\frac{1}{3}\right)^n & \text{if } n \text{ is odd} \end{cases}$$

converges.

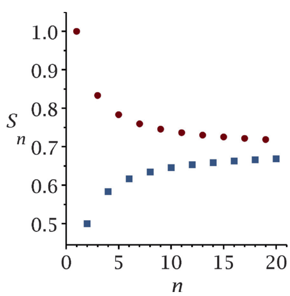


Fig. 10.4 Partial sums of an alternating series with $|a_n|$ decreasing

10.6 Alternating Series

We have mostly dealt with series that have nonnegative terms and have dealt with some series where the signs of the terms are arbitrary. In this section we mostly deal with series where the signs of the terms alternate between positive and negative. In this case we add a condition that guarantees convergence. This is different than the ratio and root tests in that some of these series converge even though the sum of the absolute values of the terms diverge. We start with a definition.

Definition 36. A series $\sum_{n=0}^{\infty} a_n$ is an *alternating series* if the signs of the terms alternate between positive and negative. This means that for some nonnegative sequence $\{b_n\}_{n=0}^{\infty}$ either $a_n = (-1)^n b_n$ or $a_n = (-1)^{n+1} b_n$.

Example 382. The series $\sum_{n=1}^{\infty} \left(-\frac{2}{3}\right)^n$ and $\sum_{k=2}^{\infty} \frac{\cos(k\pi)}{k}$ are alternating series. In the first series we have the sign of the nth term is the same as the sign of $(-1)^n$. In the second case $\cos(k\pi)$ is -1 if k is odd and 1 if k is even.

Example 383. The series $\sum_{m=1}^{\infty} \frac{\sin(m)}{\sqrt{m}}$ is not an alternating series. The first 10 terms of the series are approximately 0.8415, 0.4546, 0.04703, −0.1892, −0.1918, −0.04658, 0.09389, 0.1237, 0.04578, and −0.05440. It is clear that the signs are not alternating.

The main result of this section relies on several nice observations about certain alternating series. The major assumptions about the series $\sum_{n=1}^{\infty} a_n$ we consider are that the series is an alternating series and the sequence $\{|a_n|\}_{n=1}^{\infty}$ of absolute values of the terms decreases to 0. Although it is not necessary, to make the ideas clearer, in the discussion we also assume that the first term of the series is positive.

The partial sums with odd indices, $S_{2m+1} = \sum_{n=1}^{2m+1} a_n$, form a decreasing sequence since $S_{2m+3} - S_{2m+1} = a_{2m+2} + a_{2m+3} \leq 0$. This follows from the facts that $a_{2m+2} < 0$ and $|a_{2m+2}| \geq |a_{2m+3}|$. In a similar manner, the partial sums with even indices $S_{2m} = \sum_{n=1}^{2m} a_n$ form an increasing sequence. See Fig. 10.4 on page 362 for an illustration. In Fig. 10.4 the circles are the decreasing sequence of odd partial sums and the squares are the increasing sequence of even partial sums.

We leave it as an exercise to show that for any fixed n and all m, $S_{2n+1} > S_{2m+2}$ and $S_{2n+2} < S_{2m+1}$. This shows that the sequence of partial sums is bound above by S_1 and bounded below by S_2.

By Theorem 82 on page 345, the sequence of even partial sums converges to some L and the sequence of odd partial sums converges to some U. We must have $L = U$ since the distance between them must satisfy $U - L \leq S_{2m+1} - S_{2m+2} = |a_{2m+2}|$ where $\lim_{m\to\infty} |a_{2m+2}| = 0$. We state this result as the following theorem.

Theorem 91 (Alternating series test). *Assume that $\sum_{n=0}^{\infty} a_n$ is an alternating series whose terms have decreasing absolute values after some N. Then the series converges.*

We have shown that some of the divergent series with positive terms will converge when converted to alternating series.

Example 384. Consider the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$. This is an alternating series with terms whose absolute values, $\frac{1}{n}$, decrease to 0. This means the series converges.

It is also interesting to note that the distance from the Nth partial sum to the sum of the series is at most the absolute value of the $N+1$st term of the series. In this case the Nth partial sum, $\sum_{n=1}^{N} \frac{(-1)^n}{n}$, is at most $\frac{1}{N+1}$ away from the sum of the series.

The condition that the size of the terms decreases is essential. The next example gives a case where the series is alternating, but the terms do not decrease in size to 0.

Example 385. Consider a series whose even terms a_{2n} are $\frac{1}{n}$ for $n = 1, 2, 3, \ldots$ and whose odd terms a_{2m-1} are $-\frac{1}{2^m}$ for $m = 1, 2, 3, \ldots$. If we consider $2N$th partial sum we have

$$S_{2N} = \sum_{n=1}^{2N} a_n$$

$$= \sum_{n=1}^{N} a_{2n-1} + \sum_{m=1}^{N} a_m$$
$$= \sum_{n=1}^{N} \frac{1}{n} - \sum_{m=1}^{N} \frac{1}{2^m}. \tag{10.5}$$

The first of the two sums in Eq. 10.5 is the Nth partial sum for the harmonic series and the second sum is a partial sum for the geometric series with $r = \frac{1}{2}$. As N goes to infinity, the first sum diverges to infinity and the second sum converges to 1. This means that

$$\lim_{N\to\infty} S_{2N} = \infty.$$

The series is an alternating series whose terms go to zero that does not converge.

The idea of a series with terms of different signs goes well beyond alternating series. This is stated in Theorem 88 on page 355. The definitions relating convergence of a series with the convergence of the series whose terms are the absolute values of the first series are as follows. This was introduced on page 355 and in Theorem 88 on page 355.

Definition 37 (Absolute convergence). A convergent series $\sum_{n=0}^{\infty} a_n$ is said to be *absolutely convergent* if the series $\sum_{n=0}^{\infty} |a_n|$ converges. A convergent series $\sum_{n=0}^{\infty} a_n$ is said to be *conditionally convergent* if the series $\sum_{n=0}^{\infty} |a_n|$ diverges.

We have already seen series of both types. Here are a couple examples.

Example 386. Consider the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$. The sum of absolute values of the terms is

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Since this is a p-series with $p > 1$, the series is absolutely convergent. Recall that Theorem 88 on page 355 says that if a series converges absolutely, then it converges.

Example 387. We now look at the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$. This series converges since $\frac{1}{\sqrt{n}}$ decreases to 0 and it is an alternating series. However, the series

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

is a p-series with $p < 1$, it diverges. This means the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$ is conditionally convergent.

There is an important difference between absolutely convergent series and conditionally convergent series. It simply says that the order in which you add the terms of an absolutely convergent series does not matter. On the other hand, if a series is conditionally convergent, the order of addition can change the series from divergent to convergent. We call a series $\sum_{m=0}^{\infty} b_m$ a *rearrangement* of a series $\sum_{n=0}^{\infty} a_n$ if there is a one to one function $M(n)$ from the positive integers onto the positive integers such that $b_m = a_{M(m)}$.

Example 388. The series with terms

$$b_n = \begin{cases} \frac{1}{2^{2k}} & \text{if } n = 3k \\ \frac{1}{2^{2k+1}} & \text{if } n = 3k+1 \\ \frac{1}{3^k} & \text{if } n = 3k+2 \end{cases}$$

is a rearrangement of the series

$$a_n = \begin{cases} \frac{1}{2^k} & \text{if } n = 2k \\ \frac{1}{3^k} & \text{if } n = 2k+1 \end{cases}$$

In both series there is one term of the form $\frac{1}{2^m}$ and one term of the form $\frac{1}{3^m}$ for all nonnegative integers m.

If we write down the first few term of each series we have

$$\sum_{n=0}^{\infty} a_n = \frac{1}{2^0} + \frac{1}{3^0} + \frac{1}{2^1} + \frac{1}{3^1} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{2^4} + \frac{1}{3^2} + \dots$$

and

$$\sum_{n=0}^{\infty} b_n = \frac{1}{2^0} + \frac{1}{2^1} + \frac{1}{3^0} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{3^1} + \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{3^2} + \dots.$$

We can view the series $\sum_{n=0}^{\infty} a_n$ as putting a term of the form $\frac{1}{3^k}$ between each term of the series $\sum_{n=0}^{\infty} \frac{1}{2^n}$. The series $\sum_{n=0}^{\infty} b_n$ can be viewed as putting a term of the form $\frac{1}{3^k}$ after every second term of the series $\sum_{n=0}^{\infty} \frac{1}{2^n}$. This means that each series has one term of the form $\frac{1}{2^n}$ and one term of the form $\frac{1}{3^n}$ for $n = 0, 1, 2, \dots$.

The proof of the next result is left for later courses.

Theorem 92. *Let $\sum_{n=0}^{\infty} a_n$ be a series that converges absolutely. Then any rearrangement of this series converges and has the same sum as the original series.*

We can use Example 388 on page 364 as an application of this result.

Example 389. Consider the series $\sum_{n=0}^{\infty} a_n$ with

$$a_n = \begin{cases} \frac{1}{2^k} & \text{if } n = 2k \\ -\frac{1}{3^k} & \text{if } n = 2k+1. \end{cases}$$

The $2N$th partial sum of this series can be rewritten as

$$\begin{aligned} S_{2N} &= \sum_{n=0}^{2N} a_n \\ &= \sum_{k=0}^{N-1} \frac{1}{2^k} - \sum_{m=0}^{N-1} \frac{1}{3^m}. \end{aligned}$$

Taking the limit as $N \to \infty$ gives

$$\begin{aligned} \lim_{N\to\infty} S_{2N} &= \lim_{N\to\infty} \sum_{n=0}^{2N} a_n \\ &= \lim_{N\to\infty} \left(\sum_{k=0}^{N-1} \frac{1}{2^k} \right) - \lim_{N\to\infty} \left(\sum_{m=0}^{N-1} \frac{1}{3^m} \right) \\ &= 2 - \frac{3}{2} \\ &= \frac{1}{2}. \end{aligned}$$

Since the series $\sum_{m=0}^{\infty} b_m$ with

$$b_m = \begin{cases} \frac{1}{2^{2k}} & \text{if } m = 3k \\ \frac{1}{2^{2k+1}} & \text{if } m = 3k+1 \\ -\frac{1}{3^k} & \text{if } m = 3k+2 \end{cases}$$

is a rearrangement of the absolutely convergent series $\sum_{n=0}^{\infty} a_n$,

$$\sum_{m=0}^{\infty} b_m = \frac{1}{2}.$$

Exercises

1. Which of the following series converge absolutely, which converge conditionally, and which diverge.

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{(-2)^n}{3^n}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{(-3)^n}{2^n+10}$

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n n^3}{n^2(n^2-5)}$

(d) $\displaystyle\sum_{m=1}^{\infty} \frac{(-1)^m}{\sqrt{m^3+m}}$

(e) $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^k}{\sqrt{k+100}}$

(f) $\displaystyle\sum_{p=3}^{\infty} \frac{p\cos(p\pi)}{p^2+2p+1}$

(g) $\displaystyle\sum_{n=2}^{\infty} \frac{\cos(n\pi)n^3}{n!}$

(h) $\displaystyle\sum_{m=0}^{\infty} \frac{(-3)^m}{2^m}$

(i) $\displaystyle\sum_{n=2}^{\infty} \frac{(-3)^n}{2^{2n}}$

(j) $\displaystyle\sum_{m=0}^{\infty} \frac{m!}{(-4)^m}$

(k) $\displaystyle\sum_{n=3}^{\infty} \frac{10^n}{(-4)^n+n!}$

(l) $\displaystyle\sum_{m=1}^{\infty} (-1)^m \frac{7^{\frac{m}{3}}}{2^m}$

2. Determine if the following series converge or diverge.

(a) $\displaystyle\sum_{n=1}^{\infty} \left(\frac{1}{\ln(n+1)}\right)^n$

(b) $\displaystyle\sum_{n=3}^{\infty} \frac{1}{1+3^n}$

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{\sin(n)}{n^{\frac{3}{2}}}$

(d) $\displaystyle\sum_{m=1}^{\infty} 10m^{-m}$

(e) $\displaystyle\sum_{k=1}^{\infty} \left(\frac{k}{k+1}\right)^{k+1}$

(f) $\displaystyle\sum_{p=1}^{\infty} \frac{\ln(p)}{p}$

(g) $\displaystyle\sum_{k=2}^{\infty} \frac{\ln(k)}{k^2}$

(h) $\displaystyle\sum_{p=2}^{\infty} \sin^2\left(\frac{1}{p^2}\right)$

(i) $\displaystyle\sum_{n=0}^{\infty} \frac{n^3}{n!}$

(j) $\displaystyle\sum_{m=1}^{\infty} \frac{m!}{(m+2)!}$

(k) $\displaystyle\sum_{k=1}^{\infty} \frac{k^6+3k-2}{k^7-k^8+1}$

3. This exercise shows how to get a conditionally convergent series to diverge to $-\infty$ by rearranging the terms of the series.
Consider the conditionally convergent series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} .$$

(a) Show that $\sum_{n=1}^{\infty} \frac{1}{2n}$ diverges ti ∞.
(b) Explain why there is an increasing sequence of integers $\{k_i\}_{i=1}^{\infty}$ such that $\sum_{n=k_1}^{k_{n+1}-1} \geq \frac{1}{2}$.
(c) Consider the rearrangement of the series to $\sum_{n=1}^{\infty} b_n$. Here $b_1 = a_1$ is followed by the first k_1 even numbered $a_i's$. Then comes a_3 followed by the next k_2 even $a_i's$. This pattern continues for all $b_n's$. Show that the series using these $b_n's$ diverges.

10.7 Power Series

In the last few sections we dealt with series who's terms were numbers. We can also consider series whose terms are functions of a variable x. This can be done with general sequences of functions $f_n(x)$. Among the functions that are used for the $f_n(x)$'s are trigonometric functions of the forms $\sin(k\pi x)$ and $\cos(m\pi x)$ for integers k and m. To make things simpler we only consider the case when the functions $f_n(x)$ are monomials of the form $f_n(x) = a_n x^n$. One of the main advantages of using these monomials is that they relate nicely to geometric series. Another advantage, considered in the next few sections, is that these series are closely tied to Taylor polynomials.

Definition 38. A power series is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n$$

or of the form

$$\sum_{n=0}^{\infty} a_n (x-c)^n,$$

where x is a variable.

Example 390. Consider the series

$$\sum_{n=0}^{\infty} x^n$$

as x is allowed to change. We have already shown that this converges if $|x| < 1$ and that it diverges if $|x| \geq 1$. Since $1/1 - x = \sum_{n=0}^{\infty} x^n$ if $|x| < 1$, a graph of the values of the series are in Fig. 10.5 on page 368.

An important question for power series is for which values of the variable, x in the above, does the series converge. For these values of x the power series defines a function of x,

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n. \tag{10.6}$$

We can compare the terms of the power series with a geometric series to give a partial answer to the question of where a power series converges.

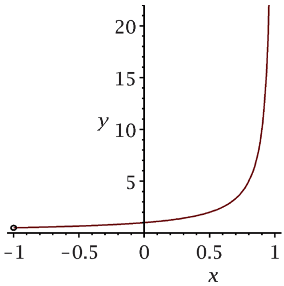


Fig. 10.5 A graph of $\sum_{n=0}^{\infty} x^n$

When we apply the ratio test for series, Theorem 89 on page 357, to a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$ we want

$$\begin{aligned} r &= \lim_{n\to\infty} \left| \frac{a_{n+1}(x-c)^{n+1}}{a_n(x-c)^n} \right| \\ &= |x-c| \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| \\ &< 1 \end{aligned}$$

to guarantee convergence. If the limit $\lim_{n\to\infty} |a_{n+1}/a_n|$ is zero, the ratio test says that the series converges for all x. When the limit $\lim_{n\to\infty} |a_{n+1}/a_n|$ is infinity, the series does not converge unless $x = c$. (Why is this true?) Finally, if the limit $\lim_{n\to\infty} |a_{n+1}/a_n|$ is finite and greater than zero, we need

$$|x-c| < \frac{1}{r} = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| = R$$

to guarantee convergence. If

$$|x-c| > \frac{1}{r} = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| = R,$$

the power series diverges. Putting these facts together gives the following theorem, the ratio test for power series.

Theorem 93 (Ratio Test for Power Series). *Let $\sum_{n=0}^{\infty} a_n(x-c)^n$ be a power series and assume the limit $\lim_{n\to\infty} |a_n/a_{n+1}|$ exists and equals R. Then*

(i) If $|x-c| < R$, the power series converges.
(ii) If $|x-c| > R$, the power series diverges.
(iii) If $|x-c| = R$, the limit gives no information on the convergence of the power series.

This R is called the radius of convergence *of the power series.*

Remark 31. We can also calculate the radius of convergence using the root test for series. If the limit exists we have

$$R = \lim_{n\to\infty} \sqrt[n]{\frac{1}{a_n}}.$$

We now proceed with several examples.

Example 391. Consider the power series

$$\sum_{n=1}^{\infty} \frac{x^n}{n^2}.$$

Here we calculate

$$\begin{aligned}
R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\
&= \lim_{n\to\infty} \left| \frac{(n+1)^2}{n^2} \right| \\
&= \lim_{n\to\infty} \left(\frac{n+1}{n} \right)^2 \\
&= 1.
\end{aligned}$$

This series converges if $|x| < 1$ or $x \in (-1,1)$.

Example 392. Consider the power series

$$\sum_{n=1}^{\infty} n(x+2)^n.$$

Here we calculate

$$\begin{aligned}
R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\
&= \lim_{n\to\infty} \frac{n}{n+1} \\
&= 1.
\end{aligned}$$

This series converges when $|x+2| < 1$ or $x \in (-3,-1)$.

Example 393. We next look at the power series

$$\sum_{n=1}^{\infty} \frac{x^n}{n!}.$$

The limit of the ratios is

$$\begin{aligned}
R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\
&= \lim_{n\to\infty} \frac{(n+1)!}{n!} \\
&= \lim_{n\to\infty} (n+1) \\
&= \infty.
\end{aligned}$$

This series converges for all x.

Example 394. Consider the power series

$$\sum_{n=1}^{\infty} n^n (x-5)^n.$$

Using the fact that $n/(n+1) < 1$, we calculate

$$\begin{aligned} R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\ &= \lim_{n\to\infty} \frac{n^n}{(n+1)^{n+1}} \\ &= \lim_{n\to\infty} \left(\frac{n}{(n+1)} \right)^n \frac{1}{n+1} \\ &\leq \lim_{n\to\infty} \frac{1}{n+1} \\ &\leq 0. \end{aligned}$$

This means that $R = 0$ and the series only converges if $x = 5$. In that case every term is 0.

Using a root test here would be simpler since

$$\begin{aligned} r &= \lim_{n\to\infty} \sqrt[n]{n^n |x-5|^n} \\ &= \lim_{n\to\infty} n|x-5| \\ &= \begin{cases} \infty & x \neq 5 \\ 0 & x = 5 \end{cases}. \end{aligned}$$

Since we need $r < 1$, this series only converges if $x = 5$.

If $R > 0$ for a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$ we know that the series converges if $|x-c| < R$ and that the series diverges if $|x-c| > R$. This means that the set where a power series converges is one of the intervals $[c-R, c+R]$, $(c-R, c+R]$, $[c-R, c+R)$ or $(c-R, c+R)$. The is the reason why we call the set where a power series converges the *interval of convergence* of the power series. It is important to note that after finding R, if we want the interval of convergence we still must consider what happens at the endpoints of the interval. *The ratio and root tests will not work at the endpoints of the interval of convergence.*

Example 395. Consider the power series $\sum_{n=1}^{\infty} x^n$, $\sum_{n=1}^{\infty} x^n/n$, and $\sum_{n=1}^{\infty} x^n/n^2$. We can show that radius of convergence of each of these series is $R = 1$. Therefore we only need to consider what happens to each of these series if $x = \pm 1$.

For the series $\sum_{n=1}^{\infty} x^n$ the terms are $(\pm 1)^n$ if $x = \pm 1$. Since these terms do not go to 0, the series does not converge if $x = \pm 1$. The interval of convergence is $(-1, 1)$.

If we put $x = \pm 1$ into the second power series we get the series $\sum_{n=1}^{\infty} 1/n$ and $\sum_{n=1}^{\infty} (-1)^n/n$. The series $\sum_{n=1}^{\infty} 1/n$ is the divergent harmonic series and the series $\sum_{n=1}^{\infty} (-1)^n/n$ converges by the alternating series test. For this power series the interval of convergence is $[-1, 1)$,

When $x = \pm 1$ is put into the third power series we get series whose n the terms have absolute value $1/n^2$. Since $\sum_{n=1}^{\infty} 1/n^2$ converges, both series converge absolutely. The interval of convergence is $[-1, 1]$.

This shows that three out of the four types of intervals of convergence happen. As was done above, we can use the series $\sum_{n=1}^{\infty} (-x)^n/n$ to get the final interval type. (Show that it is the correct form.)

There are some very nice things that can be said about power series within their interval of convergence. The justifications of these results are mostly well beyond this text and the results are stated without proof. These theorems include results on sums, products, integrals and derivatives. To simplify the statements we let the center of each series be $c = 0$ and we omit the result on products of series.

The first result concerns the sum of two series.

Theorem 94 (Sum of series). *Assume that* $f(x)=\sum_{n=0}^{\infty}a_n(x-c)^n$ *and* $g(x)=\sum_{m=0}^{\infty}b_m(x-c)^m$ *are power series that both converge on some interval* $I=(c-\alpha,c+\alpha)$. *Then the sum of the functions* $f(x)$ *and* $g(x)$ *exists on* I *with*

$$(f+g)(x)=\sum_{n=0}^{\infty}(a_n+b_n)(x-c)^n.$$

The radius of convergence of $\sum_{n=0}^{\infty}(a_n+b_n)(x-c)^n$ *is a least the minimum of the radius of convergence of* $f(x)=\sum_{n=0}^{\infty}a_n(x-c)^n$ *and the radius of convergence of* $g(x)=\sum_{m=0}^{\infty}b_m(x-c)^m$.

Example 396. Consider the series $f(x)=\sum_{m=0}^{\infty}mx^m$ and $g(x)=\sum_{m=0}^{\infty}x^m/m!$. Both series converge on $(-1,1)$ and their sum is

$$\sum_{m=0}^{\infty}\left(m+\frac{1}{m!}\right)x^m.$$

Now we consider derivatives and antiderivatives.

Theorem 95 (Calculus on series). *Let* $f(x)=\sum_{n=0}^{\infty}a_n(x-c)^n$ *be a series with radius of convergence* $R>0$. *Then* $f(x)$ *is continuous, differentiable and has an antiderivative on* $(c-R,c+R)$ *with*

$$\begin{aligned}\frac{df(x)}{dx}&=\sum_{n=0}^{\infty}\left(a_n\frac{dx^n}{dx}\right)\\&=\sum_{n=1}^{\infty}na_nx^{n-1}\\&=\sum_{n=0}^{\infty}(n+1)a_{n+1}x^n.\end{aligned}$$

and

$$\begin{aligned}\int f(x)\,dx&=\sum_{n=0}^{\infty}\left(a_n\int x^n\,dx\right)\\&=\sum_{n=0}^{\infty}\frac{a_n}{n+1}x^{n+1}+C\\&=\sum_{n=1}^{\infty}\frac{a_{n-1}}{n}x^n+C.\end{aligned}$$

Furthermore, the radius of convergence is the same for all three series.

Example 397. Let $f(x)=\sum_{m=0}^{\infty}\frac{x^m}{2^m+m}$. Using the last theorem we have

$$\begin{aligned}f'(x)&=\sum_{m=0}^{\infty}\left(\frac{d}{dx}\frac{x^m}{2^m+m}\right)\\&=\sum_{m=0}^{\infty}\frac{(m+1)x^m}{2^{m+1}+m+1}\end{aligned}$$

and

$$\int f(x)\,dx = \sum_{m=0}^{\infty}\left(\int \frac{x^m}{2^m+m}\,dx\right)$$
$$= \sum_{m=0}^{\infty} \frac{x^{m+1}}{(m+1)\,(2^m+m)} + C$$

The radius of convergence is $R = 2$ for all three series.

In many cases it is important to understand what happens to the interval of convergence when x is replaced by $A\,(y-B)$. If we assume that the series $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R, then the series

$$\sum_{n=0}^{\infty} a_n\,(A\,(y-B))^n = \sum_{n=0}^{\infty} a_n A^n\,(y-B)^n$$

will converge when $A\,(y-B) < R$ and it will diverge when $A\,(y-B) > R$. This means that the interval of convergence has center B and radius R/A. (Rewrite the inequality $A\,(y-B) < R$.) A simple example illustrates this.

Example 398. We want to find the interval of convergence of $\sum_{n=0}^{\infty} 4^n\,(x+3)^n$. Letting $z = 4\,(x+3)$, this series can be rewritten as

$$\sum_{n=0}^{\infty} 4^n\,(x+3)^n = \sum_{n=0}^{\infty} (4\,(x+3))^n$$
$$= \sum_{n=0}^{\infty} z^n.$$

Since $\sum_{n=0}^{\infty} z^n$ has interval of convergence $(-1,1)$, we only need to translate the center to -3 and change the radius to $1/4$. Therefore, the interval of convergence for the power series $\sum_{n=0}^{\infty} 4^n\,(x+3)^n$ is $\left(-3-\frac{1}{4}, -3+\frac{1}{4}\right) = \left(-\frac{7}{4}, -\frac{5}{4}\right)$.

Example 399. For the series

$$\sum_{k=2}^{\infty} \frac{1}{n^2}\left(\frac{x}{2}\right)^n$$

we can find the radius of convergence of

$$\sum_{k=2}^{\infty} \frac{1}{n^2}\,y^n$$

As was demonstrated in Example 395 on page 370, the series in y has radius of convergence 1. This gives us that the series in x converges if

$$\frac{x}{2} < 1.$$

From this equation we conclude that the radius of convergence for the x series is $R = 2$.

Consider the case, of a series of the form $\sum_{n=0}^{\infty} a_n\,(x-c)^{qn}$. Here we set $y = (x-c)^q$ and rewrite the series as $\sum_{n=0}^{\infty} a_n\,y^n$. If this series in y has radius of convergence R, the series in x converges if $|(x-c)^q| = |y| < R$, or $|x-c| < R^{1/q}$. This means the radius of convergence is $R^{1/q}$.

Example 400. Consider the series

$$\sum_{m=1}^{\infty} \frac{x^{2n}}{3^n+1}.$$

Here we can let $y = x^{2n}$ and find the radius of convergence of

$$\sum_{m=1}^{\infty} \frac{y^n}{3^n+1}.$$

This series in y has radius of convergence 3.

The series in x will converge if $\left|x^2\right| < 3$ or $|x| < \sqrt{3}$. The radius of convergence is $R = \sqrt{3}$.

Exercises

1. Find the radius of convergence for the following power series.

(a) $\sum_{n=1}^{\infty} \frac{2^n x^n}{3^n}$

(b) $\sum_{n=1}^{\infty} \frac{3^n (x+2)^n}{2^n+10}$

(c) $\sum_{n=1}^{\infty} \frac{n^3 (y-5)^n}{n^2 (n^2-5)}$

(d) $\sum_{m=1}^{\infty} \frac{x^m}{\sqrt{m^3+m}}$

(e) $\sum_{k=1}^{\infty} \frac{(z-1)^k}{\sqrt{2^{5k}+100}}$

(f) $\sum_{p=3}^{\infty} \frac{\cos(p\pi) y^p}{p^2+2p+1}$

(g) $\sum_{n=2}^{\infty} \frac{\cos(n\pi)(x+5)^{3n}}{n!}$

(h) $\sum_{m=0}^{\infty} \frac{(-3z)^m}{2^m}$

(i) $\sum_{n=2}^{\infty} \frac{(-3w)^n}{(2n)!}$

(j) $\sum_{m=0}^{\infty} \frac{m! x^m}{(-4)^m}$

(k) $\sum_{n=3}^{\infty} \frac{10^n (y+7)^n}{(-4)^n+n!}$

(l) $\sum_{m=1}^{\infty} (-1)^m \frac{7^{\frac{m}{3}} x^m}{2^m}$

2. Find the interval of convergence for the following power series.

(a) $\sum_{n=1}^{\infty} \frac{(x+3)^n}{n^3}$

(b) $\sum_{n=2}^{\infty} \frac{x^n}{\sqrt[3]{n}}$

(c) $\sum_{n=5}^{\infty} \frac{\sin(n)(z-2)^n}{n^{\frac{2}{3}}}$

(d) $\sum_{m=1}^{\infty} 10m(x-11)^m$

(e) $\sum_{k=1}^{\infty} \left(\frac{4x}{3}\right)^{k+1}$

(f) $\sum_{p=1}^{\infty} \frac{\ln(p)(z+2)^p}{p}$

(g) $\sum_{k=2}^{\infty} \frac{k^2 (y+1)^k}{\ln(k)}$

(h) $\sum_{p=2}^{\infty} \frac{z^p}{p^{\frac{1}{3}}}$

(i) $\sum_{n=0}^{\infty} \frac{x^{2n}}{4^n}$

(j) $\sum_{m=1}^{\infty} \frac{m(z-2)^m}{(m+2)!}$

(k) $\sum_{k=1}^{\infty} \frac{(k^6+3k-2) w^k}{k^7-k^8+1}$

3. Find the derivatives and integrals of the following functions on their intervals of convergence. (These are the same series as in Exercise 2.)

(a) $\sum_{n=1}^{\infty} \frac{(x+3)^n}{n^3}$

(b) $\sum_{n=2}^{\infty} \frac{x^n}{\sqrt[3]{n}}$

(c) $\sum_{n=3}^{\infty} \frac{\sin(n)\,(z-2)^n}{n^{\frac{2}{3}}}$

(d) $\sum_{m=1}^{\infty} 10m\,(x-11)^m$

(e) $\sum_{k=1}^{\infty} \left(\frac{4x}{3}\right)^{k+1}$

(f) $\sum_{p=1}^{\infty} \frac{\ln(p)\,(z+2)^p}{p}$

(g) $\sum_{k=2}^{\infty} \frac{k^2\,(y+1)^k}{\ln(k)}$

(h) $\sum_{p=2}^{\infty} \frac{z^p}{p^{\frac{1}{3}}}$

(i) $\sum_{n=0}^{\infty} \frac{x^{2n}}{4^n}$

(j) $\sum_{m=1}^{\infty} \frac{m\,(z-2)^m}{(m+2)!}$

(k) $\sum_{k=1}^{\infty} \frac{(k^6+3k-2)\,w^k}{k^7-k^8+1}$

4. Find the derivatives and integrals of the following functions.

(a) $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{(n+2)!}$

(b) $h(z) = \sum_{n=0}^{\infty} \frac{n\,z^n}{2^n}$

(c) $s(w) = \sum_{k=2}^{\infty} \left(\frac{2w}{5}\right)^k$

(d) $f(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{n!}$

(e) $h(y) = \sum_{m=1}^{\infty} \frac{(-2y)^m}{m^3}$

(f) $s(w) = \sum_{k=0}^{\infty} \frac{w^{2k}}{(2k)!}$

(g) $f(x) = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}$

(h) $h(y) = \sum_{m=0}^{\infty} \frac{(-1)^m y^{2m+1}}{(2m+1)!}$

5. Rewrite each of the following as a single power series of the form $\sum_{n=k}^{\infty} a_n x^n$.

(a) $\sum_{n=1}^{\infty} \frac{x^n}{n^2} + \sum_{n=1}^{\infty} \frac{x^n}{n^2+1}$

(b) $24 \sum_{n=0}^{\infty} \frac{x^n}{2^n}$

(c) $x \sum_{m=2}^{\infty} \frac{x^m}{2^m}$

(d) $\sum_{n=3}^{\infty} \frac{x^{n-1}}{n^2} + \sum_{n=1}^{\infty} \frac{x^{n-1}}{n^2+1}$

(e) $2 \sum_{k=1}^{\infty} \frac{x^{2k}}{k3^k} + x \sum_{m=0}^{\infty} \frac{x^{2m+1}}{3^m}$

(f) $\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} + \sum_{m=0}^{\infty} \frac{x^{2m+1}}{(2m+1)!}$

6. Consider the function defined by $H(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}$. Show that $\frac{d^2H(x)}{dx^2} = H(x)$.

7. Consider the function defined by $C(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$. Show that $\frac{d^2C(x)}{dx^2} = -C(x)$.

8. Find the radii of convergence for the series $\sum_{n=0}^{\infty} \frac{x^n}{2^n+10}$ and $\sum_{n=0}^{\infty} \frac{x^n}{3^n-5}$. Use the ratio or root test to find the radius of convergence of

$$\sum_{n=0}^{\infty} \left(\frac{1}{2^n+10} + \frac{1}{3^n-5}\right) x^n \ ?$$

Does this match Theorem 94?

10.8 Taylor Series

In the last section we considered where power series converge and some operations that we can perform on those power series. In this section we consider power series obtained from functions. These series have many uses and are used in a variety of applications. The approach here is based on Taylor polynomials, see Sect. 4.6. First we give the definition of a Taylor/Maclaurin series.

Definition 39. Assume that a function $f(x)$ has derivatives of all orders on an interval (a,b) containing a point c. The *Taylor series* for $f(x)$ centered at c is

$$\begin{aligned} f(x) &\sim f(c) + f'(c)(x-c) + \frac{f^{(2)}(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n + \ldots \\ &\sim \sum_{n=0}^{\infty} \frac{f^{(n)}}{n!}(x-c)^n. \end{aligned}$$

If $c = 0$ the series is called a *Maclaurin* series.

Remark 32. The symbol $\sim$ is used in Definition 39 rather than an equals sign because there are cases when a Taylor series does not converge to the function $f(x)$ except at the point c. In this case the series does not represent the function $f(x)$ on a nontrivial interval.

We now give a simple example to show some of the features of a Taylor series.

Example 401. Let $f(x) = e^x$. Since all of the derivatives of e^x are e^x and $e^0 = 1$, the nth degree Taylor polynomial for e^x is

$$T_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}.$$

The infinite series with kth term $\frac{x^k}{k!}$ is

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}. \tag{10.7}$$

Looking at the first few terms we have

$$e^x \sim 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots.$$

Using the ratio test this series has radius of convergence

$$\begin{aligned} R &= \lim_{k\to\infty} \frac{\frac{1}{k!}}{\frac{1}{(k+1)!}} \\ &= \lim_{k\to\infty} \frac{(k+1)!}{k!} \\ &= \lim_{k\to\infty} (k+1) \\ &= \infty. \end{aligned}$$

In the last example we saw that the series converges for all $x \in \mathbb{R}$. What was not shown is that the series converges to the function $f(x) = e^x$ for all x. To do this we usually use the following theorem. This theorem is not proven here.

Theorem 96 (Taylor polynomial error). *Assume that the function $f(x)$ has $n+1$ continuous derivatives on an interval (a,b) and that $c \in (a,b)$. For any $x \in (a,b)$*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1} \tag{10.8}$$

for some ξ between c and x.

With Theorem 96 we can now show that the series in Eq. 10.7 converges to e^x for all $x \in \mathbb{R}$.

Example 402. We show that the power series in Eq. 10.7 converges to e^x for all x. This is fairly easy since we can show that for any x and any C,

$$\lim_{k\to\infty} \frac{Cx^k}{k!} = 0$$

Rewriting Eq. 10.8 we have, for some ξ_k between x and 0,

$$\begin{aligned} \left| e^x - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k \right| &= \frac{e^{\xi_k}}{(n+1)!} x^{n+1} \\ &\le e^{|x|} \frac{|x|^{n+1}}{(n+1)!}. \end{aligned}$$

Since the last expression goes to 0 as $n \to \infty$, the power series converges to e^x for any x.

The two examples above illustrate one process for finding the Taylor series of a function $f(x)$ centered at a point c. For many series the hardest part of this procedure is finding a general pattern for the coefficient of x^n for all n. We do one more example of finding the pattern and then give the series for several common functions.

Example 403. The hyperbolic cosine function is defined by

$$\cosh(x) = \frac{e^x + e^{-x}}{2}.$$

The derivative of $\cosh(x)$ is the hyperbolic sine

$$\sinh(x) = \frac{e^x - e^{-x}}{2}.$$

From this we can find that

$$\frac{d^n}{dx^n}\cosh(x) = \begin{cases} \sinh(x) & \text{if } n \text{ is odd} \\ \cosh(x) & \text{if } n \text{ is even} \end{cases}.$$

The coefficients in the Maclaurin series for $\cosh(x)$ are

$$a_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{1}{n!} & \text{if } n \text{ is even} \end{cases}.$$

This gives us the series

$$\cosh(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}.$$

Writing out the first few terms gives

$$\cosh(x) = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \dots.$$

Using the ratio test for series we can show that this series converges for all $x \in \mathbb{R}$.

Note that the terms in the series for $\cosh(x)$ are the even terms in the series for $\exp(x)$. Why does this show that the Maclaurin series for $\cosh(x)$ converges to $\cosh(x)$ for all x?

The following common functions have nice series representations that can be found using the method above. We have already found the coefficients for two of the series. Finding the coefficients for the rest of the series is left to the exercises. All of these series converge for all real numbers.

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} \tag{10.9}$$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \tag{10.10}$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{10.11}$$

$$\cosh(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \tag{10.12}$$

$$\sinh(x) = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}. \tag{10.13}$$

There are other ways of finding the series representations of functions. An important example is the function $g(x) = \ln(x+1)$. In this case we work with the derivative of $g(x)$. If we use the sum for the geometric series, we have,

$$\begin{aligned} g'(x) &= \frac{1}{1+x} \\ &= \frac{1}{1-(-x)} \\ &= \sum_{n=0}^{\infty} (-x)^n \\ &= \sum_{n=0}^{\infty} (-1)^n x^n. \end{aligned}$$

Since the radius of convergence of this geometric series is $R = 1$, we can integrate the series within that radius to get

$$\begin{aligned} \ln(1+x) &= \int \frac{1}{1+x}\, dx \\ &= \int \left(\sum_{n=0}^{\infty} (-1)^n x^n \right) dx \\ &= \sum_{n=0}^{\infty} \left(\int (-1)^n x^n \, dx \right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \left(\frac{(-1)^n x^{n+1}}{n+1} \right) + C \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^n}{n} + C.
\end{aligned}
$$

Since $\ln(1+0) = 0$, the constant C is zero and

$$
\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^n}{n}.
$$

Using a similar series of steps we can show that the Maclaurin series for $\arctan(x)$ is

$$
\begin{aligned}
\arctan(x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \\
&= x - \frac{x^3}{3} + \frac{x^5}{5} - + \ldots.
\end{aligned}
$$

Example 404. The series for $x^2 \ln(1+x)$ can be obtained easily from the series for $\ln(1+x)$. If we fix $x \in (-1,1)$ we can write

$$
\begin{aligned}
x^2 \ln(1+x) &= x^2 \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^n}{n} \\
&= \sum_{n=1}^{\infty} \left(x^2 \frac{(-1)^{n-1} x^n}{n} \right) \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^{n+2}}{n}.
\end{aligned}
$$

Rewrite this series in the form $\sum_{m=k}^{\infty} a_m x^m$.

10.8.1 Series and Differential Equations

Another place where series can be useful is finding solutions to certain differential equations. If we assume that the differential equation has a unique solution for a given initial condition $y(x_0) = y_0$ we can try to find a series that is a solution to the differential equation and that satisfies the initial condition.

This is done by substituting a power series with unknown coefficients into the differential equation. From that we get a recursion relation for the coefficients. If we can find appropriate coefficients, we have a solution for the differential equation represented by the series. We demonstrate this idea using a simple differential equation that has already been considered, see Examples 207 and 273.

Example 405. Assume that the function $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is a solution to the following initial value problem:

$$
\frac{dy}{dx} = 4y(x) \qquad \text{and} \qquad y(0) = 1.
$$

Differentiating the series with respect to x gives

$$\frac{df}{dx}(x) = \sum_{n=1}^{\infty} n\,a_n x^{n-1} = \sum_{n=0}^{\infty} (n+1)\,a_{n+1} x^n.$$

Plugging $\sum_{n=0}^{\infty} a_n x^n$ and its derivative into the differential equation gives

$$\sum_{n=0}^{\infty} (n+1)\,a_{n+1} x^n = 4 \sum_{n=0}^{\infty} a_n x^n.$$

Taking the 4 inside the sum on the left and subtracting the right side of the equation from both sides we get

$$\begin{aligned} 0 &= \sum_{n=0}^{\infty} 4a_n x^n - \sum_{n=0}^{\infty} (n+1)\,a_{n+1} x^n \\ &= \sum_{n=0}^{\infty} \big(4a_n - (n+1)\,a_{n+1}\big) x^n. \end{aligned}$$

Since this series is always 0 we must have that all of the coefficients are 0. This simply says that

$$a_{n+1} = \frac{4}{n+1}\,a_n \tag{10.14}$$

if $n \geq 0$.

Using induction, or by back tracking, we can show that

$$a_n = \frac{4^n}{n!}\,a_0$$

when $n \geq 1$. With this the function $f(x)$ has the form

$$f(x) = a_0 \sum_{n=0}^{\infty} \frac{4^n}{n!} x^n.$$

With the initial condition $y(0) = 1$ we have $f(0) = 1 = a_0$ and the solution to the differential equation is

$$y(x) = \sum_{n=0}^{\infty} \frac{(4x)^n}{n!}.$$

This is the same function as e^{4x}, the solution we get when we solve the differential equation using separation of variables. (Demonstrate this using the series for e^z with $z = 4x$.)

Exercises

1. Using the Taylor series given in this section, find Taylor series for the following functions

(a) $f(x) = \exp(2x)$

(b) $g(y) = \cos\left(\frac{y}{2}\right)$

(c) $h(w) = \ln\left(2 + 2w^2\right)$

(d) $f(x) = \sinh(4x)$

(e) $g(y) = \cos\left(y^2\right)$

(f) $h(w) = w^3 \sin\left(\frac{w}{3}\right)$

2. Find a Maclaurin series for each of the following functions. What is the radius of convergence?

(a) $f(x) = \arctan(x)$

(b) $g(y) = \dfrac{10}{4+x}$

(c) $h(w) = w^2 \sin(\pi w)$

(d) $f(x) = \cos^2(x)$

(e) $g(y) = \tan(3y)$

(f) $h(w) = \ln(2+x^2)$

3. Find the Taylor series for the given function around the given c. What is the radius of convergence for the Taylor series?

(a) $f(x) = \sin(x)$, $c = \dfrac{\pi}{2}$

(b) $g(y) = \cos(y)$, $c = \dfrac{\pi}{4}$

(c) $h(w) = \exp(w)$, $c = 1$

(d) $f(x) = \ln(x)$, $c = 2$

(e) $g(y) = \arctan(y)$, $c = 1$

(f) $h(w) = \sec(w),\ c = \pi$

4. Find a Taylor series solution to each of the following differential equations.

(a) $\dfrac{dx}{dt} = 6x+1$, $x(0) = 2$

(b) $\dfrac{d^2y}{dx^2} = y$, $y(0) = 2$, $y'(0) = 0$

(c) $\dfrac{dz}{dx} = 2x - z$, $z(0) = -1$

5. Use Theorem 96 on page 376 to show that the series

$$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

converges to $\cos(x)$ for all x.

6. How many terms of the Maclaurin series for e^x does it take to estimate $e = e^1$ to 5 decimal places? You should make the error term in Theorem 96 on page 376 less than 0.5×10^{-5}.

7. Find the series solution for the differential equation $f'(x) = -f(x)$ and $f(0) = 1$ by finding the recurrence relation as in Eq. 10.14 on page 379.

Appendix A
Mathematical Preliminaries

A.1 Order of Operations

One of the things that many beginning calculus students in the United States do not know is a full version of the order of operations. It is very common for students to only learn the order of operations when functions, and possibly exponentiation, is not included.

The goal of having a set order of operation is to give mathematical expressions a unique meaning. This is especially important when using a computer algebra system or a calculator. These devices are not tolerant of sloppy notation since a computer can only interpret what it receives. As a example, consider the expression

$$4/10 * 5.$$

A calculator parses from left to right and will calculate $4/10 = 2/5$ and then multiply by 5 to get 2. If we parse the expression from right to left we get $10 * 5 = 50$ and then divide 4 by 50 to get $2/25$. Hopefully this example makes it clear that the order that we do computations matters.

There is a standard order of operations that is used. Since parentheses are not an operation and simply modify the order of operations by grouping operations, they are omitted from the list. We assume that students know that division is grouped with multiplication since it is multiplying by the reciprocal of a number. Also, we assume that students know that subtraction is grouped with addition since it is adding the negative of a number. Our order of operations is the following:

1. Functions
2. Exponentiation
3. Multiplication and division
4. Addition and subtraction

For those who want a mnemonic for remembering the order, one possibility is **F**red **E**ats **M**any **A**pples. In the United States one might think of the Federal Emergency Management Agency.

Example 406. Consider the expression

$$3x^{\frac{1}{2}}$$

When $x = 4$. Since we do exponentiation first, we first take $4^{1/2} = 2$ and then multiply by 3 to get $3 \cdot 4^{1/2} = 3 \cdot 2 = 6$.

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0

Example 407. The expressions

$$\frac{3x+2}{x-1} \qquad \text{and} \qquad 3x+2/x-1$$

are not the same. If we set $x=2$, the numerator in the first expression is $3\cdot 2+2=6+2=8$. The denominator of the first expression is $2-1=1$. This means that the first expression equals 8 when $x=2$.

In the second expression, when $x=2$, the multiplication and division are done first as $3\cdot 2=6$ and $2/2=1$. Applying the addition and subtraction gives $3x+2/x-1=6+1-1=6$. The two expressions are not equal.

In order to write an equivalent expression for $\frac{3x+2}{x-1}$ on a single line we must change the natural order of operation by using parentheses to force the division to be the last operation. In this case we must use $(3x+2)/(x-1)$.

It is worth noting that the operations of multiplication, division, addition and subtraction all need two expressions, one on either side. The expressions $x*y$, x/y, $x+y$ and $x-y$ all make sense. In the expression $x++y$, the addition signs are acting on each other. This does not make sense and should be avoided. If you want to subtract a negative one, -1, from 5 it is necessary to write $5-(-1)$.

For those who are interested, we can write a negative one as ${}^{-}1$. In this case, as in many calculators, we can write $5-{}^{-}1$ for five minus negative one. Since most people are not that precise in their writing, it is almost always preferable to write $5-(-1)$.

In any science or engineering field there is a common situation where the order of operations is changed according to meaning. In any physical situation the equation must make physical sense. In particular, the units of physical quantities must match. A simple example of this is the ideal gas law, $PV=nRT$. In SI the units of the quantities are, by quantity, P is pressure in Pascals or Newtons per meter squared N/m^2, V is in cubic meters m^3, n is in moles, R is the Boltzmann constant in Joules per degree Kelvin per mole and T is temperature in degrees Kelvin. This means that we can write $n=PV/RT$ in place of $n=PV/(RT)$ without confusion since the only way to have the units of PV/RT be moles is to multiply the R times the T before doing the division.

A.2 Algebra and Functions

Unfortunately, many people teaching calculus believe that large numbers of their students need to review some basic things about computing using algebra and functions. I do not have any good evidence showing how true that perception is. However, I do see students who cannot succeed in a calculus class simply because their "algebra skills" are inadequate. One very simple rule to follow is that if you are breaking the order of operations, you are almost always in trouble. The major exception is when you have additional information about a function.

An unfortunate example is computing with an expression like

$$\frac{3x+5}{5}\,.$$

Sometimes people forget that this expression is the same as $(3x+5)/5$. This means that the addition of $3x$ and 5 is done before the division by 5. Given that correct order of operations, when $x=5$ we get

$$\frac{3x+5}{5}=\frac{3\cdot 5+5}{5}=\frac{20}{5}=4.$$

Teachers get frustrated when students cancel the 5 in the numerator and the 5 in the denominator either as

$$3x + \frac{5}{5} = 3x + 1$$

or as

$$3x + 5 - 5 = 3x.$$

When $x = 5$ these expressions become 16 and 13 respectively. Both are clearly not the same as the original expression.

What we can do is factor an expression out of the numerator and then cancel that with a factor from the denominator. An example of this is the difference quotient, see Eq. 2.1. Here we calculate as follows:

$$\begin{aligned}\frac{(x+h)^2 - x^2}{h} &= \frac{(x^2 + 2xh + h^2) - x^2}{h} \\ &= \frac{2xh + h^2}{h} \\ &= \frac{h \cdot (2x+h)}{h} \\ &= 2x + h.\end{aligned}$$

There is a long list of other things that people teaching calculus want their students to know or to be able to do. The list of things that teachers find some students cannot do, do not do, or do not know is quite long. It includes using parentheses as required, adding fractions, solving quadratics using the quadratic formula, completing the square, solving by elimination, distributing and collecting to simplify, factoring cubic polynomials with an integer root, the basic algebra of functions, and the laws of exponents and logarithms. Since covering these topics usually take years of courses, we will not even attempt to cover all of the important items. Instead we look at a few topics and ask that any student who has some difficulty with a specific topic in the material that is prerequisite to calculus find other resources. There are many texts, often called precalculus texts, and an extensive set of resources on the internet.

A.2.1 Algebra with Functions

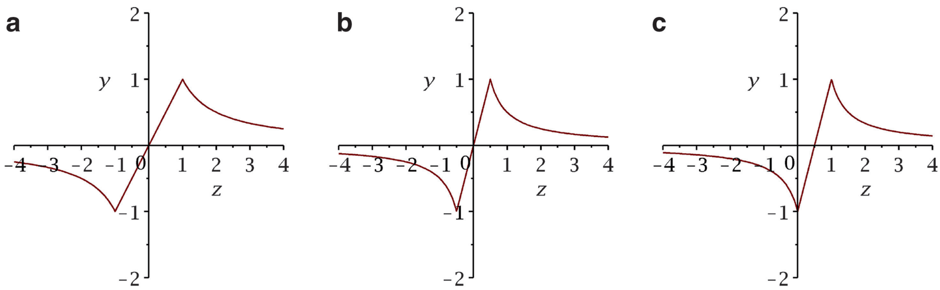
We start with how some basic algebra affects the values and graph of a function. The function we use is a piecewise-defined function that, hopefully, will make some of the changes easier to follow. That function is

$$h(z) = \begin{cases} z & \text{if } z \in [-1, 1] \\ \frac{1}{z} & \text{otherwise.} \end{cases} \tag{A.1}$$

(See Fig. A.1a on page 384.)

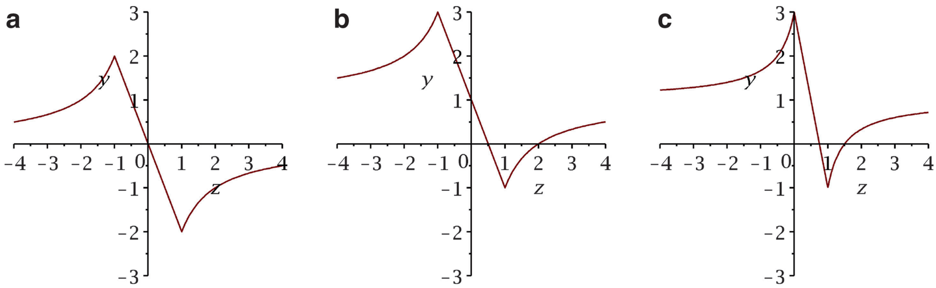
If we alter a function $f(x)$ by multiply the x inside the function by a, we compress, and possibly flip, the graph of $f(x)$ to get the graph of the new function $f(ax)$. The graph of $f(x)$ is compressed, expanded, by a factor of $1/a$. This is illustrated in the graph of $h(2z)$ in Fig. A.1b on page 384. The multiplication by 2 compresses the graph by a factor of $1/2$. This is illustrated by the movement of the corner at $(1,1)$ on the graph of $h(z)$ to the corner on the graph of $h(2z)$ at $(1/2, 1)$.

If we next add 1 to the $2z$ in $h(2z)$ we get the function, $h(2z-1)$. That function is graphed in Fig. A.1c on page 384. The graph of $h(2z-1)$ is the translate of the graph of $h(2z)$ one half



The graph of $h(z)$ from Equation A.1 — The graph of $h(2z)$ — The graph of $h(2z-1)$

Fig. A.1



The graph of $-2h(z)$ — The graph of $-2h(z)+1$ — The graph of $-2h(2z-1)+1$

Fig. A.2

unit to the right. This translation is to the right since $2z-1$ takes $z = 1/2$ to 0 and h is then applied to 0.

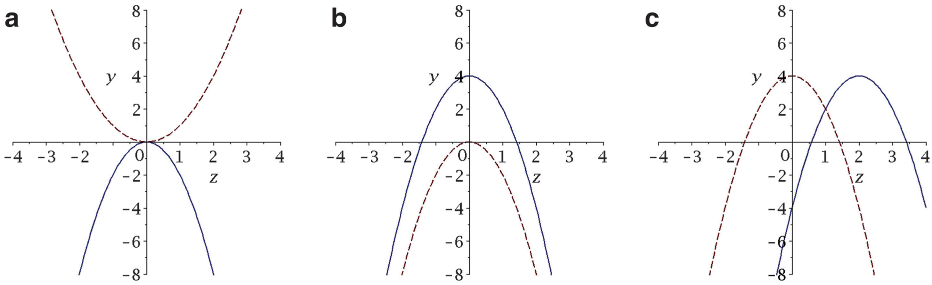
When we combine these two operations on the input of a function $f(x)$ to get $f(ax+b)$ we compress the graph of $f(x)$ by a factor of $1/a$ and then shift the graph b/a units to the left.

We now consider what happens when we manipulate the output of a function $f(x)$ to get $c \cdot f(x) + d$. The c vertically expands the graph of by a factor of c. The graph of $-2h(z)$ is in Fig. A.2a on page 384. If we add 1 to the output of $2h(z)$ we move the graph of $-2h(z)$ vertically by 1 and get the graph of $-2h(z)+1$ as in Fig. A.2b on page 384.

The general rules are that multiplying a function by c expands the graph of the function vertically by a factor of c and adding d to a function moves the graph of vertically by d. When we combine all of these operation on $h(z)$ to get $-2h(2z-1)+1$ we get the graph in Fig. A.2c on page 384.

A.2.2 Quadratic Polynomials

Since my colleagues and I find some students who do not know, or cannot use, the quadratic formula and the associated process of completing the square, that is next. The quadratic formula is used to find the zeroes of a quadratic polynomial and completing the square is important in techniques of integration.



The graphs of x^2 (dashed line) and a_1x^2

The graphs of a_1x^2 (dashed line) and $a_1x^2+c_1$

The graphs of $a_1x^2 + c_1$ (dashed line) and $a_1(x-b_1)^2 + c_1$

Fig. A.3

The graph of any quadratic function can be obtained by starting with the graph of the simplest quadratic, $f(x)=x^2$, We then multiply the quadratic by a nonzero a_1 to narrow, expand and/or flip the graph getting $f_1(x)=a_1x^2$. Next, the graph is moved up or down by adding a c_1 giving $f_2(x)=a_1x^2+c_1$. Finally the graph is moved horizontally by adding $-b_1$ to the input x. This gives the form $q(x)=a_1(x-b_1)^2+c_1$. This process is illustrated in Fig. A.3 on page 385 where $a_1=-2$, $b_1=2$ and $c_1=4$. In each graph the original graph is a dashed curve and the resulting graph is a solid curve.

The roots of the quadratic $f_2(x)=a_1x^2+c_1$ are, using some easy algebra, at $x=\pm\sqrt{-c_1/a_1}$ if the quantity c_1/a_1 is not positive. If $c_1/a_1>0$, there are no roots. (You should be able to explain this using the shift of the graph of $f_1(x)$ vertically.) Since subtracting b_1 from the input of $f_2(x)$ shifts the graph to the right b_1 units, the roots of $g(x)$ are at $x=b_1\pm\sqrt{-c_1/a_1}$ when $q(x)$ has roots. Note that when $a_1=1$, $q(x)=(x-b_1)^2+c_1$ we have $x=b_1\pm\sqrt{-c_1}$

The difficulty is in translating an arbitrary quadratic of the form ax^2+bx+c to the form $a_1(x-b_1)^2+c_1$. This is accomplished by completing the square. Expanding the quadratic $a_1\cdot(x-b_1)^2+c_1$ and setting that equal to ax^2+bx+c we get

$$ax^2+bx+c=a_1x^2-2a_1b_1x+a_1b_1^2+c_1.$$

Since polynomials are equal if and only if their coefficients are equal, the x^2 coefficients are equal, $a_0=a$. We assume that $a=1$ since we can factor a out of both quadratics without changing the roots. This is equivalent to replacing b with b/a, c with c/a, b_1 with b_1/a and c_1 with c_1/a.

This means that we want to write x^2+bx+c in the form $(x-b_1)^2+c_1$. Expanding the right side of this equation yields

$$x^2+bx+c=x^2-2b_1x+b_1^2+c_1.$$

Solving for b and c gives $b=b_1/2$ and $c=c_1+b_1^2$.

Example 408. To complete the square for the quadratic x^2+4x+5 we want to write

$$x^2+4x+5=(x-b_1)^2+c=x^2-2b_1x+b_1^2+c_1.$$

Setting the x coefficients equal gives $-2b_1=-4$ or $b_1=-2$ and setting the constants equal gives $5=(-2)^2+c_1$ or $c_1=1$. From this we get

$$x^2+4x+5=\big(x-(-2)\big)^2+1=(x+2)^2+1.$$

This quadratic has no roots since $(x+2)^2\geq 0$ for all x and $1>0$.

The quadratic formula can now be obtained using the formula $x = b_1 \pm \sqrt{-c_1}$ for the roots of the quadratic $q(x) = (x - b_1)^2 + c_1$. To obtain this form from the form $ax^2 + bx + c$ we replace b_1 with $b/(2a)$ and c_1 with

$$-\left(\frac{b}{2a}\right)^2 + \frac{c}{a} = -\frac{b^2 - 4ac}{4a^2}$$

to get the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{A.2}$$

The expression $b^2 - 4ac$ is call the *discriminant*. If the discriminant is positive, the quadratic has two real roots. If the discriminant is negative, the quadratic has no real roots. And if the discriminant is zero, the quadratic has a double root.

Example 409. Consider the quadratic function $q(t) = 3t^2 + t - 5$. Using the quadratic formula we find that the roots of this quadratic are

$$\begin{aligned} t &= \frac{-1 \pm \sqrt{1^1 - 4 \cdot 3 \cdot (-5)}}{2 \cdot 3} \\ &= \frac{-1}{6} \pm \frac{\sqrt{61}}{6}. \end{aligned}$$

Example 410. The discriminant of the quadratic $q(t) = 3t^2 + t + 1$ is $1^2 - 4 \cdot 3 \cdot 1 = -11$. Since the discriminant is negative, the quadratic has no real roots.

A.2.3 Solving Equations and Eliminating Variables

When working with many problems it is appropriate to find the solutions to an equation or to write the solutions to an equation where one variable is in terms of other variables. Examples of this can be found in the sections on related rates, Sect. 5.3, and the section on applied optimization, Sect. 5.8.

We start with some basics for factoring polynomials. There are two main theorems that are useful when factoring polynomials. The Fundamental Theorem of Algebra states that any polynomial of degree n, $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, has n complex numbers as roots. These roots may be repeated to get the total count of n.

Example 411. The polynomial $p(x) = x^6 + x^4 - x^2 - 1$ is a polynomial of degree 6 that has 6 roots, including repeated roots, $x = -i, -i, i, i, -1, 1$.

If we want to find the roots of a polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, the first thing people usually try is to use the quadratic formula or to factor the polynomial. There are formulas for finding the roots of cubic and quintic polynomials. These are used infrequently because the formulas are not easy to remember. Finding the rational roots of $p(x)$ if $a_0, a_1, \ldots, a_n$ are integers is not difficult because of the Rational Root Theorem. A proof of this result is simple but omitted.

Theorem 97 (Rational Root Theorem). *Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial with integer coefficients. If $x = q/r$ is a rational root of $p(x)$ written in lowest terms, then q is a factor of a_0 and r is a factor of a_n. These factors can be either positive or negative.*

Example 412. Consider the polynomial $p(x) = 4x^4 - 5x^2 - 10x + 6$. Since the factors of 4 are $\pm 1, \pm 2$ and ± 4 and the factors of 6 are $\pm 1, \pm 2, \pm 3$ and ± 6, the possible rational roots of $p(x)$ are

$$-6, -3, -2, -1, -\frac{3}{2}, -\frac{3}{4}, -\frac{1}{2}, -\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{3}{2}, 1, 2, 3, \text{ and } 6\ .$$

Plugging all of these values into $p(x)$ we get that $p(1/2)$ and $p(3/2)$ are both zero. None of the other possible x values are roots. Dividing $p(x)$ by $(2x-1)\cdot(2x-3)$ gives $x^2 + 2x + 2$. The discriminant of this quadratic is $4 - 4 \cdot 1 \cdot 2 < 0$. This means that the quadratic does not have any real roots and the only real factors of $p(x)$ are $x = 1/2$ and $x = 3/2$.

If we are trying to find the roots of other functions besides polynomials the techniques are mostly ad hoc. Often it is necessary to know the input values for a given function that give the desired output value. Sometimes the functions where we can do this are exponential, logarithmic or trigonometric functions. In the next section of this appendix is a list of values of trigonometric functions that are used this way. You should learn these values. In the following example we see how this can work in a simple case.

Example 413. Consider the equation $e^{\cos(x)+1/2} - 1 = 0$. Rewriting this in the form $e^{\cos(x)+1/2} = 1$ we need to have $\cos(x) + 1/2 = 0$ since the only value that makes $e^z = 1$ is $z = 0$.

Rewriting again we must have $\cos(x) = -1/2$. We have $\cos(x) = -1/2$ when either $x = 2\pi/3 + 2k\pi$ or $x = -2\pi/3 + 2k\pi$ for any integer k. These are all of the solutions of the equation.

Another useful situation is finding the zeros of a function that factors. Since a product is zero if and only if one of the factors is zero, we only need to find the zeros of the factors. The next example demonstrates how this in implemented.

Example 414. We want to find the zeros of $h(z) = \left(e^{\cos(z)-1} - 1\right)\left(z^2 - 4z + 3\right)$. To do this we find the zeros of $e^{\cos(z)-1} - 1$ and the zeros of $z^2 - 4z + 3$.

Since $e^w = 1$ only if $w = 1$, the zeros of the first factor are all points z where $\cos(z) = 1$. These are the points $z = 2k\pi$, where k is any integer.

Since $z^2 - 4z + 3 = (z-4)(z-1)$, the zeros of the second factor are $z = 4$ and $z = 1$. This means that all of the solutions of $h(z) = 0$ are $z = 1, 4$ and $2k\pi$ for $k = 0, \pm 1, \pm 2, \ldots$.

The last idea in this section is solving systems of equations by eliminating variables. This has been studied extensively and much is known about how to do this. It can be a very difficult or impossible task for a given set of equations. We restrict our study to a couple simple cases that involve only two variable.

The first case is one which most students should have seen, solving two linear equations in two unknowns. In this case, if the system can be solved, it is easy to solve one equation for one of the two variables. This expression can than be used to eliminate one variable from the second equation, which is easily solved for the second variable.

For example, to solve

$$\begin{aligned} 2x + 3z &= 2 \\ x - z &= 1 \end{aligned}$$

we can solve the second equation for x,

$$x = 1 - z.$$

Then x is replaced by $1-z$ in the first equation to give

$$2(1-z)+3z=2$$

or $z=0$.

If we set $z=0$ in either of the two original equations, we see that $x=1$.

There are three possible outcomes when we have a system of linear equation to solve: there may be a unique solution as in the example above, there may be no solutions or there may be an infinite number of solutions. Students can consult any elementary linear algebra text to learn how this works.

Using substitution for solving other systems of equations is usually more difficult, but follows the same pattern. The basic idea is to eliminate variables or expressions using substitution. We have two examples, one that is fairly simple and one that is typical of system of equations found in Lagrange multiplier problems from multivariate calculus.

Example 415. The problem is to find all pairs (x,y) such that

$$\begin{aligned} xy &= 2 \text{ and} \\ x^2+4y^2 &= 8 \ . \end{aligned}$$

First we note that neither x nor y can be zero. This means that we can solve the first equation for y to get $y=2/x$ and substitute this into the second equation to obtain

$$x^2+\frac{16}{x^2}-8=0 \ .$$

Rewriting this as the quadratic equation in x^2, $\left(x^2\right)^2-8x^2+16=0$, we can solve for x^2.

In this case the quadratic factors as $(x^2-4)^2=0$ or $x^2=4$. Here $x=\pm 2$. If $x=2$ we have $y=1$ and if $x=-2$ we have $y=-1$.

The next example is somewhat harder, but it is not very hard.

Example 416. We want to solve the following system of three equations in three unknowns:

$$\begin{aligned} 2x+zy &= 0, \\ 2y+zx &= 0 \quad \text{and} \\ xy &= 1 \ . \end{aligned}$$

There are a number of different ways to approach this problem. For example, we could multiply the first equation by x, multiply the second equation by y and then subtract the two equations. This would eliminate the z from the equations and leave us with two equations in two unknowns to solve.

The approach we take is to solve the first equation for z and substitute that into the second equation. Solving the first equation for z yields $z=-2x/y$. Replacing z in the second equation with $-2x/y$ gives

$$2y-2\frac{x^2}{y}=0$$

or $y^2=x^2$.

This last equation says that either $y=x$ or $y=-x$. In the later case, when we substitute for y in the third original equation, we have $-x^2=1$. Since there are no real numbers that satisfy this equation, we do not get any solutions where $y=-x$.

If we assume that $y = x$, the third original equation becomes $x^2 = 1$. The two solutions to this equation are $x = 1$ and $x = -1$. Putting $x = 1$ back into $xy = 1$ gives $y = 1$ and putting $x = y = 1$ back into $z = -2x/y$ gives $z = -2$. This means that one solution to the system of equations is $x = 1$, $y = 1$ and $z = -2$.

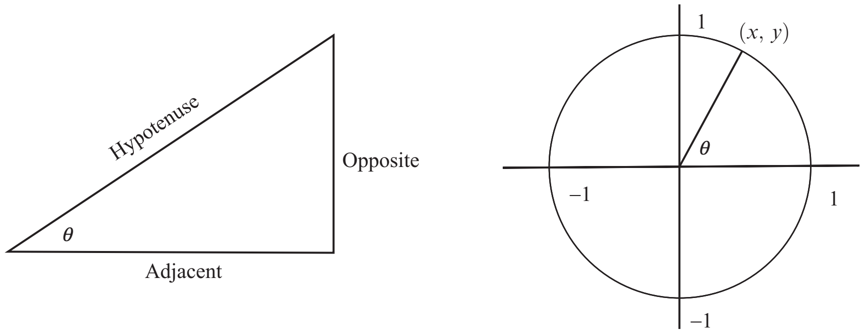


Fig. A.4 An illustration of the triangle and unit circle definitions of $\sin(\theta)$ and $\cos(\theta)$

Putting $x = -1$ back into $xy = 1$ gives $y = -1$ and putting $x = y = -1$ back into $z = -2x/y$ gives $z = -2$. This means that the second solution to the system of equations is $x = -1$, $y = -1$ and $z = -2$. Since we have not done any operations that might remove solutions to the system of equations from consideration, these are the only two solutions.

Although we have only looked at a few systems of equations to solve, the important thing is to realize that solving systems is not easy and it takes practice to master techniques for solving systems of equations.

A.3 Trigonometry

This is a short refresher on trigonometry. Trigonometry can be viewed in two different ways. One is as functions involving ratios of the lengths of sides of a right triangle with a hypotenuse of length one. The other viewpoint involves the coordinates of a point on the unit circle. This is done as functions of the counter clockwise length of the arc from the x-axis to the point on the unit circle. The angle measure use for the second viewpoint is radians, the length of the arc from the x-axis to the point divided by the radius of the circle. In this appendix it will usually be the length of the arc since we usually will have a circle of radius 1. See Fig. A.4 for an illustration of this.

The two basic trigonometric functions are the *sine* and the *cosine* of an angle. In a right triangle where the measure of an angle besides the right angle is θ, the sine of θ is the length of side opposite θ divided by the length of the hypotenuse. Similarly, the cosine of θ is the length of side adjacent to θ divided by the length of the hypotenuse. If the length of the hypotenuse is 1, they are simply the length of opposite and the adjacent sides of the triangle. This is illustrated in Fig. A.5 on page 390.

When using the unit circle to define the sine and cosine we take the same idea used for triangles and take the sine of θ, $\sin(\theta)$, to be the y coordinate of the point corresponding to θ and the cosine of θ, $\cos(\theta)$, to be the x coordinate of the point corresponding to θ. Using this definition we see that, unlike the triangle definition, $\sin(\theta)$ and $\cos(\theta)$ can take on both positive and negative values. This is illustrated in Fig. A.6 on page 390.

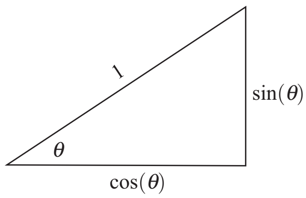


Fig. A.5 A right triangle with $\sin(\theta)$ and $\cos(\theta)$ labeled

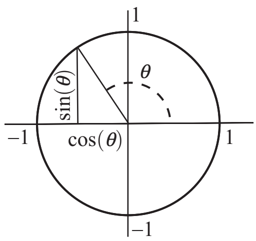


Fig. A.6 A unit circle with $\sin(\theta)$ and $\cos(\theta)$ labeled

Function	Notation	Triangle formulation	Circle formulation
tangent	$\tan(\theta)$	$\frac{\text{Opposite}}{\text{Adjacent}}$	$\frac{y}{x}$
cotangent	$\cot(\theta)$	$\frac{\text{Adjacent}}{\text{Opposite}}$	$\frac{x}{y}$
secant	$\sec(\theta)$	$\frac{\text{Hypotenuse}}{\text{Adjacent}}$	$\frac{1}{x}$
cosecant	$\csc(\theta)$	$\frac{\text{Hypotenuse}}{\text{Opposite}}$	$\frac{1}{y}$

Table A.1 The definitions of the trigonometric functions

Given these definitions for sine and cosine we define tangent, cotangent, secant and cosecant. The definitions of these functions are in Table A.1 on page 390. *From this point forward all of the trigonometric functions are assumed to be defined through the circle definition.* Unlike the functions sine and cosine, each of these four functions is not defined for all real numbers. The domain of each is restricted since the denominator of each expression can be 0.

The values of each of the six trigonometric functions are known exactly for some values of θ. We will not show how to get these values. Instead we state the values for θ in the first quadrant in Table A.2 on page 391. Using the values of sine and cosine in the first quadrant, we can find the values of sine and cosine of any angle. Figure A.7 on page 391 is the picture used to see how this is done. What the figure tells us is that $\sin(-\theta) = -\sin(\theta)$ ($\sin(\theta)$ is odd), $\cos(-\theta) = \cos(\theta)$ ($\cos(\theta)$ is even), $\cos(\pi - \theta) = -\cos(\theta)$, $\sin(\pi - \theta) = \sin(\theta)$, $\cos(\theta + \pi) = -\cos(\theta)$ and $\sin(\theta + \pi) = -\sin(\theta)$.

Example 417. Let $\theta = \pi/6$. From Fig. A.7 on page 391 we have that $\cos(5\pi/6) = -\cos(\pi/6) = -\sqrt{3}/2$ and $\sin(5\pi/6) = \sin(\pi/6) = 1/2$. This means that

$$\begin{aligned}\tan\left(\frac{5\pi}{6}\right) &= \frac{\sin\left(\frac{5\pi}{6}\right)}{\cos\left(\frac{5\pi}{6}\right)} \\ &= -\frac{1}{\sqrt{3}}.\end{aligned}$$

Angle in radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\sin(\theta)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\cos(\theta)$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\tan(\theta)$	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	–
$\cot(\theta)$	–	$\sqrt{3}$	1	$\frac{1}{\sqrt{3}}$	0
$\sec(\theta)$	1	$\frac{2}{\sqrt{3}}$	$\frac{2}{\sqrt{2}}$	2	–
$\csc(\theta)$	–	2	$\frac{2}{\sqrt{2}}$	$\frac{2}{\sqrt{3}}$	1

Table A.2 Values for the standard trigonometric functions in the first quadrant

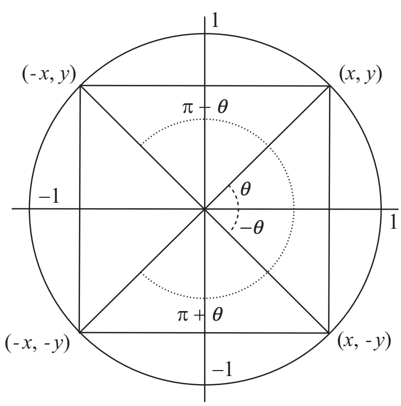


Fig. A.7 The relationships between points and quadrants for finding trigonometric functions of angles

Using Fig. A.8 on page 392 we can see that going 2π in either direction around the unit circle puts us back at the same point on the unit circle. This means that, when they are defined, all six of our trigonometric functions have the same value at θ and at $\theta + 2k\pi$ for any integer k. In other terms, each of the trigonometric functions is periodic.

The graphs of the six standard trigonometric functions are in Figs. A.9 on page 392, A.10 on page 392 and A.11 on page 393. All calculus students using this text should know all of these graphs.

To work with trigonometric functions in calculus we need to have a good number of trigonometric identities in our tool kit. A few of them will be illustrated graphically, but many of them must simply be learned and applied. The first is the Pythagorean Theorem 98. It is essential to know this theorem.

Theorem 98 (Pythagorean Theorem). *Let T be a right triangle with side lengths A, B and C with side C opposite the right angle. Then*

$$C^2 = A^2 + B^2. \tag{A.3}$$

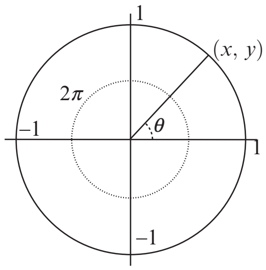


Fig. A.8 How $\sin(\theta \pm 2\pi) = \sin(\theta)$

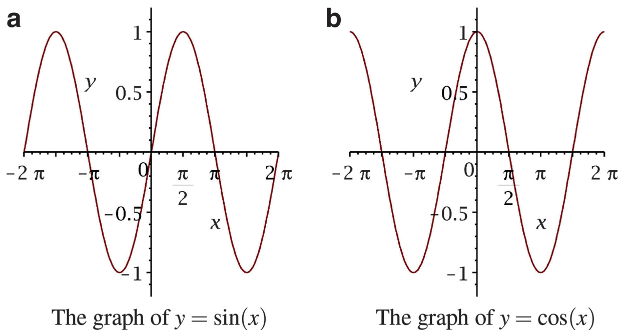


Fig. A.9

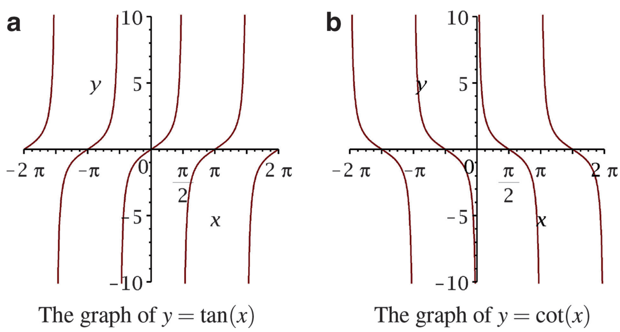


Fig. A.10

In terms of a right triangle with hypotenuse length 1 and θ *one of the angles with radian measure less than* $\pi/2$ *or in terms of* $\sin(\theta)$ *and* $\cos(\theta)$ *as defined on the unit circle*

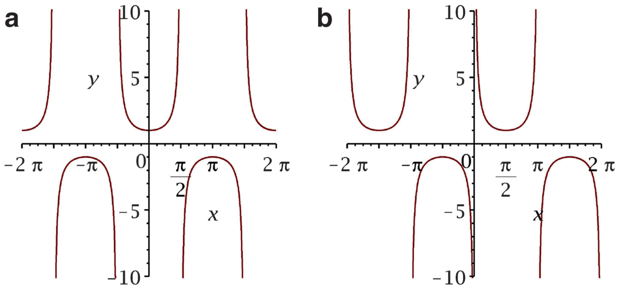
$$\sin^2(\theta) + \cos^2(\theta) = 1.$$

Example 418. The length of the hypotenuse of a right triangle with leg lengths 4 and 3 is

$$C = \sqrt{3^2 + 4^2} = 5.$$

Example 419. If the cosine of an angle θ is $1/6$ then

$$\sin^2(\theta) = 1 - \cos^2(\theta) = \frac{35}{36}.$$



The graph of $y = \sec(x)$ The graph of $y = \csc(x)$

Fig. A.11

This means that

$$\sin(\theta) = \pm\frac{\sqrt{35}}{6}.$$

The sign of $\sin(\theta)$ depends on which quadrant θ is in. If it is in the first quadrant then $\sin(\theta) = \sqrt{35}/6$. If θ is in the fourth quadrant then $\sin(\theta) = -\sqrt{35}/6$.

There are some simple identities about translations of the trigonometric functions by multiples of $\pi/2$. First we can see from Fig. A.9 on page 392 that the graph of $\cos(\theta)$ can be obtained from translating the graph of $\sin(\theta)$ $\pi/2$ units to the left. This means that

$$\cos(\theta) = \sin\left(\theta + \frac{\pi}{2}\right). \tag{A.4}$$

Replacing θ by $\theta - \pi/2$ in Eq. A.4 gives $\sin(\theta) = \cos(\theta - \pi/2)$. Using the fact that $\cos(\theta)$ is even, we have, by replacing θ with $-\theta$ in Eq. A.4, $\cos(\theta) = \sin(\pi/2 - \theta)$.

Example 420. If $\theta = \pi/6$ and $\sin(\pi/6) = 1/2$ then

$$\begin{aligned}\cos\left(\frac{2\pi}{3}\right) &= \cos\left(\frac{\pi}{2} - \frac{\pi}{6}\right)\\ &= \sin\left(\frac{\pi}{6}\right)\\ &= \frac{1}{2}.\end{aligned}$$

Law of Sines and the Law of Cosines. Both of these refer to triangles where A, B and C are lengths of the sides of a triangle and a, b and c are the angles opposite those sides. The Law of Sines states that

$$\frac{\sin(a)}{A} = \frac{\sin(b)}{B} = \frac{\sin(c)}{C} \tag{A.5}$$

and the Law of Cosines states that

$$C^2 = A^2 + B^2 - 2AB\cos(c). \tag{A.6}$$

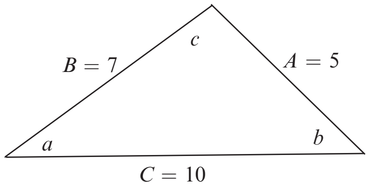


Fig. A.12

Example 421. Consider a triangle with side lengths $A = 5$, $B = 7$ and $C = 10$. We want to find information to get measures for the three angles opposite the sides: a, b and c. This triangle is illustrate in Fig. A.12 on page 394. In this case we will get the sines of the angles. By the Law of Cosines, the cosine of the angle opposite the side of length 10 satisfies

$$10^2 = 7^2 + 5^2 - 2 \cdot 7 \cdot 5 \cdot \cos(c).$$

This means that $\cos(c) = 100 - 49 - 25/70 = 13/35$. Then

$$\sin(c) = \sqrt{1 - (13/35)} = \sqrt{1{,}056/1{,}225}.$$

Using the Law of Sines, $\sin(a) = 12\sqrt{1{,}056/1{,}225}$ and $\sin(b) = 7/10\sqrt{1{,}056/1{,}225}$.

Some of the most commonly used trigonometric identities used in a calculus course are the identities involving sines and cosines of sums of angles. The following are the two basic sum identities for sine and cosine. We will not prove them here.

$$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta) \quad \text{and} \tag{A.7}$$
$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta). \tag{A.8}$$

Setting $\alpha = \beta$ in the two identities above we get the double angle formulas

$$\sin(2\alpha) = 2\sin(\alpha)\cos(\alpha) \quad \text{and} \tag{A.9}$$
$$\cos(2\alpha) = \cos^2(\alpha) - \sin^2(\alpha). \tag{A.10}$$

Replacing $\cos^2(\alpha)$ with $1 - \sin^2(\alpha)$ or replacing $\sin^2(\alpha)$ with $1 - \cos^2(\alpha)$ in Eq. A.10 we get versions of the half angle formulas

$$\sin^2(\alpha) = \frac{1 - \cos(2\alpha)}{2} \quad \text{and} \tag{A.11}$$
$$\cos^2(\alpha) = \frac{1 + \cos(2\alpha)}{2}. \tag{A.12}$$

Example 422. We can use Eq. A.12 in the form $\cos(\theta) = \pm\sqrt{(1 + \cos(2\alpha))/2}$ to find the cosine of $\pi/12$.

$$\cos\left(\frac{\pi}{12}\right) = \sqrt{\frac{1 + \cos\left(\frac{\pi}{6}\right)}{2}} = \sqrt{\frac{1 + \frac{\sqrt{3}}{2}}{2}}.$$

Appendix B
Tables

B.1 Table of Some Common Integrals

1. $\displaystyle\int a\,dx = ax$

2. $\displaystyle\int x\,dx = \frac{x^2}{2}$

3. $\displaystyle\int x^n\,dx = \frac{x^{n+1}}{n+1}$ if $n \neq -1$

4. $\displaystyle\int \frac{1}{x}\,dx = \ln|x|$

5. $\displaystyle\int \sin(x)\,dx = -\cos(x)$

6. $\displaystyle\int \cos(x)\,dx = \sin(x)$

7. $\displaystyle\int \tan(x)\,dx = -\ln|\cos(x)|$

8. $\displaystyle\int \cot(x)\,dx = \ln|\sin(x)|$

9. $\displaystyle\int \sec(x)\,dx = \ln|\sec(x)+\tan(x)|$

10. $\displaystyle\int \csc(x)\,dx = -\ln|\csc(x)+\cot(x)|$

11. $\displaystyle\int \frac{1}{\sqrt{a^2-x^2}}\,dx = \arcsin\left(\frac{x}{a}\right)$

12. $\displaystyle\int \frac{1}{x\sqrt{x^2-a^2}}\,dx = \frac{1}{a}\,\mathrm{arcsec}\left(\frac{x}{a}\right)$

13. $\displaystyle\int \frac{1}{a^2+x^2}\,dx = \frac{1}{a}\arctan\left(\frac{x}{a}\right)$

14. $\displaystyle\int \frac{1}{x^2-a^2}\,dx = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right|$

15. $\displaystyle\int e^x\,dx = e^x$

16. $\displaystyle\int a^x\,dx = \frac{a^x}{\ln(a)}$

17. $\displaystyle\int x e^x\,dx = (x-1)\,e^x$

18. $\displaystyle\int c\,f(x)\,dx = c\int f(x)\,dx$

19. $\displaystyle\int u\,dv = uv - \int v\,du$

20. $\displaystyle\int f'(g(x))\,g'(x)\,dx = f(g(x))$

21. $\displaystyle\int \mathbf{F}(x)\,dx = \left(\int f_1(x)\,dx, \int f_2(x)\,dx, \ldots, \int f_n(x)\,dx\right)$

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0

B.2 Table of Some Common Derivatives

1. $\dfrac{d}{dx}a = 0$

2. $\dfrac{d}{dx}x = 1$

3. $\dfrac{d}{dx}x^\alpha = \alpha x^{\alpha-1}$

4. $\dfrac{d}{dx}\sin(x) = \cos(x)$

5. $\dfrac{d}{dx}\cos(x) = -\sin(x)$

6. $\dfrac{d}{dx}\tan(x) = \sec^2(x)$

7. $\dfrac{d}{dx}\cot(x) = -\csc^2(x)$

8. $\dfrac{d}{dx}\sec(x) = \sec(x)\tan(x)$

9. $\dfrac{d}{dx}\csc(x) = -\csc(x)\cot(x)$

10. $\dfrac{d}{dx}\arcsin(x) = \dfrac{1}{\sqrt{1-x^2}}$

11. $\dfrac{d}{dx}\operatorname{arcsec}(x) = \dfrac{1}{x^2\sqrt{1-\frac{1}{x^2}}}$

12. $\dfrac{d}{dx}\arctan(x) = \dfrac{1}{1+x^2}$

13. $\dfrac{d}{dx}\ln(x) = \dfrac{1}{x}$

14. $\dfrac{d}{dx}e^x = e^x$

15. $\dfrac{d}{dx}a^x = \ln(a)\,a^x$

16. $\dfrac{d}{dx}c\,f(x) = c\dfrac{d}{dx}f(x)$

17. $\dfrac{d}{dx}(f(x)+g(x)) = f'(x)+g'(x)$

18. $\dfrac{d}{dx}f(g(x)) = f'(g(x))\,g'(x)$

19. $\dfrac{d}{dx}\dfrac{f(x)}{g(x)} = \dfrac{f'(x)\,g(x)-f(x)\,g'(x)}{g^2(x)}$

20. $\dfrac{d}{dx}(f(x)\,g(x)) = f'(x)\,g(x)+f(x)\,g'(x)$

21. $\dfrac{d}{dx}\mathbf{f}(x) = \left(\dfrac{d}{dx}f_1(x), \dfrac{d}{dx}f_2(x), \ldots, \dfrac{d}{dx}f_n(x)\right)$

22. $\dfrac{d}{dx}\mathbf{f}(x)\cdot\mathbf{g}(x) = \mathbf{f}'(x)\cdot\mathbf{g}(x)+\mathbf{f}(x)\cdot\mathbf{g}'(x)$

23. $\dfrac{d}{dx}\mathbf{f}(x)\times\mathbf{g}(x) = \mathbf{f}'(x)\times\mathbf{g}(x)+\mathbf{f}(x)\times\mathbf{g}'(x)$

Index

J.S. Treiman, *Calculus with Vectors*, Springer Undergraduate Texts
in Mathematics and Technology, DOI 10.1007/978-3-319-09438-0