Gerd Baumann

Mathematics for Engineers I

Basic Calculus

With CD-ROM

Oldenbourg

Mathematics for Engineers I

Basic Calculus

by
Gerd Baumann

Oldenbourg Verlag München

Prof. Dr. Gerd Baumann is head of Mathematics Department at German University Cairo (GUC). Before he was Professor at the Department of Mathematical Physics at the University of Ulm.

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
oldenbourg.de

Editor: Kathrin Mönch
Producer: Anna Grosser
Cover design: Kochan & Partner, München
Printed on acid-free and chlorine-free paper
Printing: Druckhaus „Thomas Müntzer“ GmbH, Bad Langensalza

ISBN 978-3-486-59038-8

Preface

Theory without Practice is empty,
Practice without Theory is blind.

The current text *Mathematics for Engineers* is a collection of four volumes covering the first three up to the fifth terms in undergraduate education. The text is mainly written for engineers but might be useful for students of applied mathematics and mathematical physics, too.

Students and lecturers will find more material in the volumes than a traditional lecture will be able to cover. The organization of each of the volumes is done in a systematic way so that students will find an approach to mathematics. Lecturers will select their own material for their needs and purposes to conduct their lecture to students.

For students the volumes are helpful for their studies at home and for their preparation for exams. In addition the books may be also useful for private study and continuing education in mathematics. The large number of examples, applications, and comments should help the students to strengthen their knowledge.

The volumes are organized as follows: Volume I treats basic calculus with differential and integral calculus of single valued functions. We use a systematic approach following a bottom-up strategy to introduce the different terms needed. Volume II covers series and sequences and first order differential equations as a calculus part. The second part of the volume is related to linear algebra. Volume III treats vector calculus and differential equations of higher order. In Volume IV we use the material of the previous volumes in numerical applications; it is related to numerical methods and practical calculations. Each of the volumes is accompanied by a CD containing the *Mathematica* notebooks of the book.

As prerequisites we assume that students had the basic high school education in algebra and geometry. However, the presentation of the material starts with the very elementary subjects like numbers and introduces in a systematic way step by step the concepts for functions. This allows us to repeat most of the material known from high school in a systematic way, and in a broader frame. This way the reader will be able to use and categorize his knowledge and extend his old frame work to a new one.

The numerous examples from engineering and science stress on the applications in engineering. The idea behind the text concept is summarized in a three step process:

Theory → Examples → Applications

When examples are discussed in connection with the theory then it turns out that the theory is not only valid for this specific example but useful for a broader application. In fact, usually theorems or a collection of theorems can even handle whole classes of problems. These classes are sometimes completely separated from this introductory example; e.g. the calculation of areas to motivate integration or the calculation of the power of an engine, the maximal height of a satellite in space, the moment of inertia of a wheel, or the probability of failure of an electronic component. All these problems are solvable by one and the same method, integration.

However, the three-step process is not a feature which is always used. Some times we have to introduce mathematical terms which are used later on to extend our mathematical frame. This means that the text is not organized in a historic sequence of facts as traditional mathematics texts. We introduce definitions, theorems, and corollaries in a way which is useful to create progress in the understanding of relations. This way of organizing the material allows us to use the complete set of volumes as a reference book for further studies.

The present text uses *Mathematica* as a tool to discuss and to solve examples from mathematics. The intention of this book is to demonstrate the usefulness of *Mathematica* in everyday applications and calculations. We will not give a complete description of its syntax but demonstrate by examples the use of its language. In particular, we show how this modern tool is used to solve classical problems and to represent mathematical terms.

We hope that we have created a coherent way of a first approach to mathematics for engineers.

Acknowledgments Since the first version of this text, many students made valuable suggestions. Because the number of responses are numerous, I give my thanks to all who contributed by remarks and enhancements to the text. Concerning the historical pictures used in the text, I acknowledge the support of the http://www-gapdcs.st-and.ac.uk/~history/ webserver of the University of St Andrews, Scotland. The author deeply appreciates the understanding and support of his wife, Carin, and daughter, Andrea, during the preparation of the books.

Cairo

Gerd Baumann

To Carin and Andrea

Contents

1 Outline

1.1. Introduction

During the years in circles of engineering students the opinion grew that calculus and higher mathematics is a simple collection of recipes to solve standard problems in engineering. In addition, students believe that a lecturer is responsible to convey these recipes to them in a nice and smooth way so that they can use it as it is done in cooking books. This approach of thinking has the common short-coming that with such kind of approach only standard problems are solvable of the same type which do not occur in real applications.

We believe that calculus for engineers offers a great wealth of concepts and methods which are useful in modelling engineering problems. The reader should be aware that this collection of definitions, theorems, and corollaries is not the final answer of mathematics but a first approach to organize knowledge in a systematic way. The idea is to organize methods and knowledge in a systematic way. This text was compiled with the emphasis on understanding concepts. We think that nearly everybody agrees that this should be the primary goal of calculus instruction.

This first course of Engineering Mathematics will start with the basic foundation of mathematics. The basis are numbers, relations, functions, and properties of functions. This first chapter will give you tools to attack simple engineering problems in different fields. As an engineer you first have to understand the problem you are going to tackle, and after that you will apply mathematical tools to solve the problem. These two steps are common to any kind of problem solving in engineering as well as in science. To understand a problem in engineering you need to be able to use and apply engineering knowledge and engineering procedures. To solve the related mathematical problem needs the knowledge of the basic steps how mathematics is working. Mathematics gives you the frame to handle a problem in a systematic way and to use the procedure and knowledge of mathematical methods to derive a solution. Since mathematics sets up the frame for the solution of a problem you should be able to use it efficiently. It is not to apply recipes to solve a problem but to use the appropriate concepts to solve it.

Mathematics by itself is for engineers a tool. As for all other engineering applications working with tools you must know how they act and react in applications. The same is true for mathematics. If you know how a mathematical procedure (tool) works and how the components of this tool are connected by each other you will understand its application. Like engineering tools, mathematical tools consist of components. Each component is usually divisible into other components until the basic component (elements) are found. The same idea is used in mathematics. There are basic elements from mathematics you should know as an engineer. Combining these basic elements we are able to set up a mathematical frame which incorporates all those elements which are needed to solve a problem. In other words, we use always basic ideas to derive advanced structures. All mathematical thinking follows a simple track which tries to apply fundamental ideas used to handle more complicated situations. Once you remember this simple concept you will be able to understand advanced concepts in mathematics as well as in engineering.

1.2. Concept of the Text

Concepts and conclusions are collected in definitions and theorems. The theorems are applied in examples to demonstrate their meaning. Every concept in the text is illustrated by examples, and we included more than 1,000 tested exercises for assignments, class work and home work ranging from elementary applications of methods and algorithms to generalizations and extensions of the theory. In addition, we included many applied problems from diverse areas of engineering. The applications chosen demonstrate concisely how basic calculus mathematics can be, and often must be, applied in real life situations.

During the last 25 years a number of symbolic software packages have been developed to provide symbolic mathematical computations on a computer. The standard packages widely used in academic applications are *Mathematica*®, Maple® and Derive®. The last one is a package which is used for basic calculus while the two other programs are able to handle highly sophisticated calculations. Both *Mathematica* and Maple have almost the same mathematical functionality and are very useful in symbolic calculations. However the author's preference is *Mathematica* because the experience over the last 25 years showed that *Mathematica*'s concepts are more stable than Maple's one. The author used both of the programs and it turned out during the years that programs written in *Mathematica* some years ago still work with the latest version of *Mathematica* but not with Maple. Therefore the book and its calculations are based on a package which is sustainable for the future.

Having a symbolic computer algebra program available can be very useful in the study of techniques used in calculus. The results in most of our examples and exercises have been generated using problems for which exact values can be determined, since this permits the performance of the calculus method to be monitored. Exact solutions can often be obtained quite easily using symbolic computation. In addition, for many techniques the analysis of a problem requires a high amount of laborious steps, which can be a tedious task and one that is not particularly instructive once the techniques of calculus have been mastered. Derivatives and integrals can be quickly obtained symbolically with computer algebra systems, and a little insight often permits a symbolic computation to aid in understanding the process as well.

We have chosen *Mathematica* as our standard package because of its wide distribution and reliability. Examples and exercises have been added whenever we felt that a computer algebra system would be of significant benefit.

1.3. Organization of the Text

The book is organized in chapters which will initially cover the first steps in calculus called pre-calculus. We go on with functions and their properties. Having introduced functions we will discuss how we can operate on functions and which kind of properties function will have. The application of derivatives on functions will lead us to the discussion of optimization problems which are useful in design processes in engineering. Another way to operate on functions is the integration which in its simplest application is related to the determination of areas. Finally we will deal with the problem when we leave the exact representation of functions and look for approximations. After each section there will be a test and exercise subsection divided into two parts. The first part consists of a few test questions which examines the main topics of the previous section. The second part contains exercises related to applications and advanced problems of the material discussed in the previous section. The level of the exercises ranges from simple to advanced.

The whole material is organized in six chapters where the first of this chapter is the current introduction. In Chapter 2 we will deal with the basic concepts of analytic geometry how to represent numbers in a geometric setting and how numbers can be combined with geometry. This chapter is fundamental to all other chapters of this volume and the remaining 3 volumes of this series. In Chapter 3 we deal with concepts of functions related to continuity and differentiability. We will introduce the notion of a continuous function and the meaning of limit. Based on these two concepts we define what we mean by a derivative of a function and set up rules for the differential calculus. In Chapter 4 we apply derivatives to optimization problems. These problems describe situations in which one of the variables becomes either a maximum or a minimum. Sometimes such problems are also called mini-max problems. Chapter 5 discusses the inverse operation to a differentiation or so called antiderivatives. This notion means nothing more than integration. We will integrate different types of functions and give applications in different fields. Finally, Chapter 6 deales with approximations of functions; we will introduce basic approximations by expansions around a specific point of a function. In addition, we will discuss sums, series and convergence criteria for sums and series.

1.4. Presentation of the Material

Throughout the book we will use the traditional presentation of mathematical terms using symbols, formulas, definitions, theorems, etc. to set up the working frame. This representation is the classical mathematical part. In addition to these traditional presentation tools we will use *Mathematica* as a symbolic, numeric, and graphic tool. *Mathematica* is a computer algebra system allowing hybrid calculations. This means calculations on a computer are either symbolic or/and numeric. *Mathematica* is a tool allowing us in addition to write programs and to do automatic calculations. Before you use such kind of tool it is important to understand the mathematical concepts which are used by *Mathematica* to derive symbolic or numeric results. The use of *Mathematica* allows you to minimize the calculations but you should be aware that you will only understand the concept if you do your own calculations by pencil and paper. Once you have understood the way how to avoid errors in calculations and concepts you are ready to use the symbolic calculations offered by *Mathematica.* You will never reach a a higher level of understanding if you apply the functionality of *Mathematica* as a black box solver of your problems. Therefore, I recommend to you first try to understand by using pencil and paper calculations and then to switch to the computer algebra system after having understood the concepts.

You can get a trial version of *Mathematica* directly from Wolfram Research by requesting a download address from where you can download the trial version of *Mathematica*. The corresponding web address to get *Mathematica* for free is:

http://www.wolfram.com/products/mathematica/experience/request.cgi

2

Pre-Calculus

2.1 Introduction

This lecture uses the language of mathematics to describe situations which occur in engineering. The motivation for doing this is that mathematical arguments are logical and exact, and they enable us to work out in precise detail the consequences of engineering models. For this reason, mathematical modelling has become an indispensable tool in engineering. Mathematics is not always simple for beginners but its language and its techniques enable us to frame and solve problems that cannot be attacked efficiently in other ways. Furthermore, mathematics leads not only to numerical results but also to quantitative results as well. The combination of numeric and qualitative results enables us to design products and to predict the reaction of systems. The aim of this lecture is to use mathematics in a decision process in engineering. At the end of the mathematics course you should be able to apply mathematics as a tool in your decisions as an engineer.

Much of our understanding of the world in which we live depends on our ability to describe how things change. Whether we are concerned with the motion of a pitched baseball or the path of a planet, whether the temperatures and currents of the oceans or the fluctuations of the stock market, whether the propagation of radio waves or the power produced by a chemical reaction, we are constantly forced to analyze relationships among quantities which change along with time.

Algebra and geometry, as you learned in high school, are useful tools for describing relationships among static quantities, but they do not involve concepts appropriate for describing how a quantity changes. For this we need new mathematical operations which go beyond the algebraic operations of addition, subtraction, multiplication, division and the taking of powers and roots. We require operations which measure the way related quantities change. A study of mathematics means that we have to learn new techniques and new procedures which allow us to cope with the challenges of the future. In your study it will not be sufficient to learn things by heart and recall them. In this course you will be asked for how mathematics work and what are the reasons that it works this way. So just recalling knowledge is not sufficient to handle mathematics at a University level because real situations need your decisions based on clearly defined procedures and processes. We will show you in this lecture how the procedures and processes of mathematics work. The lecture will introduce and explain them while exercises must be carried out by yourself. The procedures are accessible by every student because the structure of mathematics is logical, starting at simple relations and deriving from these the more complicated ones.

Calculus provides the tools for describing motion quantitatively. It introduces two new operations called differentiation and integration which, like addition and subtraction, are opposites of one another; what differentiation does, integration undoes. This is a simple example how mathematics work starting with an analogy and transforming it to a new situation. The main ingredient of mathematics is to combine seemingly different things to a new one by using known structures.

Calculus was invented independently and in somewhat different ways by two 17th century mathematicians, Sir Isaac Newton and Gottfried Wilhelm Leibniz.

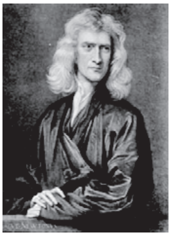


Figure 2.1. Sir Isaac Newton: Born 04.01.1643, Died 31.03.1727.

Newton's motivation was to analyze the motion of moving objects. Using his calculus he was able to formulate his laws of motion and gravitation, and to calculate from them that the planets must move around the sun in elliptical orbits, a fact that had been discovered half a century earlier by Johannes Kepler. Kepler's discovery was empirical, made from years of study of numerical data gathered by Brahe on the position of planets. In a laborious work Kepler extracted from these observations his three general planetary laws. In his famous Rudolphine tables, published in Ulm, he summarized his work which took him 20 years till the completion.

Figure 2.2. Johannes Kepler: Born 27.12.1571 in Leonberg/Württemberg. Died 15.11.1630 in Regensburg.

Many of the fundamental and important "laws of nature" are conveniently expressed as equations involving rates of change of quantities. Such equations are called differential equations, and techniques for their study and solution are at the heart of calculus.

The word "function" seems to have been introduced into mathematics by Leibniz in 1694; this concept now dominates much of mathematics and is indispensable in science and engineering. Since Leibniz' time the concept has been made precise. If y and x are two variables so related that whenever a numerical value is assigned to x there is determined a numerical value of y, then y is called a function of x, and this is symbolized by writing $y = f(x)$.

Figure 2.3. Gottfried Wilhelm von Leibniz: Born 1 July 1646 in Leipzig, Died 14 Nov 1716 in Hannover. Leibniz developed the present day notation for the differential and integral calculus. He never thought of the derivative as a limit.

Mathematics was but one of the many fields in which Leibniz showed conspicuous genius: law, religion, statecraft, history, literature, logic, metaphysics, and speculative philosophy all owe to him contributions, any one of which would have secured his fame and have preserved his memory. "Universal genius" can be applied to Leibniz without hyperbole, as it cannot to Newton, his rival in mathematics and his infinite superior in natural philosophy. Leibniz was the first-rate mind of the age, and he crystallized the calculus. But he was more than an agent for the expression of the spirit of his times, which Newton, in mathematics, was not.

Much of mathematics is related indirectly to the study of motion. We regard lines or curves as geometric objects, but the ancient Greeks thought of them as paths traced out by moving points. Nevertheless, the study of curves also involves geometric concepts such as tangency and area. The process of differentiation is closely tied to the geometric problem of finding tangent lines; similarity, integration is related to the geometric problem of finding areas of regions with curved boundaries.

2.1.1 Models in Engineering

Before we start to discuss mathematics in detail, I will show you for what mathematics is useful for. As everybody knows, in Cairo traffic is one of the most important daily experience in life. For example, you move from Heliopolis or Maadi to the university by car. On the way to the university you will meet a lot of bumpers in the street. Each time you approach a bumper you have to slow down your car and after the bumper you can speed up again. If you miss to slow down something strange will happen inside the car. You will be shaken up and down and somehow you will remember the delicious meal of yesterday evening. Now lets assume you are driving a Porsche and early in the morning the streets are empty so that you can speed up with this car to a high velocity. In your height-speed journey you suddenly realize that you are in Cairo because there is a bumper ahead. Now the question is what will happen with you and your car if you cannot slow down the Porsche. This is a simple question from real life and if you are sitting in this Porsche then you should know how the car will react.

Figure 2.4. Bright morning in the desert.

Now from an engineer's point of view we should be able to predict the behavior of the Porsche jumping over a bumper. So we need a description which allows us to make this prediction. A verbal description is somehow imprecise and does not give answers to our question how a person inside the car will feel after the Porsche passed over the bumper. To make things more precise we need a description which is reliable and measurable in a certain sense. In other words, we need a mathematical model of the car and the bumper to predict the behavior in a reproducible way. And this is the heart of engineering to make predictions based on a formulation which is reliable.

To find such a description or a model of the Porsche and the bumper lets look at the following Figure 2.5 representing the car and a simplified version of that car. First the car consist of a beautiful body which can be simplified as a box. Then the car has four wheels which are basically reduced to the suspension consisting of springs and dampers and that's it.

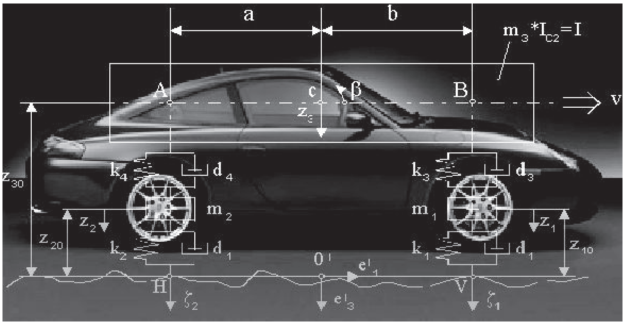


Figure 2.5. Simplified model for a car.

Having this simplification in mind the car can be represented as a set of equations describing the dynamic motion of the body and the wheels. The mathematical part consists of introducing so called differential equations which govern the motion of the car. These equations are for example given by the following set of equations:

$$m_1 z_1'' = -F_1 + F_3 \tag{2.1}$$

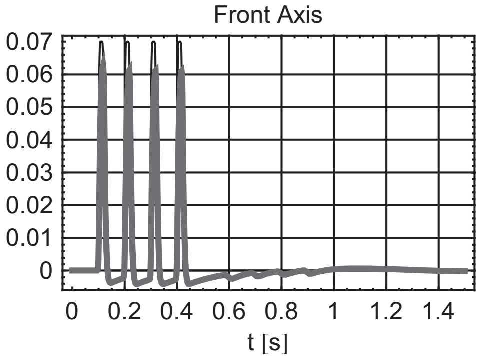
$$m_2 z_2'' = -F_2 + F_4 \tag{2.2}$$

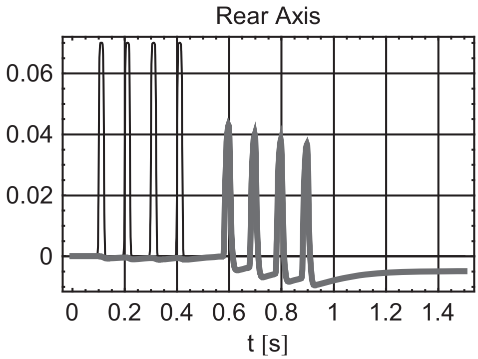
$$m_3 z_3'' = F_3 + F_4 \tag{2.3}$$

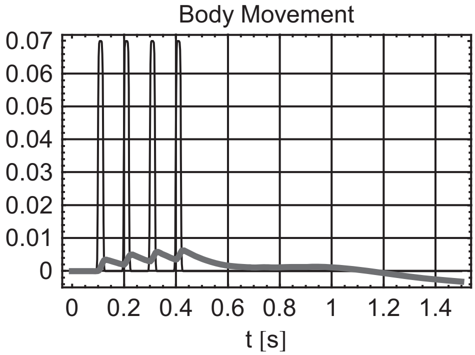
$$I \beta'' = b F_3 - a F_4 \tag{2.4}$$

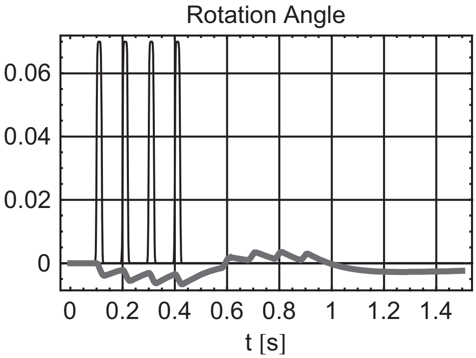
where z_1, z_2, z_3, and β are coordinates, m the mass of the car and F_1, F_2, F_3 and F_4 are acting forces. These equations should be solved and can be solved in a simulation. The results are shown next. If the car is designed the right way the driver and his mates should not experience much of the bumper. The following simulation assumes a worst case that four bumpers follow in sequence.

```
sim∘simulate[0, 1.5]
```



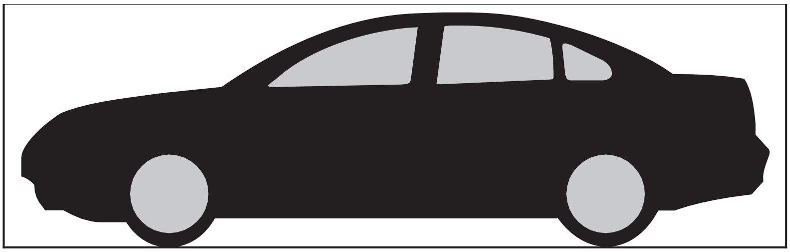






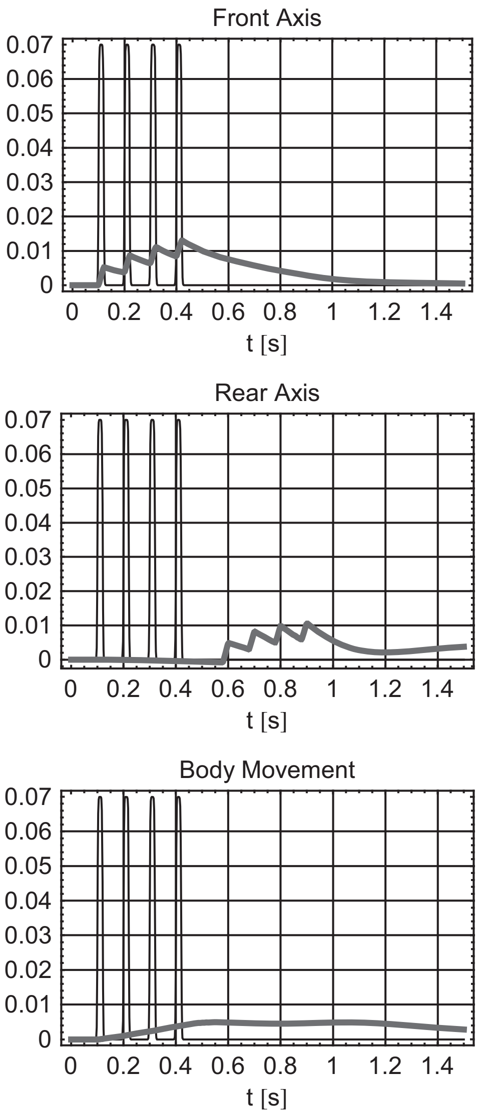
From the graphs you see that there is little shaking inside the car but the wheels take a large elongation. The animation shows the overall movement.

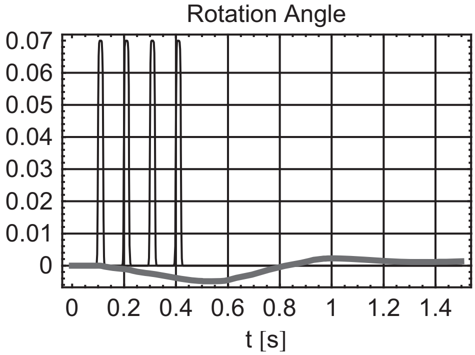
```
an = MakeAnimation∘new[];
an∘animate[sim, 0, 2, 0.05]
```



However if the car is not designed the right way then the driver and his mates get bad feelings. This is shown in the following sequence of figures.

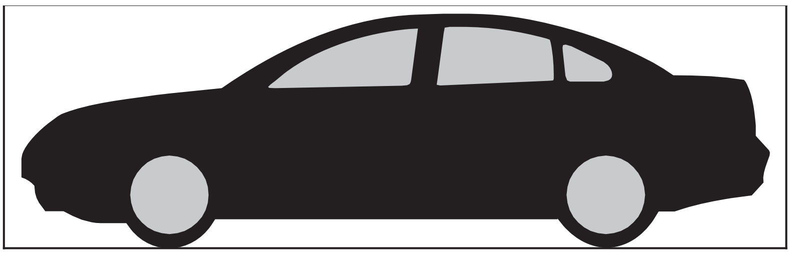
```
sim∘simulate[0, 1.5]
```


Front Axis
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 0.2 0.4 0.6 0.8 1 1.2 1.4
t [s]
Rear Axis
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 0.2 0.4 0.6 0.8 1 1.2 1.4
t [s]
Body Movement
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 0.2 0.4 0.6 0.8 1 1.2 1.4
t [s]



Here the car moves as a whole up and down while in a well designed car only the wheels move up and down. You can observe this in the following simulation

```
an = MakeAnimation∘new[];
an∘animate[sim, 0, 2, 0.05]
```



This example shows you how mathematics is used to predict the behavior of a system. This example also demonstrates that a model can be used to derive a simplified description of the real world by using elementary mathematical relations.

Now before we start to simulate the real world, we start with the simplest and basic properties in mathematics to create a system (language) to describe the real world coherently. The first step to mathematics is to introduce numbers and comparisons between these objects.

2.2 Numbers

Since numbers and their properties play a fundamental role in calculus, we will begin by reviewing some terminology and facts about them.

2.2.1 Numbers on the Real Line

The simplest numbers are the natural numbers:

```
ℕ = {1, 2, 3, 4, 5, 6, …}
{1, 2, 3, 4, 5, 6, …}
```

which are extended to all integers including zero and the negative integer numbers by:

```
ℤ = {…, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, …}
{…, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, …}
```

Such numbers are called integer numbers.

With the exception that division by zero is ruled out, ratios of integers are called rational numbers. Examples are:

$$\mathbb{Q} = \left\{\ldots, \frac{-5}{2}, \frac{-2}{3}, \frac{-1}{2}, 0, \frac{1}{2}, \frac{2}{3}, \frac{5}{2}, \ldots\right\}$$

$$\left\{\ldots, -\frac{5}{2}, -\frac{2}{3}, -\frac{1}{2}, 0, \frac{1}{2}, \frac{2}{3}, \frac{5}{2}, \ldots\right\}$$

Observe that every integer is also a rational number since an integer p can be written as the ratio

$$p == \frac{p}{1}$$

```
True
```

where p denotes any integer number. Division by zero is ruled out since we would want to be able to express the relationship

$$y == \frac{p}{0}$$

```
y == ComplexInfinity
```

in the alternative form

```
0 y == p
0 == p
```

However, if p is different from zero, this equation is contradictory. Moreover, if p is equal to zero, this equation is satisfied by any number y, which means that the ratio $p/0$ does not have a unique value — a situation that is mathematically unsatisfactory. For these reasons such symbols as

$$\frac{p}{0} \text{ and } \frac{0}{0}$$

are not assigned a value; they are said to be undefined.

The early Greeks believed that the size of every physical quantity could, in theory, be represented by a rational number. They reasoned that the size of a physical quantity must consist of a certain whole number of units plus some fraction m/n of an additional unit. This idea was shattered in the fifth century B.C. by Hippasus of Metapontum, who demonstrated by geometric methods that the hypotenuse of a triangle in Figure 2.6 cannot be expressed as a ratio of integers. This unusual discovery demonstrated the existence of irrational numbers; that is, numbers not expressible as ratios of integers.

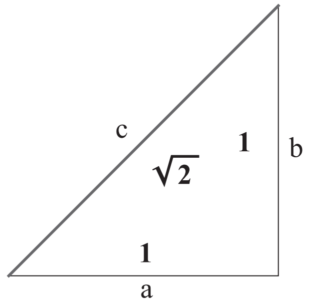


Figure 2.6. Calculation of the length of the hypotenuse by using Pythagoras formula $a^2 + b^2 = c^2$ which is for the specific case considered, is $\sqrt{1+1} = \sqrt{2}$.

Other examples of irrational numbers are

$$\mathbf{1 + \sqrt{2.}}$$

≈ 2.41421

The square root of any number; e.g.

$$\mathbf{\sqrt{3.}}$$

≈ 1.73205

The powers of numbers such as

$$\mathbf{\sqrt[3]{7.}}$$

≈ 1.91293

Special constants like π

$$N[\pi]$$

≈ 3.14159

Or the value of functions

$$N[\cos(19\,°)]$$

≈ 0.945519

Here we used the *Mathematica* function N[] for numerical conversion. The proof that π is irrational is difficult and evaded mathematicians for centuries; it was finally proven in 1761 by J.H. Lambert (see Figure 2.7).

Figure 2.7. Johann Heinrich Lambert (1728-1777), a Swiss-German scientist. Lambert taught at the Berlin Academy of Science. In addition to his work on irrational numbers, Lambert wrote landmark books on geometry, the theory on cartography, and perspective in art. His influential book on geometry foreshadowed the discovery of modern non-Euclidean geometry.

Rational and irrational numbers can be distinguished by their decimal representation. Rational numbers have repeating decimals; that is, from some point the decimals consists entirely of zeros or else a fixed finite sequence of digits are repeated over and over.

Example 2.1. Rational numbers

The following numbers demonstrate the behavior of rational numbers.

$$\boldsymbol{N\left[\frac{1}{2}\right]}$$

0.5

1 / 2 terminates after the first digit.

$$\boldsymbol{N\left[\frac{1}{3}\right]}$$

≈ 0.333333

1 / 3 has a continuous repetition of a single digit. We can check this by increasing the displayed digits of the decimal representation of the number.

$$\boldsymbol{(N[\#1, 50]\ \&)\left[\frac{1}{3}\right]}$$

≈ 0.33333333333333333333333333333333333333333333333333

The *Mathematica* notation (N[#1,50])& means that we convert the number in brackets [1/3] to a decimal number with 50 digits on the right of the decimal point. Another property of rational numbers is a finite number of digits as demonstrated with the next rational number

$$\boldsymbol{(N[\#1, 25]\ \&)\left[\frac{8}{25}\right]}$$

0.3200000000000000000000000

8 / 25 has two non-vanishing digits.

$$\boldsymbol{(N[\#1, 50]\ \&)\left[\frac{3}{11}\right]}$$

≈ 0.27272727272727272727272727272727272727272727272727

3/11 repeats the digits 2 and 7.

$$(N[\#1, 50]\ \&)\left[\frac{5}{7}\right]$$

$$\approx 0.71428571428571428571428571428571428571428571428571$$

5 / 7 repeats the 6 digits 714285

$$(N[\#1, 50]\ \&)\left[\frac{6}{13}\right]$$

$$\approx 0.46153846153846153846153846153846153846153846153846$$

6 / 13 repeats the 6 digits 461538. The repetition or termination of a finite number of digits in a decimal representation of a rational number is an important property of this kind of numbers. On the other hand there exist also numbers which do not show this property. These numbers are called irrational numbers contrary to rational numbers. Irrational numbers are represented by non repeating decimals. An example for that behavior is $\sqrt{2}$ represented with 50 digits:

$$(N[\#1, 50]\ \&)\left[\sqrt{2}\right]$$

$$\approx 1.4142135623730950488016887242096980785696718753769$$

with 250 digits this number looks as follows

$$(N[\#1, 250]\ \&)\left[\sqrt{2}\right]$$

$$\begin{aligned}\approx 1.&414213562373095048801688724209698078569671875376948073176679737990732478462107038\cdot.\\ &850387534327641572735013846230912297024924836055850737212644121497099935831413222\cdot.\\ &665927505592755799950501152782060571470109559971605970274534596862014728517418640\cdot.\\ &889199\end{aligned}$$

even within 500 digits there is no repetition as you can check by yourself

$$(N[\#1, 550]\ \&)\left[\sqrt{2}\right]$$

$$\begin{aligned}\approx 1.&414213562373095048801688724209698078569671875376948073176679737990732478462107038\cdot.\\ &850387534327641572735013846230912297024924836055850737212644121497099935831413222\cdot.\\ &665927505592755799950501152782060571470109559971605970274534596862014728517418640\cdot.\\ &889198609552329230484308714321450839762603627995251407989687253396546331808829640\cdot.\\ &620615258352395054745750287759961729835575220337531857011354374603408498847160386\cdot.\\ &899970699004815030544027790316454247823068492936918621580578463111596668713013015\cdot.\\ &618568987237235288509264861249497715421833420428568606014682472\end{aligned}$$

So our observation is that $\sqrt{2}$ does not begin to repeat from some point. The same behavior is observed for π

(N[♯1, 1000] &)[π]

≈ 3.14159265358979323846264338327950288419716939937510582097494459230781640628620899 8
6280348253421170679821480865132823066470938446095505822317253594081284811174502 84
1027019385211055596446229489549303819644288109756659334461284756482337867831652 71
2019091456485669234603486104543266482133936072602491412737245870066063155881748 81
5209209628292540917153643678925903600113305305488204665213841469519415116094330 57
2703657595919530921861173819326117931051185480744623799627495673518857527248912 27
9381830119491298336733624406566430860213949463952247371907021798609437027705392 17
1762931767523846748184676694051320005681271452635608277857713427577896091736371 78
7214684409012249534301465495853710507922796892589235420199561121290219608640344 18
1598136297747713099605187072113499999983729780499510597317328160963185950244594 55
3469083026425223082533446850352619311881710100031378387528865875332083814206171 77
6691473035982534904287554687311595628638823537875937519577818577805321712268066 13
001927876611195909216420199

which also is known as an irrational number. There is no repetition of digits within the given accuracy of 1000 digits.▪

In 1637 René Descartes published a philosophical work called *Discourse on the Method of Rightly Conducting the Reason*. In the back of that book were three appendices that purported to show how the method could be applied to specific examples. The first two appendices were minor works that endeavored to explain the behavior of lenses and the movement of shooting stars. The third appendix, however, was an inspired stroke of genius; it was described by the nineteenth century British philosopher John Stuart Mill as "The greatest single step ever made in the progress of the exact sciences". In that appendix René Descartes (Figure 2.8) linked together two branches of mathematics, algebra and geometry. Descartes' work evolved into a new subject called analytic geometry; it gave a way of describing algebraic formulas by means of geometric curves using algebraic formulas.

Figure 2.8. René Descartes (1596-1650). Descartes, a French aristocrat, was the son of a government official. He graduated from the University of Poitiers with a law degree at age 20. After a brief probe into the pleasures of Paris he became a military engineer, first for the Dutch Prince of Nassau and then for the German Duke of Bavaria. It was during his service as a soldier that Descartes began to pursue mathematics seriously and developed his analytic geometry. The story goes that Descartes coined his main philosophical point of view in a village near Ulm sitting in an oven "Cogito ergo sum" (I am thinking and therefore I exist).

In analytic geometry, the key step is to establish a correspondence between real numbers and points on a line. To accomplish this, we arbitrarily choose one direction along the line to be called positive and the other negative. It is usual to mark the positive direction with an arrowhead as shown in Figure 2.9. Next we choose an arbitrary reference point on the line to be called the origin, and select a unit of length for measuring distances.

Figure 2.9. Coordinate line allowing to measure the distance between two numbers. The origin separates the positive from the negative numbers.

With each real number we can associate a point on the line as follows:

1. Associate with each positive number r the point that is a distance of r units in the positive direction away from the origin.
2. Associate with each negative number $-r$ the point that is a distance of r units in the negative direction from the origin.
3. Associate the origin with the number 0.

The real number corresponding to a point on the line is called the coordinate of the point and the line is called a coordinate line or sometimes the real line.

Example 2.2. Numbers on the real line

In Figure 2.10 we have marked the location of the points whose coordinates are -3, -1.75, -1/2, $\sqrt{2}$, π, and 4.

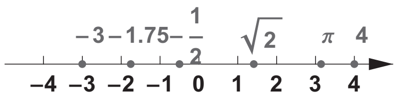


Figure 2.10. Location of some real numbers ($\in \mathbb{R}$) on the coordinate line.▲

It is evident from the way in which real numbers and points on a coordinate line are related that each real number corresponds to a single point and that each point corresponds to a single real number. This fact is described by stating that the real numbers and the point on a coordinate line are in one-to-one correspondence.

Looking at the numbers from a more general point of view, we observe that real numbers contain rational numbers and rational numbers contain integer numbers, and integer numbers contain natural numbers. This relation is graphically shown in Figure 2.11.

Figure 2.11. The relation of set of numbers. Natural numbers ℕ are included in integer numbers ℤ, integer numbers are included in rational numbers ℚ, and rational numbers are part of the real numberset ℝ. Despite of the graphical representation by circles the sets of numbers are infinite sets.

2.2.2 Complex Numbers

In the previous section we discussed different kinds of numbers which are the natural numbers ℕ, the integer numbers ℤ, the rational numbers ℚ, and the real numbers ℝ. Part of the real numbers are so called irrational numbers which are numbers without repetition on the decimal part; e.g. the square root of 2: $\sqrt{2}$. The question now arises if the square root is only defined for positive numbers as in $\sqrt{2}$ or if there are numbers which allow the calculation of the square root for negative numbers. The simplest case of such a number is

$$\sqrt{-1}$$

which we abbreviate with the new symbol i. This means

$$\sqrt{-1} = i.$$

The introduction of i as a new kind of number allows us to extend the real numbers to a second branch. So, complex numbers are ordered pairs of numbers using Descartes' idea representing the numbers on an arrow for two orthogonal directions. Complex numbers are useful in describing relations which are two-dimensional in nature but have only a single real quantity.

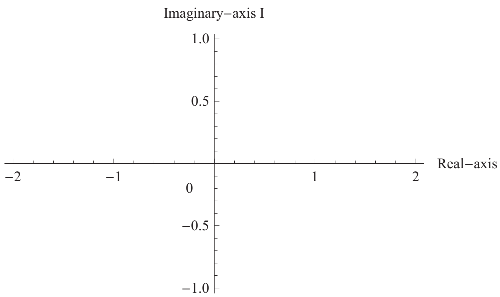


Figure 2.12. Complex plane with the real and imaginary axis.

A complex number is given by the ordered pair of numbers as

$$(a,\ b)$$

where a and b are real numbers. The first part of the ordered pair represents the real part of the number along the horizontal line while the second one represents the imaginary part on the vertical line. A complex number is thus represented as

$$z = a + i\, b$$

a sum of two real numbers where the second number is multiplied by the imaginary unit i. A typical set of complex numbers is given by

$$\{1 + i,\ 1 + 2\, i,\ 1 + 3\, i,\ 1 + 4\, i,\ 2 + i,\ 2 + 2\, i,\ 2 + 3\, i,\ 2 + 4\, i,\ 3 + i,\ 3 + 2\, i,\ 3 + 3\, i,\ 3 + 4\, i\}$$

As for real numbers a conjugate number exists on the real line which is given by the negative real number there are also complex conjugate numbers. A complex conjugate number is a complex number where the sign in front of the imaginary part is changed. We denote the complex conjugate number by $\overline{z} := z^*$ which is given as

$$z^* = (a + i\, b)^* = a - i\, b.$$

In ordered pair representation this is given by

$$(a,\ -b).$$

In *Mathematica* the conjugation is represented by the function Conjugate[] which has the short hand notation (*)

$(3 + i\, 2)^*$

$3 - 2\, i$

The overall structure of numbers within calculus is thus given as in Figure 2.13.

Figure 2.13. The sets of numbers ordered as subsets. Natural numbers ℕ are included in integer numbers ℤ, integer numbers are included in rational numbers ℚ, rational numbers are part of real number ℝ, and complex numbers ℂ cover all subsets. Despite of the graphical representation by circles, the sets of numbers are infinite sets.

2.2.3 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

2.2.3.1 Tests

T1. If $x = 3.5\,\overline{17}$. Then we have

( ) $x = \frac{3517}{999}$

( ) $x = \frac{35.99+17}{990} = \frac{3482}{990}$

( ) $x = \frac{35.99+17}{999} = \frac{3482}{999}$

T2. If $x = \frac{7}{17}$. Then we have

( ) $x = 0.4\,\overline{1}$

( ) $x = 0.411\,\overline{764}$

( ) $x = 0.411764705\,\overline{8}$

( ) $x = 0.\,\overline{4\,117\,647\,058\,823\,529}$

2.2.3.2 Exercises

E1. Convert the following decimal numbers to rational numbers: $0.47\,\overline{8}$ and $21.4\,\overline{781}$.

E2. Convert the following rational numbers to decimal numbers: $\frac{234}{999}$ and $\frac{17}{11}$.

E3. Show that the following equation holds:

$$0.\,\overline{4\,117\,647\,058\,823\,529} + \frac{179}{102} = 2.1\,\overline{6}. \tag{1}$$

E4. Among the terms integer, rational, and irrational, which one apply to the given number

a. $\frac{-3}{5}$,

b. 0,

c. $\frac{26}{9}$,

d. 0.2,

e. $-\sqrt{3}$,

f. $3^{1/3}$,

g. $0.0303\,\overline{03}$,

h. $8.00\,\overline{0}$

E5. The following ratios are all famous approximations to π.

Archimedes: 22/7
Archimedes: 223/71
Tsu Chung-Chi and others: 355/113
Adrian Athoniszoon 1582: 333/106
Ramanujan: $\frac{63}{25}\left(\frac{17+15\sqrt{5}}{7+15\sqrt{5}}\right)$

a. Use a calculator to get the decimal representations.

b. Which of your results is closest to π? Which one is larger than π?

c. Which of these approximations is most accurate?

E6. The Rhind Papyrus, which is a fragment of Egyptian mathematical writing from about 1650 B.C., is one of the oldest known examples of written Mathematics. It is stated in the papyrus that the area A of a circle is related to its diameter d by

$A = \left(\frac{8}{9}\,d\right)^2$

a. What approximation to π were the Egyptians using?

Check if the approximation to π is better by $\frac{22}{7}$ or $\frac{8}{9}$.

E7. The repeating decimal 0.1371371371 … can be expressed as ratio of integers by writing
$x = 0.1371371371 \ldots$
$1000\,x = 137.1371371 \ldots$
and subtracting to obtain $999\,x = 137$ or $x = \frac{137}{999}$. Use this idea, where needed, to express the following decimals as ratios of integers.

a. 12.7777 …,
b. 0.123123 …,
c. 0.4296000 …,
d. 38.087181818 …

E8. Can a rational number satisfy $10^x = 3$?

E9. Show that the sum or product of two irrational numbers can be rational or irrational.

E10 Prove the following results about sums of rational and irrational numbers:

a. rational + rational = rational
b. rational + irrational = irrational.

2.3 Inequalities

Of particular importance to us is the fact that the real numbers are ordered, this was in fact the idea by Descartes introducing the real line, given any two numbers a and b, exactly one of the following is true:

- a is less than b
- b is less than a
- a is equal to b.

Assuming we know what positive numbers are, we can define the order symbol $<$ (less than) and $\leq$ (less than or equal to) as follows.

Definition 2.1. *Ordering*

If a and b are real numbers, then
$a < b$ means $b - a$ is positive
$a \leq b$ means $a < b$ or $a = b$. ■

The inequality $a < b$, which is read as a is less than b can also be written as $b > a$, which is read b is greater than a. Geometrically, the inequality $a < b$ states that b falls to the right of a on a coordinate line (see Figure 2.14).

Figure 2.14. Comparing two numbers a and b on a coordinate line with $a < b$.

Similarly $a \leq b$, which is read a is less than or equal to b, can also be written as $b \geq a$, which is read, b is greater or equal to a.

Example 2.3. Inequalities

The following inequalities are correct. The result in *Mathematica* is given as a logical expression with a final value "True" or "False". Certainly 3 is less than 8 which is a true statement

3 < 8

True

and 8 is greater than or equal to 3

8 ≥ 3

True

Negative numbers are less than positive ones

−7 < 1.5

True

or the other way round, positive numbers are greater than negative numbers

1.5 > −7

True

Comparing negative numbers, then -12 is less than or equal to $-\pi$

−12 ≤ −π

True

and the converse, that $-\pi$ is smaller than 12 is wrong

−π < −12

False

Equalities are recognized as less than or equal

5 ≤ 5

True

or as greater than or equal relations.

5 ≥ 5

True

▲

A number b will be called non-negative if $0 \leq b$, and non positive if $b \leq 0$. The following properties of inequalities are frequently used in calculus. We omit the proofs.

Theorem 2.1. *Inequality Relations*

If a, b, c and d are real numbers then

a) *If* $a < b$ *and* $b < c$ *then* $a < c$.

b) *If* $a < b$ *then* $a + c < b + c$.

c) *If* $a < b$ *and* $c < d$ *then* $a + c < b + d$.

d) *If $a < b$ then $a\,c < b\,c$ when c is positive and $a\,c > b\,c$ when c is negative.*
e) *If a and b are both positive or both negative and $a < b$ then $1/a > 1/b$.*■

Remark 2.1. These five properties remain true if < and > are replaced by ≤ and ≥, respectively.

Example 2.4. Adding a constant to an inequality

To paraphrase property (b), an inequality remains true if the same number is added to both sides. For example, adding 7 to both sides of the inequality

```
-2 < 6
True
```

yields the valid inequality

```
7 - 2 < 6 + 7
True
```

Similarly, adding -5 to both sides yields the valid inequality

```
-3 - 5 < 6 - 5
True
```

Example 2.5. Addition of different values

To paraphrase property c), inequalities in the same direction can be added. For example, adding the two relations

```
2 < 6
True
```

and

```
-8 < 5
True
```

yield the valid inequality

```
2 - 8 < 6 + 5
True
```

Example 2.6. Multiplication

To paraphrase property (d), if both sides of an inequality are multiplied by a positive number, the direction of the inequality remains the same, and if both sides of an inequality are multiplied by a negative number, the direction of the inequality is reversed.

For example, multiplying the inequality

```
-4 < 8
True
```

by 3 yields

$$\mathbf{-4 \times 3 < 8 \times 3}$$

True

and multiplying by -4 yields

$$\mathbf{-4\,(-4) < 8\,(-4)}$$

False

▲

Example 2.7. Reciprocals

To paraphrase property ("e"), taking reciprocals reverses the direction of an inequality involving two positive or two negative numbers. For example, taking reciprocal of both sides of the inequality

$$\mathbf{2 < 7}$$

True

yields

$$\mathbf{\frac{1}{2} < \frac{1}{7}}$$

False

and taking reciprocals on both sides of the inequality

$$\mathbf{-6 < -4}$$

True

yields

$$\mathbf{-\frac{1}{6} < -\frac{1}{4}}$$

False

Definition 2.2. *Chaining*

If a, b and c are real numbers, we will write a < b < c when a < b and b < c. ■

Geometrically, $a < b < c$ states that on a coordinate line, b falls to the right of a, and c falls to the right of b (see Figure 2.15)

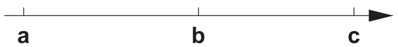


Figure 2.15. Comparing three numbers *a*, *b* and *c* on a coordinate line with $a < b < c$.

The symbol $a < b \leq c$ means $a < b$ and $b \leq c$. We leave it to the reader to deduce the meaning of such symbols as $a \leq b < c$, $a \leq b \leq c$, $a < b < c < d$, and so on.

Definition 2.3. *Set*

A set is a collection of objects; the objects are called the elements or members of the set.■

In this text we will be concerned primarily with sets whose members are real numbers. One way of describing a set is to list elements between braces. Thus, the set of all positive integers less than 6 can be written

```
integersLessThan6 = {1, 2, 3, 4, 5}
```

```
{1, 2, 3, 4, 5}
```

and the set of all positive even integers smaller than 20 can be written as

```
evenIntegersSmaller20 = Table[2 i, {i, 1, 9}]
```

```
{2, 4, 6, 8, 10, 12, 14, 16, 18}
```

Here Table[] generates a list according to the specification given in the first argument. The second argument defines the range of indices starting at 1 and ending at 9.▲

When it is inconvenient or impossible to list the members, it is sufficient to define a set by stating a property common to its members. For example,

the set of all rational numbers
the set of all real numbers x such that $2x^2 - 4x + 1 = 0$
the set of all real numbers between 2 and 3.

As an alternative to such verbal descriptions of sets, we can use the notation

{x | ___}

which is read, "the set of all x such that". Where the line is placed, we state the property that specifies the set.

Example 2.8. Sets

a. $\{x \mid x$ is a rational number$\}$ is read, "the set of all x such that x is a rational number".
b. $\{x|$ x is a real number satisfying $2x^2 - 4x + 1 = 0\}$ is read "the set of all x such that x is a real number satisfying $2x^2 - 4x + 1 = 0$".
c. $\{ x|$ x is a real number between 2 and 3$\}$ is read, " the set of all x such that x is a real number between 2 and 3".▲

Remark 2.2. When it is clear that the members of a set are real numbers, we will omit the reference to this fact. Thus the sets in example 2.8 might be written as

a. $\{x \mid x \in \mathbb{Q}\}$
b. $\left\{x \mid 2x^2 - 4x + 1 = 0\right\}$
c. $\{x \mid 2 < x < 3\}$.

To indicate that an element a is a member of a set A, we write

$a \in A$

which is read "a is an element of A" or "a belongs to A". To indicate that the element a is not a member of the set A, we write

$a \notin A$

which is read, "a is not an element of A" or "a does not belong to A".

Example 2.9. Sets

Let $A = \{x \mid x \text{ is } a \text{ rational}\} = \{x \mid x \in \mathbb{Q}\}$ and $B = \{x \mid 2 < x < 4\}$, then

$\frac{3}{4} \in A$	$-2 \in A$	$\pi \notin A$	$-\sqrt{2} \notin A$
$2.5 \in B$	$\pi \in B$	$-1 \notin B$	$4 \notin B$.▲

Sometimes it arises that a set has no members. For example

$\{x \mid x^2 < 0\}$.

A set with no member is called an empty set or a null set and is denoted by the symbol $\emptyset$. Thus,

$\emptyset = \{x \mid x^2 < 0\}$.

Of special interest are certain sets of real numbers called intervals. Geometrically, an interval is a line segment. For example, if $a < b$, then the closed interval from a to b is the set

$\{x \mid a \leq x \leq b\}$

and the open interval from a to b is the set

$\{x \mid a < x < b\}$.

These sets are pictured in Figure 2.16 and Figure 2.17.

Figure 2.16. Closed interval from a to b; i.e. $[a, b]$.

Figure 2.17. Open interval from a to b; i.e. (a, b).

Remark 2.3. A closed interval includes both its endpoints (indicated by solid dots in Figure 2.16), while an open interval excludes both endpoints (indicated by open dots in Figure 2.17). Closed and open intervals are usually denoted by the symbol $[a, b]$, and (a, b), where

$$[a, b] = \{x \mid a \leq x \leq b\}$$
$$(a, b) = \{x \mid a < x < b\}.$$

A square bracket [or] indicates that the point is included while a round bracket "(" or ")" indicates that the point is excluded.

An interval can include one endpoint and exclude the other. Such intervals are called half-open (or sometimes half-closed). For example

$[a, b) = \{x \mid a \le x < b\}$
$(a, b] = [x \mid a < x \le b\}$.

An interval can extend indefinitely in either the positive or negative direction (see Figure 2.16 and Figure 2.17). The intervals in Figure 2.18 and 2.19 are denoted by

$(a, +\infty) = \{x \mid x > a\}$

and

$(-\infty, b] = \{x \mid x \le b\}$

where the symbols $+\infty$ ("plus infinity") indicates the interval extends indefinitely in the positive direction and the symbol $-\infty$ ("minus infinity") indicates the interval extends indefinitely in the negative direction.

Figure 2.18. Open interval from *a* to $+\infty$.

Figure 2.19. Half-open interval from *b* to $-\infty$.

Definition 2.4. *Equality of Sets*

Two sets A and B are said to be equal if they have the same elements, in which case we write $A = B$. ■

Example 2.10. Equality of sets

$\{x \mid x^2 = 1\} = \{-1, 1\}$ $\{\pi, 0, 3\} = \{3, \pi, 0\}$
$\{x \mid x^2 < 9\} = \{x \mid -3 < x < 3\}$.▲

Definition 2.5. *Subset*

If every member of a set A is also a member of set B, then we say A is a subset of B and write $A \subset B$. In addition, we will agree that the empty set $\emptyset$ is a subset of every set. ■

Example 2.11. Subsets

$\{-2,4\} \subset \{-2,1,2,4\}$
$\{x \mid x \text{ is rational}\} \subset \{x \mid x \text{ is } a \text{ real number}\}$
$\emptyset \subset A$ (for every set A).▲

Remark 2.5. If $A \subset B$ and $B \subset A$ then $A = B$.

Definition 2.6. *Intersection*

If A and B are two sets, then the set of all elements belonging to both A and B is called the intersection of A and B; it is denoted by $A \cap B$. ■

Definition 2.7. *Union*

If A and B are two given sets, then the set of all elements belonging to A or B or both is called the union of A and B; it is denoted by
$A \cup B$. ■

Example 2.12. Intersection and Union

In Figure 2.20 and 2.21 A is a set of points inside the left circle and B the set of points inside the right circle. The sets $A \cap B$ and $A \cup B$ are shown as shaded regions.

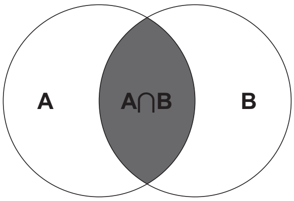


Figure 2.20. Intersecting two sets given by circles.

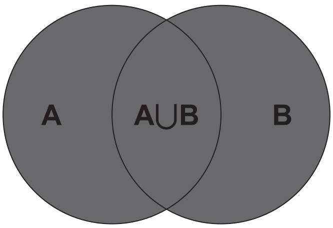


Figure 2.21. Union of two sets given by circles. ▲

2.3.1 Tests and Exercises

The following tests and exercises are related to relations and sets.

2.3.1.1 Test Problems

T1. Which of the following relation are always correct for $a, b \in \mathbb{R}$ and $a < b$.

( ) $\frac{1}{a} < \frac{1}{b}$

( ) $\frac{1}{a} > \frac{1}{b}$

( ) $\frac{1}{a} > \frac{1}{b}$ if $a > 0$

T2. Which of the following relation are always correct for $a, b \in \mathbb{R}$ and $a \leq b$.

( ) $a - 3 \leq b - 3$

( ) $3 - a \leq 3 - b$

( ) $6\,a \leq 6\,b$

( ) $a^3 \leq a^2\,b$

T3. In each part, sketch on a coordinate line all values of x

a. $x \leq 5$

b. $x \geq -1$

c. $-2 \leq x \leq 5$

d. $x^2 \leq 16$

T4. Given the sets $A = \{1, 2, 3, 4\}$, $B = \{1, 3, 5\}$, and $C = \{2, 3, 4\}$. A set $D = (A \cap C) \setminus B$ is generated by A, B, and C. Which of the following statements are true?

( ) $2 \in D$

( ) $3 - a \leq 3 - b$

( ) $\{2, 4\} \subset D$

( ) $D \cap B = \emptyset$

( ) $D \cup C = A$

2.3.1.2 Exercises

E1. Which of the following are always correct if $a, b, c, d \in \mathbb{R}$ and $a \leq b$ and $c \leq d$?

a. $a + 2\,c \leq b + 2\,d$

b. $a - 2\,c \leq b - 3\,d$

c. $a - 4\,c \geq b - 4\,d$

E2. If $a \leq b$ and $b \leq a$, what can you say about a and b?

E3. Compare the following two statements:

a. If $a < b$ is true, does it follow that $a \leq b$ must also be true?

b. If $a \leq b$ is true, does it follow that $a < b$ must also be true?

E4. In each part, list the elements in the set.

a. $\{x \mid x^2 - 5\,x = 0\}$

b. $\{x \mid x \text{ is an integer satisfying } -1 \leq x \leq 4\}$

E5. Express in interval notation.

a. $\{x \mid x^2 \leq 9\}$

b. $\{x \mid x^2 > 9\}$

E6. In each part, sketch the set on a coordinate line.

a. $[4, 5] \cup [7, 9]$

b. $(-\infty, 5) \cap [0, \infty)$

c. $(-2, 4) \cap (0, 5]$

d. $(-\infty, 4) \cup (-1, \infty)$

E7. For the following relations solve the inequality and sketch the solution on a coordinate line.

a. $3\,x - 2 < 9$

b. $\frac{x}{5} + 6 \geq 14$

c. $\frac{x}{x-3} < 4$

$x^2 > 9$

e. $x^2 \leq 5$

f. $\frac{3x+1}{x-2} < 2$

E8. Find all values of x or which the given expression yields a real number

a. $\sqrt{x^2 + x - 6}$

b. $\sqrt{\frac{x+2}{x-1}}$

E9. Every integer is either even or odd. The even integers are those that are divisible by 2, so n is even if and only if $n = 2k$ for some integer k. Each odd integer is one unit larger than an even integer, so n is odd if and only if $n = 2k + 1$ for some integer k. Now show:

a. If n is even, then so is n^2.

b. If n is odd, then so is n^2.

E10 Fahrenheit and Celsius temperatures are related by the formula $C = \frac{5}{9}(F - 32)$. If the temperature in degree Celsius ranges over the interval $25 \leq C \leq 40$ on a certain day, what is the temperature range in degree Fahrenheit that day?

E11 Solve the following relation:

$$12x^3 - 20x^2 \geq -11x + 2. \tag{1}$$

2.4 Absolute Values

So far, we introduced different sets with positive and negative numbers; e.g. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. These sets contain elements with different signs. The question now is how can we measure the magnitude of the elements in these sets. The following definition provides an answer.

Definition 2.8. *Absolute Value*

The absolute value or magnitude of a real number a is denoted by $|a|$ and is defined by

$|a| = a$ if $a \geq 0$

$|a| = -a$ if $a < 0$.■

Example 2.13. Magnitude of numbers

The following examples demonstrate the application of the absolute value on numbers. The magnitude of 5 is

$\mathbf{|5|}$

5

The magnitude of a negative rational number is

$\left|-\frac{\mathbf{4}}{\mathbf{7}}\right|$

$\frac{4}{7}$

The magnitude of zero results to

$$|0|$$

0

▲

Remark 2.6. Symbols such as $+a$ and $-a$ are deceptive since it is tempting to conclude that $+a$ is a positive and $-a$ is a negative number. However, these conclusions are unwarranted since a itself can represent either a positive or negative number. In fact, if a is negative, then $-a$ is positive and $+a$ is negative. With this comment in mind, it should be evident that $|a| \geq 0$ for any number a.

Recall that a number whose square is a, is called a square root of a. In algebra it is learned that every non-negative real number has exactly one non-negative square root; we denote this square root by $\sqrt{a}$. For example, The number 9 has two square roots, -3 and 3. Since 3 is the non-negative square root we have

$$\sqrt{9}$$

3

Remark 2.7. Readers who were previously taught to write $\sqrt{9} = \pm 3$ are advised to stop doing so, since it is incorrect.

It is another common error to write $\sqrt{a^2} = a$. Although this equality is correct when a is non-negative, it is false for negative a. For example, if $a = -4$, then

$$\sqrt{a^2}\ /.\ a \to -4$$

4

which obviously is not equal to -4. To insert the value -4 for a we used the function ReplaceAll[] (/.) and a rule $a \to -4$ defining the replacement. However, $x^2 = 9 \Longleftrightarrow (\pm 3)^2 = 9$.

A result that is correct for all a is given in the following theorem.

Theorem 2.2. *Magnitude*

For any real number a,

$\sqrt{a^2} = |a|$. ■

Proof 2.2. Since $a^2 = (+a)^2 = (-a)^2$, the numbers $+a$ and $-a$ are square roots of a^2. If $a \geq 0$, then $+a$ is the non-negative square root of a^2 and if $a < 0$, then $-a$ is the non-negative square root of a^2. Since $\sqrt{a^2}$ denotes the non-negative square root of a^2, we have

$$\sqrt{a^2} = +a \qquad \text{if} \qquad a \geq 0$$

$$\sqrt{a^2} = -a \qquad \text{if} \qquad a < 0.$$

That is, $\sqrt{a^2} = |a|$.

QED

Theorem 2.3. *Magnitude Relations*

If a and b are real numbers and n is an integer, then

a) $-|a| \le a \le |a|$

b) $|ab| = |a|\ |b|$

c) $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$

d) $|a^n| = |a|^n$. ■

Proof 2.3. We will prove a) and b) only.

a). If $a \ge 0$, then we can write

$$-a \le a \le a \qquad \text{and} \qquad |a| = a$$

from which it follows that

$$-|a| \le a \le |a|.$$

If $a < 0$, we can write

$$a \le a \le -a \qquad \text{and} \qquad -a = |a|$$

from which it again follows that

$$-|a| \le a \le |a|.$$

Thus statement a) holds in all cases.

b). Using Theorem 2.2

$$|ab| = \sqrt{(\mathrm{ab})^2} = \sqrt{a^2 b^2} = \sqrt{a^2}\ \sqrt{b^2} = |a|\ |b|.$$

QED

Remark 2.8. In words, part b) and c) of this theorem state that the absolute value of a product is the product of the absolute values and the absolute value of a ratio is the ratio of the absolute values.

The notion of absolute value arises naturally in distance problems. On a coordinate line, let *A* and *B* be points with coordinates *a* and *b*. Because distance is always non-negative, the distance *d* between *A* and *B* is

$$d = b - a$$

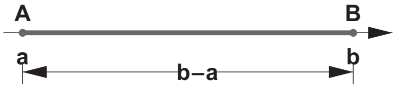


Figure 2.22. Distance on the real line between *A* and *B*.

when B is to the right of A (Figure 2.22), and

$$d = a - b$$

when B is to the left of A (Figure 2.23).

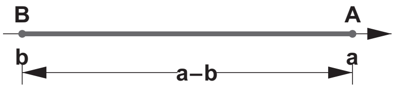


Figure 2.23. Distance on the real line between *B* and *A*.

In the first case, $b - a$ is positive, so we can write

$$d = b - a = |b - a|$$

and in the second case, $b - a$ is negative, so we can write

$$d = a - b = -(b - a) = |b - a|.$$

Thus, regardless of whether B is to the right or left of A, the distance d between A and B is

$$d = |b - a|.$$

This formula is useful when the relative positions of A and B are unknown.

For any real number b, we can write

$$|b| = |b - 0|.$$

Therefore, the absolute value of a number b can be interpreted geometrically as its distance from the origin on a coordinate line. For example, if $|b| = 5$, then b lies 5 units from the origin, that is, $b = 5$ or $b = -5$.

Example 2.14. Solution of Equations

Solve the following equation

$$|x - 3| = 4.$$

Solution 2.14. This relation expresses the fact that a number x is 4 units away from 3.

The solution can be gained geometrically and algebraically. The geometric approach uses the verbal description of the equation. The solution consists of all x that are 4 units away from the point 3. There are two such x values, $x = 7$ and $x = -1$ (Figure 2.24)

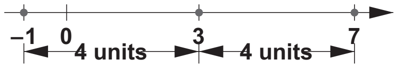


Figure 2.24. Graphical representation of the solution for the equation $|x-3|=4$.

The algebraic approach uses the following ideas. Depending on whether $x-3$ is positive or negative, the equation $|x-3|=4$ can be written as

$$x-3=4 \qquad \text{or} \qquad -(x-3)=4.$$

Solving these two equations gives

$$x=7 \qquad \text{or} \qquad x=-1$$

which agrees with the solution obtained geometrically. The *Mathematica* solution is gained by using the function Solve[] which provides

Solve[|x – 3| == 4, x]

$\{\{x \to -1\}, \{x \to 7\}\}$

The solution is given as a replacement rule for x. ▲

The example above used the equality of the expression to find the solution. In addition it is also possible to use the inequalities to determine ranges.

Example 2.15. Solution of Inequalities

Solve the following inequality

$$|x-3|<4.$$

Solution 2.15. The solution consists of all x whose distance from 3 is less than 4 units, that is, all x satisfying

$$-1<x<7.$$

This is the interval $(-1,\ 7)$ shown in Figure 2.25.

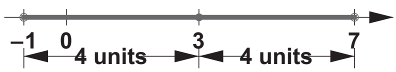


Figure 2.25. Graphical representation of the solution for the inequality $|x-3|<4$.

The interval can be determined by the *Mathematica* function Reduce[]

Reduce[|x – 3| < 4, x]

$-1 < \mathrm{Re}[x] < 7\ \&\&\ -\sqrt{7+6\,\mathrm{Re}[x]-\mathrm{Re}[x]^2} < \mathrm{Im}[x] < \sqrt{7+6\,\mathrm{Re}[x]-\mathrm{Re}[x]^2}$

The representation of the interval is given by a logic expression combining restrictions of real and imaginary parts of x. For the moment drop the second part of this logical representation of the solution including imaginary components which are not relevant for our considerations here.▲

The following general result will be of importance in later sections.

Theorem 2.4. *Magnitude Relations*

For any real number x and a and any positive number k

a) $|x| < k$ *if and only if* $-k < x < k$

b) $|x-a| < k$ *if and only if* $a-k < x < a+k$.■

Part a) states that $|x| < k$ if and only if x is within k units of the origin (Figure 2.26) and part b) states that $|x-a| < k$ if and only if x is within k units of a (Figure 2.27).

Remark 2.9. Theorem 2.4 is also true if we replace < by ≤.

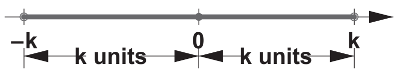


Figure 2.26. Graphical representation of relation a) of Theorem 2.4.

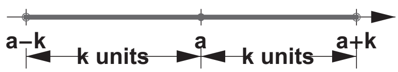


Figure 2.27. Graphical representation of relation b) of Theorem 2.4.

The following examples illustrate additional techniques for solving equations and inequalities involving absolute values.

Example 2.16. Solution of Inequalities

Solve the inequality $|x+4| > 2$.

Solution 2.16. The given inequality can be rewritten as $|x-(-4)| > 2$. Thus the solution consists of all x whose distance from -4 is greater than 2 units. This is the set $(-\infty, -6) \bigcup (-2, +\infty)$ shown in Figure 2.28.

Figure 2.28. Graphical representation of the solution for $|x+4| > 2$.

The total set is divided in two parts which are not connected. The algebraic solution is derived by

rel2 = Reduce[|x + 4| > 2, x]

```
Re[x] < -6 || (-6 ≤ Re[x] ≤ -2 && (Im[x] < -√(-12 - 8 Re[x] - Re[x]^2) ||
    Im[x] > √(-12 - 8 Re[x] - Re[x]^2))) || Re[x] > -2
```

Again we are only interested in the real part for the moment. We do not consider the imaginary part in our conclusions. The real part agrees with our graphical solution presented in Figure 2.26.▲

Example 2.17. Solution of Absolute Value Equation

Solve the following equation $|3x - 2| = |6x + 4|$.

Solution 2.17. The equation will be satisfied if either

$$3x - 2 = 6x + 4 \qquad \text{or} \qquad 3x - 2 = -(6x + 4).$$

Solving the first equation gives $x = -2$ and solving the second equation gives $x = -2/9$. Thus $x = -2$ and $x = -2/9$ are the solutions. The solutions for this equation follow using *Mathematica* by

Solve[|3 x − 2| == |6 x + 4|, x]

$$\left\{\{x \to -2\}, \left\{x \to -\frac{2}{9}\right\}\right\}$$

▲

Let us now examine the question if the magnitude operation is distributive, this mean we are looking for the following two expressions $|a + b|$ and $|a| + |b|$ and ask for the relation between them.

It is not generally true that $|a + b| = |a| + |b|$. For example, if $a = 2$ and $b = -3$, then $a + b = -1$, so that

|2 − 3| == |−3| + |2|

False

$$|a + b| = |-1| = 1$$

|2 − 3|

1

whereas

$$|a| + |b| = |2| + |-3| = 2 + 3 = 5.$$

|−3| + |2|

5

The following theorem shows that the absolute value of a sum is always less than or equal to the sum of the absolute values.

Theorem 2.5. *The Triangle Inequality*

For any real numbers a and b, $|a+b| \leq |a|+|b|$. ■

Proof 2.5. From Theorem 2.3a.

$-|a| \leq a \leq |a|$ and $-|b| \leq b \leq |b|$.

Adding these inequalities yields

$$-(|a|+|b|) \leq a+b \leq (|a|+|b|).$$

Using Theorem 2.4a with ≤ rather than <, with $x = a+b$, and with $k = |a|+|b|$, it follows that

$$|a+b| \leq |a|+|b|.$$

QED

2.4.1 Tests and Exercises

The following tests and exercises are related to the magnitude and equations containing it.

2.4.1.1 Test Problems

T1. The magnitude of $x = 7$ is given by
( ) 7
( ) −7

T2. The magnitude of $x = -\sqrt{2}$ is given by
( ) −1.41421 …
( ) 1.41421 …

T3. What is the magnitude of $a \in \mathbb{R}$
( ) a
( ) $-a$
( ) a if $a \geq 0$ and $-a$ if $a < 0$

T4. Verify the triangle inequality $|a+b| \leq |a|+|b|$ for a) $a = 3$, $b = 4$, b) $a = -4$, $b = -8$

T5. Verify $\sqrt{a^2} = |a|$ for $a = 5$ and $a = -5$.

2.4.1.2 Exercises

E1. Rewrite $\sqrt{(x-2)^2}$ without using a square root or absolute value sign.

E2. Find all values of x for which the given statement is true

- **a.** $|x-2| = 2-x$
- **b.** $|x^2+5x| = x^2+5x$
- **c.** $|2x^2+x| = x|2x+1|$
- **d.** $\sqrt{(x+5)^2} = x+5$

E3. Verify the inequalities $-|a| \leq a \leq |a|$ for $a = 1$ and for $a = -7$.

E4. Is the equality $\sqrt{a^4} = a^2$ valid for all values of a? Explain.

E5. Solve for x

- **a.** $|6x-2| = 5$
- **b.** $|6x-7| = |3+3x|$
- **c.** $2x-5 = |x+1|$
- **d.** $|3+5x| = 12$

E6. Solve for x and express the solution in terms of intervals.

- **a.** $|x+6| < 2$

$|7 - x| \leq 4$

c. $\left|\frac{1}{2}x - 2\right| \geq 2$

d. $\frac{3}{|2x-1|} \geq 4$

E7. For which values of x is $\sqrt{(x^2 - 5x + 6)^2} = x^2 - 5x + 6$ valid?

E8. Solve $2 \leq |x - 1| \leq 5$ for x.

E9. Solve $|x - 3|^2 - 4|x - 3| = 12$.

E10 Prove the following relations:

a. $|a - b| \leq |a| + |b|$

b. $|a| - |b| \leq |a - b|$

c. $||a| - |b|| \leq |a - b|$.

2.5 Cartesian Coordinate System

Just as points on a line can be placed in one-to-one correspondence with the real numbers, so points in a plane can be placed in one-to-one correspondence with pairs of real numbers by using two coordinate lines. To do this, we construct two perpendicular coordinate lines that intersect at their origins; for convenience, we make one of the lines horizontal with its positive direction to the right and the other line vertical with its positive direction upwards (see Figure 2.29). The two lines are called coordinate axes; the horizontal line is called the ***x***-axis, the vertical line is called the ***y***-axis, and the coordinate axes together form what is called a Cartesian coordinate system. The point of intersection of the coordinate axes is denoted O and is called origin of the coordinate system.

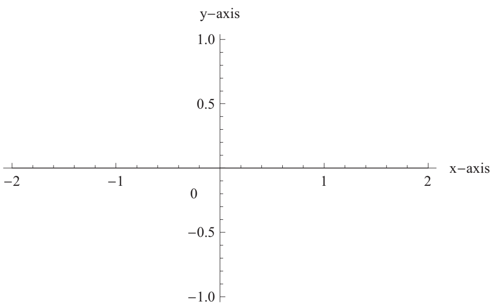


Figure 2.29. Cartesian coordinate system.

Remark 2.10. Throughout this text in the majority of cases we will assume that the same unit of measurement is used on both coordinate axes.

A plane in which a rectangular coordinate system has been introduced is called a coordinate plane or a ***x y***-plane. We will now show how to establish a one-to-one correspondence between points in a coordinate plane and pairs of real numbers. If P is a point in a coordinate plane, then we draw two lines through P, one perpendicular to the x-axis and one perpendicular to the y-axis. If the first line intersects the x-axis at the point with coordinate a and the second line intersects the y-axis at the point with coordinate b, then we associate the pair

$(a,\ b)$

with the point P (see Figure 2.30). The number a is called the ***x*-coordinate** or abscissa of P and the number b is called the ***y*-coordinate** or ordinate of P; we say that P is the point with coordinates $(a,\ b)$ and denote the point by $P(a,\ b)$.

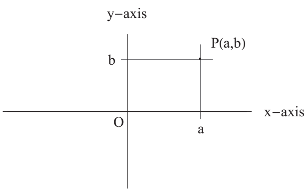


Figure 2.30. Cartesian coordinate system.

Example 2.18. Points in the Coordinate Plane

In Figure 2.31 we have located the points $P(2,\ 5)$, $Q(-4,\ 3)$, $R(-5,\ -2)$, and $S(4,\ -3)$.

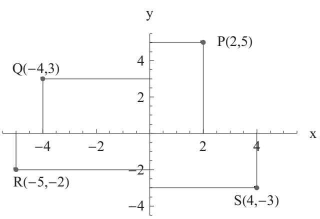


Figure 2.31. Points in the coordinate plane. ▲

From the construction above, each point in a coordinate plane determines a unique pair of numbers. Conversely, starting with a pair of numbers $(a,\ b)$, we can construct lines perpendicular to the x-axis and y-axis at the points with coordinates a and b, respectively; the intersection of these lines determine a unique point P in the plane whose coordinates are $(a,\ b)$ (Figure 2.30). Thus we have a one-to-one correspondence between pairs of real numbers and points in a coordinate plane.

Remark 2.11. Since the order in which the members of a set are listed does not matter, the set $\{a,\ b\}$ and the set $\{b,\ a\}$ are the same. However, the pair of real numbers $(a,\ b)$ and the pair $(b,\ a)$ represent different points (unless $a = b$), so that order is important. For this reason a pair $(a,\ b)$ of real numbers is often called an ordered pair.

The coordinate axes divide the plane into four parts, called quadrants. These quadrants are numbered from one to four as shown in Figure 2.32. As shown in Figure 2.33 it is easy to determine the quadrant in which a point lies from the sign of its coordinates. A point with two positive coordinates $(+,\ +)$ lies in Quadrant I, a point with negative x-coordinate and a positive y-coordinate $(-,\ +)$ lies in Quadrant II, and so on.

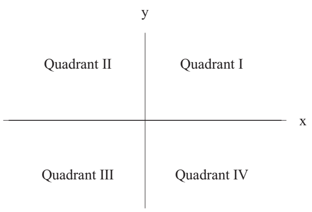


Figure 2.32. The Quadrants of a coordinate plane.

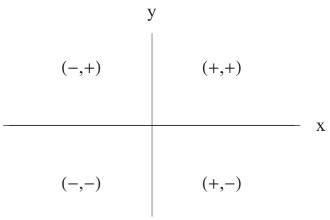


Figure 2.33. Signs of Quadrants in the coordinate plane.

Occasionally, we will use letters other than x and y to label coordinate axes. Figure 2.34 shows a uv-plane and a tQ-plane. The first letter in the name of the plane refers to the horizontal axis and the second to the vertical axis.

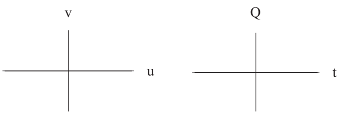


Figure 2.34. Two coordinate systems. A *uv* and a *tQ* system.

In the last section we demonstrated that the distance between points a and b on a coordinate line is $|a-b|$. It follows that the distance d between two points $A(x_1, y)$ and $B(x_2, y)$ on a horizontal line in the xy-plane is

$$d_h = |x_2 - x_1|$$

(Figure 2.35) and the distance d between two points $A(x, y_1)$ and $B(x, y_2)$ on a vertical line in the xy-plane is

$$d_v = |y_2 - y_1|$$

(see Figure 2.35).

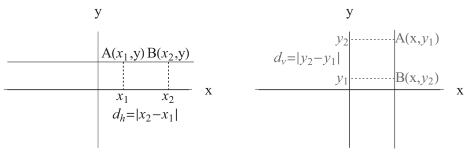


Figure 2.35. Two points on a horizontal and on a vertical line. The distances are $d_h = |x_2 - x_1|$ and $d_v = |y_2 - y_1|$.

To find the distance d between any two points $P(x_1, y_1)$ and $P(x_2, y_2)$ in the xy-plane, we apply the Theorem of Pythagoras to the triangle shown in Figure 2.36. This yields

$$d = \sqrt{|x_2 - x_1|^2 + |y_2 - y_1|^2}\,.$$

Since $|x_2 - x_1|^2 = (x_2 - x_1)^2$ and $|y_2 - y_1|^2 = (y_2 - y_1)^2$ we are led to the following result stated in Theorem 2.6.

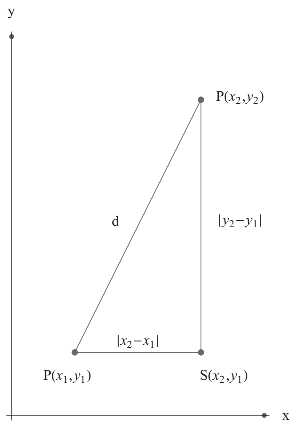


Figure 2.36. Distance *d* between two points using the Theorem of Pythagoras.

Theorem 2.6. *Distance*

The distance between two points (x_1, y_1) *and* (x_2, y_2) *in a coordinate plane is given by*
$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. ■

Example 2.19. Distance in the Coordinate Plane

The distance between $(-1, 2)$ and $(3, 4)$ is

$$\boldsymbol{d = \sqrt{(4-2)^2 + (3+1)^2}}$$

$$2\sqrt{5}$$

▲

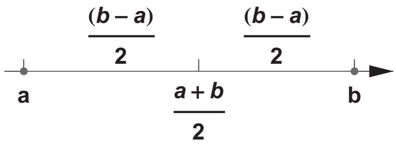


Figure 2.37. Midpoint of a line segment on the real line.

On a coordinate line, the average $\frac{1}{2}(a+b)$ is the midpoint of the line segment joining a and b (Figure 2.37). In the xy-plane the midpoint of the line segment joining points (x_1, y_1) and (x_2, y_2) is obtained by averaging the coordinates of the endpoints in each direction. More precisely: the midpoint of the line segment joining $P(x_1, y_1)$ and $P(x_2, y_2)$ is

$$m_p = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

(see Figure 2.38). The proof is left as an exercise.

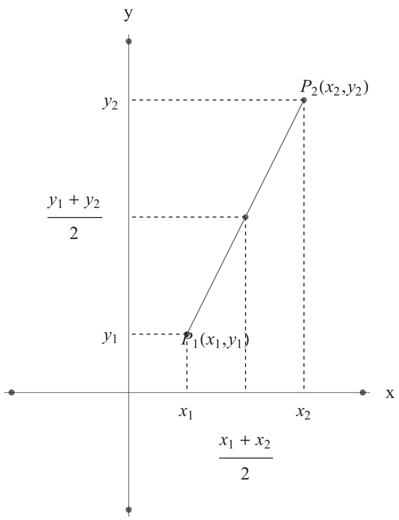


Figure 2.38. Midpoint of a line segment in the *xy*-plane.

Example 2.20. Midpoint of a Line segment

Determine the coordinates of the midpoint of the line segment between the two points (3, −4) and (7, 2).

$$\left\{\frac{3+7}{2}, \frac{-4+2}{2}\right\}$$

```
{5, -1}
```

▲

2.5.2 Tests and Exercises

The following tests and exercises are related to Cartesian coordinates.

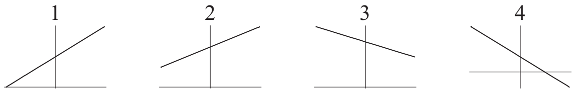
2.5.2.1 Test Problems

T1. Draw a rectangle, three of whose vertices are (6, 1), (−4, 1), and (6, 7), and find the coordinates of the fourth vertex.

T2. Draw the triangle whose vertices are (−3, 2), (5, 2), and (4, 3), and find its area.

T3. Draw a circle whose center and radius is $C = (1, 2)$ and $r = 2$. Label the circle's x- and y-intercepts, if any, with their coordinate pairs.

T4. List the lines in the accompanying figure in the order of decreasing slope.



T5. A particle initially at (2, 3), moves along a line of slope $m = -2$ to a new point (x, y). And y if $x = 4$.

2.5.2.2 Exercises

E1. Draw a rectangular coordinate system and sketch the set of points whose coordinates (x, y) satisfy the given conditions.

- **a.** $x = 1$,
- **b.** $y = -2$,
- **c.** $x > 1$,
- **d.** $y \leq -2$,
- **e.** $x \geq y$,
- **f.** $|x| \geq 3$.

E2. Find the slopes of the sides of the triangle with vertices (−2, 1), (5, 6), and (7, 2).

E3. Find the slope of the line through

- **a.** (−1, 2) and (3,4)
- **b.** (−2, −6) and (−3, 8)
- **c.** $\left(1, \sqrt{2}\right)$ and $\left(\sqrt{2}, 2\right)$.

E4. Draw the line through (3, 2) with slope

- **a.** $m = 1/2$,
- **b.** $m = -2$,
- **c.** $m = 2$,
- **d.** $m = -1/3$.

E5. Let the point $(2, k)$ lie on the line of slope $m = 5$ through (−2, 6); find k.

E6. Draw the line through (3, 2) with slope

E7. Graph the equations

- **a.** $y = 2x - 1$
- **b.** $y = -3x$
- **c.** $x = 3y - 5$
- **d.** $y = 2$

E8. Find the slope and y-intercept of

- **a.** $y = 3x + 2$
- **b.** $3x + 2y = 8$
- **c.** $5x - 2y = 13$

d. $y = 1$

E9. Find the slope and intercept form of the line satisfying the given conditions:

a. Slope $m = -2$, y-intercept $b = 4$.

b. $m = 5$, $b = -4$.

c. The line is parallel to $y = 4x - 3$ and its y-intercept is 5.

d. The line is perpendicular to $y = 3x - 8$ and its y-intercept is 7.

e. The line is parallel to $y = -7$ and passes through $(-1, -8)$

E10 Graph $F = \frac{9}{5}C + 32$ in a CF-coordinate system.

E11 In each part, classify the lines as parallel, perpendicular, or neither.

a. $y = 4x - 7$ and $y = 4x + 9$

b. $y - 1 = 2(x - 3)$ and $y - 4 = \frac{1}{2}(x + 7)$

c. $y = \frac{1}{2}x$ and $x = \frac{1}{2}y$.

2.6 Trigonometry Review

The purpose of this section is to review the fundamentals of trigonometry.

We shall be concerned with angles in the plane that are generated by rotating a ray (or half-line) about its endpoint. The starting position of the ray is called the initial side of the angle and the final position of the ray is called the terminal side. The initial and terminal sides meet at a point called the vertex of the angle (Figure 2.39)

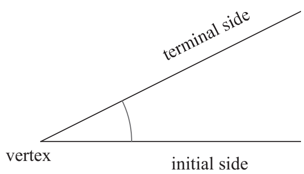


Figure 2.39. The angle as a range between two rays.

An angle is considered positive if it is generated by a counterclockwise rotation and negative if it is generated by a clockwise rotation. In a rectangular coordinate system, an angle is said to be in standard position if its vertex is at the origin and its initial side is along the positive x-axis (see Figure 2.40).

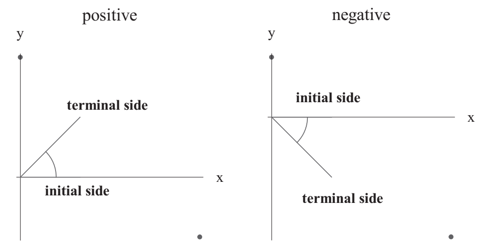


Figure 2.40. Positive (left) and negative (right) angles in standard position.

As shown in Figure 2.41, an angle may be generated by making more than one complete revolution.

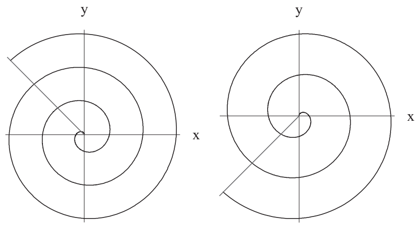


Figure 2.41. Multiple revolutions of angles.

The sine, cosine, tangent, cosecant, secant, and cotangent of a positive acute angle θ can be defined as ratio of the sides of a right triangle. Using the notation from Figure 2.42, these definitions take the form:

$$sin(\theta) = \frac{\text{side opposite } \theta}{\text{hypotenuse}} = \frac{y}{r} \tag{2.5}$$

$$cos(\theta) = \frac{\text{side adjacent to } \theta}{\text{hypotenuse}} = \frac{x}{r} \tag{2.6}$$

$$tan(\theta) = \frac{\text{side opposite } \theta}{\text{side adjacent to } \theta} = \frac{y}{x} \tag{2.7}$$

$$csc(\theta) = \frac{\text{hypotenuse}}{\text{side opposite } \theta} = \frac{r}{y} \tag{2.8}$$

$$sec(\theta) = \frac{\text{hypotenuse}}{\text{side adjacent to } \theta} = \frac{r}{x} \tag{2.9}$$

$$cot(\theta) = \frac{\text{side adjacent to } \theta}{\text{side opposite to } \theta} = \frac{x}{y} \tag{2.10}$$

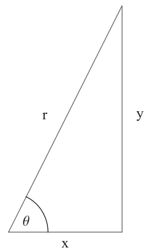


Figure 2.42. Sides and angle in a rectangular triangle.

We will call sin, cos, tan, csc, sec, and cot the trigonometric functions. Because similar triangles have proportional sides, the values of the trigonometric functions depend only on the size of θ and not on the particular right triangle used to compute the ratios.

Example 2.21. Elementary Relations

Recall from high school geometry that the two legs of a 45°-45°-90° triangle are equal of size. This fact and the Theorem of Pythagoras yield Figure 2.43. Recall also that the hypotenuse of a 30°-60°-90° triangle is twice the size of the shorter leg and the shorter leg is opposite the 30° angle. These facts and the Theorem of Pythagoras yield Figure 2.43.

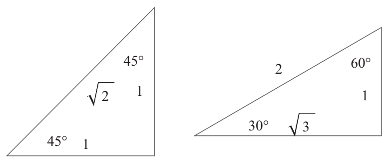


Figure 2.43. 45°-45°-90° triangle and 30°-60°-90° triangle.

From Figure 2.43 we obtain the following relations for trigonometric functions which are quite useful to remember

	30°	45°	60°
sin	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$
cos	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$
tan	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$
cot	$\sqrt{3}$	1	$\frac{1}{\sqrt{3}}$
sec	$\frac{2}{\sqrt{3}}$	$\sqrt{2}$	2
csc	2	$\sqrt{2}$	$\frac{2}{\sqrt{3}}$

Table 2.1. Collection of values for different trigonometric functions for arguments $\theta = 30\,°$, $45\,°$, and $60\,°$.▲

Since a right triangle cannot have an angle larger than 90°, formulas (2.5-2.10) are not applicable if θ is obtuse. To obtain a definition that applies to all angles, we take the following approach: Given an angle θ, introduce a coordinate system so that the angle is in standard position. Then construct a circle of arbitrary radius r centered at the origin and let $P(x, y)$ be the point where the terminal side of the angle intersects the circle (see Figure 2.44). We make the following definition:

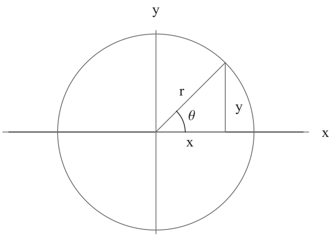


Figure 2.44. Geometric relations in a circle to define the trigonometric relations.

Definition 2.9. *Trigonometric Relations*

Trigonometric functions are defined by the following geometric relations (Figure 2.44):

$sin(\theta) = \frac{y}{r}$, $cos(\theta) = \frac{x}{r}$, $tan(\theta) = \frac{y}{x}$

$csc(\theta) = \frac{r}{y}$, $sec(\theta) = \frac{r}{x}$, $cot(\theta) = \frac{x}{y}$.■

The formulas in Definition 2.9 agree with the formulas (2.5-2.10). However, Definition 2.9 applies to all angles—positive, negative, acute, or obtuse. Note that $tan(\theta)$ and $\sec(\theta)$ are undefined if the terminal side of θ is on the y-axis (since $x = 0$), while $csc(\theta)$ and $cot(\theta)$ are undefined if the terminal side of θ is on the x-axis (since $y = 0$).

Example 2.22. Trigonometric Relations

In Figure 2.45 a point P is defined by the coordinates

$x = -3,$ $y = 4,$ and $r = 5$

so that for the angle θ shown the following relations will hold.

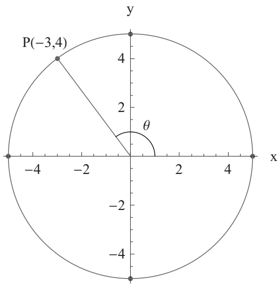


Figure 2.45. Angle and point relation.

$$\boldsymbol{\sin(\theta) = \frac{4}{5}}$$

$$\sin(\theta) = \frac{4}{5}$$

$$\boldsymbol{\cos(\theta) = -\frac{3}{5}}$$

$$\cos(\theta) = -\frac{3}{5}$$

$$\boldsymbol{\tan(\theta) = -\frac{4}{3}}$$

$$\tan(\theta) = -\frac{4}{3}$$

$$\boldsymbol{\csc(\theta) = \frac{5}{4}}$$

$$\csc(\theta) = \frac{5}{4}$$

$$\boldsymbol{\sec(\theta) = -\frac{5}{3}}$$

$$\sec(\theta) = -\frac{5}{3}$$

$$\mathbf{cot(\theta) = -\frac{3}{4}}$$

$$\cot(\theta) = -\frac{3}{4}$$

▲

Since the value of a trigonometric function is determined by the point where the terminal side of the angle intersects a circle of radius r, it follows that the value of a trigonometric function will be the same for any two angles with the same terminal side (like those in Figure 2.45).

Using properties of similar triangles, it can be shown that the values of the trigonometric functions in Definition 2.9 are the same for every value r. In particular, for the unit circle (the circle of radius $r = 1$ centered at the origin) the formulas for $sin(\theta)$ and $cos(\theta)$ in Definition 2.9 become

$x = cos(\theta)$, $y = sin(\theta)$.

Thus we have the following geometric interpretation of $sin(\theta)$ and $cos(\theta)$ (Figure 2.46)

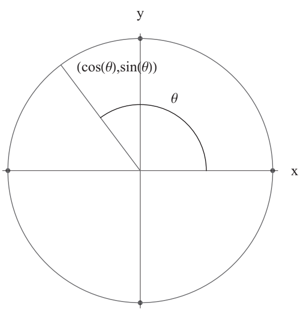


Figure 2.46. Geometric interpretation of sin and cos on a unit circle.

Theorem 2.7. *Geometric Interpretation of sin(θ) and cos(θ)*

If an angle θ is placed in standard position, then its terminal side intersects the unit circle at a point $(cos(\theta), sin(\theta))$. ■

The geometric relations of the coordinates of a point and its coordinates on a unit circle are shown in Figure 2.46.

In calculus, angles are measured in radians rather than degrees, minutes, and seconds because it simplifies many important formulas. Before we can define radial measure, we need some preliminary ideas.

Let Q be an arbitrary point on a circle of radius r and imagine that a particle, initially at Q, moves around the circle and stops at a point P (Figure 2.47). During its motion the point will travel a distance d along the circle. To distinguish between clockwise and counterclockwise motions we introduce the signed distance or signed arc length s traveled by the point; it is defined by:

$s = d$ if the motion is counterclockwise,
- $s = -d$ if the motion is clockwise,
- $s = 0$ if there is no motion.

For example, if $s = 3$, the point has traveled 3 units in the counterclockwise direction from Q; and if $s = -\pi$, the point has traveled π units in the clockwise direction from Q.

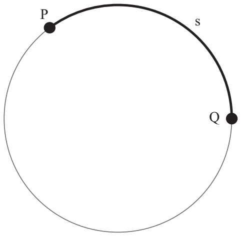


Figure 2.47. The radian measure.

Definition 2.10. *Radian Measure*

To define the radian measure of an angle, we place the vertex of the angle at the center of a circle of arbitrary radius r. As a ray sweep out the angle, the intersection of the ray with the circle moves some signed distance s along the circle (Figure 2.48). We define

$$\theta = \frac{s}{r}$$

to be the radian measure of the angle. ■

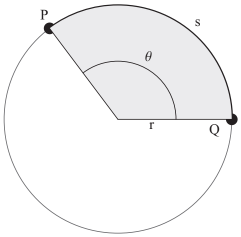


Figure 2.48. Definition of the radian measure.

Remark 2.12. We leave it as an exercise to show that the radian measure of an angle depends only on the size of the angle, and not on the radius r selected for the circle.

Example 2.23. Radian Measure I

Express the angle 90° in radian measure.

Solution 2.23. An angle 90° places with its vertex at the center of a circle of radius r intercepts one fourth of the circumference (Figure 2.48). Since the full circumference is $2\pi r$, we obtain

$$s = \frac{1}{4}(2\pi r) = \frac{1}{2}\pi r$$

Thus the radian measure of the angle is

$$\theta = \frac{s}{r} = \frac{\frac{1}{2}\pi r}{r} = \frac{\pi}{2}.▲$$

Example 2.24. Radian Measure II

From the fact that 90° corresponds to $\pi/2$, the reader should be able to obtain the relationships in the following table

$$\frac{\{30°, 45°, 90°, 120°, 135°, 150°, 180°, 270°, 360°\}}{\frac{360°}{2\pi}}$$

$$\left\{\frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{3\pi}{4}, \frac{5\pi}{6}, \pi, \frac{3\pi}{2}, 2\pi\right\}$$

The division by $360°/(2\pi)$ can be done in *Mathematica* in one step because we are using a list ({}) collecting the data which is accessible by the division operator (/) element by element.▲

Example 2.25. Radian Measure and Trigonometric Functions

Evaluate the trigonometric functions at $\theta = 5\pi/6$.

```
Map[Apply[#, {5 π / 6}] &, {Sin, Cos, Tan, Cot, Sec, Csc}]
```

$$\left\{\frac{1}{2}, -\frac{\sqrt{3}}{2}, -\frac{1}{\sqrt{3}}, -\sqrt{3}, -\frac{2}{\sqrt{3}}, 2\right\}$$

In this example we use another feature of *Mathematica* to work on lists. We map the function name to a list containing the angle $\theta = 5\pi/6$.▲

In the following we discuss some trigonometric identities. It follows from Definition 2.9 that

$$csc(\theta) = \frac{1}{sin(\theta)}, \qquad sec(\theta) = \frac{1}{cos(\theta)}, \qquad cot(\theta) = \frac{1}{tan(\theta)} \tag{2.11}$$

and

$$tan(\theta) = \frac{sin(\theta)}{cos(\theta)} \qquad cot(\theta) = \frac{cos(\theta)}{sin(\theta)}. \tag{2.12}$$

Moreover, from Figure 2.44 and the Theorem by Pythagoras, we have

$$y^2 + x^2 = r^2.$$

Divide both sides by r^2 and using the definition of $sin(\theta)$ and $cos(\theta)$, we obtain the following fundamental result

$$sin^2(\theta) + cos^2(\theta) = 1. \tag{2.13}$$

Remark 2.13. The relationships (2.11), (2.12), and (2.13) are identities, which means that the equations hold for all values of θ where both sides are defined.

Figure 2.49 shows two angles, θ and $-\theta$, having equal magnitude but opposite sign. By symmetry, the terminal sides of these angles intersect a circle centered at the origin at points whose x-coordinates are equal and whose y-coordinates are different only in sign. Thus

$$sin(-\theta) = \frac{-y}{r} = -sin(\theta) \tag{2.14}$$

$$cos(-\theta) = \frac{x}{r} = cos(\theta). \tag{2.15}$$

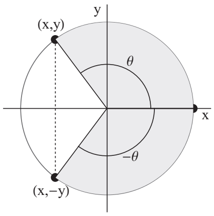


Figure 2.49. Symmetries of trigonometric relations.

In summary, we have the identities

$$sin(-\theta) = -sin(\theta) \tag{2.16}$$

$$cos(-\theta) = cos(\theta). \tag{2.17}$$

As previously noted, a trigonometric function has the same value for two angles with the same terminal sides. In radian measure, the angle θ, $\theta + 2\pi$, and $\theta - 2\pi$ all have the same terminal side, so

$$sin(\theta) = sin(\theta + 2\pi) = sin(\theta - 2\pi) \tag{2.18}$$

$$cos(\theta) = cos(\theta + 2\pi) = cos(\theta - 2\pi). \tag{2.19}$$

More generally, any multiple of 2π can be added to or subtracted from an angle θ in radians without changing the terminal side. Thus

$$sin(\theta) = sin(\theta \pm 2n\pi), \qquad n = 0, 1, 2, 3, \ldots \tag{2.20}$$

$$cos(\theta) = cos(\theta \pm 2n\pi), \qquad n = 0, 1, 2, 3 \ldots. \tag{2.21}$$

The following result, called the law of cosines, generalizes the Theorem of Pythagoras:

Theorem 2.8. *Law of Cosines*

If the sides of a triangle have lengths a, b, and c, and if θ is the angle between the sides with lengths a and b, then

$$c^2 = a^2 + b^2 - 2\,a\,b\,cos(\theta).\blacksquare \qquad (2.22)$$

Proof 2.8. Introduce a coordinate system so that θ is in standard position an the side of length a falls along the positive x-axis (Figure 2.50)

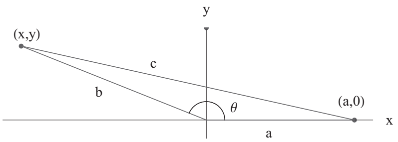


Figure 2.50. Geometry of the law of cosine.

As shown in Figure 2.50, the side of length a extends from the origin to $(a, 0)$ and the side of length b extends from the origin to some point (x, y). From the definition of $sin(\theta)$ and $cos(\theta)$ we have

$$\mathbf{eq1 = sin(\theta) = \frac{y}{b}}$$

$$\sin(\theta) = \frac{y}{b}$$

$$\mathbf{eq2 = cos(\theta) = \frac{x}{b}}$$

$$\cos(\theta) = \frac{x}{b}$$

so that the solution with respect to x and y delivers

sol1 = Flatten[Solve[{eq1, eq2}, {*x*, *y*}]]

$\{y \to b\sin(\theta),\ x \to b\cos(\theta)\}$

From the distance formula in Theorem 2.4, we obtain

$$\mathbf{eq3 = c^2 = (x - a)^2 + (y + 0)^2}$$

$$c^2 = (x - a)^2 + y^2$$

with the derived solution sol1 for the variables x and y we gain

Simplify[eq3 /. sol1]

$$a^2 + b^2 = 2\,a\,b\cos(\theta) + c^2$$

which completes the proof. The operator /. represents the replacement operator using rules to replace symbols, here x and y.

Theorem 2.9. *The Sine Law*

If the sides of a triangle have lengths a, b, and c which are incorporated in a circle of radius R and if α, β, θ are the angles opposite to the side a, b, and c, respectively, then the following *relation holds*

$$\frac{a}{\sin(\alpha)} = \frac{b}{\sin(\beta)} = \frac{c}{\sin(\theta)} = 2\,R.\blacksquare$$

Using the law of cosine we will be able to establish the following identities, called the addition formula for sin and cosine.

$$sin(\alpha-\beta) = sin(\alpha)\,cos(\beta) - cos(\alpha)\,sin(\beta) \tag{2.23}$$

$$cos(\alpha-\beta) = cos(\alpha)\,cos(\beta) + sin(\alpha)\,sin(\beta) \tag{2.24}$$

$$sin(\alpha+\beta) = sin(\alpha)\,cos(\beta) + cos(\alpha)\,sin(\beta) \tag{2.25}$$

$$cos(\alpha+\beta) = cos(\alpha)\,cos(\beta) - sin(\alpha)\,sin(\beta) \tag{2.26}$$

Identities (2.25) and (2.26) can be obtained by substituting $-\beta$ for β in (2.23) and (2.24) and using the identities

$$\sin(-\beta) = -\sin(\beta), \qquad \cos(-\beta) = \cos(\beta). \tag{2.27}$$

We shall derive (2.24) and then deduce (2.23). In our derivation of (2.24), we will assume that $0 \leq \beta < \alpha < 2\,\pi$. The proofs for the remaining cases are left to the reader as exercises. These proves follow the same line of arguments as we will present it.

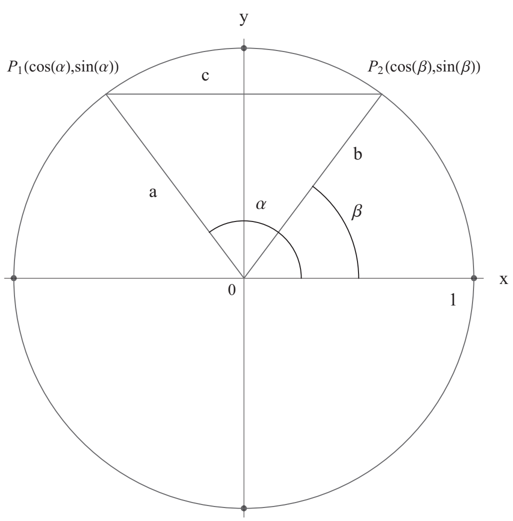


Figure 2.51. Geometry for the addition formulas.

As shown in Figure 2.51, the terminal sides of α and β intersect the unit circle at the points $P_1(cos(\alpha),\ sin(\alpha))$ and $P_2(cos(\beta),\ sin(\beta))$. If we denote the lengths of the sides of triangle O P_1 P_2 by a, c, and b, then $a = b = 1$ and from the distance formula in Theorem 2.8 we gain

eq1 = c^2 == Expand[$(\sin(\beta)-\sin(\alpha))^2+(\cos(\beta)-\cos(\alpha))^2$]

$c^2 = -2\sin(\alpha)\sin(\beta)-2\cos(\alpha)\cos(\beta)+\sin^2(\alpha)+\cos^2(\alpha)+\sin^2(\beta)+\cos^2(\beta)$

In the next line we replace $\sin^2(a)$ by $1-\cos(a)$

eq1 = eq1 /. $\sin^2$(a_) :→ $1-\cos^2(a)$

$c^2 = -2\sin(\alpha)\sin(\beta)-2\cos(\alpha)\cos(\beta)+2$

But angle $P_1\,O\,P_2=\alpha-\beta$, so that the law of cosine yields

eq2 = $c^2 == a^2-2\,a\,b\cos(\alpha-\beta)+b^2$

$c^2 = a^2-2\,a\,b\cos(\alpha-\beta)+b^2$

where $a=b=1$. These relations are introduced by the rules $\{a\to 1,\ b\to 1\}$ and the replacement /. in *Mathematica* notation

eq2 = eq2 /. {$a\to 1$, $b\to 1$}

$c^2 = 2-2\cos(\alpha-\beta)$

Equating the two expressions for c^2 and simplifying, we obtain

Thread[$-\frac{1}{2}$ Thread[(eq2[[2]] == eq1[[2]]) − 2, Equal], Equal]

$\cos(\alpha-\beta) = \frac{1}{2}(2\sin(\alpha)\sin(\beta)+2\cos(\alpha)\cos(\beta))$

which completes the derivation. The same relation is gained by the *Mathematica* function TrigExpand[] applied to the trigonometric expression

TrigExpand[cos($\alpha-\beta$)]

$\sin(\alpha)\sin(\beta)+\cos(\alpha)\cos(\beta)$

TrigExpand[sin($\alpha-\beta$)]

$\sin(\alpha)\cos(\beta)-\cos(\alpha)\sin(\beta)$

To derive identity (2.23), we will need the identities

$$cos\left(\frac{\pi}{2}-\alpha\right) = sin(\alpha) \tag{2.28}$$

$$\sin\left(\frac{\pi}{2}-\alpha\right) = \cos(\alpha) \tag{2.29}$$

which states that the sine of an angle is the cosine of its complement and the cosine of an angle is the sine of its complement. The derivations of (2.28) and (2.29) are left as exercises.

We can now derive (2.23) as follows:

$$
\begin{aligned}
sin(\alpha - \beta) &= cos\left(\frac{\pi}{2} - (\alpha - \beta)\right) \\
&= cos\left(\left(\frac{\pi}{2} - \alpha\right) + \beta\right) \\
&= cos\left(\left(\frac{\pi}{2} - \alpha\right) - (-\beta)\right) \\
&= cos\left(\frac{\pi}{2} - \alpha\right) cos(-\beta) + sin\left(\frac{\pi}{2} - \alpha\right) sin(-\beta) \\
&= cos\left(\frac{\pi}{2} - \alpha\right) cos(\beta) - sin\left(\frac{\pi}{2} - \alpha\right) sin(\beta) \\
&= sin(\alpha)\, cos(\beta) - cos(\alpha)\, sin(\beta).
\end{aligned}
$$

In the special case where $\alpha = \beta$, (2.24) and (2.25) reduce to the identities

$$\sin(2\,\alpha) = 2\sin(\alpha)\cos(\alpha) \tag{2.30}$$

$$\cos(2\,\alpha) = \cos^2(\alpha) - \sin^2(\alpha). \tag{2.31}$$

These are called the double-angle formulas. By using the identity $\sin^2(\alpha) + \cos^2(\alpha) = 1$ (2.31) can be rewritten in the alternate forms

$$\cos(2\,\alpha) = 2\cos^2(\alpha) - 1 \tag{2.32}$$

$$\cos(2\,\alpha) = 1 - 2\sin^2(\alpha). \tag{2.33}$$

If we replace α by $\alpha/2$ in (2.32) and (2.33) and use some algebra, we obtain the half-angle formulas.

$$\cos^2\left(\frac{\alpha}{2}\right) = \frac{1 + \cos(\alpha)}{2} \tag{2.34}$$

$$\sin^2\left(\frac{\alpha}{2}\right) = \frac{1 - \cos(\alpha)}{2}. \tag{2.35}$$

Identities (2.23) and (2.25) and some algebra yield the first of the following product formulas.

$$\sin(\alpha)\cos(\beta) = \frac{1}{2}\,(\sin(\alpha - \beta) + \sin(\alpha + \beta)) \tag{2.36}$$

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}\,(\cos(\alpha - \beta) - \cos(\alpha + \beta)) \tag{2.37}$$

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}\,(\cos(\alpha - \beta) + \cos(\alpha + \beta)) \tag{2.38}$$

The derivations of (2.37) and (2.38) are left as exercises.

We can deduce some important identities involving tangent, cotangent, secant, and cosecant from (2.12) and (2.13). The first of the identities

$$\tan^2(\theta) + 1 = \sec^2(\theta) \tag{2.39}$$

$$1 + \cot^2(\theta) = \csc^2(\theta) \tag{2.40}$$

can be obtained by dividing

$$\sin^2(\theta) + \cos^2(\theta) = 1 \tag{2.41}$$

by $\cos^2(\theta)$ and the second by dividing by $\sin^2(\theta)$.

As an exercise we ask the reader to deduce the first of the identities

$$\tan(\alpha + \beta) = \frac{\tan(\alpha) + \tan(\beta)}{1 - \tan(\alpha)\tan(\beta)} \tag{2.42}$$

$$\tan(\alpha - \beta) = \frac{\tan(\alpha) - \tan(\beta)}{1 + \tan(\alpha)\tan(\beta)} \tag{2.43}$$

from the relation $\tan(\alpha + \beta) = \sin(\alpha + \beta) / \cos(\alpha + \beta)$. Identity (2.43) is obtained from (2.42) by substituting $-\beta$ for β and using the relationship

$$\tan(-\beta) = \frac{\sin(-\beta)}{\cos(-\beta)} = \frac{-\sin(\beta)}{\cos(\beta)} = -\tan(\beta). \tag{2.44}$$

If we let $\alpha = \beta$ in (2.42), we obtain the double-angle formula

$$\tan(2\,\alpha) = \frac{2\tan(\alpha)}{1 - \tan^2(\alpha)}. \tag{2.45}$$

If we let $\beta = \pi$ in (2.42) and (2.43) and use the fact that $tan(\pi) = 0$, we obtain the identities

$$\tan(\alpha + \pi) = \tan(\alpha) \tag{2.46}$$

$$\tan(\alpha - \pi) = \tan(\alpha). \tag{2.47}$$

The trigonometric identities obtained in this section are the ones most commonly used.

Even in the days of symbolic calculations it is necessary to know these identities. The following example demonstrates that a symbolic calculation only delivers preliminary results.

TrigExpand[tan(2 α)]

$$\frac{2\sin(\alpha)\cos(\alpha)}{\cos^2(\alpha) - \sin^2(\alpha)}$$

TrigFactor[tan(2 α)]

$$\frac{2\sin(\alpha)\cos(\alpha)}{(\cos(\alpha) - \sin(\alpha))(\sin(\alpha) + \cos(\alpha))}$$

2.6.3 Tests and Exercises

The following tests and exercises are related to trigonometric functions.

2.6.3.1 Test Problems

T1. Which of the following statements is true.

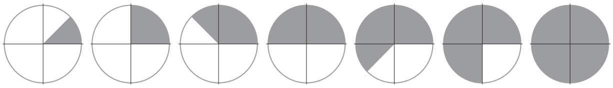
( ) To convert x in radians to an angle in degree we use $\alpha = \frac{180^\circ}{\pi}\, x$.

( ) $\sin(x)$, $x \in \mathbb{R}$, is periodic with period $2\,\pi$.

( ) $\sin(x)$, $x \in \mathbb{R}$, is periodic with period $4\,\pi$.

( ) $\sin\left(\frac{2\,\pi}{T}\,x\right)$, $x \in \mathbb{R}$, with $T > 0$ is periodic with period T.

T2. Use the terms acute, right, obtuse, strait, reflex, and full to denote the angles in the following figure.



T3. Are the following identities correct?

( ) $\sin\left(\frac{\pi}{2} + \theta\right) = \cos(\theta)$

( ) $\cos\left(\frac{\pi}{2} + \theta\right) = -\sin(\theta)$

( ) $\sin\left(\frac{3\,\pi}{2} - \theta\right) = -\cos(\theta)$

( ) $\cos\left(\frac{3\,\pi}{2} - \theta\right) = -\sin(\theta)$

T4. Which of the following identities are true.

( ) $\cos(\theta) = \cos\left(\frac{\pi}{3} + \theta\right) + \cos\left(\frac{\pi}{3} - \theta\right)$

( ) $\tan\left(\frac{\theta}{2}\right) = \frac{1+\cos(\theta)}{\sin(\theta)}$

( ) $\sin(3\,\theta) + \sin(\theta) = 2\,\sin(2\,\theta)\cos(\theta)$

( ) $\tan\left(\frac{\theta}{2}\right) = \frac{\sin(\theta)}{1+\cos(\theta)}$

T5. Express the angles in degree or radians.

a. $75\,^\circ$

b. $376\,^\circ$

c. 2

d. $3\,\pi/2$

e. $\pi/13$.

2.6.3.2 Exercises

E1. In each part, let θ be an acute angle of a right triangle. Express the remaining five trigonometric functions in terms of a.

a. $\sin(\theta) = a/3$

b. $\tan(\theta) = a/7$

c. $\sec(\theta) = a/2$

E2. Use the information to find the exact values of the remaining five trigonometric functions of θ.

a. $\cos(\theta) = 2/5$, $0 < \theta < \pi/2$

b. $\tan(\theta) = 1\big/\sqrt{2}$, $\pi/2 < \theta < \pi$

c. $\sin(\theta) = 1/5$, $0 < \theta < \pi/2$

d. $\cot(\theta) = 1/3$, $0 < \theta < \pi/2$

E3. Find all values of θ (in radians) such that

a. $\sin(\theta) = 1$

b. $\cos(\theta) = 1$

c. $\tan(\theta) = 1$

d. $\cot(\theta) = 1$

E4. How could you use a ruler and protractor to approximate $\sin(17\,^\circ)$ and $\cos(17\,^\circ)$.

E5. Find the length of the circular arc on a circle of radius 4 cm subtended by an angle of

a. $\pi/6$,

b. $150\,°$.

E6. Find a formula for the area A of a circular sector in terms of its radius r and arc length s.

E7. Two sides of a triangle have lengths of 3 cm and 7 cm and meet at an angle of $60\,°$. Find the area of the triangle.

E8. An observer on level ground is at a distance d from a building. The angles of elevation to the bottoms of the windows on the second and third floor are α and β, respectively. Find the distance h between the bottoms of the Windows in terms of α, β, and d.

E9. Prove: The area A of a triangle ABC can be written as

$$A = \frac{1}{2}\, b\, c \sin(\alpha) \tag{1}$$

E10 Prove the law of sines: In any triangle, the ratio of the sides to the sines of the opposite angle are equal; that is

$$\frac{a}{\sin(\alpha)} = \frac{b}{\sin(\beta)} = \frac{c}{\sin(\gamma)}. \tag{2}$$

2.7 Complex Numbers

As we introduced in section 2.2 complex numbers originate from a desire to extract square roots of negative numbers. Since we now are familiar with trigonometric functions we can discuss a link to complex numbers and trigonometric functions. A very important practical application of complex numbers is the calculation of alternating currents in electrical engineering applications. Alternating currents are usually represented by means of trigonometric functions. The relation between trigonometric functions and complex numbers were first taken seriously in the eighteenth century by mathematicians such as de Moivre, who proved the first theorem in the subject in 1722. Also Euler, who introduced the notation for $\sqrt{-1}$, and who discovered the mysterious formula $e^{i\theta} = \cos(\theta) + i \sin(\theta)$ in 1748. And third Gauss, who was the first to prove the fundamental theorem of algebra concerning existence of roots of polynomial equations in 1799. The nineteenth century saw the construction of the first model for the complex numbers by Argand in 1806, later known as the Argand diagram, and more recently as the complex plane. Also the first attempts to do calculus with complex numbers by Cauchy in 1825. Complex numbers were first so called by Gauss in 1831. Previously they were known as imaginary numbers, or impossible numbers. It was not until the twentieth century that complex numbers found applications in science and technology, particularly to electrical engineering and fluid dynamics.

If we want square roots of negative numbers it is sufficient to introduce $i = \sqrt{-1}$ because then, for example, $\sqrt{-2} = \sqrt{-1}\,\sqrt{2}$. Combining i with real numbers by addition and multiplication cannot produce anything more general than $x + i\,y$ where x, y are real. This is because the sum and product of any two numbers of this form are also of this form. For example,

$$\mathbf{(1 + 2\, i) + (3 + 4\, i)}$$

$$4 + 6\, i$$

and the product of two complex numbers results to

$$(1 + 2\,i)\,(3 + 4\,i)$$

$$-5 + 10\,i$$

Subtraction produces nothing new since, for example,

$$(1 + 2\,i) - (3 + 4\,i)$$

$$-2 - 2\,i$$

Neither does division since, for example,

$$\frac{1 + 2\,i}{3 + 3\,i}$$

$$\frac{1}{2} + \frac{i}{6}$$

The number $3 - 3\,i$ is called the conjugate of $3 + 3\,i$ which is used in the denominator to generate a real number. For any $x + i\,y$ we have

ComplexExpand[(x + i y) (x − i y)]

$$x^2 + y^2$$

so division can always be done except when $x = y = 0$, that is, when $x + i\,y = 0$.

It is also possible to extract square roots of numbers of the form $x + i\,y$ as numbers of the same form. For example, suppose

$$\sqrt{1 + 2\,i} = A + i\,B$$

$$\sqrt{1 + 2\,i} = A + i\,B$$

then we have

ComplexExpand$\left[\sqrt{1 + 2\,i}^{\,2} = (A + i\,B)^2\right]$

$$1 + 2\,i = A^2 + 2\,i\,A\,B - B^2$$

So we require

$$A^2 - B^2 = 1$$

$$A^2 - B^2 = 1$$

and

$$A\,B = 1$$

$$A\,B = 1$$

The second equation gives $B = 1 / A$, which on substitution in the first equation gives

equation1 = $A^2 - B^2 = 1$ /. $B \to \frac{1}{A}$

$$A^2 - \frac{1}{A^2} = 1$$

Solving this quartic equation in A we obtain

solution = Solve[equation1, A]

$$\left\{\left\{A \to -\sqrt{\frac{1}{2}+\frac{\sqrt{5}}{2}}\right\}, \left\{A \to \sqrt{\frac{1}{2}+\frac{\sqrt{5}}{2}}\right\}, \left\{A \to -i\sqrt{\frac{1}{2}\left(\sqrt{5}-1\right)}\right\}, \left\{A \to i\sqrt{\frac{1}{2}\left(\sqrt{5}-1\right)}\right\}\right\}$$

Since we assumed that A and B are real quantities we can neglect the two imaginary solutions.

Simplify[$\sqrt{1+2i} = A + \frac{i}{A}$ /. solution⟦2⟧]

$$\frac{\sqrt{5}+(1+2i)}{\sqrt{2\left(1+\sqrt{5}\right)}} = \sqrt{1+2i}$$

This last property of numbers of the form $x + i\,y$ represents a bonus over what might reasonably have been expected. Introducing square roots of negative real numbers is one thing. Creating a number system in which square roots can always be taken is asking rather more. But this is precisely what we have achieved. Existence of square roots means that quadratic equations can always be solved. We shall see shortly that much more is true, namely that polynomial equations of any degree can be solved with numbers of the form $x + i\,y$. This is the fundamental theorem of algebra.

2.7.1 Notation and Terminology

If $i = \sqrt{-1}$, then numbers of the form $x + i\,y$ are called complex numbers. We write $z = x + i\,y$ and call x the real part of z which we abbreviate to $\mathcal{Re}(z)$, and y the imaginary part of z which we abbreviate to $\mathcal{Im}(z)$; where both $\mathcal{Re}(z)$ and $\mathcal{Im}(z)$ are real.

For $z = x + i\,y$ we write (by definition) $\overline{z} = x - i\,y$, and call $\overline{z} = z^*$ the conjugate of z. For $z = x + i\,y$ we write (by definition) $|z| = \sqrt{x^2 + y^2}$, and call $|z|$ the modulus of z.

For example, if $z = 2 + 5\,i$ we have $\mathcal{Re}(z) = 2$, $\mathcal{Im}(z) = 5$, $\overline{z} = 2 - 5\,i$, and

$$\left|z\right| = \sqrt{4+25} = \sqrt{29}.$$

2.7.2 Properties of $|z|$, $\overline{z}$

We list the fundamental properties of $\overline{z}$, $|z|$.

1. $z\,\overline{z} = |z|^2$. To see this, observe that if $z = x + i\,y$, then

$z\,\overline{z} = (x + i\,y)\,(x - i\,z) = x^2 + y^2 = \,|z|^2\,.$

2. $\mathcal{Re}\,(z) = (z + \overline{z})/2$, $\mathcal{Im}\,(z) = (z - \overline{z})/2\,i$. To see this observe that if $z = x + i\,y$, then

$z + \overline{z} = (x + i\,z) + (x - i\,y) = 2\,x$

$z - \overline{z} = (x + i\,y) - (x - i\,y) = 2\,i\,y$

3. $\overline{z + w} = \overline{z} + \overline{w}$. To see this observe that if $z = x + i\,y$, $w = u + i\,v$, then

$z + w = x + i\,y + u + i\,v = (x + u) + i\,(y + v)$

therefore we have

$\overline{z + w} = (x + u) - i\,(y + v) = x - i\,y + u - i\,v = \overline{z} + \overline{w}.$

4. $\overline{z\,w} = \overline{z}\,\overline{w}$. To see this observe that if $z = x + i\,y$, $w = u + i\,v$, then

$z\,w = (x + i\,y)\,(u + i\,v) = x\,u - y\,v + i(y\,u + x\,v)$

$\overline{z}\,\overline{w} = (x - i\,y)\,(u - i\,v) = (x\,u - y\,v) - i\,(y\,u + x\,v)$

which proves

$\overline{z\,w} = \overline{z}\,\overline{w}.$

5. $|z\,w| = |z||w|$ To see this observe that if $z = x + i\,y$, $w = u + i\,v$, then

$|(x + i\,y)\,(u + i\,v)|^2 = \;|x\,u - y\,v + i(y\,u + x\,v)|^2 = \;(x\,u - y\,v)^2 + (y\,u + x\,v)^2 = \left(x^2 + y^2\right)\left(u^2 + v^2\right)$

$|z|^2\;\;|w|^2 = \;\left(x^2 + y^2\right)\left(u^2 + v^2\right)$

6. $|z + w| \;\leq\; |z| + |w|$ To see this observe that if $z = x + i\,y$, $w = u + i\,v$, then

$|z + w|^2 = \;(x + u)^2 + (y + v)^2$

$|z|^2 + |w|^2 = \;x^2 + y^2 + u^2 + v^2$

2.7.3 The Argand Diagram

We obtain a geometric model for the complex numbers by representing the complex number $z = x + i\,y$ by the point $(x,\;y)$ in the real plane with coordinates x and y. Observe that the horizontal x-axis represents complex numbers $x + i\,y$ with $y = 0$, that is, the real numbers. We therefore call the horizontal axis the real axis. The vertical y-axis represents complex numbers $x + i\,y$ with $x = 0$, that is, numbers of the form $i\,y$ where y is real. We call these numbers pure imaginary, and we call the vertical axis the imaginary axis. The origin O represents the number zero which is of course real (Figure 2.52).

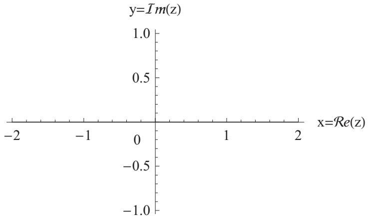


Figure 2.52. Complex plane with the real and imaginary axis.

2.7.4 Geometric Interpretation of Addition

If we have two complex numbers $z = x + i\, y$, $w = u + i\, v$, then their sum $z + w$ is given by

$$z + w = (x + i\, y) + (u + i\, v) = x + u + i(y + v)$$

and therefore appears on the Argand diagram as the sum of their real and imaginary coordinates (see Figure 2.53).

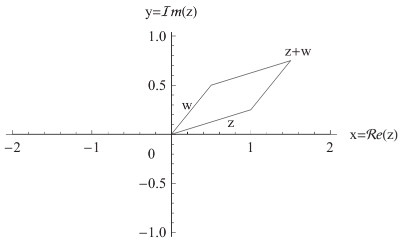


Figure 2.53. Addition of two complex numbers z and w to a new complex number $z + w$.

2.7.5 Polar Form and Euler's Formula

An alternative representation of points in the plane is by polar coordinates θ. The coordinate r represents the distance of the point from the origin O. The coordinate θ represents the angle the line joining the point to O makes with the positive direction of the x-axis measured counterclockwise (see Figure 2.54). Suppose the complex number $z = x + i\, y$ on the Argand diagram has polar coordinates r, θ. We call r the modulus of z, and denote it by $|z|$. Pythagoras' theorem gives

$$|z| = \sqrt{x^2 + y^2}$$

consistent with the definition of $|z|$ given in section 2.7.1.

We call θ argument of z which we abbreviate to arg(z). A little bit trigonometry on Figure 2.54 gives

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

Observe that whilst $|z|$ is single valued, arg(z) is many-valued. This is because for any given value of θ we could take instead $\theta + 2\pi$ (in radians) and arrive at the same complex number z. For example, suppose $z = 1 + i$. Then $|z| = \sqrt{2}$, but arg(z) can be taken to be any of the values $\pi/4$, $5\pi/4$, $9\pi/4$, etc., also $-3\pi/4$, $-7\pi/4$ etc. Equivalently, arg(z) = $\pi/4 + 2\pi n$ for any integer n.

We define the principal value (PV) of arg(z) to be that value of θ which satisfies $-\pi < \theta \leq \pi$. For example, the principal value of arg($1 + i$) is $\pi/4$ (see Figure 2.54). We write arg($1 + i$) = $\pi/4$ (PV) to denote that we mean the principal value of the argument.

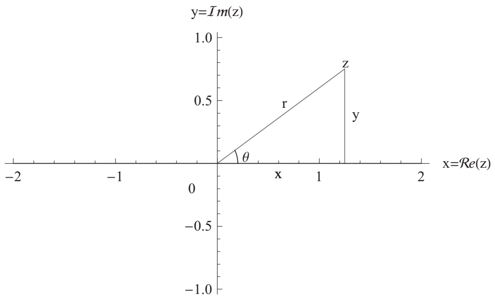


Figure 2.54. Polar representation of a complex number $z = x + i\,y$.

For general $z = x + i\,y$ we have $\cos(\theta) = x/r$, $\sin(\theta) = y/r$. Therefore

$$\begin{aligned} z &= x + i\,y \\ &= r\cos(\theta) + i\,r\sin(\theta) \\ &= r\,(\cos(\theta) + i\sin(\theta)) \\ &= r\,e^{i\theta} \end{aligned}$$

We call the formula

$$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

Euler's formula. We call the representation $z = r\,e^{i\theta}$ a polar form for z, and $z = x + i\,y$ is called a Cartesian representation. Since both representations are equivalent we have for example $1 + i = \sqrt{2}\;e^{i\pi/4}$.

Having Euler's formula available and knowing that the conjugate complex of a number is given by

$$\left(e^{i\theta}\right)^* = e^{-i\theta} = \cos(\theta) - i\sin(\theta)$$

we are able to use these two representations to derive

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

and

$$\sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2\,i}$$

which allows to represent trigonometric functions by exponential functions using complex numbers.

An immediate consequence of Euler's formula is the result known as de Moivre's theorem which states in short

$$(\cos(\theta) + i\sin(\theta))^n = \left(e^{i\theta}\right)^n = e^{i\,n\,\theta} = \cos(n\,\theta) + i\sin(n\,\theta).$$

This theorem is useful to derive several of the relations for trigonometric functions as discussed in Section 2.6.

2.7.6 Tests and Exercises

The following tests and exercises are related to complex numbers.

2.7.6.1 Test Problems

T1. Is there a difference between a real and a complex number?

T2. Explain the relation between the real numbers and the complex numbers.

T3. How are complex numbers graphically represented?

T4. How is the argument and the modulus of a complex number defined?

T5. Is the argument a unique property of a complex number?

T6. Why do we use the principal value for the argument?

2.7.6.2 Exercises

E1. Express the following complex numbers in the form $x + i\,y$.

- **a.** $(1+3\,i)+(6+9\,i)$
- **b.** $(1+3\,i)-(6+9\,i)$
- **c.** $(1+3\,i)(6+9\,i)$
- **d.** $\frac{1+3\,i}{6+9\,i}$
- **e.** $\sqrt{3+2\,i}$
- **f.** $\log(1+2\,i)$

E2. Find the value of $\sqrt{1+i}$

E3. Expand the relation $(\cos(x) + i\sin(x))^3$ to find relations for $\cos(3\,x)$ and $\sin(3\,x)$ in terms of $\cos(x)$ and $\sin(x)$.

E4. Use the results of E3 to verify the relations

- **a.** $\cos(3\,x) = 4\cos(x)^3 - 3\cos(x)$
- **b.** $\sin(3\,x) = 3\sin(x) - 4\sin(x)^3$.

E5. Expand $\left(e^{i\theta} \pm e^{-i\theta}\right)^3$ to show

- **a.** $\cos(\theta)^3 = \frac{1}{4}(3\cos(\theta) + \cos(3\,\theta))$,
- **b.** $\sin(\theta)^3 = \frac{1}{4}(3\sin(\theta) - \sin(3\,\theta))$.

2.8 Equations of Lines

In this section we will discuss ways to measure the steepness or slope of a line in the plane. The ideas we developed here will be important when we discuss equations and graphs of straight lines.

2.8.1 Slope of a Line

The steepness of a line segment connecting two distinct points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ can be measured by traveling from one end of the line segment to the other and relating the vertical change in height to the horizontal change in distance. To illustrate, compare the line segment in Figure 2.55 as we travel from P_1 to P_2. In the two cases, the horizontal change $x_2 - x_1$ is the same. However, the corresponding vertical change $y_2 - y_1$ is larger for the steeper line segment, which means that the quotient,

$$\frac{y_2 - y_1}{x_2 - x_1}$$

is larger for the steeper line segment. This suggests that we can use this quotient as a numerical measure of steepness.

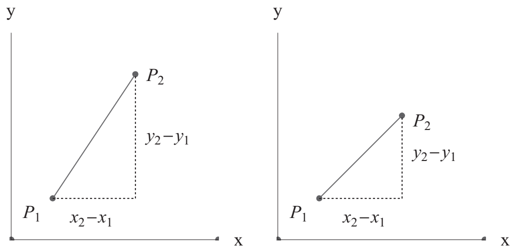


Figure 2.55. Comparison of two line segments with equal horizontal extension $x_2 - x_1$. The vertical extension $y_2 - y_1$ differs and determines the slope.

Definition 2.11. *Slope*

If $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ are distinct points such that $x_1 \neq x_2$, then the number m given by the ratio

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

is called the ***slope of a line segment*** *connecting P_1 and P_2.* ■

The formula for the slope m given in Definition 2.11 does not apply to line segments parallel to the y-axis. For such segments $x_2 = x_1$ and m would involve a division by zero; therefore, for line segments parallel to the y-axis we say the slope is undefined, or the line segment has infinite slope.

Remark 2.14. In light of the equality

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}$$

it does not matter which endpoint is labeled P_1 and which is labeled P_2 when we apply the slope m to compute the slope of a line segment.

Example 2.26. Slope between two Points I

Let L be the line segment connecting the two points (6, 2) and (9, 8). If we label the first point P_1 and the second P_2 then from the slope equation in Definition 2.11 we find

$$\boldsymbol{m = \frac{8-2}{9-6}}$$

2

Similarly, the slope of the line segment connecting the two points (2, 9) and (4, 3) is

$$\mathbf{m2 = \frac{3-9}{4-2}}$$

-3

▲

It is clear geometrically that two line segments on the same straight line will be equally steep and consequently have the same slope. This is the content of the following theorem.

Theorem 2.10. *Slope of a Line*

On a non-vertical line, all line segments have the same slope. ■

Proof 2.10. Let $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ be distinct points on a non-vertical line L and let $P_1^*(x_1^*, y_1^*)$ and $P_2^*(x_2^*, y_2^*)$ be another pair of distinct points on L. We will show that the slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

of the line segment joining P_1 and P_2 is equal to the slope

$$m^* = \frac{y_2^* - y_1^*}{x_2^* - x_1^*}$$

of the line segment joining P_1^* and P_2^*. (We will assume the points are ordered as in Figure 2.56. The proof of the remaining case are similar and will be omitted for brevity.)

QED

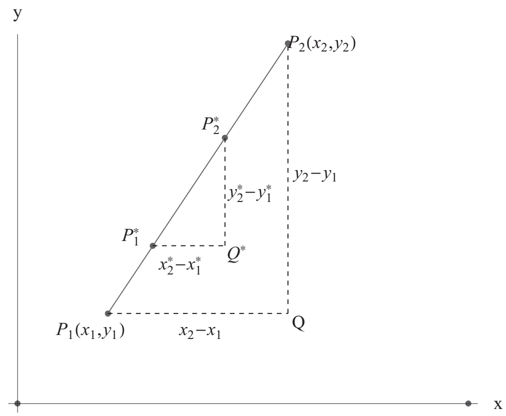


Figure 2.56. Common slope of a line.

The triangles $P_1\,Q\,P_2$ and $P_1^*\,Q\,P_2^*$ in Figure 2.56 are similar, so that the lengths of corresponding sides are proportional. Therefore

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_2^* - y_1^*}{x_2^* - x_1^*}$$

or, equivalently

$$m = m^*$$

We define the slope of a non-vertical line L to be the common slope of all line segments on L. The slope of a vertical line is undefined.

The slope of a line has a useful interpretation. Imagine a particle moving left to the right along a line L from a point $P_1(x_1,\ y_1)$ to a point $P_2(x_2,\ y_2)$. As shown in Figure 2.56, the particle moves $y_2 - y_1$ units in the y-direction as it travels $x_2 - x_1$ units in the x-direction. By Definition 2.11 these quantities are related by

$$y_2 - y_1 = m\,(x_2 - x_1) \tag{2.48}$$

where m is the slope of the line. Thus (2.48) states that the movement in the y-direction is proportional to the movement in the x-direction and the slope m is the constant of proportionality. For this reason, the slope m is said to measure the rate at which y changes with respect to x along L.

Example 2.27. Slope between two Points II

In Example 2.26, we showed that the line determined by the two points $P_1(6,\ 2)$ and $P_2(9,\ 8)$ has slope $m = 2$. This means that a particle traveling left to right along this line rises two units for every unit it moves in the positive x-direction (Figure 2.57).

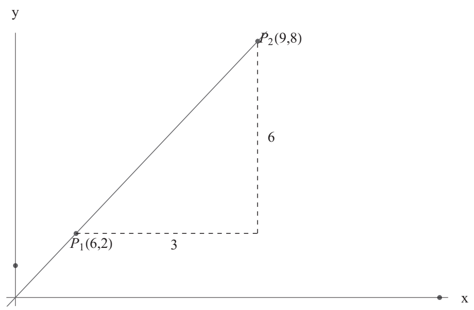


Figure 2.57. Slope of a line defined by two points.▲

Example 2.28. Slope between two Points III

In Example 2.26, we showed that the line determined by the points $P_1(2, 9)$ and $P_2(4, 3)$ has slope -3. This means that a particle traveling to right along this line falls three units for every unit it moves in the positive x-direction (Figure 2.58)

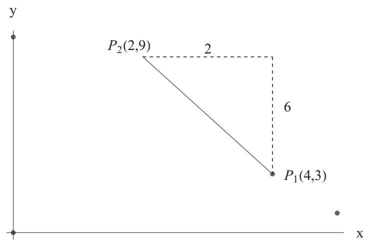


Figure 2.58. Negative slope of a line defined by two points.▲

In Figure 2.59 we have sketched several lines with varying slopes. Observe that lines with negative slope are inclined downward to the right.

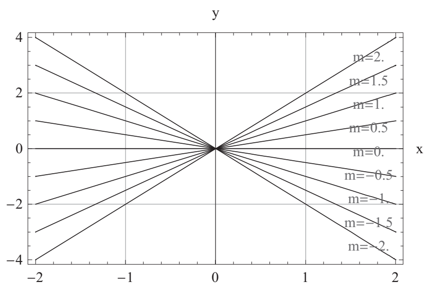


Figure 2.59. Negative slope of a line defined by two points.▲

In engineering slopes are usually measured with devices especially designed for this purpose. There are two measuring devices which are widely used to determine the slope. These devices are sine bars and so called bevel protractors. Samples of the two devices are shown in Figures 2.60 and

2.61. Sine bars are tools to measure angles in metalworking. It consists of a hardened, precision ground body with two precision ground cylinders fixed at the ends. The distance between the centers of the cylinders is precisely controlled, and the top of the bar is parallel to a line through the centers of the two rollers. The dimension between the two rollers is chosen to be a whole number (for ease of later calculations) and forms the hypotenuse of a triangle when in use.

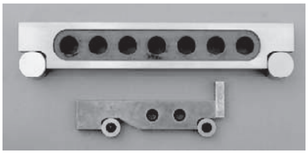


Figure 2.60. Two versions of sine bars. The rolls are used to define a fixed distance. Typical lengths are 100 mm.

When a sine bar is placed on a level surface the top edge will be parallel to that surface. If one roller is raised by a known distance, usually using gauge blocks, then the top edge of the bar will be tilted by the same amount forming an angle that may be calculated by the application of the sine rule (2.5).

In geometry a protractor is a circular or semicircular tool for measuring an angle. The units of measurement utilized are usually degrees. Some protractors are simple half-discs; these have existed since ancient times. More advanced protractors, such as the Bevel Protractor have one or two swinging arms, which can be used measuring the angle.

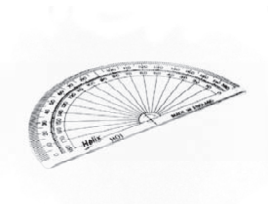


Figure 2.61. A half circle protractor marked in degrees.

The slope of a line is related to the angle the line makes with the positive *x*-axis. To establish this relationship we need the following definition.

Definition 2.12. *Angle of Inclination*

For a line L not parallel to the x-axis, the angle of inclination is the smallest angle ϕ measured counterclockwise from the direction of the positive x-axis to L (Figure 2.62). For a line parallel to the x-axis, we take $\phi = 0$. ■

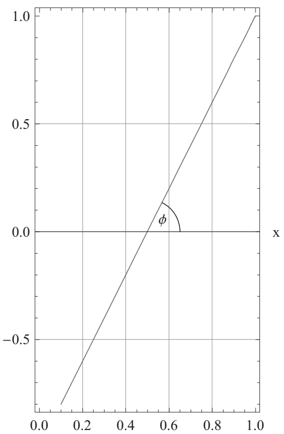


Figure 2.62. Angle of inclination for a line L.▲

Remark 2.16. In degree measure the angle of inclination satisfies $0° \leq \phi \leq 180°$ and in radian measure it satisfies $0 \leq \phi \leq \pi$.

Theorem 2.11. *Slope of a Line*

For a line not parallel to the y-axis, the slope and angle of inclination are related by

$$m = tan(\phi).\blacksquare \tag{2.49}$$

Proof 2.11. If the line is horizontal then $m = 0$ and $\phi = 0°$. Thus $\tan(\phi) = \tan(0) = 0 = m$ so (2.49) holds in this case. If the line is not horizontal, let $(x_0, 0)$ be its point of intersection with the x-axis, and construct a circle of radius 1 centered at this point. Because the angle of inclination of the line is ϕ, the line will intersect this circle at the point $(x_0 + \cos(\phi), \sin(\phi))$ (Figure 2.63).

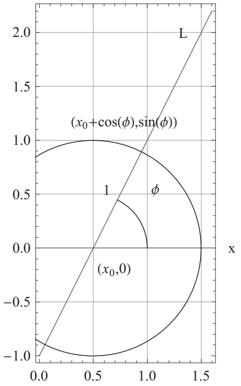


Figure 2.63. Relation between the angle of inclination ϕ and the slope m for a line L located at an arbitrary point $(x_0, 0)$ along the x-axis.▲

Since the points $(x_0, 0)$ and $(x_0 + \cos(\phi), \sin(\phi))$ lie on L, we can use them to compute the slope m. This gives

$$m = \frac{\sin(\phi) - 0}{x_0 + \cos(\phi) - x_0} = \frac{\sin(\phi)}{\cos(\phi)} = \tan(\phi)$$

which proves (2.49).

Remark 2.17. If the line L is parallel to the y-axis, then $\phi = \pi / 2$ and $\tan(\phi)$ in (2.49) is undefined. This agrees with the fact that the slope m is also undefined in this case.

Example 2.29. Angle of Inclination

Find the angle of inclination for a line of slope $m = 1$ and also for a line of slope $m = -1$.

If $m = 1$, then from (2.49), $\tan(\phi) = 1$ so that $\phi = \pi / 4$ (or in degree measure 45°). If $m = -1$, then from (2.49), $\tan(\phi) = -1$; from this equality and the fact that $0 \le \phi \le \pi$, we obtain $\phi = 3\pi / 4$ or, in degree measure, $\phi = 135$ °(Figure 2.64).

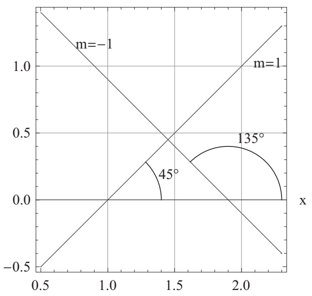


Figure 2.64. Two lines with slope $m = 1$ and $m = -1$.▲

As a consequence of Theorem 2.11, obtain the following basic result.

Theorem 2.12. *Parallel Lines*

Two non-vertical lines are parallel if and only if they have the same slope. ■

Proof 2.12. If L_1 and L_2 are non-vertical parallel lines, then their angles of inclination ϕ_1 and ϕ_2 are equal since two parallel lines cut by a transverse have equal corresponding angles (Figure 2.65). Thus

$$\text{slope}(L_1) = \tan(\phi_1) = \tan(\phi_2) = \text{slope}(L_2).$$

Conversely, if L_1 and L_2 have the same slope m, then

$$m = \tan(\phi_1) = \tan(\phi_2).$$

This relationship and the fact that

$$0 \leq \phi_1 \leq \pi, \qquad 0 \leq \phi_2 \leq \pi$$

imply that ϕ_1 and ϕ_2 are equal. Thus L_1 and L_2 are parallel (Figure 2.65).

QED

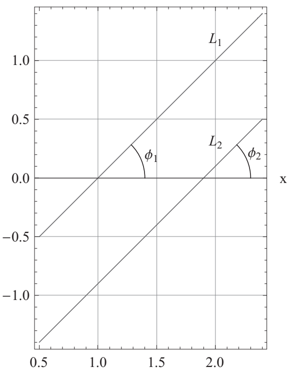


Figure 2.65. Parallel lines have the same slope and thus the same angle of inclination.

The next theorem shows how slopes can be used to determine whether two lines are perpendicular.

Theorem 2.13. *Perpendicular Lines*

Two non vertical lines are perpendicular if and only if the product of their slopes is −1; equivalently, lines with slopes m_1 and m_2 are perpendicular if and only if

$$m_1 = \frac{-1}{m_2}. \blacksquare \tag{2.50}$$

Proof 2.13. Suppose two non-vertical perpendicular lines L_1 and L_2, have angles of inclination ϕ_1 and ϕ_2 and slopes m_1 and m_2, respectively. Assume L_1 has the smaller angle of inclination

(Figure 2.65), so that

$$\phi_2 = \phi_1 + \frac{\pi}{2}.$$

Thus

$$m = \tan(\phi_2) = \tan\left(\phi_1 + \frac{\pi}{2}\right) = \frac{\sin\left(\phi_1 + \frac{\pi}{2}\right)}{\cos\left(\phi_1 + \frac{\pi}{2}\right)} = \frac{\sin(\phi_1)\cos\left(\frac{\pi}{2}\right) + \cos(\phi_1)\sin\left(\frac{\pi}{2}\right)}{\cos(\phi_1)\cos\left(\frac{\pi}{2}\right) - \sin(\phi_1)\sin\left(\frac{\pi}{2}\right)} =$$

$$\frac{-\cos(\phi_1)}{\sin(\phi_1)} = -\frac{1}{\frac{\sin(\phi_1)}{\cos(\phi_1)}} = -\frac{1}{\tan(\phi_1)} = -\frac{1}{m_1}$$

which establishes (2.50). In *Mathematica* this proof simplifies to

```
m2 = Tan[ϕ2] /. ϕ2 -> ϕ1 + π/2
```

```
-Cot[ϕ1]
```

where $\cot(\phi_1) = 1/\tan(\phi_1)$. The proof of the converse, that is that L_1 and L_2 are perpendicular if (2.50) holds is left as an exercise.

QED

Example 2.30. Perpendicular Lines

Use slopes to show that points $A(1, 3)$, $B(3, 7)$, and $C(7, 5)$ are vertices of a right triangle.

Solution 2.30. The line through A and B has slope

$$\mathbf{m1} = \frac{\mathbf{7-3}}{\mathbf{3-1}}$$

2

and the line through B and C has slope

$$\mathbf{m2} = \frac{\mathbf{5-7}}{\mathbf{7-3}}$$

$$-\frac{1}{2}$$

Since

m1 m2

-1

the line through A and B is perpendicular to the line through B and C; thus ABC is a right triangle.▲

Example 2.31. Temperature Gradient

Suppose that a uniform rod of length 0.4 *m* is thermally insulated around the lateral surface (heat pipe) and that the exposed ends of the rod are held at constant temperatures of 25°C and 5°C, respectively. It is shown in physics that under appropriate conditions the graph of the temperature *T* versus the distance *x* from the lower to the higher temperature will be a straight line. Determine the slope of the temperature for this rod.

Solution 2.31. The slope of the temperature is determined by the temperature drop over the distance by

$$\boldsymbol{m = \frac{5-25}{0.4+0}}$$

$$-50,$$

where we assumed that the origin of the length measurement is located at the higher temperature reservoir.▲

2.8.2 Equations of Straight Lines

In this section, we are concerned with recognizing those equations whose graphs are straight lines and finding equations for lines specified geometrically. This section is of some importance for later applications when linear regression is used to model some data sets from engineering.

2.8.2.1 Vertical Lines

A line parallel to the y-axis meets the x-axis at some point $(a, 0)$. Every point on this line has an x-coordinate of a and conversely every point with an x-coordinate of a lies on the line (Figure 2.66). Thus

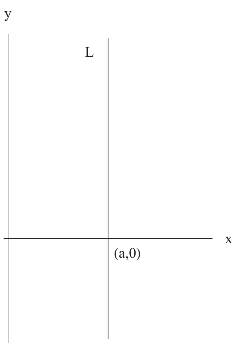


Figure 2.66. Vertical line intersecting the x-axis in $(a, 0)$.

Theorem 2.14. *Vertical Line*

The vertical line passing through $(a, 0)$ *is represented by the equation* $x = a$.■

Example 2.32. Vertical Line

The graph of $x = -2$ is the vertical line passing through $(-2, 0)$.▲

2.8.2.2 Lines Determined by Point and Slope

A non-vertical line can be determined by specifying a point on the line and the slope; the slope determines how the line is tilted and the point pins down the location of the line in the plane. Let us try to find an equation for the line L passing through $P_1(x_1, y_1)$ and having slope m.

If $P(x, y)$ is any point on L, other than P_1, then the slope of L can be obtained from the point $P(x, y)$ and $P_1(x_1, y_1)$; this gives

$$m = \frac{y - y_1}{x - x_1} \tag{2.51}$$

which can be rewritten as

$$y - y_1 = m\,(x - x_1) \tag{2.52}$$

With the possible exception of (x_1, y_1) we have shown that every point on L satisfies (2.49). But $x = x_1$, $y = y_1$ satisfies (2.49), so that all points on L satisfy (2.49). To summarize:

Theorem 2.15. *Line given by a Slope and a Point*

The line passing through $P_1(x_1, y_1)$ and having slope m is given by the equation

$$y - y_1 = m\,(x - x_1). \tag{2.53}$$

This is called the ***point-slope form*** *of a line.* ■

Example 2.33. **Point-Slope form** of a Line

Find the point-slope form of the line through $(4, -3)$ with slope 5.

Solution 2.33. Substituting the values $x_1 = 4$, $y_1 = -3$, and $m = 5$ in (2.53) yields the point-slope form

$$\boldsymbol{y + 3 = 5\,(x - 4)}$$

$$y + 3 = 5\,(x - 4)$$

▲

Example 2.34. Spring Scale

An old-fashioned spring scale shows the following elongations in cm if weights in kg are put on it. The elongations and weights are listed in the following table:

Weight [kg]	Elongation [cm]
1.5	0.3
2.5	1.1
3.5	1.9
5.5	3.5
9.5	6.7

Table 2.2. Weight w and elongation y for a spring

a) Explain why a linear function is appropriate for the data in the table.

b) Find a linear equation that relates the elongation y and the weight w by calculating the slope of the line.

Solution 2.34.

a. The five data points lie on a line, since each 1-unit increase in w produces a corresponding 0.8-unit increase in y.

b. The slope of the line segment joining any two successive data points is

$$m = \frac{1.1 - 0.3}{2.5 - 1.5}$$

```
0.8
```

A linear equation relating y and w can be obtained from the point slope form of a line using the slope $m = 0.8$ by

eq1 = *y* − 0.3 == *m* (*w* − 1.5)

```
-0.3 + y == 0.8 (-1.5 + w)
```

which can be solved with respect to the weight by using the *Mathematica* function Solve[]

sol1 = Simplify[Flatten[Solve[eq1, *y*]]]

```
{y → -0.9 + 0.8 w}
```

meaning that the elongation y follows the relation $y = 0.8\,w - 0.9$ if the weight is changed. The graph of this relation is shown in the following figure

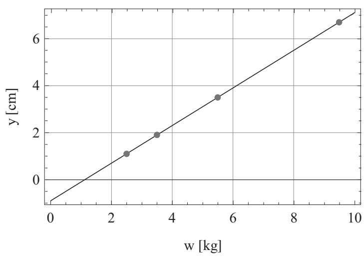


Figure 2.67. Graph of the data points and the theoretical model for a spring. The data are listen in Table 2.2.

The points represent the measurements of the elongations▲

2.8.2.3 Lines Determined by Slope and *y*-Intercept

A non-vertical line crosses the y-axis at some point $(0, b)$. If we use this point in the point slope form of its equation, we obtain

$$y - b = m\,(x - 0) \tag{2.54}$$

which we can write as

$$y = m\,x + b. \tag{2.55}$$

The number b in this equation is called the y-intercept of the line; it is the y-coordinate of the point where the line crosses the y-axis. To summarize:

Theorem 2.16. *Slope-Intercept form of a Line*

The line with y-intercept b and slope m is given by the equation

$$y = m\,x + b. \tag{2.56}$$

*This is called the **slope-intercept form** of a line.* ■

Remark 2.18. The slope-intercept form is of particular importance because it expresses y in terms of x.

Example 2.35. **Slope-Intercept form** of a Line

Comparing

$$y = 4\,x + 7$$

to (2.55), we have $m = 4$ and $b = 7$, so that the equation represents a line crossing the y-axis at (0, 7) with slope 4.▲

2.8.2.4 Horizontal Lines

By definition, the angle of inclination of a horizontal line is $\phi = 0$. Thus the slope of a horizontal line is

$$m = \tan(\phi) = \tan(0) = 0.$$

Substituting $m = 0$ in (2.55) yields the following result.

Theorem 2.17. *Horizontal Line*

The horizontal line passing through (0, b) is represented by the equation

$$y = b. \blacksquare \tag{2.57}$$

Example 2.36. Horizontal Line

The equation $y = -7$ represents the horizontal line passing through $(0, -7)$.▲

2.8.2.5 Lines Determined by Two Points

If $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ are distinct points on a non-vertical line, then the slope of the line is

$$m = \frac{y_2 - y_1}{x_2 - x_1}. \tag{2.58}$$

Substituting this expression in (2.53) we obtain the following result.

Theorem 2.18. *Two-Point Form*

The non-vertical line determined by the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ can be represented by the equation

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1). \tag{2.59}$$

*This is called the **two-point form** of a line.*■

Example 2.37. Two-Point Form I

Find the slope-intercept form of the line passing through the points (3, 4) and (2, −5).

Solution 2.34. Letting $(x_1, y_1) = (3, 4)$ and $(x_2, y_2) = (2, -5)$ and substituting in (2.59), we obtain the two-point form

$$\mathbf{eq1} = y - 4 == \frac{(-5-4)(x-3)}{2-3}$$

$$y - 4 == 9(x-3)$$

Solving for y yields the slope-intercept form as

Solve[eq1, *y*]

$\{\{y \to 9x - 23\}\}$

▲

2.8.2.6 The General Equation of a Line

An equation expressible in the form

$$A x + B y + C = 0 \tag{2.60}$$

where A, B, and C are constants and A and B are not both zero, is called a first degree equation in x and y.

Example 2.38. Two-Point Form II

Comparing $7x + 2y - 2 = 0$ to (2.59), we see that this is a first degree equation in x and y with $A = 7$, $B = 2$, and $C = -3$.▲

The following theorem shows that the first degree equations in x and y are precisely the equations whose graphs in the xy-plane are straight lines.

Theorem 2.19. *First Degree Equation*

Every first degree equation in x and y has a straight line as its graph and, conversely, every straight line can be represented by a first degree equation in x and y.■

Proof 2.19. Let

$$A\,x + B\,y + C = 0 \tag{2.61}$$

We introduce this equation in *Mathematica* by

```
eq2 = C + A x + B y == 0
```

```
C + A x + B y == 0
```

be any first degree equation in x and y. To prove that the graph of this equation is a straight line, we distinguish between two cases, $B = 0$ and $B \neq 0$.

If $B \neq 0$, we can solve (2.61) for y in terms of x to obtain

```
Solve[eq2, y]
```

$$\left\{\left\{y \to \frac{-A\,x - C}{B}\right\}\right\}$$

But this is the slope-intercept form (2.56) of the line with

$$m = -\frac{A}{B} \qquad \text{and} \qquad b = -\frac{C}{B}.$$

If $B = 0$, then $A \neq 0$ since A and B are not both zero in a first degree equation in x and y. Thus (2.61) reduces to

```
eq3 = eq2 /. B → 0
```

$$A\,x + C = 0$$

which can be rewritten as

```
Solve[eq3, x]
```

$$\left\{\left\{x \to -\frac{C}{A}\right\}\right\}$$

But this is the equation of a line parallel to the y-axis. Thus, in either case, the graph of (2.61) is a straight line.

Conversely, consider any straight line in the xy-plane. If the line is vertical, then it has an equation of the form $x = a$. This can be rewritten as $x - a = 0$, which is of the form (2.60) with

$$A = 1, \qquad B = 0, \qquad \text{and} \qquad C = -a.$$

If the line is not vertical, it can be expressed in slope-intercept form

$$y = m\,x + b.$$

This can be rewritten as

$$m\,x - y + b = 0$$

which is of form (2.60) with

$$A = m, \qquad B = -1, \qquad \text{and} \qquad C = b.$$

Because every straight line has a first degree equation in x and y and every first degree equation in x and y has a straight line for its graph, (2.60) is sometimes called a general equation of a linear equation in x and y.

Example 2.39. Linear Equation

Graph the equation

$$3\,x - 4\,y + 12 = 0. \tag{2.62}$$

Solution 2.39. First let us define the equation as a *Mathematica* expression

```
eq4 = 3 x − 4 y + 12 == 0
```

```
12 + 3 x - 4 y == 0
```

Since this is a linear equation in x and y, its graph is a straight line. Thus to sketch the graph, we need only two points on the graph and draw a line through them. It is particularly convenient to plot the points where the line crosses the x-axis when $y = 0$. Substituting $x = 0$ in (2.62) and solving for y, we obtain the intersection (0, 3) with the y-axis.

```
Solve[eq4 /. x → 0, y]
```

```
{{y → 3}}
```

Substituting $y = 0$ in (2.62) and solving for x, we obtain the intersection $(-4, 0)$ with the x-axis. The graph of (2.62) is given in Figure 2.68. The steps described are contained in the following *Mathematica* line.

```
Solve[eq4 /. y -> 0, x]
```

```
{{x → -4}}
```

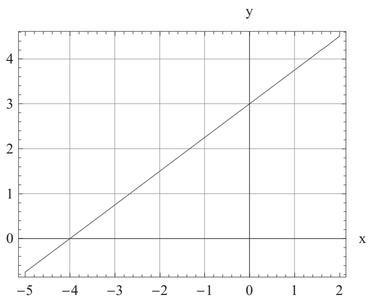


Figure 2.68. Graph of the equation $3\,x - 4\,y + 12 = 0$. The graph is generated as described in the text above.▲

2.8.3 Tests and Exercises

The following tests and exercises are related to lines and slopes.

2.8.3.1 Test Problems

T1. Is a two point form of a line:
( ) A list of two points?
( ) A formula which relates two points in a linear equation?
( ) A set with two coordinates?

T2. A first degree equation has the following structure:
( ) $A\,x + B\,y + \mathrm{C}x\,y + D = 0$

( ) $A x + B y + C = 0$
( ) $\mathrm{A\,x + B\,y + C + D = 0}$

T3. The slope intercept form has the structure
( ) $y - y_1 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1)$
( ) $A x + B y + C = 0$
( ) $m x + b = y$

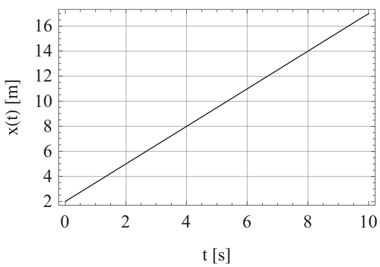
2.8.3.2 Exercises

E1. Find the slopes of the sides of the triangle with vertices (0, 3), (2, 0), and (6, 8/3).

E2. Find the angle of inclination of the line with slope m to the nearest degree.
- **a.** $m = 1/2$,
- **b.** $m = -2$,
- **c.** $m = 3/2$,
- **d.** $m = -1/4$,
- **e.** $m = 1.76$.

E3. In each part, classify the lines as parallel, perpendicular, or neither.
- **a.** $y = 4x - 9$ and $y = 4x + 5$,
- **b.** $5x - 3y = -6$ and $10x - 6y = -7$,
- **c.** $y = 4x - 5$ and $y = -\frac{1}{4} x + 2$,
- **d.** $x - y = 3x + 5$ and $y = \frac{1}{2} x - 7$.

E4. The accompanying figure shows the position versus time curve for the motion of a car moving along the x-axis.
- **a.** What is the velocity of the car?
- **b.** What is the x-coordinate of the car at $t = 0$?
- **c.** What is the x-coordinate of the car at $t = 3$?
- **d.** At what time does the car have an x-coordinate of $x = 4$?



E5. A locomotive travels on a track at a constant speed of 40 km/h, then reverses direction and returns to its starting point, traveling at a constant speed of 60 km/h.
- **a.** What is the average velocity for the round-trip?
- **b.** What is the total distance traveled by the train if the total trip took 5 h?

E6. A car is stopped at a toll booth on a straight highway. Starting at time $t = 0$ it accelerates at a constant rate of $10\, m/s^2$ for $10\, s$. It then travels at a constant speed of $100\, m/s$ for $90\, s$. At that time it begins to decelerate at a constant rate of $5\, m/s^2$ for $20\, s$, at which point in time it reaches a full stop at a traffic light.
- **a.** Sketch the velocity versus time curve.
- **b.** Express v as a piecewise function of t.

E7. Find the slope-intercept form of the equation of the line satisfying the stated conditions and check answer by plotting your result.
- **a.** Slope $m = -2$, y-intercept $b = 5$.

b. The line is parallel to $4x + 3y = 7$ and passes through $(-1, 2)$.

c. The line is perpendicular to $4x - y = 8$ and passes through $(3, 4)$.

E8. Find the angel of inclination of the line to the nearest degree for:

a. $5y = 2 - \sqrt{15}\ x$

b. $y + 7x = -2y + 5$

c. Find equations for the y- and x- axes.

E9. Thermometers are calibrated using the so-called "triple point" of water, which is $T = 273.16\,K$ on the Kelvin scale and 0.01 °C on the Celsius scale. A one-degree difference on the Celsius scale is the same as a one-degree difference on the Kelvin scale, so there is a linear relationship between the temperature $T_{[C]}$ in degree Celsius and the temperature $T_{[K]}$ in Kelvin.

a. Find an equation that relates $T_{[C]}$ and $T_{[K]}$.

b. Absolute zero ($0\,K$ on the Kelvin scale) is the temperature below which a body's temperature cannot be lowered. Express absolute zero in °C.

E10 A resistance thermometer is a device that determines temperature by measuring the resistance of a fine wire whose resistance varies with temperature. Suppose that the resistance R in Ohms (Ω) varies linearly with the temperature T in °C and that $R = 123.4\,\Omega$ when $T = 20$ °C and that $R = 133.9\,\Omega$ when $T = 45$ °C.

a. Find an equation for R in terms of T.

b. If R is measured experimentally as $128.6\,\Omega$, what is the temperature?

2.9 Functions and their Graphs

In this section we consider one of the most fundamental concepts in all of mathematics, the notion of a function. Historically, the term function was first used by Leibniz in 1673 to denote the dependence of a one quantity on another. To illustrate:

- The area A of a circle depends on its radius r by the equation $A = \pi r^2$; we say that *A is a function of r*.
- The price P of a stock depends on time t; we say P is a function of t.
- At a fixed point on earth, the wind speed w varies with the time t. Thus w is a function of time.

2.9.1 General Properties of Functions

In order to discuss functions or relationships between quantities without stating specific formulas, the Swiss mathematician, Leonhard Euler, developed the ingenious idea of using a letter of the alphabet such as f to denote a function or relationship. Thus by writing

$$y = f(x) \tag{2.63}$$

(read y equals f of x), we convey the idea that y is a function of x, that is, the value of y depends on the value of x. To indicate that the wind speed w at a fixed point on earth depends on the time t at which it is measured, we can write

$$w = f(t). \tag{2.64}$$

There is nothing special about the letter f; any symbol can be used to denote a function; thus (2.64) could have been written as

$$w = F(t), \qquad w = f_1(t), \qquad w = g(t), \text{ or } \qquad w = \phi(t). \tag{2.65}$$

By having a variety of symbols available to denote functions, we are able to distinguish between different possible relationships. For example, at a point in Alexandria the wind speed w will have a certain dependence on the time t, say,

$$w = g_1(t) \tag{2.66}$$

while at a point in Cairo the wind speed w will have a different dependence on the time t, say

$$w = g_2(t). \tag{2.67}$$

The idea of using a symbol to denote a function is a relatively sophisticated concept. In (2.64), for example, the letters w and t denote measurable physical quantities, wind speed and time. The letter f, however, does not denote a physical quantity; it stands for a dependence of one physical quantity on another; this is a completely abstract idea.

The notation of a function $f(t)$ used in *Mathematica* is a little bit different. Since a computer cannot distinguish between the arguments of a function and the algebraic operation of multiplication, such as $f(a+b)$. The developers decided to use square brackets [] instead of round braces to denote functional arguments. In *Mathematica* any function is denoted by $f[t]$. In fact our mathematical notation is not precise at this point distinguishing multiplication and functional arguments. The meaning of $f(a+b)$ can be either f times the sum of a plus b or the function f applied to the sum of a plus b.

Since the time of Euler and Leibniz the original idea of a function has evolved into the following more precise and general mathematical concept:

Definition 2.13. *Function*

Given two non-empty sets X and Y. A rule f which correlates to each element x of a subset $A = D(f) \subset X$ exactly one element $y = f(x) \in Y$ is called a mapping of X to Y. The subset $D(f)$ is called the domain f and $B(f) = \{y \mid y = f(x) \; \forall \; x \in D(f)\}$ is called the range of f. In short a function is a rule that assigns to each element in a set A one and only one element in a set B (see Figure 2.69). ■

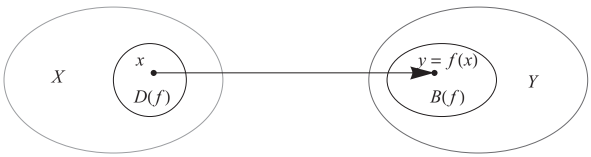


Figure 2.69. Graphical representation of a function in terms of sets X and Y. The subsets $D(f)$ and $B(f)$ represent the domain and range of the function, respectively.

In general the set A and B need not be sets of real numbers; however, for the time being we will only be concerned with functions for which A and B are both subsets of the real numbers.

To see how Definition 2.13 relates to the historical concept of a function, consider the relationship

$$w = f(t) \tag{2.68}$$

where w denotes the wind speed at a point and t denotes the time of measurement. Let A be the set of all possible t values and let B be the set of all possible w values. With each instance t in time there is associated by (2.68) a unique wind speed w; that is, the function f in (2.68) assigns to each element t in A a unique element w in B. This is precisely the definition of a function as stated in Definition 2.13.

The set A in Definition 2.13 is called the domain of the function. If x is an element in the domain of a function f, then the element that f associates with x is denoted by the symbol $f(x)$ (read f of x) and is called the image of $\boldsymbol{x}$ under $\boldsymbol{f}$ or the value of $\boldsymbol{f}$ at $\boldsymbol{x}$ (Figure 2.69). The set of all possible values of $f(x)$ as x varies over the domain is called the range of f.

Most often the rule for obtaining the value of a function is given by means of a formula.

Example 2.40. Functions and Formulas

The formula

$$f(x) = x^3$$

tells us that the value of f at x is x^3. In *Mathematica* we define such a function as

```
f[x_] := x^3
```

where f[x_] represents the function with its name f and its argument x. The x_ in the argument is used as a pattern for any kind of input which can be used as an argument for this function. The left hand side is representing in fact the notation of the function while the right hand side of the definition sign (:=) is specifying the functional value or how a function is evaluated. In the present case we rise x to the power of three. Thus if we use this definition by typing

```
f[2]
```

```
8
```

we get in return the cube of the argument which is $2^3 = 8$. Now we can replace the argument value by any number such as -4

```
f[-4]
```

```
-64
```

The result is -64 taking the sign into account. Finally we test the function by zero

```
f[0]
```

```
0
```

delivering the expected value 0.▲

Example 2.41. Functions and Formulas and Evaluation I

Again we define the function by

```
f[x_] := 2 x^2 - 1
```

then this function delivers the following results for the argument $x = 2$

```
f[2]
```

```
7
```

for $x = 3$ we get

```
f[3]
```

```
17
```

for $x = a$ where a is an arbitrary symbol we find the definition of the function in terms of a

```
f[a]
```

$-1 + 2\,a^2$

Even if we use algebraic expressions, we find after expanding the result (application // of Expand[]) the correct expression

```
f[k + 1] // Expand
```

$1 + 4\,k + 2\,k^2$

▲

Remark 2.19. If x is an element in the domain of a function f, the domain of a function (Definition 2.13) requires that f assigns one and only one value to x. This means that a function can not be multivalued. For example the expression $\pm\sqrt{x}$ does not define a function of x since it assigns two values to each positive x.

If the rule for evaluating a function is given by a formula, and there is no mention of the domain, it is understood that the domain consists of all real numbers for which the formula makes sense and yields a real value.

Example 2.42. Functions and Formulas and Evaluation II

Let

$$g(\text{x\_}) := \sqrt{x - 1}$$

when $x < 1$, $g(x)$ is imaginary; thus the domain is $[1, +\infty)$. As x varies over the interval, $g(x)$ varies from 0 to $+\infty$. Therefore the range of g is $[0, +\infty)$. In this simple case (monotonic function), we can check the evaluation of this function by inserting the two boundaries 1 and $+\infty$ to get the boundaries for the range.▲

Example 2.43. Functions and Formulas and Evaluation III

Let

$$h(\text{x\_}) := \frac{1}{(x - 1)\,(x - 3)}$$

$$h(1)$$

ComplexInfinity

$h(3)$

ComplexInfinity

When $x = 1$ or $x = 3$, $h(x)$ is undefined since division by zero is not allowed; otherwise $h(x)$ yields a real values. Thus the domain of h consists of all real numbers other than $x = 1$ or $x = 3$. In interval notion the domain is $(-\infty, 1) \cup (1,3) \cup (3, +\infty)$. ▲

Sometimes it is necessary to specify restrictions on the domain even though the formula for the function makes sense elsewhere.

Example 2.44. Functions and Restrictions

Assume that a machine can produce 12 parts per minute. Then the output of the machine as a function of time is $f(t)$. The formula for the produced parts $f(t)$ is

$$f(t) = 12\, t$$

If we were to specify no restrictions on the domain of this function, then the domain would be $(-\infty, +\infty)$ since the formula above makes sense for all real t. Physically, however, t must be a non negative integer (one minute or an integer multiple of one minute); thus we must restrict the domain to the set $\{0, 1, 2, 3, \ldots\}$. We do this by writing

$$f(t) = 12\, t \qquad \text{with } t = 0, 1, 2, 3, 4, \ldots$$

or in *Mathematica*

f(t_Integer /; $t \geq 0$) := 12 t

The restriction stated above as an with clauses have two components in *Mathematica.* First the argument t is restricted to integer numbers by specifying the domain of values as Integer by t_Integer. Second the values of t should be positive which is introduced in the argument of the function by a clause stating that $t \geq 0$. Only for values satisfying these constraints are accepted by the function f. From this definition it is understood that formula $f(t) = 12\, t$ only applies when $t = 0, 1, 2, 3, 4, \ldots$; for other values of t, $f(t)$ is undefined. Thus

$f(2)$

24

while

$f(1.4)$

$f(1.4)$

is undefined also for rational numbers

$$\boldsymbol{f\left(\frac{1}{2}\right)}$$

$$f\left(\frac{1}{2}\right)$$

we get as a return value the function unevaluated.▲

As the next example shows, functions are sometimes specified by formulas that have been pieced together.

Example 2.45. Test Drive Cycle of a Car

In testing cars, so called test drive cycles are defined by the car manufacturer to make sure that the car satisfies certain requirements. A typical test drive cycle consists of 4 segments: 1. regular driving with constant velocity at a lower level e.g. $v = 0.75$, 2. acceleration of the car to a high speed velocity with a standard acceleration of $a = 0.5$, 3. driving at constant high velocity of $v = 1.25$, and 4. deceleration of the car till to stall with $a = -2$. This cycle can be defined for unique time intervals as follows

$$0.75 \qquad \text{if} \qquad 0 < t \le 1$$

and

$$0.75 + 0.5\,(t-1) \qquad \text{if} \qquad 1 < t \le 2$$

and

$$1.25 \qquad \text{if} \qquad 2 < t \le 3$$

and

$$1.25 - 2\,(t-3) \qquad \text{if} \qquad t > 3$$

This list of conditions is cumbersome if we deal with the different time domains. Therefore we express this in mathematics by writing

$$f(t) = \begin{cases} 0.75 & 0 < t \le 1 \\ 0.75 + 0.5\,(x-1) & 1 < t \le 2 \\ 1.25 & 2 < t \le 3 \\ 1.25 - 2\,(t-3) & t > 3 \end{cases}$$

In *Mathematica* this is defined by using a piecewise function. We use the *Mathematica* function Piecewise[] in the following definition

```
f[t_] := Piecewise[{{0.75, 0 < t ≤ 1}, {0.75 + 0.5 (t - 1), 1 < t ≤ 2},
   {1.25, 2 < t ≤ 3}, {1.25 - 2 (t - 3), t > 3}}]
```

The given *Mathematica* definition follows exactly the same pattern as the mathematical definition. The elements of the function are collected in sublists by a common list. The sublists contain as first element the function value definition and as a second element the domain of validity of this definition. The drive cycle function can be used for a ride of a duration of 3.5 *s* as follows

```
f[3.5]
```

```
0.25
```

The specified time value will be converted to a function value using the different constraints and selecting the appropriate one. A graphical representation of this function is given in Figure 2.70.

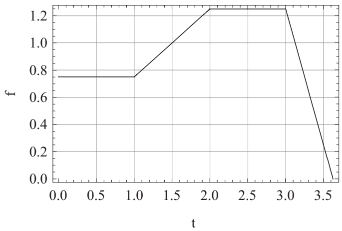


Figure 2.70. Total drive cycle of a car starting at $t = 0$ and ending at $t = 3.625$.▲

Although most of the functions are specified by formulas, this is not the only possibility. Any description that tells us what values f assigns to the points in its domain will suffice. One possibility is to use a table relating the values of x and $f(x)$. Such definitions of functions are useful if we have only a finite number of points available like for example in experiments.

Example 2.46. Function given by a Table (Forces in Crash-elements)

Each well designed car is based on a crash element incorporated in the front structure. This element ideally should have the following characteristic. At low crash forces the element should behave elastic (Hook's law). If the force acting is above a first threshold the crash element should deform plastically. After a defined crash-length the element becomes stiff with a very high reaction force. Since this description is an idealized description a designer should have the freedom to change the nature and characteristic of the crash-element depending on the needed requirements. To find a proper design it is common practice to represent the crash-forces of the crash element by means of a force table. The following table specifies a crash element by means of coordinates in the force elongation diagram.

x [mm]	F_c [N]
0	0
0.5	5
3	5.5
3.5	12

Table 2.3. Crash characteristic defined by points in the force elongation diagram.

The graph of these points is generated by connecting each point by a straight line. The result is shown in Figure 2.71.

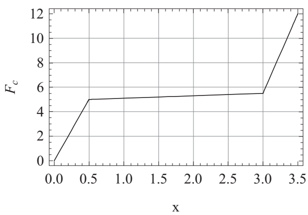


Figure 2.71. Crash force F_c as a function of the elongation x as given in Table 2.3.▲

If x is an element in the domain of a function f, then by forming the ordered pair of numbers

$$(x,\ f(x)) \tag{2.69}$$

we match each x with the value of f at x. For the function

```
f[x_] := x^3
```

a few such ordered pairs would be

```
TableForm[Table[{x, f[x]}, {x, 0, 7, 1.4}],
 TableHeadings → {{}, {"x", "f(x)=x^3"}}]
```

x	f(x)=x³
0.	0.
1.4	2.744
2.8	21.952
4.2	74.088
5.6	175.616
7.	343.

In the symbol $f(x)$, we can think of x as a quantity that can be varied arbitrarily over the domain of f. As x varies, so does the number $f(x)$ but with one difference; where the value of x can be varied arbitrarily within the domain of f, the value of $f(x)$ is determined once the value of x is specified. For this reason, x is sometimes called the independent variable, and $f(x)$ the dependent variable. In many problems it is convenient to introduce a single letter such as y to stand for the dependent variable. Thus if we write

$$y = f(x) \tag{2.70}$$

then x is the independent variable and y the dependent variable.

Example 2.47. Dependent and Independent Variables

If

$$f(x\_) := 9\,x^3 - 2\,x + 7$$

then the independent variable is x and

$$y = f(x)$$

is the dependent variable.▲

The following definition will enable us to study functions geometrically.

Definition 2.14. *Graph*

We define the graph of a function f to be the graph of the equation

$$y = f(x).\blacksquare \tag{2.71}$$

Example 2.48. Graph

Sketch the graph of

$$f(x\_) := x + 3$$

Solution 2.48. By definition, the graph of the function $x+3$ is the graph of the equation

$$y = f(x)$$

```
3 + x
```

which is a line of slope 1 with a y-intercept 3 Figure 2.72. We generated this graph in *Mathematica* by using the function Plot[]. Plot uses two main arguments the function as a first argument and the domain as a second argument. The function is either specified as discussed above or a symbol is used which contains the assigned function. The second argument consists of a list containing the independent variable and the lower and upper boundary of the domain. The elements in Plot are separated by commas. The result is a graph showing the specified function. Options like FrameLabel allow you to label the axes or frame of the plot.

```
Plot[y, {x, -2, 2}, FrameLabel → {"x", "y"}]
```

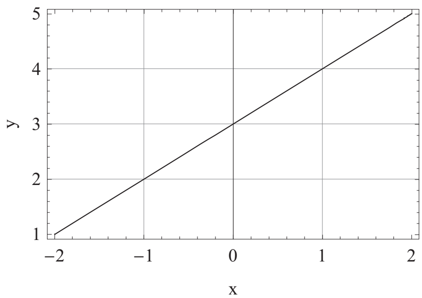


Figure 2.72. Graph of the equation $y = x + 3$.▲

Example 2.49. Graph of $|x|$

Sketch the graph of

$$f(x\_) := |x|$$

Solution 2.49. By definition, the graph of $f(x) = |x|$ is the graph of the equation

$$y = f(x)$$

```
Abs[x]
```

or equivalently

$$y = \begin{cases} x & x \geq 0 \\ -x & x < 0. \end{cases}$$

```
Plot[y, {x, -2, 2}, FrameLabel → {"x", "y"}]
```

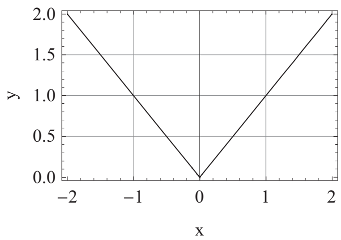


Figure 2.73. Graph of the function $f(x) = |x|$.

The graph coincides with the line $y = x$ for $x \geq 0$ and with the line $y = -x$ for $x < 0$.▲

Example 2.50. Graph of Rational Functions

Sketch the graph of

$$\boldsymbol{h}(\mathbf{x\_}) := \frac{\boldsymbol{x^2 - 4}}{\boldsymbol{x - 2}}$$

The function h can be written as

$$h = \frac{(x-2)(x+2)}{(x-2)} \tag{2.72}$$

algebraically it is desirable to cancel the factor $(x-2)$, but we must exercise some care. If we just cancel the factor and write

$$h = x + 2 \tag{2.73}$$

then we would erroneously alter the domain of h. To see this, observe that the domain of the function h in (2.72) consists of all x other than $x = 2$, whereas the domain of the function h in (2.73) consists of all x. In fact, when $x = 2$ the function h in (2.73) has the value

$$h(2) = 2 + 2 = 4$$

whereas for the original function h in (2.72),

$$h = \frac{2^2 - 4}{2 - 2} = \frac{0}{0}$$

which is undefined.

To cancel the factor $x - 2$ in (2.72) and not alter the domain of h, we must restrict the domain in (2.73) and write

$$h(x) = x + 2 \quad \text{and } x \neq 2.$$

Thus the plot does not differ for the two functions.

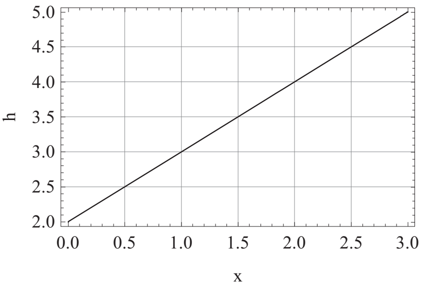


Figure 2.74. Graph of the function $h(x) = \frac{x^2-4}{x-2}$ and $h(x) = x + 2$.

▲

Example 2.51. Graph of a Stepwise Defined Functions

Sketch the graph of

$$g(x) = \begin{cases} 1 & \text{for } x \leq 2 \\ x+1 & \text{for } x > 2. \end{cases}$$

By definition, the graph of g is the graph of the relation

$$y = \begin{cases} 1 & \text{for } x \leq 2 \\ x+1 & \text{for } x > 2. \end{cases}$$

Thus for $x \leq 2$, the value of y is always 1 and for $x > 2$, y is given by $y = x + 1$

$$\mathbf{g(x\_) := Piecewise}\left[\begin{pmatrix} \mathbf{1} & \boldsymbol{x \leq 2} \\ \boldsymbol{x+1} & \boldsymbol{x > 2} \end{pmatrix}\right];$$

$$\boldsymbol{y = g(x);}$$

The graph thus follows by

```
Plot[y, {x, -1, 4}, FrameLabel → {"x", "y"}]
```

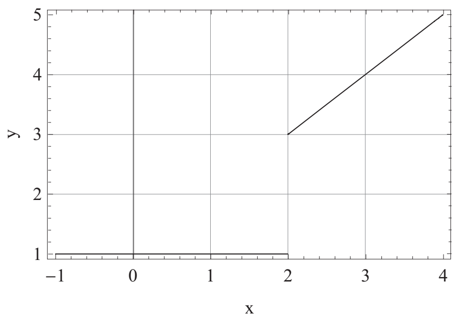


Figure 2.75. Graph of a stepwise defined function.▲

Remark 2.20. Note that the function value at $x = 2$ belongs to the horizontal line with $y = 1$

and not to the inclined line.

In each of the last examples we found the graph of a given function. We shall now consider the inverse problem. Given a graph in the xy-plane, does there exist a function f whose graph is the given curve? This problem is investigated in the following examples.

Example 2.52. Given a Graph Look for the Function

Show that the graph of

```
graph1 = 3 x^2 - 2 y == 1
```

$$3x^2 - 2y = 1$$

is also a graph of $f(x)$ for some function f.

The equation given above can be rewritten by solving this relation with respect to y

```
f1 = Flatten[Solve[graph1, y]]
```

$$\left\{y \to \frac{1}{2}\left(3x^2 - 1\right)\right\}$$

will give us the function representing the relation in the xy-plane

```
fFunction = y /. f1
```

$$\frac{1}{2}\left(3x^2 - 1\right)$$

▲

Example 2.53. Vertical Line Test

The curve in Figure 2.76 cannot be the graph of any function of x. To see why, consider the vertical line in Figure 2.76. This line intersects the curve at the two points (a, b) and (a, c). If the cure were the graph of

$$y = f(x) \tag{2.74}$$

for some function f, then, since (a, b) and (a, c) lie on the curve, (2.74) would imply that

$$b = f(a) \qquad \text{and} \qquad c = f(a).$$

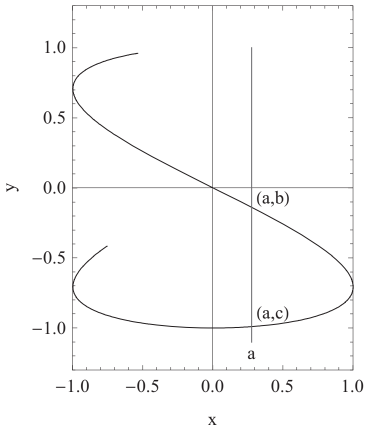


Figure 2.76. Graph of a relation.

But this is impossible since f cannot assign two different values of f at a. Thus there is no function f whose graph is the curve in Figure 2.76. This example illustrates the following general result, which we will call the **vertical line test**.

Theorem 2.20. *Vertical Line Test*

A curve is the graph of $f(x)$ for some function f if and only if no vertical line intersects the curve more than once. ■

Example 2.54. Circle

Is the graph of the circle

$$\mathbf{circle} = \boldsymbol{x^2 + y^2 = 16}$$

$$x^2 + y^2 = 16$$

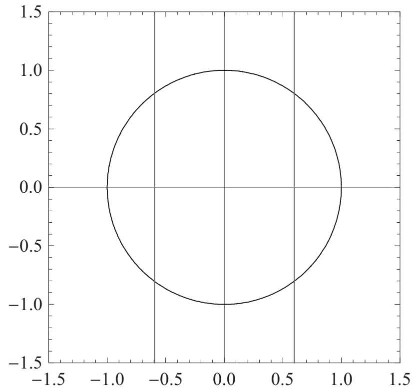


Figure 2.77. Graph of a circle.

also the graph of $f(x)$ for some function f?

Since some vertical lines intersect the circle more than once (see Figure 2.77), the circle is not the graph of any function.

We can also deduce this result algebraically by observing that the equation for the circle can be written as

solution1 = Solve[circle, *y*]

$$\left\{\left\{y \to -\sqrt{16-x^2}\right\}, \left\{y \to \sqrt{16-x^2}\right\}\right\}$$

But the right side of these solutions is not a single function of x and thus it is multi-valued. Actually, the relation becomes clear if we insert the solutions to the single variable y

***y* /. solution1**

$$\left\{-\sqrt{16-x^2}, \sqrt{16-x^2}\right\}$$

This demonstrates that the circle cannot be represented by a one-to-one relation since we get two different representations. Thus the algebraic relations found do not represent a function as

$y = f(x)$.▲

We conclude this subsection by some useful terminology. If the graph of an equation in x and y is also the graph of some function f, then we say that the equation defines y as a function of x; if in addition, the equation is written in the form $y = f(x)$ we say that the equation defines y explicitly as a function of x. If an equation defines y as a function of x, but is written in some form other than $y = f(x)$, then we say that the equation defines y implicitly as a function of x.

Example 2.55. Implicit versus Explicit Function

From Example 2.52 we know the following relation

graph1 = 3 x^2 − 2 y = 1

$3\,x^2 - 2\,y = 1$

which defines y implicitly as a function of x. When we rewrite this relation by solving with respect to y

***y* = (*y* /. Flatten[Solve[graph1, *y*]])**

$$y = \frac{1}{2}\left(3\,x^2 - 1\right)$$

we gain an equation that defines y explicitly as function of x.▲

2.9.2 Tests and Exercises

The following tests and exercises are related to functions and relations.

2.9.2.1 Test Problems

T1. What is a function?
( ) A collection of points in the plane.
( ) A mapping of a set into another set.
( ) A set of coordinate points in the xy-plane.

T2. What is a graph?
( ) A picture.
( ) A graphical representation of a function.
( ) A collection of points in a coordinate system.

T3. The practical meaning of the vertical line test is:
( ) To identify a relation?
( ) To identify a unique function.
( ) To select one or more points in a graph.

T4. What is a piecewise defined function?
( ) A function with different graphs.
( ) A function with different slopes.
( ) A function defined on subsets of the total domain.

T5. What is the difference between explicit and implicit functions?
( ) There is no difference.
( ) Explicit functions are the inverse of implicit functions.
( ) An explicit function can be represented as a unique algebraic expression while implicit functions are not solvable with respect to the dependent variable.

2.9.2.2 Exercises

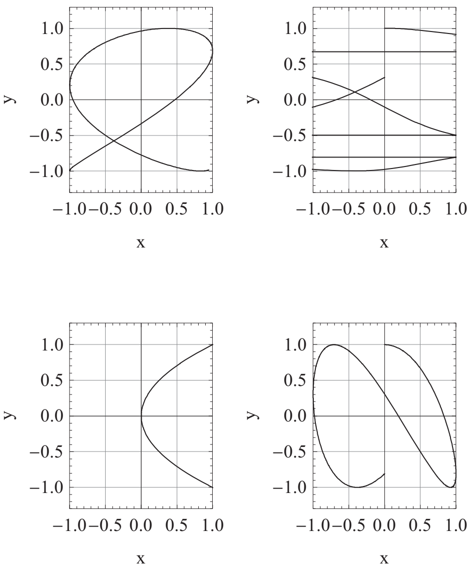
E1. You put some ice cubes in a glass, fill the glass with cold water, and then let the glass sit on a table. Describe how the
temperature of the water changes as time passes. Then sketch a rough graph of the temperature of the water as a function of the elapsed time.

E2. Sketch a rough graph of the outdoor temperature as a function of time during a typical spring day.

E3. Sketch the graph of the amount of a particular brand of coffee sold by a store as a function of the price of the coffee.

E4. In this section we discussed examples of ordinary, everyday functions: Population is a function of time, postage cost is a
function of weight, water temperature is a function of time. Give three other examples of functions from everyday life that are described verbally. What can you say about the domain and range of each of your functions? If possible, sketch a rough graph of each function.

E5. Determine whether the curves given below are graphs of a function of x. If it is, state the domain and range of the function.



E6. A homeowner mows the lawn every Wednesday afternoon. Sketch a rough graph of the height of the grass as a function of time over the course of a four-week period.

E7. Find the domain and range and sketch the graph of the function

$h(x) = \sqrt{6 - x^2}$

E8. Find the domain and sketch the graph of the following functions function:

a. $f(x) = -2$

b. $f(x) = x^2 - 7$

c. $f(x) = \frac{|x|}{x^2}$

d. $f(x) = |x^2 - 7x|$

e. $f(x) = \begin{cases} x + 2 & x \le -1 \\ x^2 & x > -1 \end{cases}$

E9. A spherical balloon with radius r inches has volume $V(r) = \frac{4}{3}\pi r^3$. Find a function that represents the amount of air required to inflate the balloon from a radius of *r meter* to a radius of $r + 1$ *meter*.

E10 Find the domain of the following functions:

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

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

c. $f(x) = \frac{5x - 7}{x^2 - 3x + 2}$

d. $f(x) = \sqrt{x} - \sqrt{5 - x}$

2.9.3 Operations on Functions

Just as a number can be added, subtracted, multiplied, and divided to produce other numbers, there is a useful way of adding subtracting, multiplying, and dividing functions to produce other functions. These operations are defined as follows:

Definition 2.15. *Operations on Functions*

Given functions f and g, their sum $f+g$, difference $f-g$, product $f\,g$ and quotient $f\,/\,g$ are defined by

$$(f+g)\,(x) \;=\; f(x)+g(x) \tag{2.75}$$

$$(f-g)\,(x) \;=\; f(x)-g(x) \tag{2.76}$$

$$(f\,g)\,(x) \;=\; f(x)\,g(x) \tag{2.77}$$

$$(f\;/\,g)\,(x) \;=\; f(x)\,/\,g(x).\blacksquare \tag{2.78}$$

For the functions $f+g$, $f-g$, and $f\,g$ the domain is defined to be the intersection of the domains of f and g, and for $f\,/\,g$ the domain is this intersection with the points where $g(x)=0$ excluded.

Example 2.56. Functions and Operations

Let f and g be the functions

$$\boldsymbol{f(\mathrm{x\_}) := \sqrt{5-x}}$$

and

$$\boldsymbol{g(\mathrm{x\_}) := \sqrt{x-2}}$$

Then the formulas for $f+g$, $f-g$, $f\,g$, and $f\,/\,g$ are

$$\boldsymbol{f(x)+g(x)}$$

$$\sqrt{5-x}+\sqrt{x-2}$$

For the difference

$$\boldsymbol{f(x)-g(x)}$$

$$\sqrt{5-x}-\sqrt{x-2}$$

For the product

$$\boldsymbol{f(x)\,g(x)}$$

$$\sqrt{5-x}\,\sqrt{x-2}$$

and the quotient

$$\frac{f(x)}{g(x)}$$

$$\frac{\sqrt{5-x}}{\sqrt{x-2}}$$

Since the domain of f is $(-\infty, 5]$ and the domain of g is $[2, \infty)$, the domain of $f+g$, $f-g$, and $f\,g$ is $[2, 5]$ because this is the intersection of both intervals. Since $g(x) = 0$ when $x = 2$, we must exclude this point to obtain the domain for $f\,/\,g$. Thus the domain for this relation is $(2, 5]$.▲

Remark 2.21. In the last section we agreed that when a function is given by a formula and there is no mention of the domain, it is understood that the domain consists of all real numbers where the formula makes sense and yields real values. Since the set [2,5] is precisely where formula (2.75), (2.76), and (2.77) makes sense and yields real values, the formulas themselves specify the domain for us. Similarly formula in (2.78) specifies that $(2, 5]$ is the domain of $f\,/\,g$ since this is precisely where the formulas makes sense and yield real values. However, as the next example shows, the formula for $f+g$, $f-g$, $f\,g$, ad $f\,/\,g$ does not always suffice to specify the domain.

Example 2.57. Domain

Let

$$f(\mathrm{x\_}) := 3\sqrt{x}$$

and

$$g(\mathrm{x\_}) := \sqrt{x}$$

find $f\,g$.

Since the domain of f is $[0, +\infty)$ and the domain of g is $[0, +\infty)$, the domain of $f\,g$ is also $[0, +\infty)$ since this is the intersection of the domains of f and g. The formula for f and g combined deliver

$$f(x)\,g(x)$$

$$3\,x$$

Since the formula derived makes sense and yields real values for all x the formula alone will not correctly describe the domain of $f\,g$ for us; we must write

$(f\,g)\,(x) = 3\,x$ with $x \geq 0$.▲

Definition 2.16. *Function Multiplied by a Real Number*

If f is a function and k is a real number, the function $k\,f$ is defined by

$$(k\,f)\,(x) = k\,f(x) \tag{2.79}$$

and the domain of $k\,f$ is the same as the domain of f.■

Example 2.58. Multiplication by a Constant

Let

$$\boldsymbol{f(\mathrm{x\_}) := x^2 + \sqrt{x}}$$

Then the multiplication by a constant k is just given by

$$\boldsymbol{k\, f(x)}$$

$$k\left(x^2 + \sqrt{x}\right)$$

▲

Sometimes a given function has the same effect as two simple functions evaluated in succession. For example consider the function h given by

$$\boldsymbol{h(\mathrm{x\_}) := (x+1)^2}$$

To evaluate $h(x)$ we first compute $x+1$ and then square the result. In other words, if we consider the functions g and f given by

$$\boldsymbol{g(\mathrm{x\_}) := x+1}$$

and

$$\boldsymbol{f(\mathrm{x\_}) := x^2}$$

then

$$\boldsymbol{h(x)}$$
$$\boldsymbol{f(g(x))}$$

$$(x+1)^2$$

$$(x+1)^2$$

Thus h has the same effect as g and f evaluated successively. Loosely speaking, h is composed of the two functions f and g. This idea is formulated in the following definition.

Definition 2.17. *Composition of Functions*

Given two functions f and g, the composition of f with g, denoted by $f \circ g$ is the function defined by

$$(f \circ g)(x) = f(g(x)) \tag{2.80}$$

where the domain of $f \circ g$ consists of all x in the domain of g for which $g(x)$ is in the domain of f. ■

Example 2.59. Composition

Find $f \circ g$ if

$$\boldsymbol{f(\mathrm{x\_}) := x^2 + 3}$$

and

g(x_) := $\sqrt{x}$

Since the domain of g is $[0, +\infty)$ and the domain of f is $(-\infty, +\infty)$, the domain of $f \circ g$ consists of all x in $[0, +\infty)$ such that $g(x) = \sqrt{x}$ lies in $(-\infty, +\infty)$; thus the domain of $f \circ g$ is $[0, +\infty)$. The result of the composition is

$f(g(x))$

```
3 + x
```

▲

Example 2.60. Composition

Let

f(x_) := $x - 1$

and

g(x_) := $\sqrt{x}$

find $f \circ g$ and $g \circ f$.

Solution 2.60. Since the domain of g is $[0, +\infty)$ and the domain of f is $(-\infty, +\infty)$, the domain of $f \circ g$ consists of all x in $[0, +\infty)$ such that $g(x) = \sqrt{x}$ lies in $(-\infty, +\infty)$, that is, all x in $[0, +\infty)$. Since

$f(g(x))$

$\sqrt{x} - 1$

For this result there is no need to indicate that the domain is $[0, +\infty)$ since this is precisely the set where $\sqrt{x} - 1$ is defined and yields real values.

The domain of $g \circ f$ consists of all x in the domain of f such that $f(x)$ lies in the domain of g. Since the domain of f is $(-\infty, +\infty)$ and the domain of g is $[0, +\infty)$, the domain of $g \circ f$ consists of all x in $(-\infty, +\infty)$ such that $f = x - 1$ lies in $[0, +\infty)$, Thus the domain is $[1, +\infty)$. Since

$g(f(x))$

$\sqrt{x - 1}$

▲

Generally, $f(g(x)) \neq g(f(x))$.

The idea of solving an equation $y = f(x)$ for x as a function of y, say $x = g(y)$, is an important idea in mathematics. Sometimes, solving an equation is a simple process; for example, using the basic algebra the equation

$f = y == x^3 + 1$

$y = x^3 + 1$

can be solved for x as a function of y

g = First[x /. Solve[f, x]]

$\sqrt[3]{y-1}$

The first equation is better for computing y if x is known, and the second is better for computing x if y is known.

Our primary interest in this section is to identify relationships that may exist between the function f and g when an equation

$$y = f(x) \tag{2.81}$$

is expressed as

$$x = g(y) \tag{2.82}$$

or conversely.

Before we can use the same function names in *Mathematica* we should clear the values and the symbols. This cleaning up is done by using the function Remove[]

Remove[f, g]

Now let us, for example consider the function

f(x_) := $x^3 + 1$

and

g(y_) := $\sqrt[3]{y-1}$

discussed above. When these functions are composed in either order they cancel out the effect of one another in the sense that

PowerExpand[g(f(x))]

x

f(g(y))

y

The first of these equations states that each output of the composition $g(f(x))$ is the same as the input, and the second states that each output of the composition $f(g(y))$ is the same as the input. Pairs of functions with these two properties are so important that there is some terminology for them.

Definition 2.18. *Inverse Functions*

If the functions f and g satisfy the two conditions

$g(f(x)) = x$ *for every x in the domain of f and*

$f(g(y)) = y$ *for every y in the domain of g*

then we say that f and g are inverse. Moreover, we call f an inverse function of g and g an *inverse function of* f.■

The inverse of a function f is commonly denoted by f^{-1} (read f inverse).

The domain and range of functions and it inverse are related by

$$\text{domain of } f^{-1} = \text{range of } f$$

$$\text{range of } f^{-1} = \text{domain of } f.$$

If any equation $y = f(x)$ can be solved for x as a function of y, then f has an inverse function and the resulting equation is

$$x = f^{-1}(y) \tag{2.83}$$

Example 2.61. Inverse Function

Find the inverse function of $f(x) = \sqrt{3\,x-2}$.

Solution 2.61. From the discussion above we can find a formula for $f^{-1}(y)$ by solving the equation

eq = y = $\sqrt{3\,x-2}$

$y = \sqrt{3\,x-2}$

for x as a function of y. The computation is

g = x /. Flatten[Solve[eq, x]]

$\frac{1}{3}\left(y^2+2\right)$

At this point we have successfully produced a formula for f^{-1}; however, we are not quite done, since there is no guarantee that the natural domain associated with this formula is the correct domain for f^{-1}. To determine whether this is so, we will examine the range of

$$y = f(x) = \sqrt{3\,x-2}\,.$$

The range consists of all y in the interval $[0, +\infty)$. This interval is also the domain of $f^{-1}(y)$; thus the inverse of f is given by the formula

g

$\frac{1}{3}\left(y^2+2\right)$

▲.

2.9.4 Tests and Exercises

The following tests and exercises are related to operations on functions.

2.9.4.1 Test Problems

T1. What is a scaled function?
( ) A set of points in the plane shifted to the right.
( ) A set of points in the plane reflected on the x-axis.
( ) A set of coordinate points multiplied by a constant number.

T2. How can two functions be composed?
( ) The two functions are chained.
( ) We add the functions together.
( ) We devise the two functions by each other.

T3. What is an inverse function?
( ) The reciprocal of a function?
( ) The exchange of x and y coordinates.
( ) The division by another function.

2.9.4.2 Exercises

E1. What do all members of the family of linear functions $f(x) = c + x$ have in common? Sketch several members of the family.

E2. Find an equation for the family of linear functions with slope 2 and sketch several members of the family.

E3. Find an equation for the family of linear functions such that $f(1) = 3$ and sketch several members of the family.

E4. The relationship between the Fahrenheit (F) and Celsius (C) temperature scales is given by the linear function

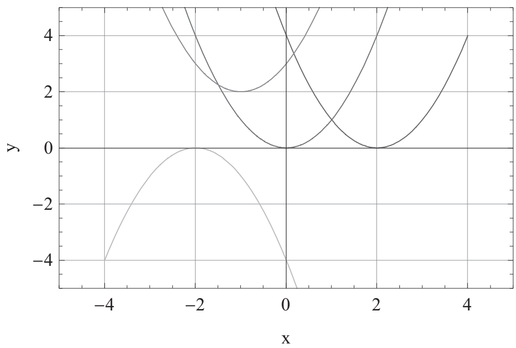
$$F = \frac{9}{5} C + 32$$

a. Sketch a graph of this function.
b. What is the slope of the graph and what does it represent?
c. What is the F-intercept and what does it represent?

E5. Explain how each graph is obtained from the graph of $y = f(x)$.

a. $y = 5 f(x)$
b. $y = f(x - 6)$
c. $y = -f(x)$
d. $y = 5 f(x) - 8$
e. $y = f(6 x)$
f. $y = f(-x)$

E6. The graph of $y = f(x)$ is given. Match each equation with its graph and give reasons for your choices.



E7. Find the functions $f \circ g$, $g \circ f$, $f \circ f$, and $g \circ g$ and their domains.

a. $f(x) = 2 x^2 - 2$, $g(x) = 4 x - 1$

$f(x) = x - 2$, $g(x) = x^2 + 3x - 1$

c. $f(x) = \sqrt{x-1}$, $g(x) = 4x^3 - 1$

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

e. $f(x) = \cos(x)$, $g(x) = \sin(3x - 2)$

E8. Find $f \circ g \circ h$ for the following functions

a. $f(x) = 2x^2 - 2$, $g(x) = 4x - 1$, $h(x) = \sqrt{x}$

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

c. $f(x) = \sqrt{x-1}$, $g(x) = 4x^3 - 1$, $h(x) = \frac{1}{\sqrt{x}} + x - 2$

d. $f(x) = \frac{1}{x} - x^2$, $g(x) = \sqrt{x} - \frac{1}{x+3}$, $h(x) = \log(x) - 2$

e. $f(x) = \cos(x)$, $g(x) = \sin(3x - 2)$, $h(x) = \cos^{-1}(x)$

E9. Find the inverse function of the following functions:

a. $f(x) = 2x^2 - 2$,

b. $f(x) = x - 2$,

c. $f(x) = 4x^3 - 1$,

d. $f(x) = \frac{1}{x} - x^2$,

e. $f(x) = \cos(x)$.

E10 The **Heaviside function** H is defined by

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

It is used in the study of electric circuits to represent the sudden surge of electric current, or voltage, when a switch is instantaneously turned on.

a. Sketch the graph of the Heaviside function.

b. Sketch the graph of the voltage $V(t)$ in a circuit if the switch is turned on at time $t = 0$ and 120 volts are applied instantaneously to the circuit. Write a formula for $V(t)$ in terms of $H(t)$.

c. Sketch the graph of the voltage $V(t)$ in a circuit if the switch is turned on at time $t = 5$ seconds and 240 volts are applied instantaneously to the circuit. Write a formula for $V(t)$ in terms of $H(t)$. (Note that starting at $t = 5$ corresponds to a translation.)

2.9.5 Symmetry

Figure 2.78 shows the graph of three curves that have certain obvious symmetries. The graph on the left is symmetric about the x-axis in the sense that for each point (x, y) on the graph the point $(x, -y)$ is also on the graph; the graph in the middle is symmetric about the y-axis in the sense that for each point (x, y) on the graph the point $(-x, y)$ is also on the graph; and the graph to the right in Figure 2.78 is symmetric about the origin in the sense that for each point (x, y) on the graph the point $(-x, -y)$ is also on the graph. Geometrically, symmetry about the origin occurs if rotating the graph $180°$ about the origin leaves the graph unchanged.

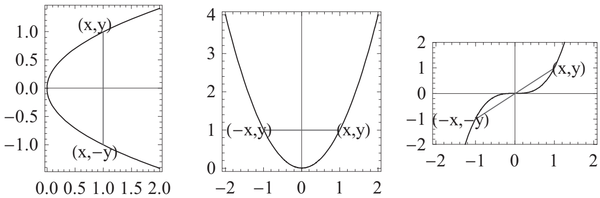


Figure 2.78. Graph of symmetric functions.

Symmetries can be often detected from the equation of a curve. For example, the graph of

$$y = x^3 \tag{2.84}$$

must be symmetric about the origin because for any point (x, y) whose coordinates satisfy (2.78), the conditions of the point $(-x, -y)$ also satisfy (2.84), since substituting these coordinates in (2.84) yields

$$-y = (-x)^3 \tag{2.85}$$

which simplifies to (2.84)

Theorem 2.21. *Symmetry Test*

a) A planar curve is symmetric about the y-axis if and only if replacing x by $-x$ in its equation produces an equivalent equation.

b) A planar curve is symmetric about the x-axis if and only if replacing y by $-y$ in its equation produces an equivalent equation.

c) A planar curve is symmetric about the origin if and only if replacing both x by $-x$ and y by $-y$ in its equation produces an equivalent equation. ∎

2.9.5.1 Even and Odd Functions

For the graph of a function f to be symmetric about the y-axis, the equations $y = f(x)$ and $y = f(-x)$ must be equivalent; for this to happen we must have

$$f(x) = f(-x). \tag{2.86}$$

A function with this property is called an **even function**. Some examples are x^2, x^4, x^6, and $\cos(x)$. Similarly, for the graph of a function f to be symmetric about the origin, the equations $y = f(x)$ and $-y = f(-x)$ must be equivalent; for this to happen we must have

$$f(x) = -f(-x). \tag{2.87}$$

A function with this property is called an odd function. Some examples are x, x^3, x^5, and $\sin(x)$.

2.9.5.2 Translations

Once you know the graph of an equation $y = f(x)$, there are some techniques that can be used to help visualize the graph of the equations

$$y = f(x) + c, \qquad y = f(x) - c, \qquad y = f(x + c), \qquad y = f(x - c) \tag{2.88}$$

where c is any positive constant.

If a positive constant is added to a structured form $f(x)$, the geometric effect is to translate the graph of $y = f(x)$ along the y-axis; addition translates the graph in the positive direction and subtraction translates it in the negative direction. This is illustrated in the following animations:

Out[9]=

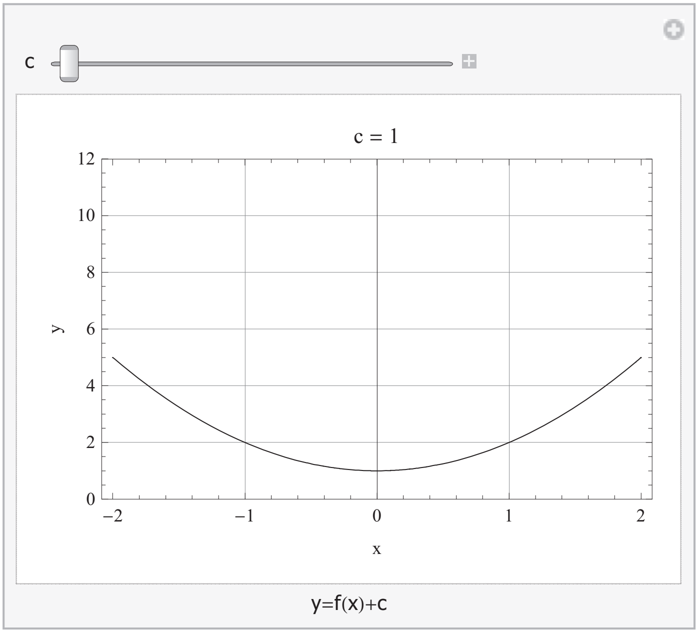


Figure 2.79. Translation of the graph $y = f(x)$ by adding a constant c to the function $f(x)$; $y = f(x) + c$.

Out[10]=

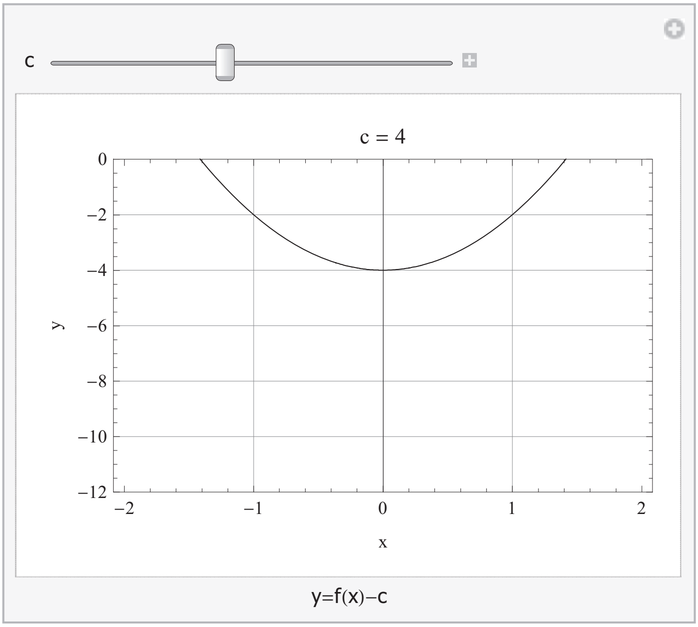


Figure 2.80. Translation of the graph $y = f(x)$ by subtracting a constant c from the function $f(x)$; $y = f(x) - c$.

Similarly, if a positive constant is added to or subtracted from the independent variable x, the geometric effect is to translate the graph of the function along the x-axis; subtraction translate the graph in the positive direction and addition translates it in the negative direction. Examples are given in the following animations:

Out[11]=

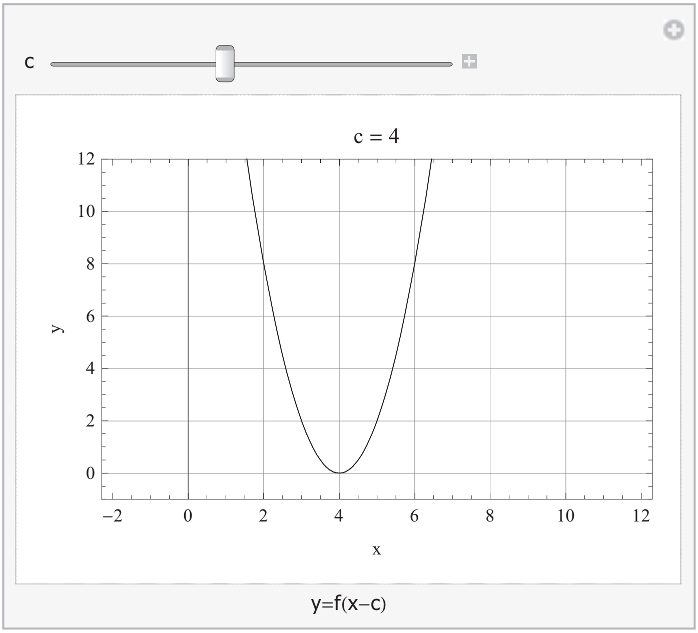


Figure 2.81. Translation of the graph $y = f(x)$ by subtracting a constant c from the independent variable x; $y = f(x - c)$.

Out[12]=

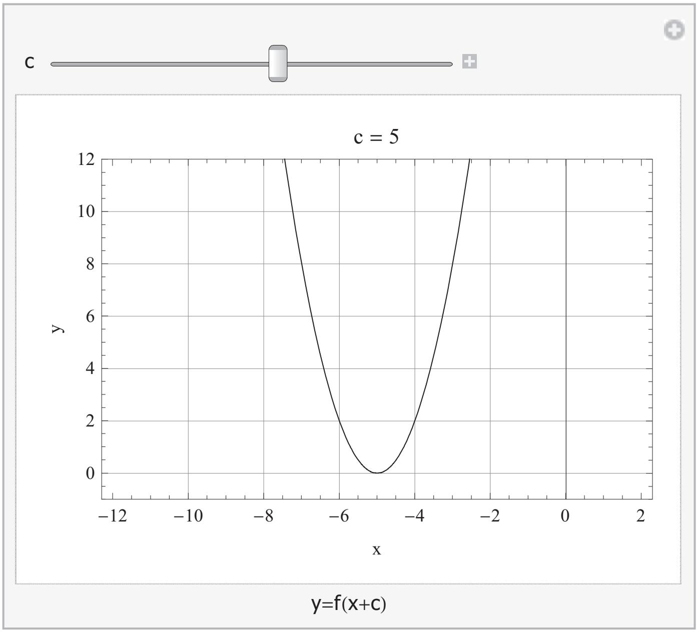


Figure 2.82. Translation of the graph $y = f(x)$ by adding a constant c from the independent variable x; $y = f(x + c)$.

Example 2.62. Translation I

Sketch the graph of

$$\boldsymbol{f}(\mathbf{x\_}) := \sqrt{\boldsymbol{x} - \mathbf{3}}$$

and

$$\boldsymbol{g}(\mathbf{x\_}) := \sqrt{\boldsymbol{x} + \mathbf{3}}$$

Solution 2.62. The graph of the equation $y = \sqrt{x-3}$ can be obtained by translating the graph of $y = \sqrt{x}$ to the right by 3 units, and the graph of $y = \sqrt{x+3}$ by translating the graph of $y = \sqrt{x}$ left 3 units.

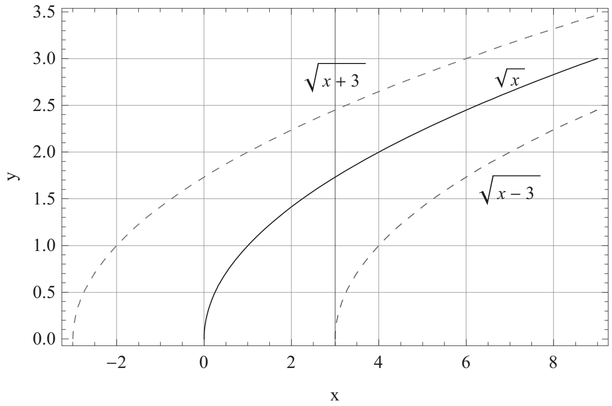


Figure 2.83. Translation of the graph $y = \sqrt{x}$ by adding and subtracting a constant $c = 3$.

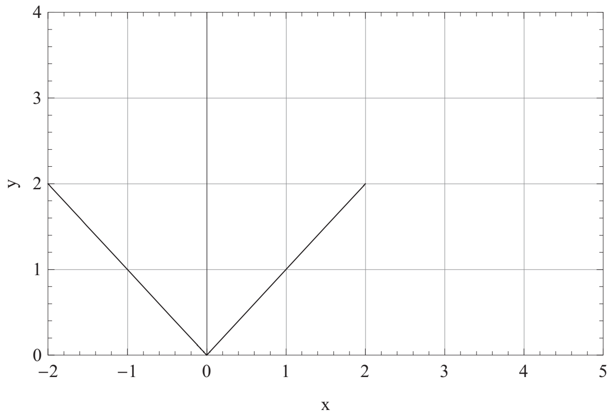
Example 2.63. Translation II

Sketch the graph of

$$f(x) := |x - 3| + 2$$

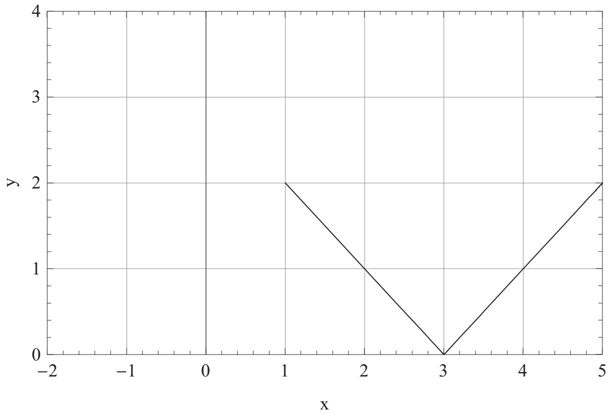
Solution 2.63. In a first step we start from the graph $y = |x|$; this helps us to identify the principal graph

```
pl1 = Plot[Abs[x], {x, -2, 2},
  FrameLabel → {"x", "y"}, PlotRange → {{-2, 5}, {0, 4}}]
```



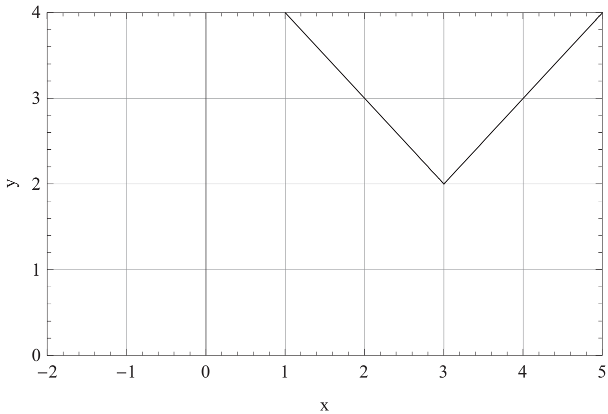
The next step is to move the plot 3 steps to the right by $y = |x - 3|$:

```
pl2 = Plot[Abs[x - 3], {x, 1, 5},
  FrameLabel → {"x", "y"}, PlotRange → {{-2, 5}, {0, 4}}]
```

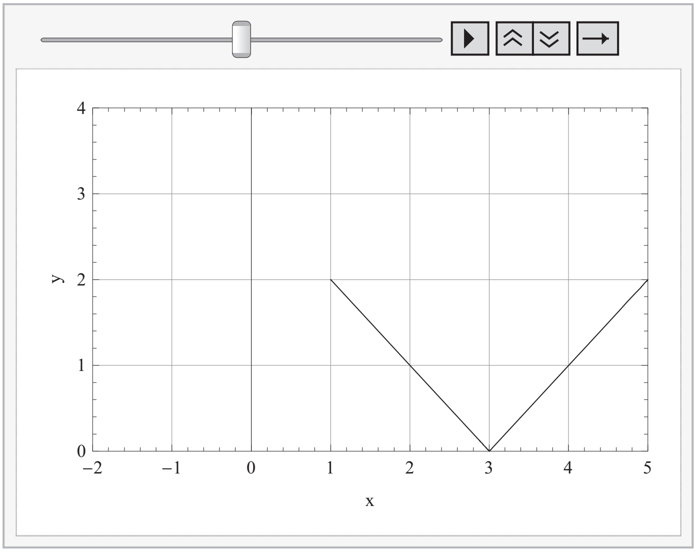


The last step translates the graph 2 units in the y direction $y = |x-3|+2$:

```
pl3 = Plot[Abs[x - 3] + 2, {x, 1, 5},
  FrameLabel → {"x", "y"}, PlotRange → {{-2, 5}, {0, 4}}]
```



The sequence can be animated by:

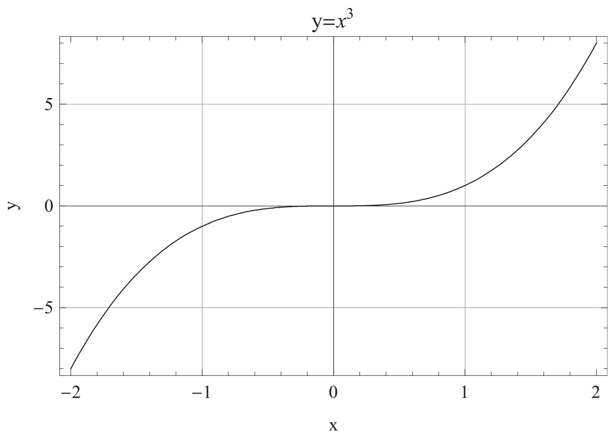


2.9.5.3 Scaling and Reflection

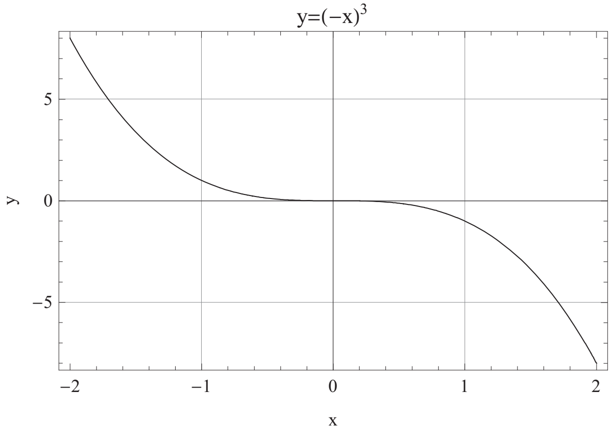
Scaling and reflections are useful operations to transform a function into a new state.

The graph of $y = f(-x)$ is the reflection of the graph of $y = f(x)$ about the y-axis, and the graph of $y = -f(x)$ is the reflection of the graph of $y = f(x)$ about the x-axis. Thus, if you know what the graph of $y = f(x)$ looks like, you can obtain the graphs of $y = f(-x)$ and $y = -f(x)$ by making appropriate reflections. This is illustrated in the following figures and in the animation.

First we plot the function $y = x^3$



In the next step we change the sign of x so that we have to plot $y = (-x)^3$.



The following animation shows you that the change of sign changes the graph in the way discussed above.

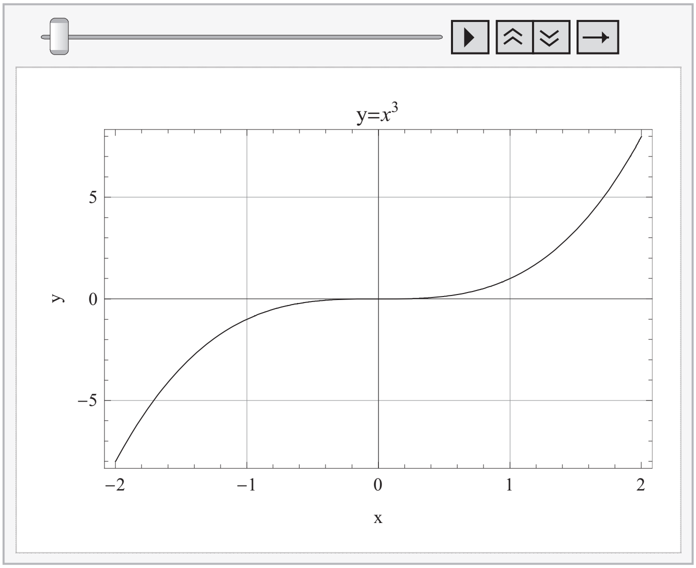


Figure 2.84. Animation of the reflection process of the graph $y = x^3$ by changing the sign in front of x.

The receipt to generate reflections about the y-axis is to replace the original x value by the negative x value is a short hand notation; this reads

$$f(x) \text{ /. } x \to -x$$

$$f(-x)$$

where the notation /. represents the replacement operation and $x \to -x$ gives the rule for the replacement.

The following example shows how reflections about the x-axis take place. Lets assume that the function

$$f(\text{x\_}) := \sqrt{x}$$

is related about the x-axis. The graph of the function is given by

$$y = f(x)$$

$$\sqrt{x}$$

and the plot of the graph is shown in the following figure

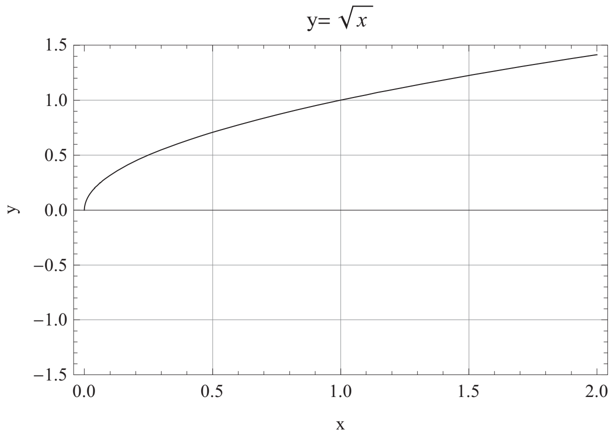


Figure 2.85. Plot of the graph $y = \sqrt{x}$.

The reflection about the x-axis is generated by replacing the sign of the function $f(x)$ such that the graph changes to $y = -\sqrt{x}$

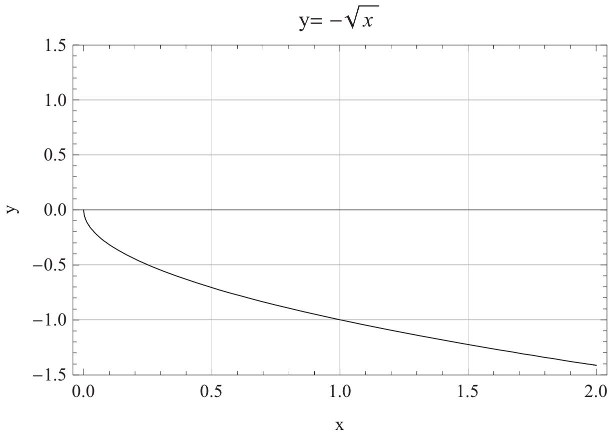


Figure 2.86. Plot of the graph $y = -\sqrt{x}$.

The following animation shows how the operation of changing the sign takes action on the graph

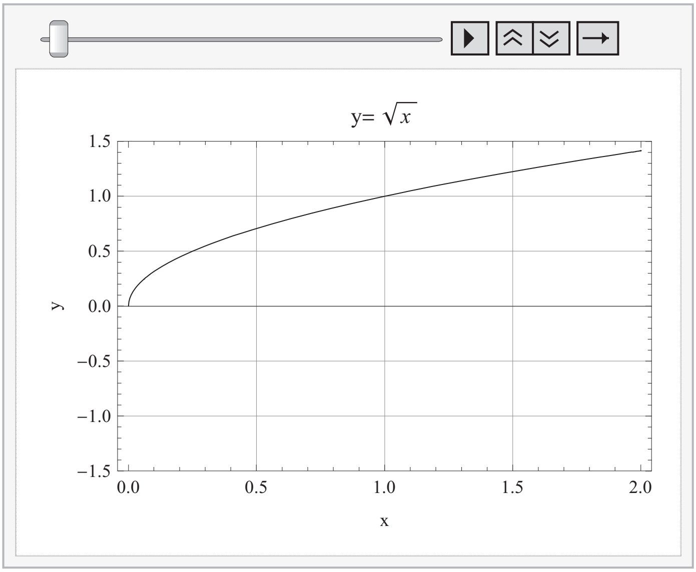


Figure 2.87. Animation of the reflection process of the graph $y = \sqrt{x}$ by changing the sign in front of $\sqrt{x}$.

The overall receipt to make reflection about the x-axis is the replacement of the total function by the negative function

```
Remove[f];
```

$f(x)$ /. $f(x) \to -f(x)$

$-f(x)$

The term scaling describes a seemingly very simple situation where the arguments of a function or the function itself is stretched or squeezed. This behavior often appears in the mathematical modeling of various phenomena from economics and engineering.

Scaling along the horizontal axis means that the graph of a function $y = f(a\,x)$ is the scaled representation of the function $f(x)$. Scaling actually means that we stretch or squeeze the units of the x-axis depending on the amount of the stretching factor a. For $a > 1$ we stretch the x-axis and for $a < 1$, we squeeze the x-axis.

The following example demonstrates that. Let us assume the function is given by

f(x_) := cos(x)

Then the graph of this function is generated by

$y = f(x)$

`Cos[x]`

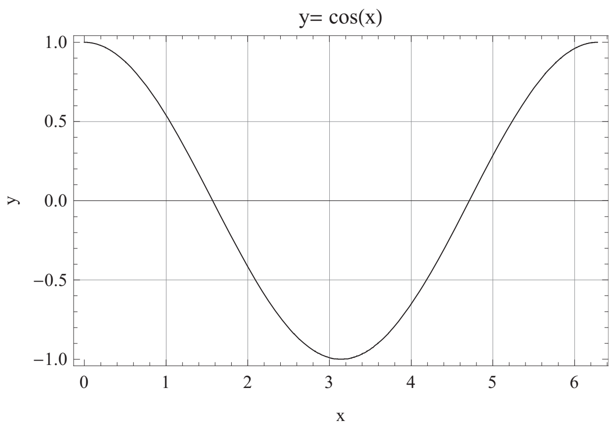


Figure 2.88. Plot of the graph $y = \cos(x)$.

The scaled version of the plot is generated by replacing the argument of the function with the stretched argument. Lets assume the scaling factor $a = 2$ then the scaled graph is

$y = y$ /. $x \to 2\,x$

$\cos(2\,x)$

The plot of the graph is

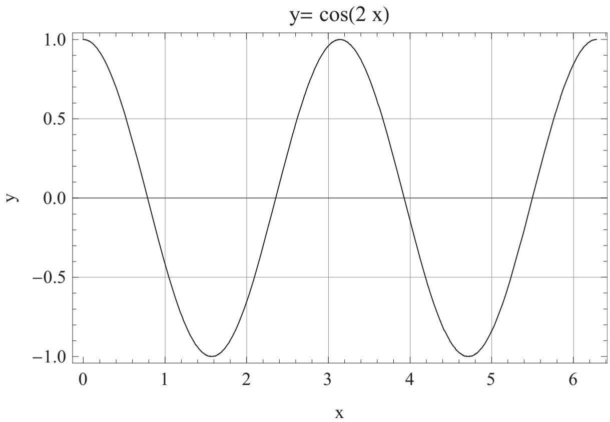


Figure 2.89. Plot of the graph $y = \cos(2\,x)$.

The action of the scaling process can be seen in the following animation

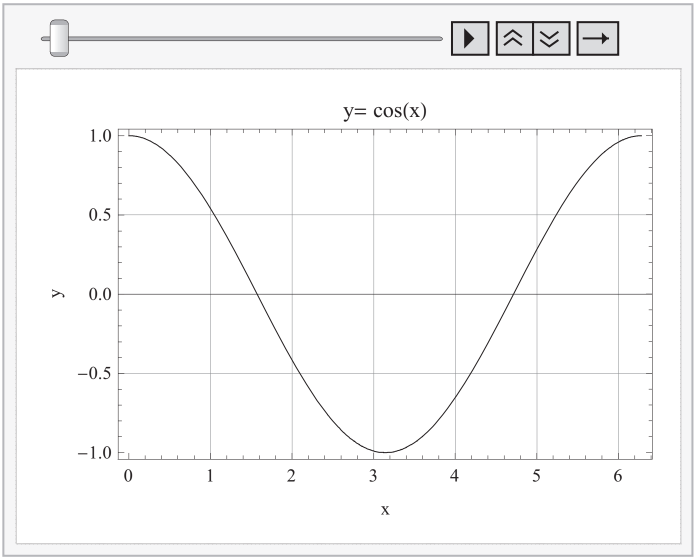


Figure 2.90. Animation of the scaling process of the graph $y = \cos(x)$ by changing the argument of the function from $x \to 2\,x$.

The receipt of scaling along the horizontal axis is to replace the argument of a function by a scaled variable

```
Remove[f];
```

$f(x)$ /. $x \to a\,x$

$f(a\,x)$

The vertical scaling of a function is generated by amplifying or compressing the function $f(x)$ by a factor a which can be greater or smaller than 1. The amplification of the function $f(x)$ by a factor a is given by the graph $y = a\,f(x)$ of the function $f(x)$ with $a > 1$ or $a < 1$.

The following example shows the scaling process along the vertical axis of the function

$f(x\_) := \sin(x)$

The graph of he function is

$y = f(x)$

$\sin(x)$

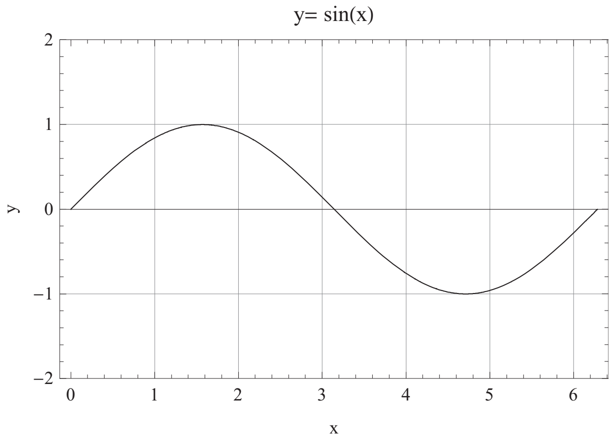


Figure 2.91. Plot of the graph $y = \sin(x)$.

The scaled version of the plot is generated by replacing the function with the amplified function. Lets assume the scaling factor $a = 2$ then the scaled graph is

$y = y$ /. $f(x) \to 2\,f(x)$

$2\sin(x)$

The plot of the graph is

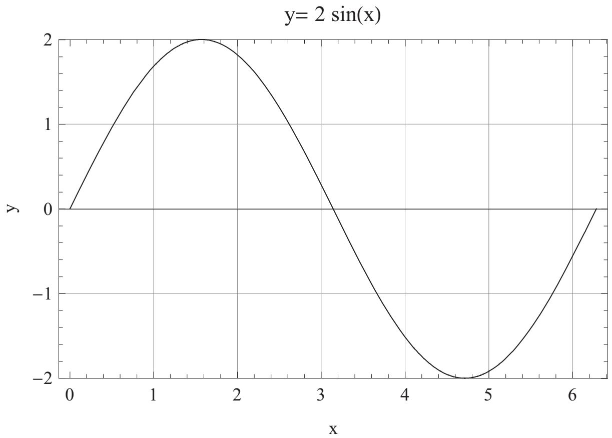


Figure 2.92. Plot of the graph $y = 2\sin(x)$.

The action of the amplification process can be seen in the following animation

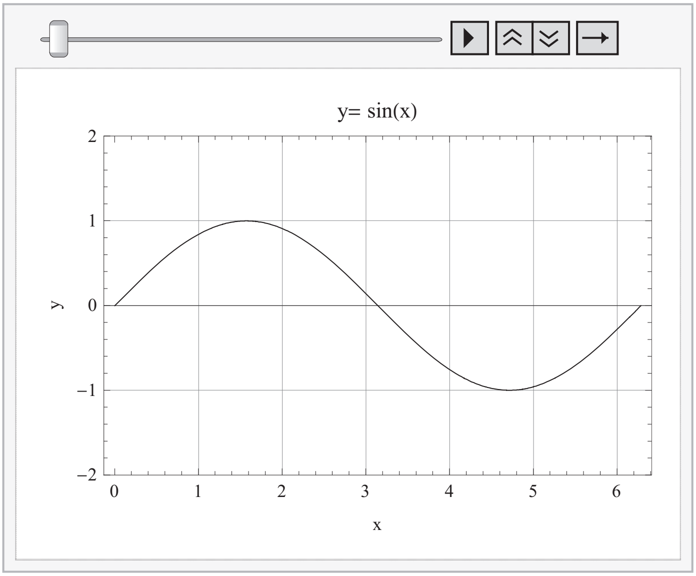


Figure 2.93. Animation of the amplification process of the graph $y = \sin(x)$ by changing the function from $\sin(x) \to 2\sin(x)$.

The receipt of scaling along the vertical axis is to replace the function by a scaled version of the function

```
Remove[f];
```

$$f(x) \,/.\, f(x) \to a\, f(x)$$

$$a\, f(x)$$

This is known as scaling of the vertical axis, where $a > 1$ corresponds to an amplification and $a < 1$ to a compression along the vertical axis. The following animation shows the change of the function if we scale it.

Out[13]=

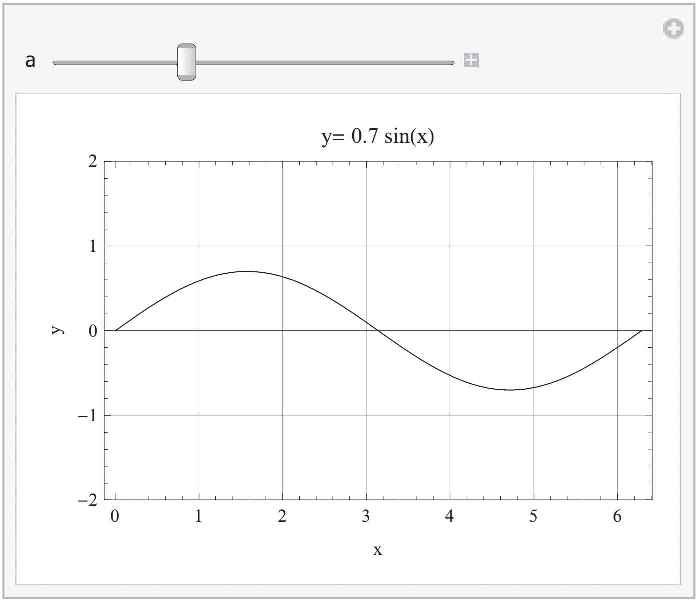


Figure 2.94. Scaling of the function $f(x) = a \sin(x)$.

2.9.6 Tests and Exercises

The following tests and exercises are related to symmetry properties of functions.

2.9.6.1 Test Problems

T1. How do you translate a function?
( ) A function is translated by shifting its argument.
( ) A function is translated by shifting its function value.
( ) A function is translated by multiplying its argument by a constant.

T2. What is a reflection of a function?
( ) A multiplication by one.
( ) A multiplication by minus one.
( ) A division by a negative constant.

T3. What are symmetries of a function?
( ) Translations.
( ) Reflections.
( ) Scaling.

T4. What are the symmetry properties of even functions?
( ) $f(x) = -f(x-1)$

( ) $f(x) = f(-x)$
( ) $f(x) = -f(-x)$

T5. What are the symmetry properties of odd functions?
( ) $f(x) = -f(x)$
( ) $f(x) = f(-x)$
( ) $-f(x) = f(-x)$

2.9.6.2 Exercises

E1. Assume that f is an even function, g is an odd function, and both f and g are defined on the entire real line $\mathbb{R}$. Which of the following expressions (where defined) are even? odd?

a. $f\,g$
b. $g^2 = g\,g$
c. g/f
d. $f \circ g$
e. $g \circ f$
f. $f^2 = f\,f$

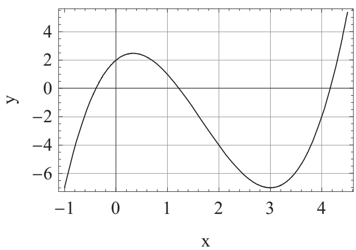
E2. Graph the function $f(x) = x^4 + c\,x^2 + x$ for several values of c. How does the graph change when c changes?

E3. Use graphs to determine which of the functions $f(x) = x^4 - 100\,x$ and $g(x) = x^3$ is eventually larger.

E4. Graph the functions given below by hand. Do not use plotting points, but by starting with the graph of one of the standard functions, and then applying the appropriate transformations.

a. $y = (x+1)^2$
b. $y = -x^3$
c. $y = 1 - x^2$
d. $y = (1-x)^2$
e. $y = (1+x)^2$
f. $y = -(1+x)^2$
g. $y = (x+2)^4 + 5$
h. $\frac{1}{2}\left(x^2 + 8\,x\right)$

E5. Use the given graph of to sketch the graph of $y = 1/f(x)$. Which features of f are the most important in sketching $y = 1/f(x)$? Explain how they are used.



E6. Explain how each graph is obtained from the graph of $y = f(x)$.

a. $y = 4\,f(x)$
b. $y = -f(x)$
c. $y = -4\,f(x) - 2$
d. $y = f(2\,x) + 4$
e. $y = f(x-1)$

E7. Suppose the graph of $f(x)$ given. Write equations for the graphs that are obtained from the graph of $f(x)$ as follows.

a. Shift 4 units upward.
b. Shift 4 units downward.
c. Shift 4 units to the right.
d. Shift 4 units to the left.
e. Reflect about the x-axis.
f. Reflect about the y-axis.
g. Stretch vertically by a factor of 4.
h. Shrink vertically by a factor of 4.

E8. Suppose g is an even function and let $h = f \circ g$. Is h always an even function?

E9. Suppose g is an odd function and let $h = f \circ g$. Is h always an odd function? What if f is odd? What if f is even?

E10 Suppose f is even and g is odd. What can you say about $f\, g$?

2.9.7 Classification of Functions

A function that assigns the same value to every member of its domain is called a **constant function**.

Example 2.64. Constant Function

In national-income models, when investment (I) is exogenously determined, we may have an investment function of the form

f(x_) := 1000

or in general $f(x) = f_0$ with f_0 a real number. By definition the graph of f is the graph of the equation $y = f(x)$; that is

y = f(x)

```
1000
```

This function is given in the following plot.

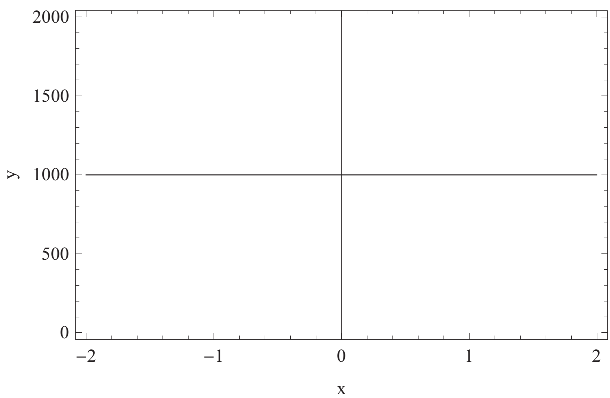


Figure 2.95. Graph of the constant function $f(x) = 1000$.▲

2.9.7.1 Monomials

A function of the form $c\,x^n$, where c is a constant and n is a non-negative integer is called a **monomial in x**. Examples for monomial are for power 3

$$\mathbf{3\,x^3}$$

$$3\,x^3$$

For power 7 we may have

$$\boldsymbol{\pi\,x^7}$$

$$\pi\,x^7$$

Even for power $n = 0$ we talk about monomials which are in fact constants

$$\mathbf{4\,x^0}$$

$$4$$

and power 5 with negative coefficient

$$\mathbf{-6\,x^5}$$

$$-6\,x^5$$

and a big power like 34

$$\mathbf{x^{34}}$$

$$x^{34}$$

The following functions

$$\mathbf{4\,\sqrt{x}}$$

$$4\,\sqrt{x}$$

and

$$\mathbf{\frac{1}{x^6}}$$

$$\frac{1}{x^6}$$

are not monomials because the powers of x are not non-negative integers.

2.9.7.2 Polynomials

A function that is expressible as the sum of finitely many monomials in x is called a **polynomial in *x***. Examples are cubic polynomials

$$\mathbf{x^3 + 4\,x + 15}$$

$$x^3 + 4\,x + 15$$

A polynomial of order 4 is

$$\mathbf{17 - \frac{3\,x^4}{5}}$$

$$17 - \frac{3\,x^4}{5}$$

A constant, a polynomial which represents a polynomial of order 0, is a special type of polynomials

9

9

and

$\boldsymbol{x^9}$

x^9

Also algebraic expressions can result in a polynomial. If we expand the algebraic expression $(x-7)^5$, we get

$\mathbf{(x-7)^5}$ **// Expand**

$-16\,807 + 12\,005\,x - 3430\,x^2 + 490\,x^3 - 35\,x^4 + x^5$

In fact this is a polynomial in x because it is expressible as a sum of monomials in x. In general a function f is a polynomial in x if it is expressible in the form

$$f(x) = a_0 + a_1\,x + a_2\,x^2 + a_3\,x^3 + \ldots + a_n\,x^n \tag{2.89}$$

where n is a non-negative integer and $a_0, a_1, a_2, \ldots, a_n$ are real constants. The domain of any polynomial is $(-\infty, +\infty)$.

A polynomial is called **linear** if it has the simple form

$\boldsymbol{a_1\,x + a_0}$

$a_1\,x + a_0$

where $a_1 \neq 0$. We call it **quadratic** if it has the form

$\boldsymbol{a_2\,x^2 + a_1\,x + a_0}$

$a_2\,x^2 + a_1\,x + a_0$ where $a_2 \neq 0$

and **cubic** if it has the form

$\boldsymbol{a_3\,x^3 + a_2\,x^2 + a_1\,x + a_0}$

$a_3\,x^3 + a_2\,x^2 + a_1\,x + a_0$ where $a_3 \neq 0$

Binomials

Sometimes it is useful to reformulate a polynomial in a binomial representation. Binomials are polynomials which can be reduced to a power function including two terms combined either by addition or subtraction. They can be reformulated as polynomials. This reformulation is governed by the following theorem.

Theorem 2.22. *Binomial Theorem*

For any positive integer n the following expression holds

$$(x+a)^n = \\ x^n + n\,x^{n-1}\,a + \frac{n(n-1)}{1*2}\,x^{n-2}\,a^2 + \frac{n(n-1)\,(n-2)}{1*2*3}\,x^{n-3}\,a^3 + \ldots + n\,x\,a^{n-1} + a^n. \blacksquare \tag{2.90}$$

Applications of this theorem are

Expand$\left[(a+x)^2\right]$

$a^2 + 2\,a\,x + x^2$

and

Expand$\left[(a+x)^3\right]$

$a^3 + 3\,a^2\,x + 3\,a\,x^2 + x^3$

or

Expand$\left[(x-a)^2\right]$

$a^2 - 2\,a\,x + x^2$

Expand$\left[(x-a)^3\right]$

$-a^3 + 3\,a^2\,x - 3\,a\,x^2 + x^3$

Quadratic Completion

This procedure is especially useful for generating closed expressions of quadratic polynomials. Lets assume that we have given an expression like

$$a\,x^2 + b\,x + c \tag{2.91}$$

where $a \neq 0$. We are looking for an expression which represents the second order polynomial in a binomial form like

$$a\,(x+C)^2 + K. \tag{2.92}$$

This representation can be generated by the following steps. First, we extract the common factor a from the terms including x

$$a\,x^2 + b\,x + c = a\left(x^2 + \frac{b}{a}\,x\right) + c \tag{2.93}$$

Then we add and subtract to the terms in braces the term $b^2 \big/ \left(4\,a^2\right)$ which gives

$$a\,x^2 + b\,x + c = a\left(x^2 + \frac{b}{a}\,x + \frac{b^2}{4\,a^2} - \frac{b^2}{4\,a^2}\right) + c. \tag{2.94}$$

Grouping terms together we find

$$a\,x^2 + b\,x + c = a\left(x^2 + \frac{b}{a}\,x + \frac{b^2}{4\,a^2} - \frac{b^2}{4\,a^2}\right) + c = a\left(x^2 + \frac{b}{a}\,x + \frac{b^2}{4\,a^2}\right) + a\left(-\frac{b^2}{4\,a^2}\right) + c. \tag{2.95}$$

which simplifies to

$$a\,x^2 + b\,x + c = a\left(x + \frac{b}{2\,a}\right)^2 + c - \frac{b^2}{4\,a}. \tag{2.96}$$

Now we can identify $C = b/(2\,a)$ and $K = c - b^2\big/(4\,a)$. This means that

$$a\,x^2 + b\,x + c = a\left(x + \frac{b}{2\,a}\right)^2 + c - \frac{b^2}{4\,a} = a(x + C)^2 + K. \tag{2.97}$$

Polynomial Division

When two positive integers are divided, the numerator can be expressed as the quotient plus the remainder over the divisor, where the remainder is less than the divisor. For example,

$$\frac{17}{5} = 3 + \frac{2}{5}. \tag{2.98}$$

If we multiply this equation by 5, we obtain

$$17 = 5 \times 3 + 2 \tag{2.99}$$

which states that the numerator is the divisor times the quotient plus the remainder.

The following theorem, which is stated without proof, is an analogous result for division of polynomials.

Theorem 2.23. *Divisor Theorem*

If $p(x)$ and $s(x)$ are polynomials, and if $s(x)$ is not the zero polynomial, then $p(x)$ can be expressed as

$$p(x) = s(x)\,q(x) + r(x) \tag{2.100}$$

where $q(x)$ and $r(x)$ is the quotient and remainder that result when $p(x)$ is divided by $s(x)$, and either $r(x)$ is the zero polynomial or the degree of $r(x)$ is less than the degree of $s(x)$. ■

In the special case where $p(x)$ is divided by a first-degree polynomial of the form $x - c$, the remainder must be some constant r, since it is either zero or has degree less than 1. Thus, Theorem 2.23 implies that

$$p(x) = (x - c)\,q(x) + r \tag{2.101}$$

and this in turn implies that $p(c) = r$. In summary, we have the following theorem.

Theorem 2.24. *Remainder Theorem*

If a polynomial $p(x)$ is divided by $x - c$, then the remainder is $p(c)$. ■

Example 2.65. Remainder Theorem

According to the Remainder Theorem, the reminder on dividing

$$p(x) = 2\,x^3 + 3\,x^2 - 4\,x - 3$$

by $x + 4$ should be

$$p(-4) = 2\,(-4)^3 + 3\,(-4)^2 - 4\,(-4) - 3 = -67.$$

Show that this is so.

Solution 2.65 By means of a long division

$$\begin{array}{l}
\quad 2\,x^3 + 3\,x^2 - 4\,x - 3 : x + 4 = 2\,x^2 - 5\,x + 16 \\
-\left(2\,x^3 + 8\,x^2\right) \\
\quad ------ \\
\qquad\qquad -5\,x^2 - 4\,x \\
\qquad\quad -\left(-5\,x^2 - 20\,x\right) \\
\qquad\qquad ---------- \\
\qquad\qquad\qquad\quad 16\,x - 3 \\
\qquad\qquad\qquad -(16\,x + 64) \\
\qquad\qquad\qquad -------- \\
\qquad\qquad\qquad\qquad -67
\end{array}$$

which shows that the remainder is -67.▲

The Factor Theorem

To factor a polynomial $p(x)$ is to write it as a product of lower-degree polynomials, called factors of $p(x)$. For $s(x)$ to be a factor of $p(x)$ there must be no reminder when $p(x)$ is divided by $s(x)$. For example, if $p(x)$ can be factored as

$$p(x) = s(x)\,q(x) \tag{2.102}$$

then

$$\frac{p(x)}{s(x)} = q(x) \tag{2.103}$$

so dividing $p(x)$ by $s(x)$ produces a quotient $q(x)$ with no remainder. Conversely, the last relation implies the previous one, so $s(x)$ is a factor of $p(x)$ if there is no remainder when $p(x)$ is divided by $s(x)$.

In the special case where $x - c$ is a factor of $p(x)$, the polynomial $p(x)$ can be expressed as

$$p(x) = (x - c)\,q(x) \tag{2.104}$$

which implies that $p(c) = 0$. Conversely, if $p(c) = 0$, then the Remainder Theorem implies that $x - c$ is a factor of $p(x)$, since the remainder is 0 when $p(x)$ is divided by $x - c$. These results are summarized in the following theorem.

Theorem 2.25. *Factor Theorem*

A polynomial $p(x)$ has a factor $x - c$ if and only if $p(c) = 0$. ■

It follows from this theorem that the statements below say the same thing in different ways:

- $x - c$ is a factor of $p(x)$
- $p(c) = 0$.
- c is a zero of $p(x)$.
- c is a root of the equation $p(x) = 0$.
- c is a solution of the equation $p(x) = 0$.
- c is an x-intercept of $y = p(x)$.

Example 2.66. Factor Theorem

Confirm that $x - 1$ is a factor of

$$p(x) = x^3 - 3x^2 - 13x + 15$$

use one of the above methods suggested to check this.

Solution 2.66 By inserting the value $c = 1$ into the polynomial

$$\mathbf{p(x\_) := x^3 - 3x^2 - 13x + 15}$$

we find that

$$\mathbf{p(1)}$$

$$0$$

This means that $x - 1$ is a factor of the above polynomial.▲

Using one Factor to Find Other Factors

If $x - c$ is a factor of $p(x)$, and if $q(x) = p(x)/(x - c)$, then

$$p(x) = (x - c)\, q(x) \tag{2.105}$$

so that additional linear factors of $p(x)$ can be obtained by factoring the quotient $q(x)$.

Example 2.67. Factoring Polynomials

Factor the polynomial

$$\mathbf{p = x^3 - 3x^2 - 13x + 15}$$

$$x^3 - 3x^2 - 13x + 15$$

Solution 2.67 One factor of this polynomial is $x - 1$ which allows us to write this polynomial as

$$x^3 - 3x^2 - 13x + 15 = (x-1)\left(x^2 - 2x - 15\right)$$

$$x^3 - 3x^2 - 13x + 15 = (x-1)\left(x^2 - 2x - 15\right)$$

Factoring the polynomial $q(x)$ by inspection we find

$$x^2 - 2x - 15 = (x-5)(x+3)$$

$$x^2 - 2x - 15 = (x-5)(x+3)$$

So that in total the polynomial can be written as

$$x^3 - 3x^2 - 13x + 15 = (x-1)(x-5)(x+3)$$

$$x^3 - 3x^2 - 13x + 15 = (x-5)(x-1)(x+3)$$

This calculation can be done automatically by using the factoring function of *Mathematica*

Factor[p]

$(x-5)(x-1)(x+3)$

which agrees with the derived result.▲

Methods for Finding Roots

A general quadratic equation $a x^2 + b x + c = 0$ can be solved by using the quadratic formula to express the solutions of the equation in terms of the coefficients. Versions of this formula were known since Babylonian times, and by the seventeenth century formulas had been obtained for solving general cubic and quartic equations. However, attempts to find formulas for the solutions of general fifth-degree equations and higher proved fruitless. The reason for this became clear in 1829 when the French mathematician Evariste Galois (1811-1832) proved that it is impossible to express the solutions of a general fifth-degree equation or higher in terms of its coefficients using algebraic operations.

Today, we have powerful computer programs for finding the zeros of specific polynomials. This means that the proof of Galois still holds but in special cases there are ways to find solutions by simple operations. These operations are used by those programs. For example, it takes only seconds for a computer algebra system, such as *Mathematica* or Maple, to show that the zeros of a polynomial

$$p(\text{x\_}) := 10x^4 - 23x^3 - 10x^2 + 29x + 6$$

are

Solve[$p(x) = 0$, x]

$$\left\{\{x \to -1\}, \left\{x \to -\frac{1}{5}\right\}, \left\{x \to \frac{3}{2}\right\}, \{x \to 2\}\right\}$$

The algorithm that these programs use to find the integer and rational zeros of a polynomial, if any, are based on the following theorem.

Theorem 2.26. *Roots of a Polynomial*

Suppose that

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \ldots + c_1 x + c_0 \tag{2.106}$$

is a polynomial with integer coefficients.

a) If r is an integer zero of $p(x)$, then r must be a divisor of the constant term c_0.

b) If $r = a / b$ is a rational zero of $p(x)$ in which all common factors if a and b have been canceled, then a must be a divisor of the constant term c_0, and b must be a divisor of the leading coefficient c_n. ■

For example in the polynomial above the constant term is 6 (which has divisors ± 1, ± 2, ± 3, and ± 6) and the leading coefficient is 10 (which has divisors ± 1, ± 2, ± 5, and ± 10). Thus, the only possible integer zeros of $p(x)$ are

$$\pm 1,\ \pm 2,\ \pm 3,\ \pm 6$$

and the only possible non integer rational zeros are

$$\pm \frac{1}{2},\ \pm \frac{1}{5},\ \pm \frac{1}{10},\ \pm \frac{2}{5},\ \pm \frac{3}{2},\ \pm \frac{3}{5},\ \pm \frac{3}{10},\ \pm \frac{6}{5}.$$

Using a computer, it is a simple matter to evaluate $p(x)$ at each of the numbers in these lists to show that its only rational zeros are the numbers found.

2.9.7.3 Rational Functions

A function that is expressible as a ratio of two polynomials is called a **rational function**. For example

$$\boldsymbol{\frac{x^5 + 7 x^3 + c x + b}{x^3 - 5}}$$

$$\frac{x^5 + 7 x^3 + c x + b}{x^3 - 5}$$

defines a rational function. In general a rational function if it is expressible in the form

$$f = \frac{a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n}{b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_m x^m}. \tag{2.107}$$

The domain of the function f consists of all x where the denominator differs from zero.

Example 2.68. The Van-der-Waals Equation

For a real gas the pressure p, the volume V, and the Temperature T are related by the Van-der-Waals state equation

$$\left(p + \frac{a}{V^2}\right)(V - b) = R\,T$$

where a, b and R are constants. At constant temperature T the pressure p will be a function of the volume V. The representation of the pressure can be derived by solving the VdW-equation with respect to p

$$\mathbf{Flatten}\left[\mathbf{Solve}\left[(V-b)\left(\frac{a}{V^2}+p\right)=R\,T,\,p\right]\right]$$

$$\left\{p \to \frac{-a\,b + a\,V - R\,T\,V^2}{V^2\,(b-V)}\right\}$$

which represents a rational function in V.▲

2.9.7.4 Power Functions

Power functions are functions which can be represented as special monomials consisting of a single variable with arbitrary exponents. A function $f(x) = x^a$, where a is a constant, is called a **power function**. We consider the most important cases.

When $\boldsymbol{a = n}$ **and** $\boldsymbol{n}$ **is a positive integer** the graph of $f(x) = x^n$ for $n = 1, 2, 3, 4, \ldots$. In this case power functions are equivalent with monomials. The graphs of monomials up to $n = 5$ are shown in the following plots

Out[14]=

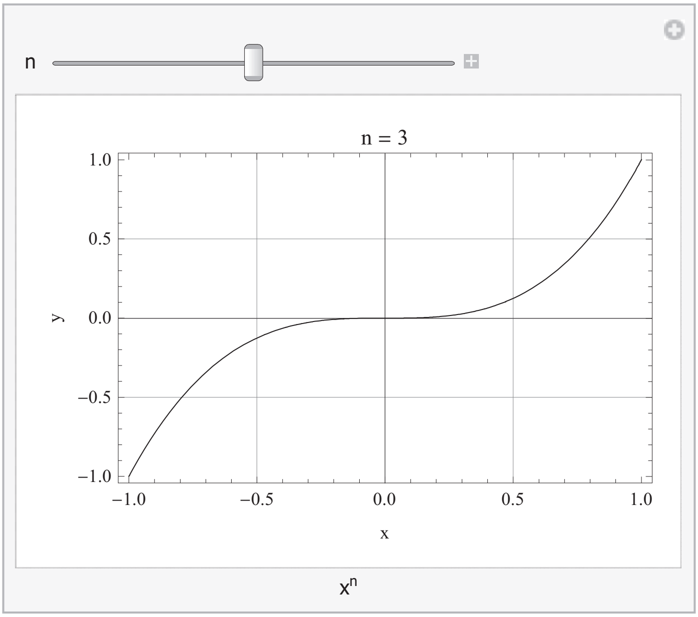


Figure 2.96. Graph of the power function $f(x) = x^n$ with $n = 1, 2, 3, 4, 5$, respectively.

The general shape of the graph of $f(x) = x^n$ depend on whether n is even or odd. If n is even, then $f(x) = x^n$ is an even function (symmetric with respect to the vertical coordinate line). The graph of even functions is similar to the parabola $y = x^2$. If n is odd, then $f(x) = x^n$ is an odd function and its graph is similar to that of $y = x^3$. As shown in Figure 2.97 the graph of even power functions becomes flatter near 0 and steeper when $|x| \geq 1$.

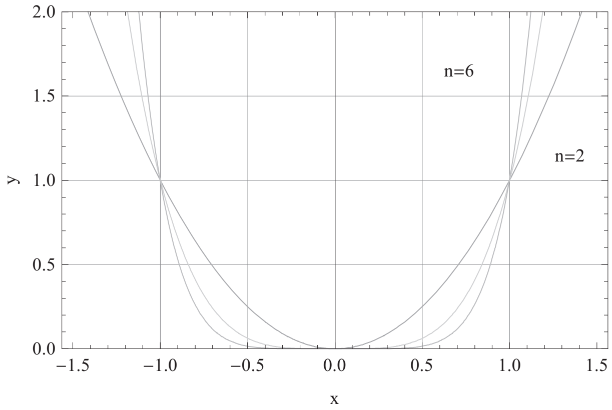


Figure 2.97. Graph of even power function $f(x) = x^{2n}$ with $n = 1, 2, 3$. For increasing n the function becomes flatter near 0 and steeper for $|x| \geq 1$.

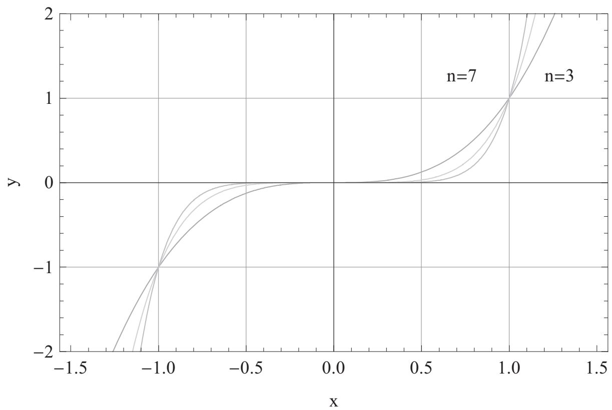


Figure 2.98. Graph of odd power function $f(x) = x^{2n+1}$ with $n = 1, 2, 3$. For increasing n the function becomes flatter near 0 and steeper for $|x| \geq 1$.

When $\boldsymbol{a = 1/n}$ and n is a positive integer then the function $f(x) = x^{1/n} = \sqrt[n]{x}$ is a root function. For $n = 2$ it is a square root function $f(x) = \sqrt{x}$, whose domain is $[0, \infty)$ and whose graph is the upper half of the parabola $x = y^2$. A similar behavior is observed for larger n. The function becomes steeper near $x = 0$ and flatter for $x \geq 1$ for increasing n (see Figure 2.99).

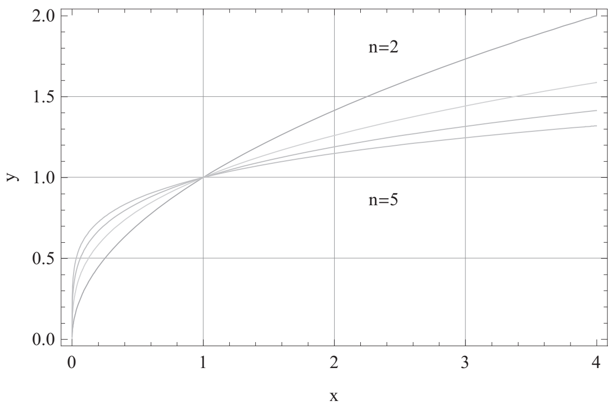


Figure 2.99. Graph of power function $f(x) = x^{1/n}$ with $n = 1, 2, 3$. For increasing n the function becomes steeper near 0 and flatter for $x \geq 1$.

When $\boldsymbol{a = -1/n}$ and n is a positive integer then the function $f(x) = x^{-1/n} = 1 / x^{1/n}$. In the special case when $n = 1$ a hyperbola is generated. The graph $y = 1 / x^{1/n}$ is shown in Figure 2.100 for different values of n.

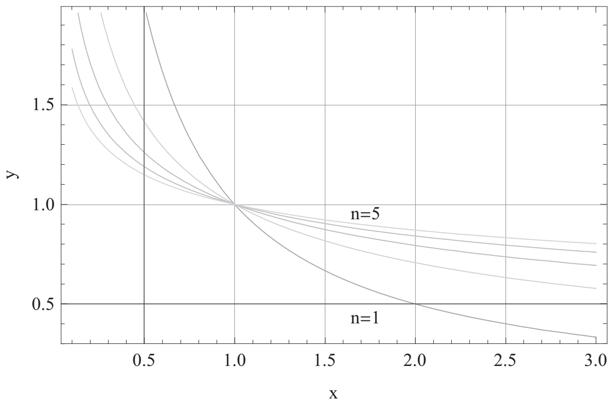


Figure 2.100. Graph of power function $f(x) = 1 / x^{1/n}$ with $n = 1, 2, 3, 4, 5$. For increasing n the function becomes steeper near 0 and flatter for $x \geq 1$. This kind of function is interesting to represent natural growth relations.

2.9.7.5 Explicit Algebraic Functions

An explicit algebraic function is a function that can be evaluated using finitely many additions, subtractions, multiplications, divisions, and root extractions. For example

$$x^{2/3} == \sqrt[3]{x^2} == \left(\sqrt[3]{x}\right)^2;$$

and

$$\frac{(x-3)\sqrt[4]{x^7}}{x^7+\sqrt{x^7+1}}$$

define explicit algebraic functions of x.

2.9.7.6 Implicit Algebraic Functions and Transcendental Functions

The rest of the function can be divided into two categories, implicit algebraic functions and transcendental functions.

Transcendental functions are not algebraic. The set of transcendental functions includes the trigonometric, exponential, and logarithmic functions, but it also includes the hyperbolic functions. These functions are called transcendental because they cannot be defined directly by algebraic formulas. In other words, a function which "transcends," i.e., cannot be expressed in terms of, algebra. This means that the only way to work with these functions is to learn to use their algebraic and geometric properties. The collection of all explicit and transcendental functions are so called explicit elementary functions. All other functions that have never been named are non elementary functions. An overview on the different classes of functions is given in Figure 2.106 below.

2.9.7.7 Trigonometric Functions

In calculus the convention is that radian measure is always used. For example, when we use the function $f(x) = \sin(x)$ it is understood that $\sin(x)$ means the sine of the angle whose radian measure is x. Thus the graphs of the sine and cosine functions are as shown in Figure 2.101.

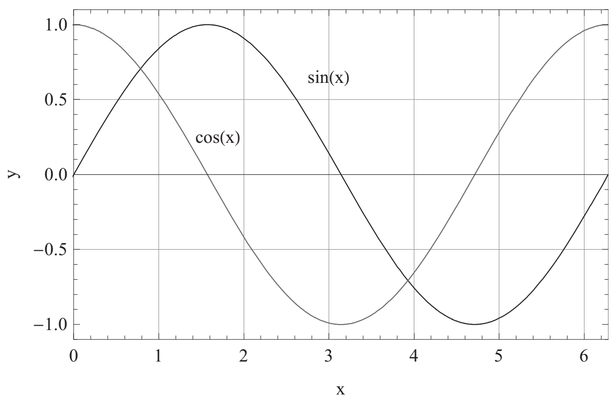


Figure 2.101. Graph of trigonometric function $f(x) = \sin(x)$ and $f(x) = \cos(x)$. The functions shows an oscillating behavior with period 2π.

The periodic nature of these functions makes them suitable for modeling repetitive phenomena such as tides, stock fluctuations, market up and downs, and price fluctuations.

The tangent function is related to the sine and cosine functions by the equation

$$\tan(x) = \frac{\sin(x)}{\cos(x)}.$$

Its graph is shown in Figure 2.102.

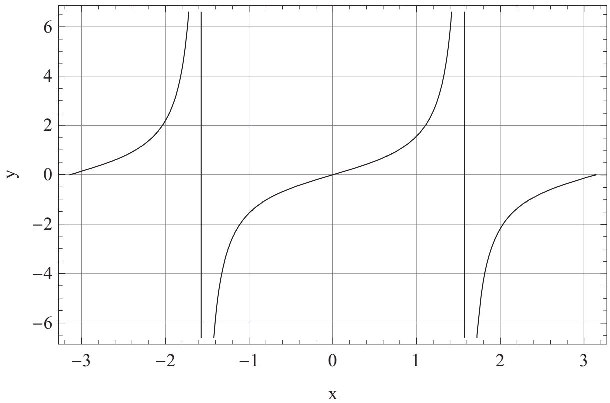


Figure 2.102. Graph of power function $f(x) = \tan(x)$. The function shows a periodic behavior with period 2π.

2.9.7.8 Exponential Functions

The exponential functions are the functions of the form $f(x) = a^x$, where the base a is a positive constant. The graph of different functions of this kind is shown in Figure 2.103.

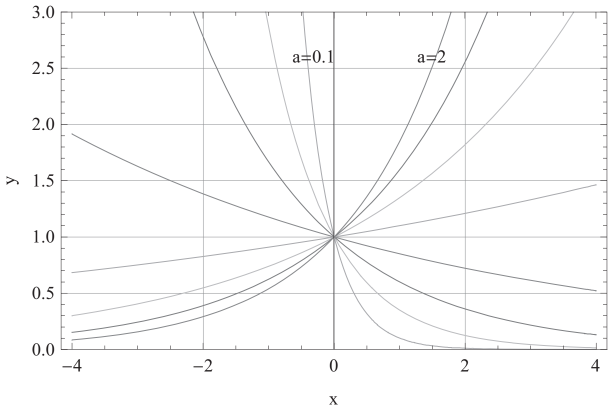


Figure 2.103. Graph of exponential function $f(x) = a^x$ with $1 \leq a \leq 2$.

Exponential functions are useful for modelling many natural and engineering problems, such as population growth (if $a > 1$) and decays (if $a < 1$).

A distinguished exponential function is the function with $a = e$ where $e \approx 2.71828$ is a natural constant which is called in honor of the Swiss mathematician Leonhard Euler, Euler's number. This constant arises as the horizontal asymptote of the graph of the equation

$$f(\text{x\_}) := \left(1 + \frac{1}{x}\right)^x$$

A graph of this function is given in Figure 2.104.

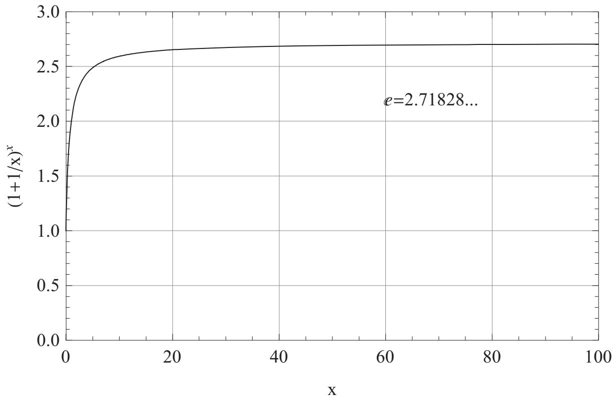


Figure 2.104. Representation of Euler's number e as an asymptotic number for the function $f(x) = (1 + 1/x)^x$.

The exponential function itself is defined by

$$f(x\_) := e^x$$

which can be represented by an infinite number of terms in a polynomial like representation as

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720} + \ldots$$

This polynomial can be written in a compact way by a sigma notation as

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

True

From which Euler's number follows again by setting $x = 1$ in the above relation

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

True

The logarithmic function $f(x) = \log_a(x)$ where the basis a is a positive constant, is the inverse function of an exponential function. Figure 2.105 shows the graph of logarithmic functions with various bases. In each case, the domain is $(0, \infty)$, the range is $(-\infty, \infty)$, and the function increases slowly when $x > 1$.

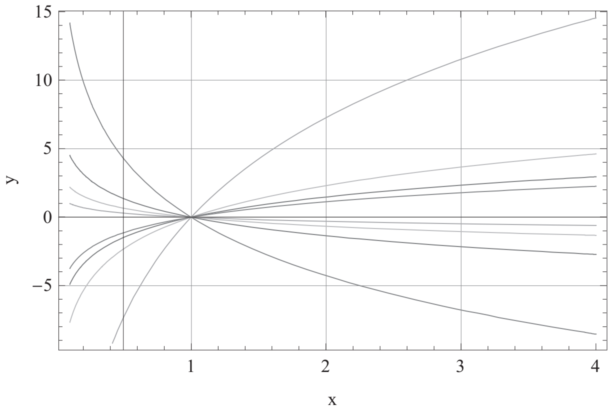


Figure 2.105. Graph of logarithmic function $f(x) = \log_a x$ with $1 \leq a \leq 2$.

Example 2.69. Estimation of Energy Release in Earthquakes

On the Richter scale, the magnitude M of an earthquake is related to the released energy E in joules (J) by the equation

$$E = 25\,084.4\, e^{3.45\, M}.$$

Find the released E in the 1906 San Francisco earthquake that registered $M = 8.2$ on the Richter scale and the 1992 Cairo earthquake with $M = 5.8$. Compare the energies released in the two earthquakes.

Solution 2.69 The application of the formula given above to the San Francisco earthquake in 1906 is

$$\mathbf{EnergySanFrancisco = 25\,084.4\, e^{3.45 \times 8.2}}$$

$$4.84836 \times 10^{16}$$

This means that in this earthquake approximately 5×10^{10} MJ were released. This corresponds to nearly 600 times more energy than in the first atomic explosion which released about 8.2×10^{13} J.

The 1992 Cairo earthquake was registered as a small incident with $M = 5.8$. The energy released at that time was

$$\mathbf{EnergyCairo = 25\,084.4\, e^{3.45 \times 5.8}}$$

$$1.22924 \times 10^{13}$$

$$\mathbf{\frac{EnergySanFrancisco}{EnergyCairo}}$$

```
3944.19
```

The energy release in the Cairo earthquake was about 4000 times weaker than the San Francisco earthquake.▲

2.9.7.9 Summary of Elementary functions

The following Figure 2.106 collects all the information on elementary functions. The idea of this classification is that the use of the basic algebraic operations like addition $(+)$, subtraction $(-)$, multiplication $(*)$, division $(/)$, and composition $(\circ)$ is used to combine and operate on this functions. The tree of functions starts with the most simplest function, the constant function, and continues up to explicit elementary functions.

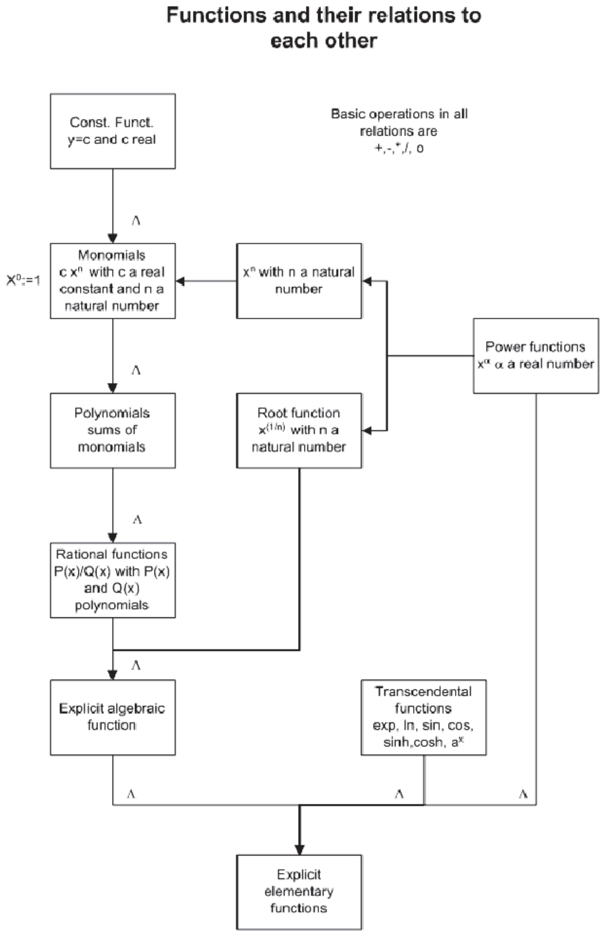


Figure 2.106. Overview of different classes of functions which are collected in the class of explicit elementary functions.

2.9.8 Tests and Exercises

The following tests and exercises are related to the different types of functions.

2.9.8.1 Test Problems

T1. What are the important types of functions frequently encountered in calculus? Give an example of each type.

T2. In terms of its graph, what is meant by an increasing function? A decreasing function? Give an example of each.

T3. How do you change the equation $y = f(x)$ to compress or stretch the graph by $k > 0$? Reflect the graph across a coordinate axis?

T4. When is it possible to compose one function with another? Give examples of composites and their values at various points. Does the order in which functions are composed ever matter?

2.9.8.2 Exercises

E1. Find the domain and sketch the graph of the function.

- **a.** $f(x) = 2 - \frac{x^2}{2-x}$,
- **b.** $f(x) = \frac{2-x^2}{2+x^3}$,
- **c.** $f(x) = \frac{|x|}{x^3}$,
- **d.** $f(x) = \frac{5x+|x|}{x}$,
- **e.** $f(x) = x^{-2} + \frac{x}{4-x^2}$,
- **f.** $f(x) = e^{-x^2}$.

E2. Find the domain of the function.

- **a.** $f(x) = \frac{x}{2x-2}$,
- **b.** $f(x) = \sqrt{x} + \sqrt[3]{x}$,
- **c.** $f(x) = \frac{3x+2}{x^2+3x+2}$,
- **d.** $f(x) = \sqrt{5x} - \sqrt{2-x}$,
- **e.** $f(x) = \frac{e^x+e^{-x}}{2}$,
- **f.** $f(x) = \frac{\ln(x)-e^{-x}}{e^{-x^2}}$.

E3. Classify each function as a power function, root function, polynomial with its degree, rational function, algebraic function, trigonometric function, exponential function, or logarithmic function.

- **a.** $f(x) = x^6$,
- **b.** $f(x) = 2^x$,
- **c.** $f(x) = x^2 + 4x - 3x^4$,
- **d.** $f(x) = \log_{10}(x)$,
- **e.** $f(x) = \sin(x) - \cos(x)$,
- **f.** $f(x) = \frac{x^2-2}{x^3+x-1}$.

E4. Graph $y = \tan(x)$ and $y = \cot(x)$ together for $-7 \leq x \leq 7$. Comment on the behavior of $\cot(x)$ in relation to the signs and values of $\tan(x)$.

E5. Graph the following functions. What is the period of each function?

- **a.** $f(x) = \sin(2x)$,
- **b.** $f(x) = \sin(x/2)$,
- **c.** $f(x) = \cos(\pi x)$,
- **d.** $f(x) = \cos(\pi x/2)$,
- **e.** $f(x) = \sin(x - \pi/2)$,
- **f.** $f(x) = \cos(x + \pi/4)$.

E6. Derive the tangent sum formula. The standard formula for the tangent of the sum of two angles is

$$\tan(x+y) = \frac{\tan(x) + \tan(y)}{1 - \tan(x)\tan(y)}. \tag{1}$$

Derive also a formula for $\tan(x - y)$.

E7. Use a computer algebra system (CAS) to examine the behavior of the following general sin function

$$f(x) = \alpha \sin\left(\frac{2\pi}{\beta}(x - \delta)\right) + \gamma. \tag{2}$$

Change the parameters α, β, δ, and γ in the following ranges $\alpha \in [1, 3]$, $\beta \in [1, 4]$, $\delta \in [0, 2\pi]$, and $\gamma \in [0, 3]$.

E8. The curves with equations

$$f(x) = \frac{|x|}{\sqrt{c - x^2}} \tag{3}$$

are called **bullet-nose curves**. Graph some of these curves to see why. What happens as c increases?

E9. Graph the given functions on a common screen. How are these graphs related?

a. $y = 2^x$, $y = e^x$, $y = 5^x$, $y = 15^x$,

b. $y = e^x$, $y = e^{-x}$, $y = 7^x$, $y = 7^{-x}$,

c. $y = 3^x$, $y = 10^x$, $y = \left(\frac{1}{3}\right)^x$, $y = \left(\frac{1}{10}\right)^x$,

d. $y = 0.9^x$, $y = 0.7^x$, $y = 0.3^x$, $y = 0.15^x$.

E10 Find the domain of each function and graph the function.

a. $f(x) = \frac{1}{1+e^x}$,

b. $f(x) = \frac{1}{1-e^x}$,

c. $f(x) = \sin(e^{-x})$,

d. $f(x) = \sqrt{1 - 2^{t-1}}$.

3

Limits, Continuity and Derivatives

3.1 Introduction to Single Variable Calculus

The development of calculus in the seventeenth century by Newton and Leibniz provided scientists with their first real understanding of what is meant by an instantaneous rate of change such as velocity and acceleration. Once the idea was understood conceptually, efficient computational methods followed, and science took a quantum leap forward. The fundamental building block on which rates of change rest is the concept of a limit, an idea that is so important that all other calculus concepts are now based on it.

In this chapter we will examine different properties such as limits, continuity, and derivatives. Limits will show us how functions behave if we approach a specific value in the domain of the function. Limits are also useful to judge the slope of a function in such a case when a point in the domain is fixed and another point approaches this distinguished point. Continuity of functions will provide a tool to measure the smoothness of the function. We will develop the concept of a limit in stages, proceeding from an informal, intuitive notion to a precise mathematical definition. We will also develop theorems and procedures for calculating limits.

3.2 The Limit of a Function

The following story tells you on the importance of limits. When the Old Kingdom in Egypt was formed, there were initially 32 000 tonnes of grain in the state granary. Each year half of the existing stock of grain was consumed and another 8000 tonnes of grain were produced. The question at that time was how long will be grain available to survive. Another question was how many tonnes of grain did the state granary contain after twelve years.

We could solve this problem by calculating y_1, y_2, y_3 and so on, up to y_{12}, but a better way is to find a general formula for y_t, the amount of grain in the granary after t years. This approach has the advantage that if we then asked for the contents of the granary after, 20 years, we need not work out

y_{13}, y_{14} and so on. We need only substitute $t = 20$ in the general formula for y_t. The initial condition for the granary problem is $y_0 = 32\,000$, and the recurrence equation for y_t when $t \geq 1$ is

y(t_) := 0.5 y(t − 1) + 8000

y(0) = 32 000

```
32 000
```

The sequence of grain stock can be calculated by starting with y_0 and inserting this value in the first equation deriving the next result for the current time step. The following table shows how the numbers are generated

year	amount
1	24 000.
2	20 000.
3	18 000.
4	17 000.
5	16 500.
6	16 250.
7	16 125.
8	16 062.5
9	16 031.3
10	16 015.6
11	16 007.8
12	16 003.9

A representation of the data are give in Figure 3.1.

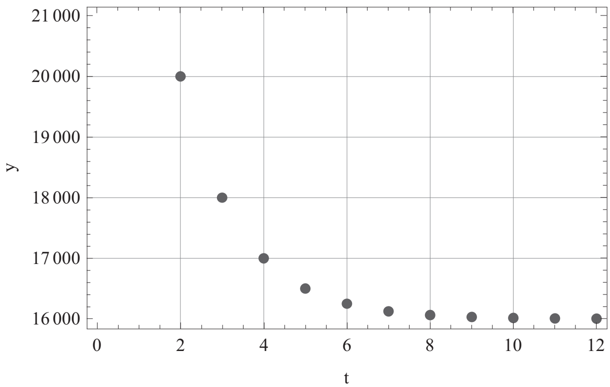


Figure 3.1. Decrease of grain stock in the Old Kingdom granary.

This example shows how a sequence of figures approaches a certain value and stays at this value. Figure 3.1 also shows how a limiting value is approached from above.

The general solution for the stock in the granary is given by

$$y(t\_) := 16\,000\,(0.5^t + 1)$$

Since the factor 0.5^t vanishes for $t \to +\infty$, we find that the limiting value for $t \to +\infty$ is

$$\lim_{t\to\infty} y(t)$$

```
16 000.
```

Thus the Old Kingdom had a chance to survive.

3.2.1 Standard Situation

The most basic use of limits is to describe how a function behaves as the independent variable approaches a given value. For example, let us examine the behavior of the function

$$f(x\_) := x^2 - x + 1$$

for x-values closer and closer to 2. It is evident from the graph in Figure 3.2 that the values of $f(x)$ get closer and closer to 3 as values of x are selected closer and closer to 2 on either the left or right side of 2. We describe this by saying that the limit of $x^2 - x + 1$ is 3 as x approaches 2 from either side, and we write

$$\lim_{x\to 2}\left(x^2 - x + 1\right) = 3 \tag{3.1}$$

Observe that in our investigation of the limit we are only concerned with the value of $f(x)$ near $x = 2$ and not the value of $f(x)$ at $x = 2$.

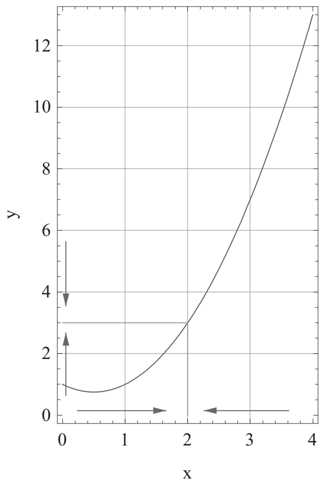


Figure 3.2. Graph of the function $f(x) = x^2 - x + 1$.

This leads us to the following general idea. If the value of $f(x)$ can be made as close as we like to L by taking values of x sufficiently close to a (but not equal to a), then we write

$$\lim_{x\to a} f(x) = L \tag{3.2}$$

which is read the limit of $f(x)$ as x approaches a is L.

Equation (3.2) is also commonly written as

$$f(x) \to L \qquad \text{as} \qquad x \to a. \tag{3.3}$$

With this notation we can express (3.1) as

$$x^2 - x + 1 \to 3 \qquad \text{as} \qquad x \to 2. \tag{3.4}$$

In order to investigate $\lim_{x\to a} f(x)$ we ask ourselves the question: If x is close to, but different from, a, is there a particular number to which f(x) is close? This question presumes that the function f is defined everywhere near a, in other words, that f is defined at all points x in some open interval containing a, except possibly at $x = a$. The value of f at a, if it exists at all, is not relevant to the determination of $\lim_{x\to a} f(x)$. Many important applications of the limit concept involve constants in which the domain of the function excludes a.

The process of determining a limit generally includes a discovery phase, followed by a verification phase. The discovery phase begins with sampled x-values, and ends with a conjecture for the limit.The following tables represents the discovery phase of a limit calculation. First the approach from the left toward 2 gives

x	f(x)
1	1
1.5	1.75
1.8	2.44
1.9	2.71
1.95	2.8525
1.99	2.9701
1.995	2.98503
1.999	2.997

The approach from the right gives

x	f(x)
2.0001	3.0003
2.001	3.003
2.01	3.0301
2.1	3.31
2.2	3.64
2.3	3.99
2.5	4.75
2.8	6.04

The sampled values show that for x near to 2 the corresponding values of $f(x)$ are close to 3. Our conjecture is that $\lim_{x\to 2} x^2 - x + 1 = 3$. We can only sample a relatively few values of x near 2. We cannot sample all values of x near 2, for no matter how close to 2 we take a x-value, there are infinitely many values of x near yet to 2.

Example 3.1. Conjecture and Verification

Make a conjecture about the value of the limit

$$\lim_{x \to 0} \frac{\sin(x)}{x}. \tag{3.5}$$

Solution 3.1. The function $f(x) = \frac{\sin(x)}{x}$ is not defined at $x = 0$, but, as discussed previously this has no bearing on the limit. The plot of the function shows that $f(x)$ approaches 1 if $x \to 0$.

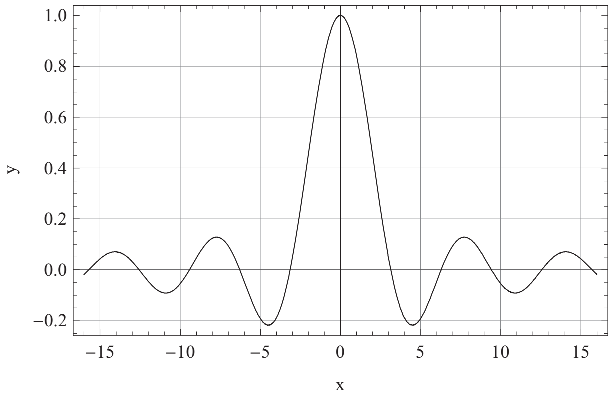


Figure 3.3. Graph of the function $f(x) = \frac{\sin(x)}{x}$.

The graph suggests that

$\lim_{x \to 0} \frac{\sin(x)}{x} = 1$. Later in this chapter we will give a geometric argument that this conjecture is correct.▲

3.2.2 Sampling Problems

Although numerical and graphical evidence is helpful for guessing at limits, we can be mislead by an insufficient or poorly selected sample. For example, the following table for

$$\lim_{x \to 0} \sin\left(\frac{\pi}{x}\right) = 0? \tag{3.6}$$

shows

x	sin(π/x)
-3.14159	-0.841471
-0.1	0
-0.01	0
-0.0001	0
-0.00001	0
0.00001	0
0.0001	0
0.001	0
0.01	0
0.1	0
3.14159	0.841471

However the graph of the function shows the following

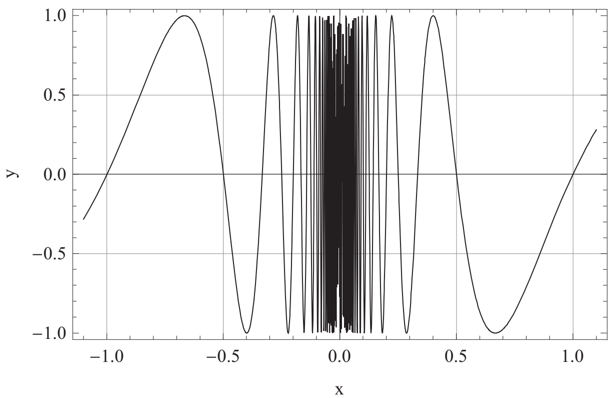


Figure 3.4. Graph of the function $f(x) = \sin\left(\frac{\pi}{x}\right)$.

The plot demonstrates that the values are by far not zero. The graph indicates that as $x \to 0$, the values of f oscillates between -1 and 1 with increasing rapidity, and hence do not approach a limit.

Thus numerical evidence can led to incorrect conclusions about limits because of roundoff errors or because the sample of values used is not extensive enough to give a good indication of the behavior of the function. Thus when a limit is conjectured from a table of values, it is important to look corroborating evidence to support the conjecture.

3.2.3 One-Sided Limits

The limit in (3.1) is commonly called a two-sided limit because it requires the values of $f(x)$ to get closer and closer to L as values of x are taken from either side of $x = a$. However, some functions exhibit different behavior on the two sides of an x-value a, in which case it is necessary to distinguish whether values of x near a are on the left side or on the right side of a for purposes of investigating limiting behavior. For example, consider the function

$$f(\mathbf{x\_}) := \frac{|x|}{x}$$

Figure 3.5 shows the graph of this function.

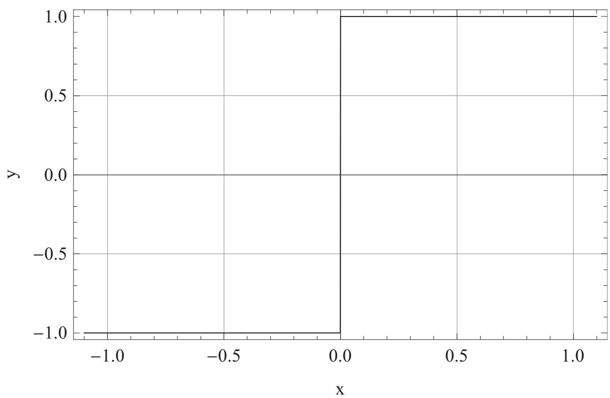


Figure 3.5. Graph of the function $f(x) = |x| / x$.

Note that x-values approaching 0 and to the right of 0 produces $f(x)$ values that approach 1 (in fact, they are exactly 1 for all such values of x). On the other hand x-values approaching 0 and to the left of 0 produce $f(x)$ values that approach -1. We describe these two statements by saying that the limit of $f(x) = |x| / x$ is 1 as x approaches 0 from the right and that the limit of $f(x) = |x| / x$ is -1 as x approaches 0 from the left. We denote these limits by writing

$$\lim_{x \to 0^+} \frac{|x|}{x} = 1 \qquad \text{and} \qquad \lim_{x \to 0^-} \frac{|x|}{x} = -1. \tag{3.7}$$

With this notation the superscript + indicates a limit from the right and a superscript - indicates a limit from the left.

This leads to the following general idea:

If the value of $f(x)$ can be made as close as we like to L by taking the value of x sufficiently close to a (but greater than a), then we write

$$\lim_{x \to a^+} f(x) = L \tag{3.8}$$

which is read the limit of $f(x)$ as x approaches a from the right is L. Similarly, if the values of $f(x)$ can be made as close as we like to L by taking values of x sufficiently close to a (but less than a), then we write

$$\lim_{x \to a^-} f(x) = L \tag{3.9}$$

which read the limit of $f(x)$ as x approaches a from the left is L.

Equations (3.8) and (3.9), which are allowed one-sided limits, are also commonly written as

$$f(x) \to L \qquad \text{as} \qquad x \to a^+ \tag{3.10}$$

and

$$f(x) \to L \qquad \text{as} \qquad x \to a^-, \tag{3.11}$$

respectively. With this notation (3.7) can be expressed as

$$\frac{|x|}{x} \to 1 \qquad \text{as} \qquad x \to 0^+ \tag{3.12}$$

and

$$\frac{|x|}{x} \to -1 \qquad \text{as} \qquad x \to 0^-. \tag{3.13}$$

In general there is no guarantee that a function will have a limit at a specified location. If the values of $f(x)$ do not get closer and closer to some single number L as $x \to a$, then we say that the limit of $f(x)$ as x approaches a **does not exist** (and similarly for one-sided limits). For example, the two-sided limit $\lim_{x\to 0} |x|/x$ does not exist because the values of $f(x)$ do not approach a single number as $x \to 0$; the values approach -1 from the left and 1 from the right.

In general, the following condition must be satisfied for the two-sided limit of a function to exist.

Theorem 3.1. *Relationship between one-sided and two-sided limits*

The two-sided limit of a function $f(x)$ exists at a if and only if both of the one-sided limits exist at a and have the same value; that is,

$lim_{x\to a} f(x) = L$ *if and only if* $lim_{x\to a^+} f(x) = L = lim_{x\to a^-} f(x)$.■

Remark 3.1. Sometimes, one or both of a one-sided limits may fail to exist (which, in turn, implies that the two-sided limit does not exist). For example, we saw earlier that the one-sided limits of $f(x) = \sin(\pi/x)$ does not exist as x approaches 0 because the function keeps oscillating between -1 and 1, failing to settle on a single value. This implies that the two-sided limit does not exist as x approaches 0.

Example 3.2. One-Sided Limit

For the functions in Figure 3.6, find the one-sided and two-sided limits at $x = 1/2$ if they exist.

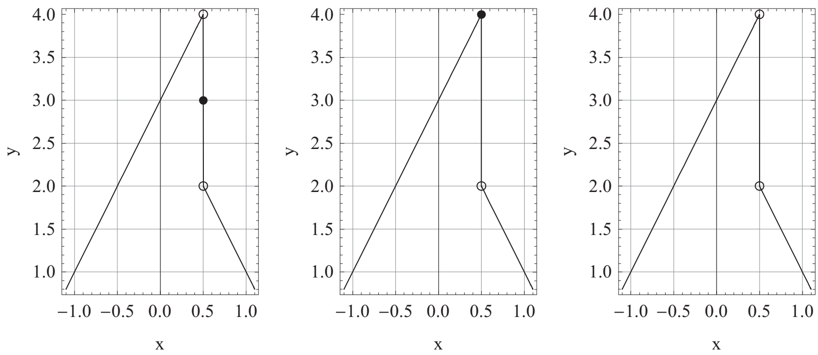


Figure 3.6. Three different examples for a one-sided limit for the function $f(x)$.

Solution 3.2. The functions in all three figures have the same one-sided limits as $x \to \frac{1}{2}$, since the functions are identical except at $x = 1/2$. These limits are

$\lim_{x\to 1/2^+} f(x) = 2.$ and $\lim_{x\to 1/2^-} f(x) = 4.$

In all three cases the two-sided limit does not exist as $x \to 1/2$ because the one-sided limits are not equal.▲

Example 3.3. Two-Sided Limit

For the functions in Figure 3.7 find the one-sided and two-sided limits at $x = 1/2$ if they exist.

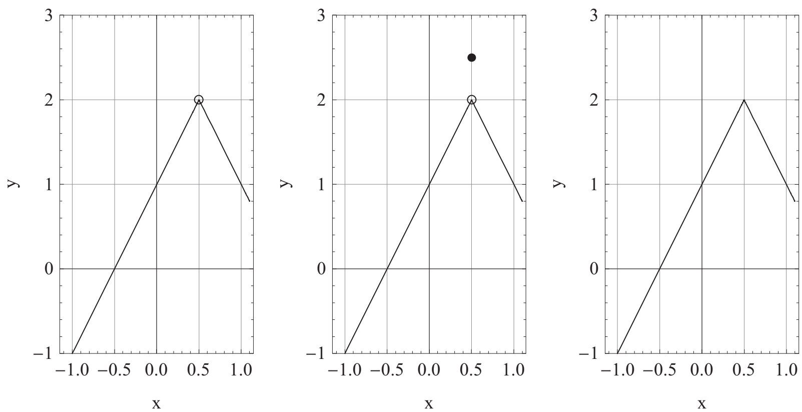


Figure 3.7. Three different examples for a two-sided limit for the function $f(x)$.

Solution 3.3. As in the preceding example, the value of f at $x = a$ has no bearing on the limits as $x \to 1/2$, so that in all three cases we have

$$\lim_{x\to 1/2^+} f(x) = 2. \qquad \text{and} \qquad \lim_{x\to 1/2^-} f(x) = 2.$$

Since the one-sided limits are equal, the two-sided limit exists and

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

3.2.4 Asymptotes and Infinite Limits

Sometimes one-sided or two-sided limits will fail to exist because the values of the function increases or decreases infinitely. For example, consider the behavior of the function $f(x) = 1/x$ for values of x near 0. It is evident from the graph in Figure 3.8 that the x-values are taken closer and closer to 0 from the right, the value of $f(x) = 1/x$ are positive and increase infinitely; and as x-values are taken close to 0 from the left, the values of $f(x) = 1/x$ are negative and decrease infinitely. We describe these limiting behaviors by writing

$$\lim_{x\to 0^+} \frac{1}{x} = +\infty \qquad \text{and} \qquad \lim_{x\to 0^-} \frac{1}{x} = -\infty. \tag{3.14}$$

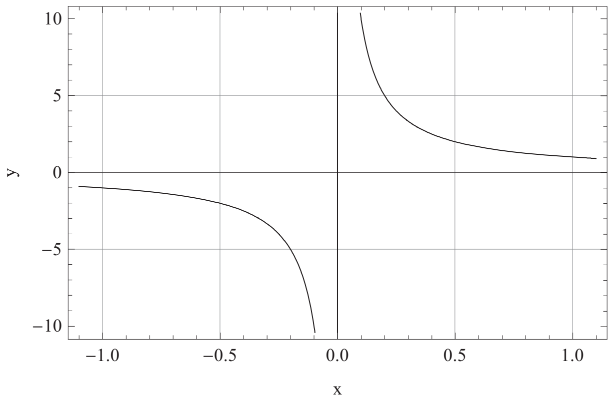


Figure 3.8. Behavior of the function $f(x) = 1/x$ near the origin $x = 0$.

Theorem 3.2. *Infinite Limits*

If the values of $f(x)$ increase infinitely as x approaches a from the right or left, then we write

$$\lim_{x\to a^+} f(x) = +\infty \qquad \text{or} \qquad \lim_{x\to a^-} f(x) = +\infty$$

as appropriate, and we say that $f(x)$ increases without bound, or $f(x)$ approaches $+\infty$, as $x \to a^+$ or as $x \to a^-$. Similarly, if the values of $f(x)$ decreases infinitely as x approaches a from the right or left, then we write

$$\lim_{x\to a^+} f(x) = -\infty \qquad \text{or} \qquad \lim_{x\to a^-} f(x) = -\infty$$

as appropriate, and say that $f(x)$ decreases without bound, or $f(x)$ approaches $-\infty$, as $x \to a^+$ or as $x \to a^-$. Moreover, if both one-sided limits are $+\infty$, the we write

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

and if both one-sided limits are $-\infty$, then we write

$lim_{x \to a} f(x) = -\infty$.■

Remark 3.2. It should be emphasized that the symbol $+\infty$ and $-\infty$ are not real numbers. The phrase $f(x)$ approaches $+\infty$ is akin to saying that $f(x)$ approaches the unapproachable; it is a colloquialism for $f(x)$ increases without bound. The symbols $+\infty$ and $-\infty$ are used here to encapsulate a particular way in which limits fail to exist. To say, for example, that $f(x) \to +\infty$ as $x \to a^+$ is to indicate that $\lim_{x \to a^+} f(x)$ dos not exist and that it fails to exist because the values of $f(x)$ increases without bound as x approaches a from the right. Furthermore, since $+\infty$ and $-\infty$ are not numbers, it is inappropriate to manipulate these symbols using rules of algebra. For example, it is not correct to equate $(+\infty) - (+\infty)$ with 0.

Example 3.4. Infinite Limit

Determining the limiting value for the functions in Figure 3.9

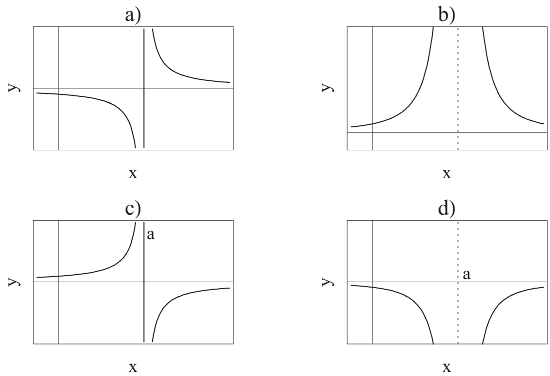


Figure 3.9. Four different examples showing infinite limits for the following functions $f(x) = 1/(x-a)$, $f(x) = 1/(x-a)^2$, $f(x) = -1/(x-a)$, $f(x) = -1/(x-a)^2$ which corresponds to the cases a), b), c), and d), respectively.

Solution 3.4. In Figure 3.9a the function infinitely as x approaches a from the right and decreases infinitely as x approaches a from the left. Thus

$$\lim_{x \to a^+} \frac{1}{x-a} = +\infty \qquad \text{and} \qquad \lim_{x \to a^-} \frac{1}{x-a} = -\infty.$$

In Figure 3.9b the function increases infinitely as x approaches a from the right and increases infinitely as x approaches a from the left. Thus,

$$\lim_{x \to a} \frac{1}{(x-a)^2} = \lim_{x \to a^+} \frac{1}{(x-a)^2} = \lim_{x \to a^-} \frac{1}{(x-a)^2} = +\infty.$$

In Figure 3.9c the function decreases indefinitely as x approaches a from the right and increases indefinitely as x approaches a from the left. Thus,

$$\lim_{x\to a^+} \frac{-1}{x-a} = -\infty \qquad \text{and} \qquad \lim_{x\to a^-} \frac{-1}{x-a} = +\infty.$$

In Figure 3.9d the function decreases indefinitely as x approaches a from both the left and the right. Thus

$$\lim_{x\to a} \frac{-1}{(x-a)^2} = \lim_{x\to a^+} \frac{-1}{(x-a)^2} = \lim_{x\to a^-} \frac{-1}{(x-a)^2} = -\infty.\blacktriangle$$

Geometrically if $f(x) \to +\infty$ as $x \to a^-$ or as $x \to a^+$, then the graph of $y = f(x)$ rises without bound and squeezes closer to the vertical line $x = a$ on the indicated side of $x = a$. If $f(x) \to -\infty$ as $x \to a^-$ or as $x \to a^+$, then the graph of $y = f(x)$ fall without bound and squeezes closer to the vertical line $x = a$ on the indicated side of $x = a$. In these cases, we call the line $x = a$ a vertical asymptote. Asymptote comes from the Greek asymptotos, meaning non-intersecting. We will see soon that taking "asymptote" to be synonymous with "non-intersecting" is a bit misleading.

Definition 3.1. *Vertical Asymptote*

A line $x = a$ is called a vertical asymptote of the graph of a function f if $f(x) \to +\infty$ or $f(x) \to -\infty$ as x approaches a from the left or right. $\blacksquare$

Example 3.5. Asymptote

The four functions shown in Figure 3.9 all have a vertical asymptote at $x = a$, which is indicated in the figure by a full or dashed line at $x = a$.$\blacktriangle$

3.2.5 Horizontal Asymptotes and Limits at Infinity

Thus far, we have used limits to describe the behavior of $f(x)$ as x approaches a. However, sometimes we will not be concerned with the behavior of $f(x)$ near a specific x-value, but rather with how the values of $f(x)$ behaves as x increases without bound or decreases without bound. This is sometimes called the end behavior of the function because it describes how the function behaves for values of x that are far from the origin. For example, it is evident from Figure 3.10 that as x increases without bound, the values of $f(x) = 1/x$ are positive, but get closer and closer to 0; and as x decreases without bound, the values of $f(x) = 1/x$ are negative, and also get closer to 0. We indicate these limiting behaviors by writing

$$\lim_{x\to+\infty} \frac{1}{x} = 0 \qquad \text{and} \qquad \lim_{x\to-\infty} \frac{1}{x} = 0. \tag{3.15}$$

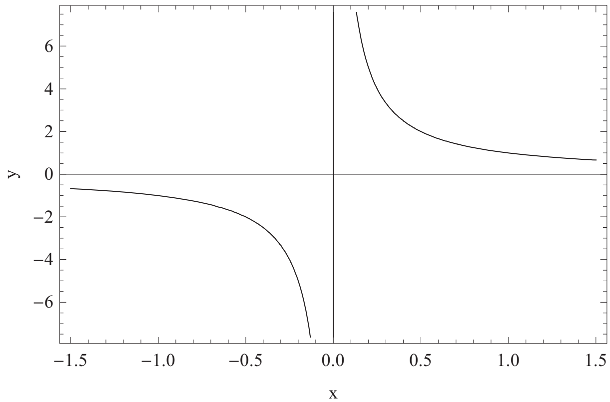


Figure 3.10. Asymptotic behavior of the function $f(x) = 1/x$.

Theorem 3.3. *Limits at Infinity*

If the values of $f(x)$ eventually get closer and closer to the number L as x increases without bound, then we write

$\lim_{x\to+\infty} f(x) = L$ *or* $f(x) \to L$ *as* $x \to +\infty$.

Similarly, if the values of $f(x)$ eventually get closer and closer to the number L as x decreases without bound, then we write

$\lim_{x\to-\infty} f(x) = L$ *or* $f(x) \to L$ *as* $x \to -\infty$.■

Geometrically, if $f(x) \to L$ as $x \to +\infty$, then the graph of $y = f(x)$ eventually gets closer and closer to the line $y = L$ as the graph is traversed in the positive direction (Figure 3.11); and if $f(x) \to L$ as $x \to -\infty$, then the graph of $y = f(x)$ eventually gets closer and closer to the line $y = L$ as the graph is traversed in the negative x-direction (Figure 3.11). In either case we call the line $y = L$ a horizontal asymptote of the graph of f. For example the function in Figure 3.10 has $y = 0$ as a horizontal asymptote.

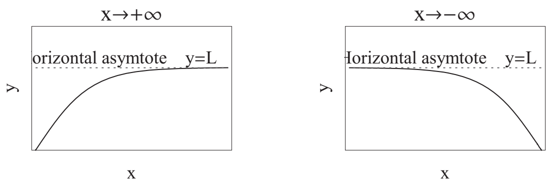


Figure 3.11. Horizontal asymptotes.

Definition 3.2. *Horizontal Asymptote*

A line $y = L$ is called a horizontal asymptote of the graph of a function f if

$lim_{x \to +\infty} f(x) = L$ *or* $lim_{x \to -\infty} f(x) = L$.■

Sometimes the existence of a horizontal asymptote of a function f will be readily apparent from the formula for f. For example, it is evident that the function

$$f(x) = \frac{3x+1}{x} = 3 + \frac{1}{x}$$

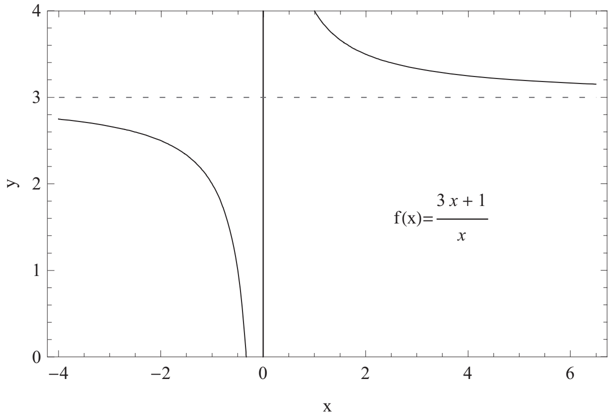


Figure 3.12. Horizontal asymptotes for the function $f(x) = \frac{3x+1}{x}$.

has a horizontal asymptote at $y = 3$ (Figure 3.12), since the value of $1/x$ approaches 0 as $x \to +\infty$ or as $x \to -\infty$. More complicated functions, algebraic manipulations, or special techniques we will study in the next section. They may have to be applied to confirm the existence of horizontal asymptotes.

3.2.6 Limits do not Exist at Infinity

Limits at infinity can fail to exist for various reasons. One possibility is that the value of $f(x)$ may increase or decrease without bound as $x \to +\infty$ or as $x \to -\infty$. For example, the values of $f(x) = x^3$ increases without bound as $x \to +\infty$ and decreases without bound as $x \to -\infty$; and for $f(x) = -x^3$ the values decrease without bound as $x \to +\infty$ and increases without bound as $x \to -\infty$ (Figure 3.13). We denote this by writing

$$\lim_{x \to +\infty} x^3 = +\infty, \qquad \lim_{x \to -\infty} x^3 = -\infty \tag{3.16}$$

$$\lim_{x \to +\infty} \left(-x^3\right) = -\infty, \qquad \lim_{x \to -\infty} \left(-x^3\right) = +\infty. \tag{3.17}$$

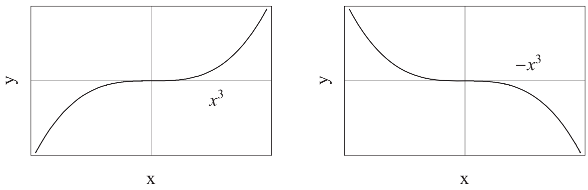


Figure 3.13. Infinite limits for the tow functions $f(x) = x^3$ and $f(x) = -x^3$.

Theorem 3.4. *Infinite Limit at Infinity*

If the value of $f(x)$ increases without bound as $x \to +\infty$ or as $x \to -\infty$, then we write

$lim_{x \to +\infty} f(x) = +\infty$ *or* $\quad lim_{x \to -\infty} f(x) = +\infty$

as appropriate; and if the values of $f(x)$ decrease without bound as $x \to +\infty$ or as $x \to -\infty$ then we write

$lim_{x \to +\infty} f(x) = -\infty$ *or* $\quad lim_{x \to -\infty} f(x) = -\infty.$ ■

Limits at infinity can also fail to exist because the graph of the function oscillates indefinitely in such a way that the values of the function do not approach a fixed number and do not increase or decrease without bound; the trigonometric functions $\sin(x)$ and $\cos(x)$ have this property, for example (Figure 2.101). In such a case we say that the limit fails to exist because of oscillation.

3.2.7 Tests and Exercises

The following two subsections serve to test your understanding of the material discussed in the last sections. Work first on the test problems and then try to solve the exercises. The material serves to make you familiar with limits and their use.

3.2.7.1 Test Problems

T1. What is an informal or intuitive definition of the limit

$$\lim_{x \to x_0} f(x) = L \tag{1}$$

Why is the definition "informal"? Give examples.

T2. What function behaviors might occur for which the limit may fail to exist? Give examples.

T3. What theorems are available for calculating limits? Give examples of how the theorems are used.

T4. Does the existence and value of the limit of a function $f(x)$ as x approaches x_0 ever depend on what happens at $x = x_0$? Explain and give examples.

T5. How are one-sided limits related to limits? How can this relationship sometimes be used to calculate a limit or prove it does not exist? Give examples.

3.2.7.2 Exercises

E1. Use a table of values to estimate the value of the limit. If you have a graphing device, use it to confirm your result graphically.

a. $\lim_{x \to 0} \frac{\sqrt{x+4} - 2}{x}$

b. $\lim_{x \to 0} \frac{\tan(4x)}{2\tan(7x)}$

$\lim_{x\to 0} \frac{9^x-6^x}{x}$

d. $\lim_{x\to 1} \frac{x^7-1}{x^8-1}$

E2. Sketch the graph of the following function and use it to determine the values of a for which $\lim_{x\to a} f(x)$ exists:

$$f(x) = \begin{cases} 2-x & \text{if } x < -1 \\ x & \text{if } -1 \le x \le 1 \\ (x-1)^2 & \text{if } x > 1 \end{cases} \tag{1}$$

E3. Determine $\lim_{x\to 1^-} \frac{1}{x^2-1}$ and $\lim_{x\to 1^+} \frac{1}{x^2-1}$

a. by evaluating $f(x) = 1/(x^2-1)$ for values of x that approach 1 from the left and from the right,

b. from a graph of f.

E4. Determine the infinite limits.

a. $\lim_{x\to 2^+} \frac{x+3}{x+2}$,

b. $\lim_{x\to 2^-} \frac{x+3}{x+2}$,

c. $\lim_{x\to 1} \frac{3-x}{(x-1)^3}$,

d. $\lim_{x\to 2\pi^-} x \csc(x)$,

e. $\lim_{x\to 2^-} \frac{x^2-3x}{x^2-4x+4}$,

f. $\lim_{x\to 5^-} \frac{e^x}{(x-5)^4}$,

g. $\lim_{x\to 3^+} \ln(x^2-9)$.

E5. Find the vertical asymptotes of the function

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

Confirm your answer by graphing the function with *Mathematica.*

E6. In the theory of relativity, the mass of a particle with velocity v is

$$m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3}$$

where m_0 is the mass of the particle at rest and c is the speed of light. What happens as $v \to c^-$?

E7. Find the limits by substitution. Support your answers with a computer algebra system.

a. $\lim_{x\to 2} 2x$,

b. $\lim_{x\to -1} \frac{3x^2}{2x-1}$,

c. $\lim_{x\to \pi} \frac{\cos(x)}{1-\pi}$,

d. $\lim_{x\to -1} 3x(2x-1)$,

e. $\lim_{x\to 1} \frac{-1}{3x-1}$,

f. $\lim_{x\to \pi/2} x \sin(x)$.

E8. Use a CAS to perform the estimation of the limit and generate a graph of the function to verify your estimation for the functions:

a. $\lim_{x\to 0} \frac{2x^2}{4-4\cos(x)}$,

b. $\lim_{x\to 3} \frac{x^2-9}{\sqrt{x^2+7}-4}$,

c. $\lim_{x\to 0} \frac{\sqrt[3]{1+x}-1}{x}$,

$\lim_{x\to 2} \frac{x^4-16}{x-2}$,

e. $\lim_{x\to -1} \frac{x^3-x^2-5\,x-3}{(x+1)^2}$.

E9. Graph the function $f(x) = \sin(\pi/x)$ in the interval $x \in [-1, 1]$. Use a CAS to zoom in toward $x \to 0$. Describe your observations.

E10 Use numerical and graphical evidence to guess the value of the limit

$$\lim_{x\to 1} \frac{x^3 - 1}{\sqrt{x} - 1} \tag{4}$$

How close to 1 does x have to be to ensure that the function is within a distance 0.5 of its limit?

3.3 Calculating Limits Using the Limit Laws

In this section we will discuss algebraic techniques for computing limits of many functions. We base these results on the informal development of the limit concept discussed in the preceding section.

3.3.1 Basic Limits

Our strategy for finding limits algebraically has two parts:

- First we will obtain the limits of some simple functions
- Then we will develop a repertoire of theorems that will enable us to use the limits of those simple functions as building blocks for finding limits of more complicated functions.

We start with the cases of constant function $f(x) = k$, the identity function $f(x) = x$, and the reciprocal function $f(x) = 1/x$.

Theorem 3.5. *Simple Limits*

Let a and k be real numbers

$lim_{x\to a}\, k = k$

$lim_{x\to a}\, x = a$

$lim_{x\to 0^-} \frac{1}{x} = -\infty$

$lim_{x\to 0^+} \frac{1}{x} = +\infty.$ ■

The four limits in Theorem 3.5 should be evident from inspection of the function graphs shown in Figure 3.14.

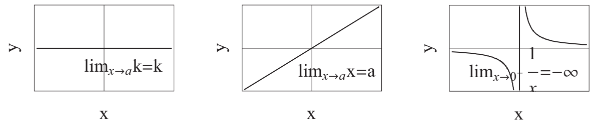


Figure 3.14. Simple limits for functions $f(x)=k$, $f(x)=a$, and $f(x)=1/x$.

In the case of the constant function $f(x)=k$, the value of $f(x)$ do not change as x varies, so the limit of $f(x)$ is k, regardless of at which number a the limit is taken. For example,

$$\lim_{x\to -25} 3 = 3, \qquad \lim_{x\to 0} 3 = 3, \qquad \lim_{x\to \pi} 3 = 3.$$

Sine the identity function $f(x)=x$ just echoes its input, it is clear that $f(x)=x\to a$ as $x\to a$. In terms of our informal definition of limits, if we decide just how close to a we would like the value of $f(x)=x$ to be, we need only restrict its input x to be just as close to a.

The one side limits of the reciprocal function $f(x)=1/x$ about 0 should conform with your experience with fractions: making the denominator closer to zero increases the magnitude of the fraction; i.e. increases its absolute value.

The following theorem will be our basic tool for finding limits algebraically.

Theorem 3.6. *Limits Algebraically Calculated*

Let a be a real number, and suppose that

$$lim_{x\to a} f(x) = L_1 \qquad \text{and} \qquad lim_{x\to a} g(x) = L_2.$$

That is, the limits exist and have values L_1 and L_2, respectively. Then:

a) $lim_{x\to a}(f(x)+g(x)) = lim_{x\to a} f(x) + lim_{x\to a} g(x) = L_1 + L_2$

b) $lim_{x\to a}(f(x)-g(x)) = lim_{x\to a} f(x) - lim_{x\to a} g(x) = L_1 - L_2$

c) $lim_{x\to a}(f(x)\,g(x)) = (lim_{x\to a} f(x))\,(lim_{x\to a} g(x)) = L_1\,L_2$

d) $lim_{x\to a}\frac{f(x)}{g(x)} = \frac{lim_{x\to a} f(x)}{lim_{x\to a} g(x)} = \frac{L_1}{L_2}$ *provided* $L_2 \neq 0$

e) $lim_{x\to a}\sqrt[n]{f(x)} = \sqrt[n]{lim_{x\to a} f(x)} = \sqrt[n]{L_1}$, *provided* $L_1 > 0$ *if n is even.*

Moreover, these statement are also true for the one-sided limits as $x\to a^-$ or as $x\to a^+$.■

The proof of these statements are omitted. A causal restatement of this theorem is as follows:

- The limit of a sum is the sum of the limits.
- The limit of a difference is the difference of the limits.
- The limit of a product is the product of the limits.
- The limits of the quotient is the quotient of the limits, provided the limit of the denominator is not zero.
- The limit of an *n*th root is the *n*th root of the limit.

Remark 3.3. Although results a) and c) in Theorem 3.6 are stated for two functions, they hold for any finite number of functions. For example, if the limits of $f(x)$, g(x), and $h(x)$ exist as $x \to a$, then the limit of their product also exist as $x \to a$ and are given by the formulas.

$$\lim_{x\to a}(f(x)+g(x)+h(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) + \lim_{x\to a} h(x) \tag{3.18}$$

$$\lim_{x\to a}(f(x)\, g(x)\, h(x)) = (\lim_{x\to a} f(x))\,(\lim_{x\to a} g(x))\,(\lim_{x\to a} h(x)) \tag{3.19}$$

In particular, if $f(x) = g(x)\, h(x)$, then this yields

$$\lim_{x\to a}(f(x))^3 = (\lim_{x\to a} f(x))^3. \tag{3.20}$$

More generally, if n is a positive integer, then the limit of the nth power of a function is the nth power of the function's limit. Thus,

$$\lim_{x\to a} x^n = (\lim_{x\to a} x)^n = a^n. \tag{3.21}$$

For example

$$\lim_{x\to 3} x^4 = 3^4 = 81.$$

Another useful result follows from part c) of Theorem 3.6 in the special case when one of the factors is a constant k:

$$\lim_{x\to a}(k\, f(x)) = (\lim_{x\to a} k)\,(\lim_{x\to a} f(x)) = k\,(\lim_{x\to a} f(x)) \tag{3.22}$$

and similar for $\lim_{x\to a}$ replaced by a one-sided limit, $\lim_{x\to a^+}$ or $\lim_{x\to a^-}$. Rephrased, this last statement says: A constant factor can be moved through a limit symbol.

3.3.2 Limits of Polynomials and Rational Functions

This section covers the calculation of limits for polynomials and rational functions.

Example 3.6. Limit of a Polynomial

Find the limit of

$$\lim_{x\to 5}\left(x^2 - 4\,x + 3\right)$$

and adjust each step. For our calculations we use the *Mathematica* function Limit[] which is designed in such a way that all the rules of Theorem 3.6 are incorporated The syntax of this function is Limit[f[x],x→a] corresponding the mathematical definition $\lim_{x\to a} f(x)$.

Solution 3.6. First note that by equation (3.21)

$$\lim_{x\to 5} x^2$$

```
25
```

Also from equation (3.22)

$$\lim_{x\to 5} 4\,x$$

```
20
```

Since

$$\lim_{x\to 5} 3$$

3

By Theorem 3.6 we may find

$$\lim_{x\to 5} 3 - \lim_{x\to 5} 4\,x + \lim_{x\to 5} x^2$$

8

However, for conciseness, it is common to reverse the order of this argument and simply write

$$\left(\lim_{x\to 5} x\right)^2 - 4\left(\lim_{x\to 5} x\right) + \lim_{x\to 5} 3$$

8

▲

Remark 3.4. In our presentation of limit arguments, we will adopt the convention of providing just a concise, reverse argument, bearing in mind that the validity of each equality may be conditional upon the successful resolution of the remaining limits.

Theorem 3.7. *Limit for a Polynomial*

For any polynomial

$p(x) = c_0 + c_1\, x + \ldots + c_n\, x^n$

and any real number a

$\lim_{x\to a} p(x) = c_0 + c_1\, a + \ldots + c_n\, a^n = p(a).$ ■

The proof is left as an exercise. Recall that a rational function is a ratio of two polynomials. Theorem 3.7 and Theorem 3.6 can often be used in combination to compute limits of rational functions. We illustrate this in the following example and prove a more general result in Theorem 3.8.

Example 3.7. Limit of a Rational Polynomial

Find the limit of the function

$$f(\mathrm{x\_}) := \frac{5\,x^3 + 4}{x - 3}$$

for $x \to 2$.

Solution 3.7. Applying Theorem 3.6 provides us with the rule that denominator and numerator can be separated and for that case we have to treat polynomials

$$\frac{\lim_{x\to 2}\left(5\,x^3+4\right)}{\lim_{x\to 2}\,(x-3)}$$

-44

which is equal to

$$\lim_{x\to 2} f(x)$$

-44

▲

Here is the general theorem on limits of rational functions.

Theorem 3.8. *Limits of Rational Functions*

Consider the rational function

$f(x) = \frac{p(x)}{q(x)}$

where $p(x)$ *and* $q(x)$ *are polynomials. For any real number a,*

a) *if* $q(a) \neq 0$*, then* $\lim_{x\to a} f(x) = f(a)$*,*

b) *if* $q(a) = 0$ *but* $p(a) \neq 0$*, then* $\lim_{x\to a} f(x)$ *does not exist.* ■

The proof of this theorem is generated by applying Theorem 3.6 and 3.7 to the polynomials. As an illustration of part b) of Theorem 3.8 consider again the function from Example 3.7

$$f(x)$$

$$\frac{5\,x^3+4}{x-3}$$

and calculate the limit for $x \to 3$.

$$\lim_{x\to 3} f(x)$$

∞

showing that the limit is indeterminate. This happens because the limit for denominator and numerator are

$$\lim_{x\to 3}\,(x-3)$$

0

and

$$\lim_{x\to 3}\left(5\,x^2+4\right)$$

49

Since the denominator of the limit equals zero, we find that the total limit of the function due to

Theorem 3.8 is undetermined.

3.3.3 Limits of Indeterminate Forms of Type 0/0 and Cancellations

The missing case in Theorem 3.8 is when both the numerator and the denominator of a rational function $f(x) = p(x)/q(x)$ have a zero at $x = a$. In this case, $p(x)$ and $q(x)$ will each have a factor of $x - a$, and canceling this factor may result in a rational function to which Theorem 3.8 applies.

Example 3.8. Cancellations in Limits

Find the limit of the function

$$f(\text{x\_}) := \frac{x^2 - 4}{x - 2}$$

Solution 3.8. Since 2 is a zero of both the numerator and the denominator, they share a common factor of $x - 2$. The limit can be obtained as follows:

$$\lim_{x \to 2} \frac{(x-2)(x+2)}{x-2}$$

4

which is equivalent to

$$\lim_{x \to 2} (x+2)$$

4

▲

Remark 3.5. Although correct, the second representation of the limit in the preceding computation needs some justification, since canceling the factor $x - 2$ alters the function by expanding its domain. However, as discussed in Example 2.50 in section 2.9.1 the two functions are identical, except at $x = 2$. From our discussion in the last section, we know that this difference has no effect on the limit as x approaches 2.

Example 3.9. Cancellations in Limits

Find the limit of the following functions

$$f(\text{x\_}) := \frac{x^2 - 6x + 9}{x - 3}$$

$$g(\text{x\_}) := \frac{2x + 8}{x^2 + x - 12}$$

and

$$h(\text{x\_}) := \frac{x^2 - 3x - 10}{x^2 - 10x + 25}$$

Solution 3.9. The numerator and the denominator both have a zero at $x = 3$, so there is a common factor of $x - 3$. We can factor the numerator by using the *Mathematica* function Numerator[] and Factor[] applied to the function. Numerator[] extracts the numerator of $f(x)$ and Factor[] calculates a factorization.

numf = Factor[Numerator[$f(x)$]]

$(x-3)^2$

The same can be done with the denominator using the functions Denominator[] and Factor[]

denomf = Factor[Denominator[$f(x)$]]

$x - 3$

Then the limit after cancellation of the common factor

$$\textbf{cancel} = \frac{\textbf{numf}}{\textbf{denomf}}$$

$x - 3$

$$\lim_{x\to 3} \textbf{cancel}$$

0

For the function g the numerator and the denominator both have a zero at $x = -4$, so there is a common factor of $x - (-4) = x + 4$. Then the numerator is given by

numg = Factor[Numerator[$g(x)$]]

$2(x+4)$

The denominator calculates to

denomg = Factor[Denominator[$g(x)$]]

$(x-3)(x+4)$

Canceling the common factor results to

$$\textbf{cancel} = \frac{\textbf{numg}}{\textbf{denomg}}$$

$$\frac{2}{x-3}$$

The limit of the resulting function reduces to

$$\lim_{x\to 4} \textbf{cancel}$$

2

For function h the numerator and denominator have a zero at $x = 5$.

numh = Factor[Numerator[$h(x)$]]

$(x-5)(x+2)$

denomh = Factor[Denominator[*h*(*x*)]]

$(x-5)^2$

But one of the term cancels as

$$\textbf{cancel} = \frac{\textbf{numh}}{\textbf{denomh}}$$

$$\frac{x+2}{x-5}$$

In this case due to Theorem 3.8 the limit does not exist.

$$\lim_{x\to 5} \textbf{cancel}$$

∞

▲

The case of a limit of a quotient,

$$\lim_{x\to a} \frac{f(x)}{g(x)} \tag{3.23}$$

where $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, is called an indeterminate form of type 0/0. Note that the limits in the two examples above produced a variety of answers. The word indeterminate here refers to the fact that the limiting behavior of the quotient cannot be determined without further study. The expression 0/0 is just a mnemonic device to describe the circumstance of a limit of a quotient in which both the numerator and the denominator approaches 0.

3.3.4 Limits of Piecewise-Defined Functions

For functions that are defined piecewise, a two-sided limit at x-values where the formula changes is best obtained by first finding the one-sided limits at that number.

Example 3.10. Piecewise-Defined Functions and their Limits

Let

$$f(x) = \begin{cases} 1/(x+2) & x < -2 \\ x^2 - 5 & -2 < x \le 3 \\ \sqrt{x+13} & x > 3. \end{cases}$$

Find the limits $\lim_{x\to -2} f(x)$, $\lim_{x\to 0} f(x)$, and $\lim_{x\to 3} f(x)$.

Solution 3.10. The graph of the function is shown in Figure 3.15.

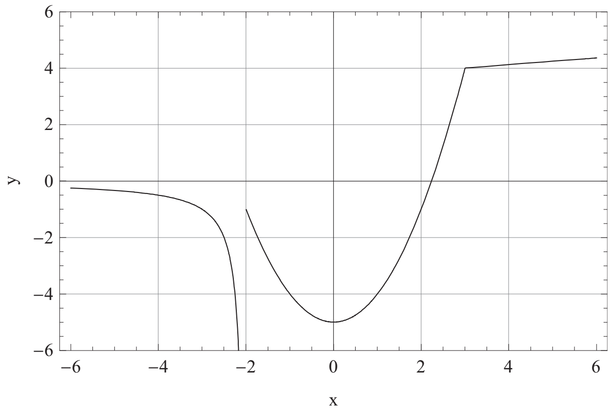


Figure 3.15. Graph of the function $f(x)$.

As x approaches -2 from the left the limit is given by

$$\lim_{x \to (-2)^-} f(x)$$

$-\infty$

As x approaches -2 from the right, the value is

$$\lim_{x \to (-2)^+} f(x)$$

-1

Thus the limit $\lim_{x \to -2} f(x)$ does not exist. As x approaches 0 from either the left or the right the limiting values are

$$\lim_{x \to (0)^-} f(x)$$

-5

$$\lim_{x \to (0)^+} f(x)$$

-5

demonstrating that the limit value is equal to -5. As x approaches 3 from the left and from the right the values are

$$\lim_{x \to (3)^-} f(x)$$

4

$$\lim_{x \to (3)^+} f(x)$$

4

Thus the limit exists for $x \to 3$.▲

3.3.5 Limits for Large Arguments

The behavior of a function towards the extremes of its domain is sometimes called its end behavior. Here we will use limits to investigate the end behavior of a function as $x \to -\infty$ or as $x \to +\infty$. As in the section 3.3.1, we will begin by obtaining limits of some simple functions and then use these as building blocks for finding limits of more complicated functions.

Theorem 3.9. *Simple Limits at* $\pm\infty$

Let k be a real number.

$lim_{x\to-\infty} k = k \qquad lim_{x\to+\infty} k = k$

$lim_{x\to-\infty} x = -\infty \qquad lim_{x\to+\infty} x = +\infty$

$lim_{x\to-\infty} \frac{1}{x} = 0 \qquad lim_{x\to+\infty} \frac{1}{x} = 0.$ ■

The six limits in Theorem 3.9 should be evident from inspection of the function graph in Figure 3.16

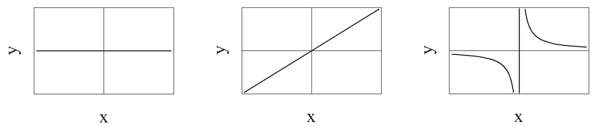


Figure 3.16. Limits for functions $f(x) = k$, $f(x) = x$, and $f(x) = 1/x$ for $x \to \pm\infty$.

The limits of the reciprocal function $f(x) = 1/x$ should make sense to you intuitively, based on your experience with fractions: increasing the magnitude of x makes its reciprocal closer to zero.

The following theorem mirrors Theorem 3.6 as our tool for finding limits at $\pm\infty$ algebraically

Theorem 3.10. *Limit Algebra for* $x \to \pm\infty$

Suppose that the two limits exist

$lim_{x\to+\infty} f(x) = L_1$ *and* $lim_{x\to+\infty} g(x) = L_2$.

That is, the limits exist and have values L_1 *and* L_2*, respectively. Then:*

a) $lim_{x\to+\infty} (f(x) + g(x)) = lim_{x\to+\infty} f(x) + lim_{x\to+\infty} g(x) = L_1 + L_2$

b) $lim_{x\to+\infty} (f(x) - g(x)) = lim_{x\to+\infty} f(x) - lim_{x\to+\infty} g(x) = L_1 - L_2$

c) $lim_{x\to+\infty} (f(x)\, g(x)) = (lim_{x\to+\infty} f(x))\, (lim_{x\to+\infty} g(x)) = L_1\, L_2$

d) $lim_{x\to+\infty} \frac{f(x)}{g(x)} = \frac{lim_{x\to+\infty} f(x)}{lim_{x\to+\infty} g(x)} = \frac{L_1}{L_2}$ *provided* $L_2 \neq 0$

e) $lim_{x\to+\infty} \sqrt[n]{f(x)} = \sqrt[n]{lim_{x\to+\infty} f(x)} = \sqrt[n]{L_1}$, *provided* $L_1 > 0$ *if n is even.*

Moreover, these statements are also true for limits as $x \to -\infty$. ■

Remark 3.6. Results a) and c) can be extended to sums or products of any finite number of functions. In particular, for any positive integer n there exists the valid relation

$$\lim_{x\to+\infty} (f(x))^n = (\lim_{x\to+\infty} f(x))^n \quad \text{and} \quad \lim_{x\to-\infty} (f(x))^n = (\lim_{x\to-\infty} f(x))^n$$

Also, since $\lim_{x\to+\infty} 1/x = 0$, if n is a positive integer, then

$$\lim_{x\to+\infty} \left(\frac{1}{x}\right)^n = \left(\lim_{x\to+\infty} \frac{1}{x}\right)^n = 0 \quad \text{and} \quad \lim_{x\to-\infty} \left(\frac{1}{x}\right)^n = \left(\lim_{x\to-\infty} \frac{1}{x}\right)^n = 0$$

Another useful result follows from part c) of Theorem 3.10 in the special case where one of the factors is a constant k:

$$\lim_{x\to+\infty} (k\, f(x)) = (\lim_{x\to+\infty} k)\, (\lim_{x\to+\infty} f(x)) = k \lim_{x\to+\infty} f(x)$$

and similarly, for $\lim_{x\to+\infty}$ replaced by $\lim_{x\to-\infty}$. Rephrasing this last statement says: A constant factor can be moved through a limit symbol.

3.3.5.1 Limits of x^n as $x \to \pm\infty$

In Figure 3.17 we have graphed the monomials of the form x^n for $n = 1, 2, 3$, and 4. In each figure we have indicated the limit as $x \to \pm\infty$.

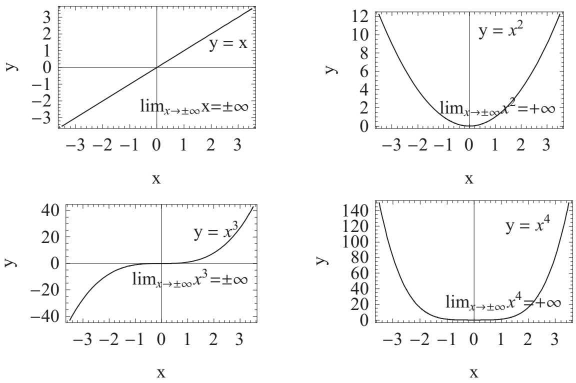


Figure 3.17. Limits for monomials x^n with $n = 1, 2, 3$, and 4 for $x \to \pm\infty$.

The results in the figure are special cases of the following general results:

$$\lim_{x\to+\infty} x^n = +\infty \qquad \text{for} \qquad n = 1, 2, 3, 4, \ldots \tag{3.24}$$

and

$$\lim_{x\to-\infty} x^n = \begin{cases} -\infty & \text{with } n = 1, 3, 5, \ldots \\ +\infty & \text{with } n = 2, 4, 6, \ldots \end{cases} \tag{3.25}$$

Multiplying x^n by a positive real number does not affect (3.24) and (3.25), but multiplying by a negative real number reverses the sign.

Example 3.11. Limits of Monomials

Here are some examples for limits of monomials

$$\lim_{x\to\infty} 2\,x^5$$

$$\infty$$

The same monomial but with the variables approaching $-\infty$.

$$\lim_{x\to-\infty} 2\,x^5$$

$$-\infty$$

A similar monomial with a negative coefficient for $x \to \infty$

$$\lim_{x\to\infty} -2\,x^8$$

$$-\infty$$

The same monomial but for negative values of x

$$\lim_{x\to-\infty} -2\,x^8$$

$$-\infty$$

▲

3.3.5.2 Limits of Polynomials as $x \to \pm\infty$

There is a useful principle about polynomials which, expresses informally, states that:

The end behavior of a polynomial matches the end behavior of its highest degree term.

More precisely, if $c_n \neq 0$ then

$$\lim_{x\to+\infty}\left(c_0 + c_1\,x + c_2\,x^2 + \ldots + c_n\,x^n\right) = \lim_{x\to+\infty} c_n\,x^n \tag{3.26}$$

and

$$\lim_{x\to-\infty}\left(c_0 + c_1\,x + c_2\,x^2 + \ldots + c_n\,x^n\right) = \lim_{x\to-\infty} c_n\,x^n. \tag{3.27}$$

We can motivate these results by factoring out the highest power of x from the polynomial and examining the limit of the factored expression. Thus

$$\left(c_0 + c_1\,x + c_2\,x^2 + \ldots + c_n\,x^n\right) = x^n\left(\frac{c_0}{x^n} + \frac{c_1}{x^{n-1}} + \frac{c_2}{x^{n-2}} + \ldots + c_n\right). \tag{3.28}$$

As $x \to -\infty$ or $x \to +\infty$, it follows from the remarks to Theorem 3.10 that all of the terms with positive powers of x in the denominator approaches zero, so that (3.26) and (3.27) are satisfied.

Example 3.12. Limits of Polynomials

Lets assume we have a polynomial of degree 15 and are interested in the limiting behavior for large values of $x \to \pm\infty$. The polynomial examined shows

$$\lim_{x \to -\infty} \left(9\, x^{15} - 4\, x^4 + 3\, x + 8\right)$$

$$-\infty$$

The limiting behavior of the largest monomial in the polynomial shows the behavior

$$\lim_{x \to -\infty} \left(9\, x^{15}\right)$$

$$-\infty$$

The other direction is examined by

$$\lim_{x \to \infty} \left(9\, x^{15} - 4\, x^4 + 3\, x + 8\right)$$

$$\infty$$

and the largest monomial delivers

$$\lim_{x \to \infty} \left(9\, x^{15}\right)$$

$$\infty$$

▲

3.3.5.3 Limits of Rational Functions as $x \to \pm\infty$

A useful technique for determining the end behavior of a rational function $f(x) = p(x)/q(x)$ is to factor and cancel the highest power of x that occurs in the denominator $q(x)$ from both $p(x)$ and $q(x)$. The denominator of the resulting fraction then has a (nonzero) limit equal to the leading coefficient of $q(x)$, so the limit of the resulting fraction can be quickly determined using (3.24), (3.25), and (3.26). The following example illustrates this technique.

Example 3.13. Limits of Rational Functions

Find the limit of the following rational function

$$f(\text{x\_}) := \frac{3\, x + 5}{6\, x - 8}$$

Solution 3.13. First determine the denominator and the numerator of the function

den = Denominator[$f(x)$]

$6\, x - 8$

num = Numerator[$f(x)$]

$3\, x + 5$

Then find the common monomial to both the denominator and numerator. In this case this monomial is given by

mon = x

x

Then factor out this monomial in the numerator and denominator we use the *Mathematica* function Expand[]

$$\textbf{denom1} = \textbf{Expand}\left[\frac{\textbf{den}}{\textbf{mon}}\right]$$

$$6 - \frac{8}{x}$$

and for the numerator we get

$$\textbf{num1} = \textbf{Expand}\left[\frac{\textbf{num}}{\textbf{mon}}\right]$$

$$\frac{5}{x} + 3$$

Theorem 3.10 states that the numerator and the denominator of a rational function can be examined separately, we apply this theorem now

$$\textbf{limdenom} = \lim_{x\to\infty} \textbf{denom1}$$

6

$$\textbf{limnum} = \lim_{x\to\infty} \textbf{num1}$$

3

The division of both results provides

$$\frac{\textbf{limnum}}{\textbf{limdenom}}$$

$$\frac{1}{2}$$

which in fact is the limiting value of the rational function

$$\lim_{x\to\infty} f(x)$$

$$\frac{1}{2}$$

▲

Example 3.14. Limits of Rational Functions

Find the limit of the two functions

$$f(\text{x\_}) := \frac{4x^2 - x}{2x^3 - 5}$$

and

$$g(x\_) := \frac{5\,x^3 - 6\,x^2 + 2}{3\,x^3 + 7}$$

Solution 3.14. Extracting numerator and denominator from the function allows us to apply Theorem 3.10.

num = Numerator[$f(x)$]

$4\,x^2 - x$

denom = Denominator[$f(x)$]

$2\,x^3 - 5$

The largest monomial of the numerator and denominator is

mon = x^3

x^3

Division of numerator and denominator gives

$$\textbf{num1 = Expand}\left[\frac{\textbf{num}}{\textbf{mon}}\right]$$

$$\frac{4}{x} - \frac{1}{x^2}$$

$$\textbf{denom1 = Expand}\left[\frac{\textbf{denom}}{\textbf{mon}}\right]$$

$$2 - \frac{5}{x^3}$$

The limiting values of numerator and denominator are

$$\textbf{limNum} = \lim_{x\to\infty} \textbf{num1}$$

0

$$\textbf{limDenom} = \lim_{x\to\infty} \textbf{denom1}$$

2

and the ratio delivers

$$\frac{\textbf{limNum}}{\textbf{limDenom}}$$

0

which agrees with the limit of the rational function

$$\lim_{x\to\infty} f(x)$$

0

for the second function $g(x)$ we take the same procedure

num = Numerator[$g(x)$]

$5\,x^3 - 6\,x^2 + 2$

denom = Denominator[$g(x)$]

$3\,x^3 + 7$

The largest monomial of the numerator and denominator is

mon = x^3

x^3

Division of numerator and denominator by this term results to

$$\textbf{num1 = Expand}\left[\frac{\textbf{num}}{\textbf{mon}}\right]$$

$$\frac{2}{x^3} - \frac{6}{x} + 5$$

$$\textbf{denom1 = Expand}\left[\frac{\textbf{denom}}{\textbf{mon}}\right]$$

$$\frac{7}{x^3} + 3$$

The limiting values of numerator and denominator are

$$\textbf{limNum} = \lim_{x\to\infty} \textbf{num1}$$

5

$$\textbf{limDenom} = \lim_{x\to\infty} \textbf{denom1}$$

3

and the ratio gives

$$\frac{\textbf{limNum}}{\textbf{limDenom}}$$

$$\frac{5}{3}$$

which agrees with the limit of the function

$$\lim_{x\to\infty} g(x)$$

$$\frac{5}{3}$$

▲

3.3.5.4 Limits Involving Radicals

If functions contain radicals, we can apply section e) of Theorem 3.10. It may happen that we also need the other parts of this theorem to evaluate the limits. The following examples demonstrate how radicals are treated in a limit process.

Example 3.15. Limits of Radicals I

Find the limit of the following function

$$f(\mathrm{x\_}) := \sqrt[3]{\frac{3x-5}{6x-8}}$$

Solution 3.15. The radicand can be extracted from the function by

radicand = First[$f(x)$]

$$\frac{3x-5}{6x-8}$$

Applying Theorem 3.10 to this expression we can write

$$\sqrt[3]{\lim_{x\to\infty} \text{radicand}}$$

$$\frac{1}{2^{1/3}}$$

which gives the same limiting value as

$$\lim_{x\to\infty} f(x)$$

$$\frac{1}{2^{1/3}}$$

▲

Example 3.16. Limits of Radicals II

Find the limits for $x \to \pm\infty$ of the following function

$$f(\mathrm{x\_}) := \frac{\sqrt{x^2+5}}{3x-7}$$

Solution 3.15. To calculate the limits extract numerator and denominator

num = Numerator[$f(x)$]

$$\sqrt{x^2+5}$$

denom = Denominator[$f(x)$]

$$3x-7$$

From the numeration extract the polynomial by

pol = First[num]

$x^2 + 5$

Set the largest monomial to

mon = $\boldsymbol{x^2}$

x^2

Divide the polynomial and take the square root

pol1 = Expand$\left[\dfrac{\textbf{pol}}{\textbf{mon}}\right]$

$$\frac{5}{x^2} + 1$$

num1 = $\sqrt{\textbf{pol1}}$

$$\sqrt{\frac{5}{x^2} + 1}$$

The limit of the numerator then is

limNum = $\lim\limits_{x \to \infty}$ **num1**

```
1
```

Since the largest monomial in the denominator is x which is also the common factor on the numerator, we find

denom1 = Expand$\left[\dfrac{\textbf{denom}}{\boldsymbol{x}}\right]$

$$3 - \frac{7}{x}$$

The limit of the denominator now is

limDenom = $\lim\limits_{x \to \infty}$ **denom1**

```
3
```

Combining both limits due to Theorem 3.10, we find

$$\frac{\textbf{limNum}}{\textbf{limDenom}}$$

$$\frac{1}{3}$$

which agrees with

$$\lim_{x\to\infty} f(x)$$

$$\frac{1}{3}$$

The graph of the function $f(x)$ is shown in the following Figure 3.18

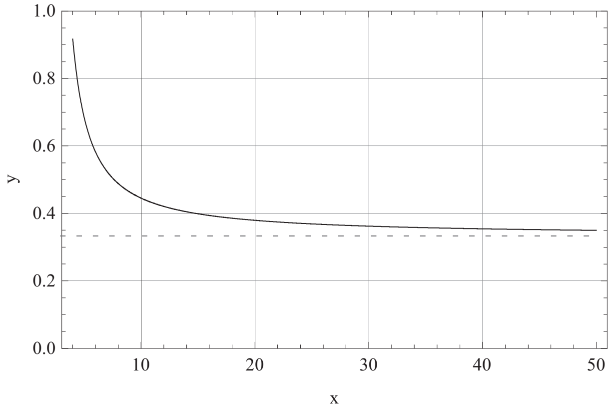


Figure 3.18. Limit for large $x \to +\infty$ values of the function $f(x) = \frac{\sqrt{x^2+5}}{3x-7}$.

The limit for the other direction to infinity is given by a similar calculation starting with the separation of numerator and denominator

num = Numerator[$f(x)$]

$$\sqrt{x^2+5}$$

denom = Denominator[$f(x)$]

$$3x-7$$

From the numeration extract the polynomial by

pol = First[num]

$$x^2+5$$

Set the largest monomial as

mon = x^2

$$x^2$$

Divide the polynomial and take the square root

$$\textbf{pol1} = \textbf{Expand}\left[\frac{\textbf{pol}}{\textbf{mon}}\right]$$

$$\frac{5}{x^2} + 1$$

$$\textbf{num1} = \sqrt{\textbf{pol1}}$$

$$\sqrt{\frac{5}{x^2} + 1}$$

The limit of the numerator then is

$$\textbf{limNum} = \lim_{x \to -\infty} \textbf{num1}$$

$$1$$

Since the largest monomial in the denominator is -x (we are working in the negative x-direction) which is also the common factor on the numerator, we find

$$\textbf{denom1} = \textbf{Expand}\left[-\frac{\textbf{denom}}{\boldsymbol{x}}\right]$$

$$\frac{7}{x} - 3$$

The limit of the denominator now is

$$\textbf{limDenom} = \lim_{x \to -\infty} \textbf{denom1}$$

$$-3$$

Combining both limits due to Theorem 3.10, we find

$$\frac{\textbf{limNum}}{\textbf{limDenom}}$$

$$-\frac{1}{3}$$

which agrees with

$$\lim_{x \to -\infty} f(x)$$

$$-\frac{1}{3}$$

The following figure shows how the function behaves for large negative x values. Surprisingly the intermediate behavior of the function around $x \approx -10$ is not a part of our analysis. Before the function reaches the limit $y = -1/3$ it traverses the limit from below and approaches the actual limit $-1/3$ from above.

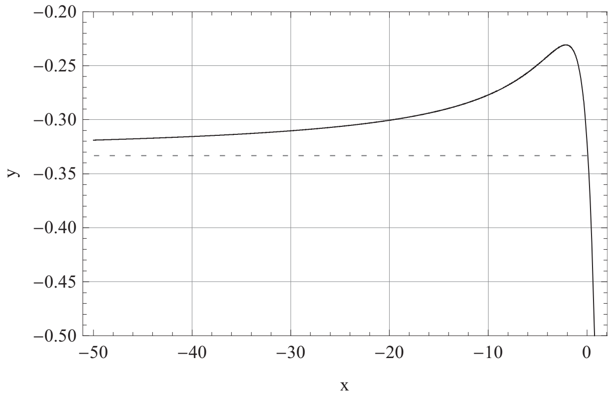


Figure 3.19. Limit for large $x \to -\infty$ values of the function $f(x) = \frac{\sqrt{x^2+5}}{3x-7}$.

It is also obvious from the following figure that the function has a singularity at $x = 7/3$.

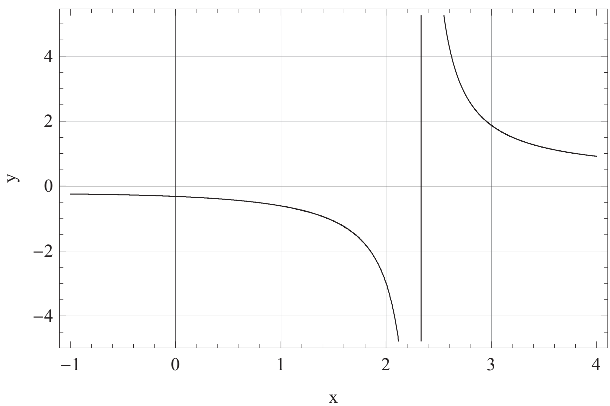


Figure 3.20. Behavior of $f(x) = \frac{\sqrt{x^2+5}}{3x-7}$a round $x = 7/3$.

This can be examined by calculating the right and left hand limit at that position.

$$\lim_{x \to \frac{7}{3}^+} f(x)$$

∞

$$\lim_{x \to \frac{7}{3}^-} f(x)$$

$-\infty$

▲

3.3.6 Tests and Exercises

The following two subsections serve to test your understanding of the material on limits and limit laws.

3.3.6.1 Test Problems

T1. Define what it means to say that $\lim_{x\to 0} f(x) = k$.

T2. What are $\lim_{x\to\pm\infty} k$ (k a constant) and $\lim_{x\to\pm\infty} 1/x$? How do you extend these results to other functions? Give examples.

T3. What are horizontal, vertical, and oblique asymptotes? Give examples.

T4. How do you find the limit of a rational function as $x \to \pm\infty$? Give examples.

3.3.6.2 Exercises

E1. Evaluate the limit and justify each step by indicating the appropriate Limit Law(s).

a. $\lim_{x\to -2}\left(3\,x^4 + 2\,x^2 - x + 2\right)$,

b. $\lim_{x\to}\left(x^2 - 1\right)^2 (x+7)^3$,

c. $\lim_{x\to -2}\sqrt{x^4 + 3\,x + 6}$,

d. $\lim_{x\to 1}\left(\frac{1+3\,x}{1+4\,x^2+3\,x^4}\right)^3$,

e. $\lim_{x\to 4^-}\sqrt{16 - x^2}$,

f. $\lim_{x\to 8}\left(1 + \sqrt[3]{x}\right)\left(2 - \log(x-8) + x^2 - 5\,x^4\right)$.

E2. What is wrong with the following equation?

$$\frac{x^2 + x - 6}{x - 2} = x + 3 \tag{1}$$

Explain why the following equation is correct

$$\lim_{x\to 2}\frac{x^2 + x - 6}{x - 2} = \lim_{x\to 2}(x + 3). \tag{2}$$

E3. Use the Squeezing Theorem to show that

$$\lim_{x\to 0}\sqrt{x^3 + x^2}\;\sin\left(\frac{\pi}{x}\right) = 0 \tag{3}$$

Illustrate by graphing the three functions involved in the Squeezing Theorem on the same screen.

E4. Prove that

a. $\lim_{x\to 0} x^4 \cos\left(\frac{2}{x}\right) = 0$,

b. $\lim_{x\to 0^+}\sqrt{x}\;e^{\sin(\pi/x)} = 0$.

E5. Find the limit, if it exists. If the limit does not exist, explain why.

a. $\lim_{x\to 0^-}\left(\frac{1}{x} - \frac{1}{|x|}\right)$,

b. $\lim_{x\to 0^+}\left(\frac{1}{x} - \frac{1}{|x|}\right)$,

c. $\lim_{x\to -2}\frac{2-|x|}{2+x}$,

d. $\lim_{x\to 3}(2\,x + |\,x-3\,|)$.

E6. Grinding engine cylinders. Before contracting to grind engine cylinders to a cross sectional area of 9 cm^2 you need to know how much deviation from the ideal cylinder diameter of $r_0 = 3.385$ cm you can allow and still have the area come within 0.01 cm^2 of the required 9 cm^2. To find out, you let $A = \pi(r/2)^2$ and look for the interval in which you must hold x to make $|\,A - 9\,| \le 0.01$. What interval do you find?

E7. Evaluate the limit, if it exists.

a. $\lim_{x\to 1} \frac{\sqrt{1+x}-1}{x-1}$,

b. $\lim_{x\to 0} \frac{(2+x)^3-7}{x}$,

c. $\lim_{x\to -2} \frac{x+2}{x^3+8}$,

d. $\lim_{x\to -3} \frac{x^2-9}{2x^2+7x+3}$,

e. $\lim_{x\to -4} \frac{x^2+5x+4}{x^2+3x-4}$,

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

g. $\lim_{x\to -4} \frac{\frac{1}{4}+\frac{1}{x}}{4+x} - \frac{\frac{2}{x}+4}{3x+12}$.

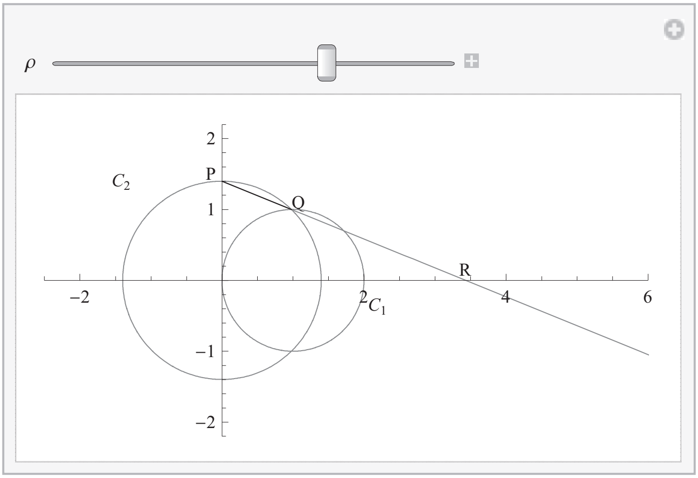
E8. Let $f(x) = \frac{x^2-1}{|x-1|}$. Find the limits

a. $\lim_{x\to 1^-} f(x)$

b. $\lim_{x\to 1^+} f(x)$.

E9. The figure shows a fixed circle C_1 with equation $(x-1)^2 + y^2 = 1$ and a shrinking circle C_2 with radius r and center the origin. P is the point $(0, r)$, Q is the upper point of intersection of the two circles, and R is the point of intersection of the line PQ and the x-axis. What happens to R as C_2 shrinks, that is, as $r \to 0^+$?

Out[20]=



E10 Prove that $\lim_{x\to 2} \frac{1}{x} = \frac{1}{2}$.

3.4 Continuity

A moving object cannot vanish at some point and reappear somewhere else to continue in its motion. Thus, we perceive the path of a moving object as an unbroken curve, without gaps, breaks, or holes. In this section, we translate "unbroken curve" into a precise mathematical formulation called continuity, and develop some fundamental properties of continuous curves.

Recall from Theorem 3.7 that if $p(x)$ is a polynomial and c is a real number, then the

$\lim_{x \to c} p(x) = p(c)$ (see Figure 3.21).

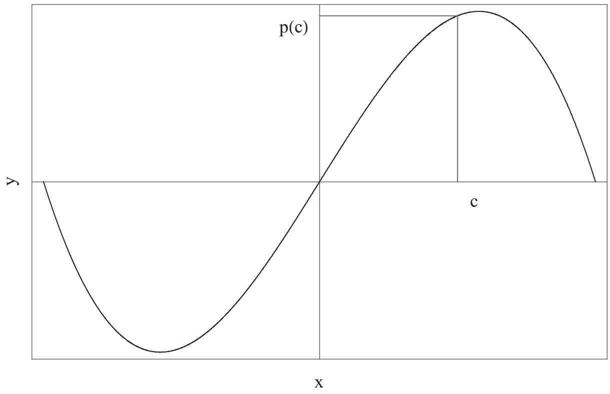


Figure 3.21. Limiting behavior of a polynomial for $x \to c$.

Together with Theorem 3.6, we are able to calculate limits of a variety of combinations of functions by evaluating their combinations. That is, we saw many examples of functions $f(x)$ such that $\lim_{x \to c} f(x) = f(c)$ if $f(x)$ is defined on an interval containing a number c. In this case, function values $f(x)$ can be guaranteed to be near $f(c)$ for any x-value selected close enough to c.

On the other hand, we have also seen functions for which this nice property is not true. For example, the function

$$f(x) = \begin{cases} \sin(\pi / x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

does not satisfy $\lim_{x \to 0} f(x) = f(0)$, since $\lim_{x \to 0} f(x)$ fails to exist (Figure 3.22)

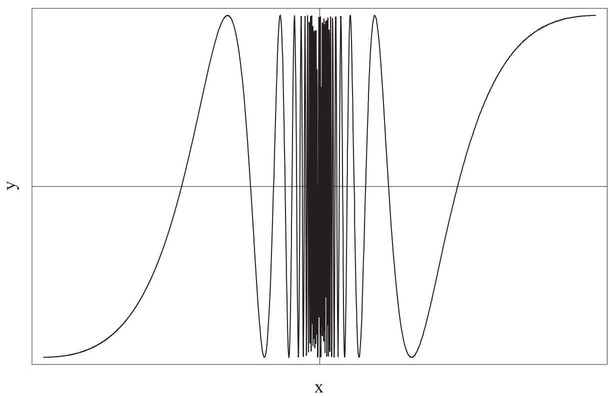


Figure 3.22. Graph of the function $f(x) = \begin{cases} \sin(\pi / x) & x \neq 0 \\ 0 & x = 0 \end{cases}$.

The term continuous is used to describe the useful circumstances where the calculation of a limit can be accomplished by mere evaluation of the function.

Definition 3.3. *Continuity*

A function f is said to be continuous at $x = c$ provided the following conditions are satisfied:

1. *$f(c)$ is defined*
2. *$lim_{x \to c} f(x)$ exists*
3. *$lim_{x \to c} f(x) = f(c)$.* ■

If one or more of the conditions of this definition fails to hold, then we will say that f has a discontinuity at $x = c$. Each function drawn in Figure 3.23 illustrates a discontinuity at $x = c$. In Figure 3.23a the function is not defined at $x = c$, violating the first condition of Definition 3.3. In Figure 3.23b and 3.23c. $\lim_{x \to c} f(x)$ does not exist, violating the second condition of Definition 3.3. In Figure 3.23d, the function is defined at $x = c$ and the limit $\lim_{x \to c} f(x)$ exists, but these two values are not equal, violating the third condition of Definition 3.3.

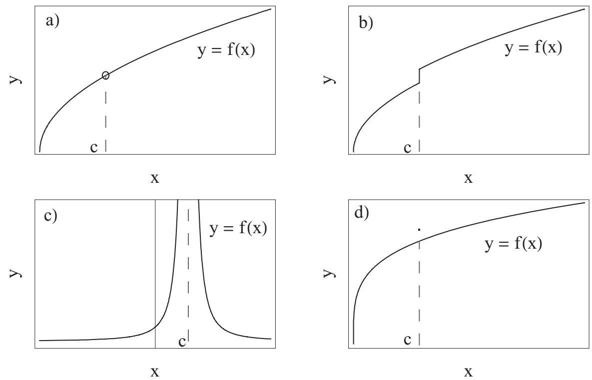


Figure 3.23. Different situations when no continuity for a function $f(x)$ is given .

From such graphs, we can develop an intuitive, geometric feeling for where a function is continuous and where it is discontinuous. Observe that continuity at c may fail due to a break in the graph of the function, either due to a hole or a jump as in Figure 3.23. Although the intuitive interpretation of "f is continuous at c" as "the graph is unbroken at c" lacks precision, it is a useful guide in most circumstances.

Remark 3.7. Note that the third condition of Definition 3.3 really implies the first two conditions, since it is understood in the statement $\lim_{x \to c} f(x) = f(c)$ that the limit on the left exists, the expression $f(c)$ on the right is defined and has a finite value and that quantities on the two sides are equal. Thus, when we want to establish continuity of a function at a point, our usual procedure will be to establish the validity of the third condition only.

Example 3.17. Continuity

Determine whether the following functions are continuous at $x = 2$.

$$f(\text{x\_}) := \frac{x^2 - 4}{x - 2}$$

$$g(\text{x\_}) := \text{Piecewise}\left[\begin{pmatrix} \frac{x^2-4}{x-2} & x \neq 2 \\ 3 & x = 2 \end{pmatrix}\right]$$

$$h(\text{x\_}) := \text{Piecewise}\left[\begin{pmatrix} \frac{x^2-4}{x-2} & x \neq 2 \\ 4 & x = 2 \end{pmatrix}\right]$$

Solution 3.17. In each case, we must determine whether the limit of the function as $x \to 2$ is the same as the value of the function at $x = 2$. In all three cases the functions are identical except at $x = 2$, and hence all three have the same limit at $x = 2$, namely

$$\lim_{x\to 2} f(x)$$

4

and

$$\lim_{x\to 2} g(x)$$

4

as well as

$$\lim_{x\to 2} h(x)$$

4

The function f is undefined at $x = 2$, and hence is not continuous at $x = 2$ (Figure 3.24a). The function g is defined at $x = 2$, but its value here is $g(2) = 3$, which is not the same as the limit as x approaches 2; hence, g is also not continuous at $x = 2$ (Figure 3.24b). The value of the function h at $x = 2$ is $h(2) = 4$, which is the same as the limit as x approaches 2; hence, h is continuous at $x = 2$ (Figure 3.24c).

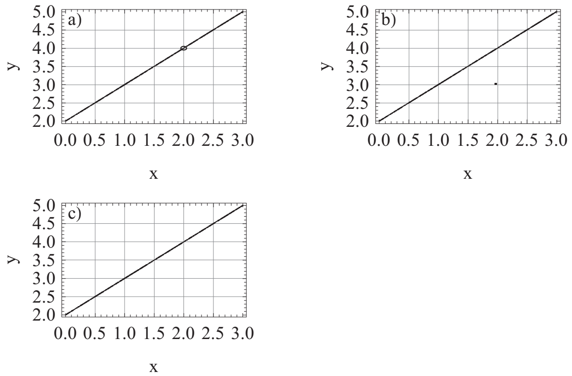


Figure 3.24. Different functions and their status of continuity. a) and b) are discontinuous and c) is continuous.▲

3.4.1 Continuity in Applications

In applications, discontinuities often signal the occurrence of important phenomena. For example, Figure 3.25 is a graph of voltage versus time for an underground cable that is accidentally cut by a work crew at time $t = t_0$. The voltage drops to zero when the line is cut.

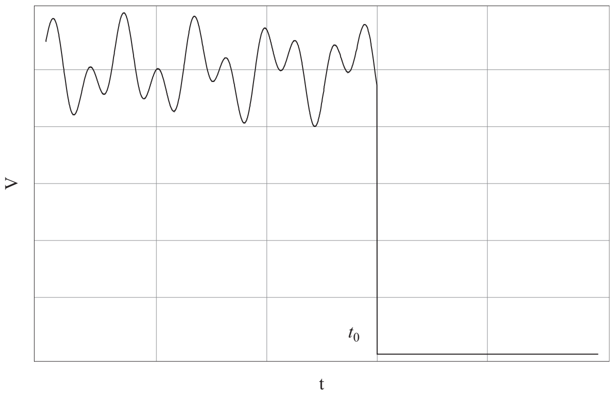


Figure 3.25. Voltage drop on a cable which was cut.

Figure 3.26 shows the graph of inventory versus time for a company that restocks its warehouse to n_1 units when the inventory falls to n_0 units. The discontinuities occur at those times where restocking occurs.

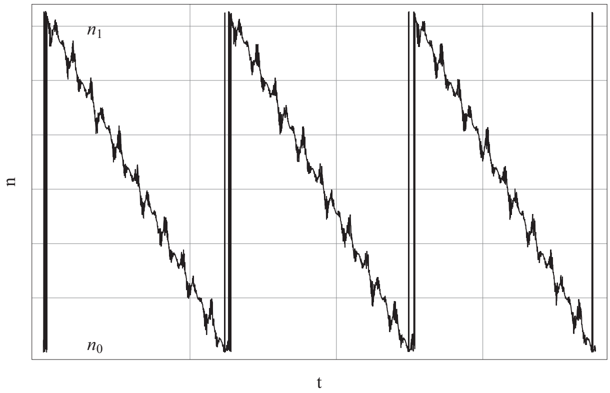


Figure 3.26. Restocking of a warehouse if the inventory reaches a lower level n_0.

Given the possible significance of discontinuity it is important to be able to identify discontinuities for specific functions and to be able to make general statements about the continuous properties of entire families of functions.

3.4.2 Continuity of Polynomials

If a function f is continuous at each number in an open interval (a, b), then we say that f is continuous on (a, b). This definition applies to infinite open intervals of the form $(a, +\infty)$, $(-\infty, b)$, and $(-\infty, +\infty)$. In the case where f is continuous on $(-\infty, +\infty)$, we will say that f is continuous everywhere.

The general procedure for showing that a function is continuous everywhere is to show that it is continuous at an arbitrary real number. For example, we showed in Theorem 3.7 that if $p(x)$ is a polynomial and a is any real number, then

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

Thus we have the following result

Theorem 3.11. *Continuity of a Polynomial*

Polynomials are continuous everywhere.∎

Example 3.18. Continuity of a Polynomial

Show that $|x|$ is continuous everywhere. The graph of the magnitude of x is shown in Figure 3.27

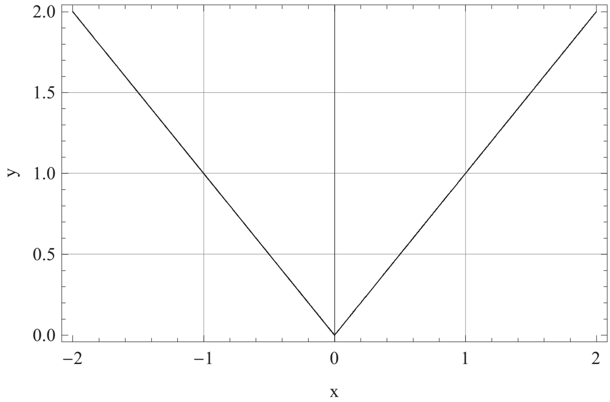


Figure 3.27. Graph of the function $f(x) = |x|$.

Solution 3.18. We can write $|x|$ as

$$|x| = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -x & \text{if } x < 0 \end{cases} \tag{3.29}$$

so $|x|$ is the same as the polynomial x on the interval $(0, +\infty)$ and is the same as the polynomial $-x$ on the interval $(-\infty, 0)$. But polynomials are continuous everywhere, so $x = 0$ is the only possible discontinuity for $|x|$. Since $|0| = 0$, to prove the continuity at $x = 0$, we must show that

$$\lim_{x\to 0} |x| = 0. \tag{3.30}$$

Because the formula for $|x|$ changes at 0, it will be helpful to consider the one-sided limits at 0 rather than the two-sided limit. We obtain

$$\lim_{x\to 0^+} |x|$$

0

and

$$\lim_{x\to 0^-} |x|$$

0

Thus (3.30) holds and $|x|$ is continuous at $x = 0$.▲

3.4.3 Properties of Continuous Functions

The following Theorem, which is a consequence of Theorem 3.6, will enable us to reach conclusions about the continuity of functions that are obtained by adding, subtracting, multiplying, and dividing continuous functions.

Theorem 3.12. *Continuity of Algebraic Operations*

If the functions f and g are continuous at c, then

a) $f + g$ *is continuous at* c

b) $f - g$ *is continuous at* c

c) $f\, g$ *is continuous at* c

d) f / g *is continuous at* c *if* $g(c) \neq 0$ *and has a discontinuity at* c *if* $g(c) = 0$.■

We will prove part d). The remaining proofs are similar and will be omitted.

Proof 3.12. First, consider the case where $g(c) = 0$. In this case $f(c)/g(c)$ is undefined, so the function f / g has a discontinuity at c.

Next, consider the case where $g(c) \neq 0$. To prove that f / g is continuous at c, we must show that

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{f(c)}{g(c)}.$$

Since f and g are continuous at c,

$$\lim_{x\to c} f(x) = f(c) \qquad \text{and} \qquad \lim_{x\to c} g(x) = g(c).$$

Thus by Theorem 3.6d

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)} = \frac{f(c)}{g(c)}$$

which proves the above relation.

QED

3.4.4 Continuity of Rational Functions

Since polynomials are continuous everywhere, and since rational functions are ratios of polynomials, part d) of Theorem 3.12 yields the following result.

Theorem 3.13. *Continuity of Rational Functions*

A rational function is continuous at every number where the denominator is nonzero.■

Example 3.19. Continuity of Rational Functions

For what values of x is there a hole or gap in the graph of

$$f(\text{x\_}) := \frac{x^2 - 9}{x^2 - 5x + 6}$$

Solution 3.19. The function being graphed is a rational function, and hence is continuous at every number where the denominator is nonzero. Solving the denominator equation

```
Solve[Denominator[f(x)] == 0, x]
```

$\{\{x \to 2\}, \{x \to 3\}\}$

yields discontinuities at $x = 2$ and $x = 3$. However, the graph of the function shows a discontinuity at $x = 2$ only (see Figure 3.28).

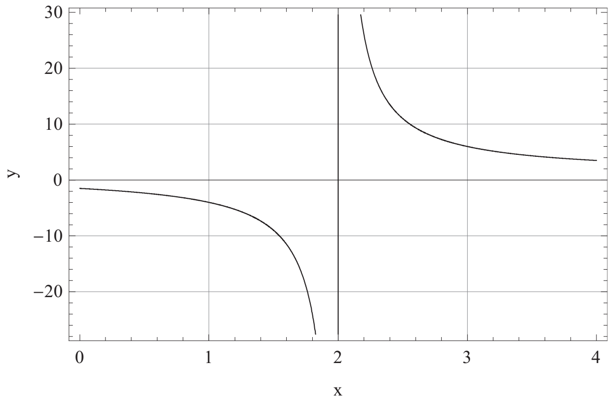


Figure 3.28. Graph of the function $f(x) = \frac{x^2-9}{x^2-5x+6}$.

The reason why this happen becomes clear if we decompose numerator and denominator into factors

factoredNumerator = Factor[Numerator[$f(x)$]]

$(x-3)(x+3)$

factoredDenominator = Factor[Denominator[$f(x)$]]

$(x-3)(x-2)$

So algebraically the singularity at $x = 3$ can be removed by dividing out the term $(x-3)$ in the algebraic expression

$$\frac{\textbf{factoredNumerator}}{\textbf{factoredDenominator}}$$

$$\frac{x+3}{x-2}$$

▲

3.4.5 Continuity of Compositions

The following theorem, whose proof is omitted, will be useful for calculating limits of compositions of functions.

Theorem 3.14. *Continuity of Limits in Compositions*

If $\lim_{x\to c} g(x) = L$ *and if the function* f *is continuous at L, then* $\lim_{x\to c} f(g(x)) = f(L)$. *That is,*

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

This equality remains valid if $\lim_{x\to c}$ is replaced everywhere by one of $\lim_{x\to c^+}$, $\lim_{x\to c^-}$, $\lim_{x\to +\infty}$, or $\lim_{x\to -\infty}$. ■

In words, this theorem states: A limit symbol can be moved through a function sign provided the limit of the expression inside the function sign exists and the function is continuous at this limit.

We know from Example 3.18 that the function $|x|$ is continuous everywhere; thus, it follows that if $\lim_{x\to a} g(x)$ exists, then

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

That is, a limit symbol can be moved through an absolute value sign, provided the limit of the expression inside the absolute value sign exists. For example,

$$\lim_{x\to 3} |5-x^2| = \left|\lim_{x\to 3} \left(5-x^2\right)\right| = |-4| = 4.$$

The following theorem is concerned with the continuity of compositions of functions; the first part deals with continuity at a specific number, and the second part with continuity everywhere.

Theorem 3.15. *Continuity of Compositions*

a) If the function g is continuous at c, and the function f is continuous at g(c), then the composition $f \circ g$ is continuous at c.

b) If the function g is continuous everywhere and the function f is continuous everywhere, then the composition $f \circ g$ is continuous everywhere. ■

Proof 3.15. We will prove part a) only; the proof of part b) can be obtained by applying part a) at an arbitrary number c. To prove that $f \circ g$ is continuous at c, we must show that the value of $f \circ g$ and the value of its limit are the same at $x = c$. But this is so, since we can write

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

QED

We know from Example 3.18 that the function $|x|$ is continuous everywhere. Thus, if $g(x)$ is continuous at c, then by part a) of Theorem 3.15 the function $|g(x)|$ must also be continuous at c; and more generally, if $g(x)$ is continuous everywhere, then so is $|g(x)|$. Stated informally: The absolute value of a continuous function is continuous.

For example, the polynomial

g(x_) := 4 − x^4

is continuous everywhere, so we can conclude that the function $|4-x^4|$ is also continuous everywhere (Figure 3.29)

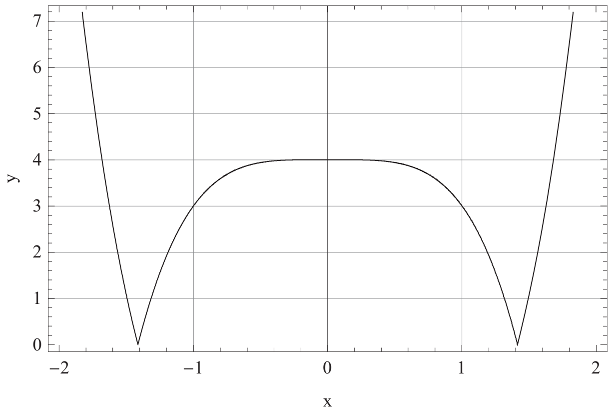


Figure 3.29. Graph of the function $f(x) = |4 - x^4|$.

3.4.6 Continuity from the Right and Left

Because Definition 3.3 involves a two-sided limit, that definition does not generally apply at the endpoints of a closed interval $[a, b]$ or at the endpoint of an interval of the form $[a, b)$, $(a, b]$, $(-\infty, b]$, or $[a, +\infty)$. To remedy this problem, we will agree that a function is continuous at an endpoint. For example, the function graphed in Figure 3.30 is continuous at the right endpoint of the interval $[a, b]$ because

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

but is not continuous at the left endpoint because

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

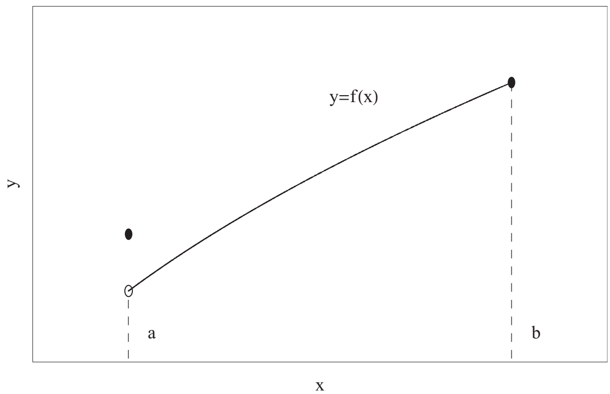


Figure 3.30. Graph of a function with continuous behavior on the right and discontinuous behavior on the left of the closed interval $[a, b]$.

In general, we will say a function f is continuous from the left at c if

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

and is continuous from the right at c if

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

Using this terminology, we define continuity on a closed interval as follows.

Definition 3.4. *Continuity on a Closed Interval*

A function f is said to be continuous on a closed interval $[a, b]$ if the following conditions are satisfied:

1. *f is continuous on (a, b).*
2. *f is continuous from the right at a.*
3. *f is continuous from the left at b.* ■

Remark 3.8. the definition also hold if the interval is replaced by either of the following cases $[a, +\infty)$, $(-\infty, b]$, $(b, a]$, and $[a, b)$.

Example 3.20. Continuity at Endpoints

What can we say about the continuity of the function

$$f(\mathrm{x\_}) := \sqrt{9 - x^2}$$

Solution 3.20. Because the natural domain of this function is the closed interval $[-3, 3]$, we will need to investigate the continuity of f on the open interval $(-3, 3)$ and at the two endpoints. If c is any number in the interval $(-3, 3)$, then it follows from Theorem 3.6e that

$$\lim_{x\to c} f(x) = \lim_{x\to c} \sqrt{9 - x^2} = \sqrt{\lim_{x\to c} \left(9 - x^2\right)} = \sqrt{9 - c^2} = f(c)$$

which proves that f is continuous at each point in the interval $(-3, 3)$. The function f is also continuous at the endpoints since

$$\lim_{x\to 3^-} f(x)$$

0

and

$$\lim_{x\to(-3)^+} f(x)$$

0

Thus f is continuous on the closed interval $[-3, 3]$.▲

3.4.7 Continuity and Limits of Trigonometric Functions

In this section we will investigate the continuity properties of the trigonometric functions, and we will discuss some important limits involving these functions.

Before we begin, recall that in the expressions $\sin(x)$, $\cos(x)$, $\tan(x)$, $\cot(x)$, $\sec(x)$, and $\csc(x)$ it is understood that x is in radian measure.

In trigonometry, the graphs of $\sin(x)$ and $\cos(x)$ are drawn as continuous curves (Figure 3.31). To actually prove that these functions are continuous everywhere, we must show that the following equalities hold for every real number c:

$$\lim_{x\to c} \sin(x) = \sin(c) \quad \text{and} \quad \lim_{x\to c} \cos(x) = \cos(c). \tag{3.31}$$

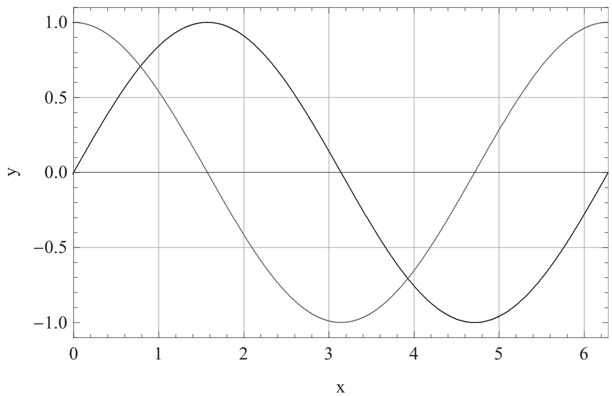
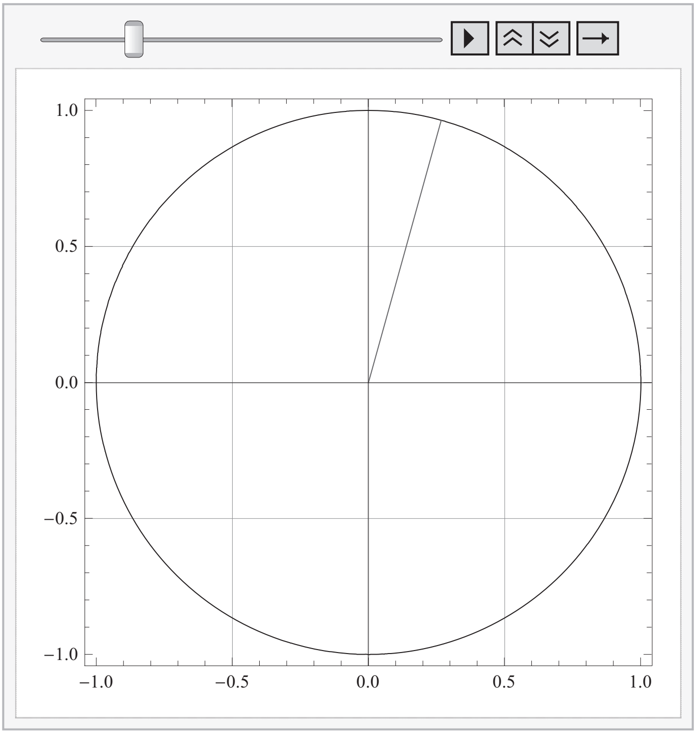


Figure 3.31. Graph of trigonometric function $f(x) = \sin(x)$ and $f(x) = \cos(x)$. The functions shows an oscillating behavior with period 2π.

Although we will not formally prove these results, we can make them plausible by considering the behavior of the point $P(\cos(x), \sin(x))$ as it moves around the unit circle. For this purpose, view c as a fixed angle in radian measure, and let $Q(\cos(c), \sin(c))$ be the corresponding point on the unit circle. As $x \to c$; i.e. as the angle x approaches the angle c, the point P moves along the circle toward Q, and this implies that the coordinates of P approach the corresponding coordinates of Q; that is, $\cos(x) \to \cos(c)$, and $\sin(x) \to \sin(c)$



Formulas (3.31) can be used to find limits of the remaining trigonometric functions by expressing them in terms of $\sin(x)$ and $\cos(x)$; for example, if $\cos(c) \neq 0$, then

$$\lim_{x\to c} \tan(x) = \lim_{x\to c} \frac{\sin(x)}{\cos(x)} = \frac{\sin(c)}{\cos(c)} = \tan(c). \tag{3.32}$$

Thus, we are led to the following theorem.

Theorem 3.16. *Continuity of Trigonometric Functions*

If c is any number in the natural domain of the standard trigonometric functions, then

$\lim_{x\to c} \sin(x) = \sin(c)$

$\lim_{x\to c} \cos(x) = \cos(c)$

$\lim_{x\to c} \tan(x) = \tan(c)$

$\lim_{x\to c} \csc(x) = \csc(c)$

$\lim_{x\to c} \sec(x) = \sec(c)$

$\lim_{x\to c} \cot(x) = \cot(c).$ ■

It follows from this theorem, that $\sin(x)$ and $\cos(x)$ are continuous everywhere and that $\tan(x)$ is continuous, except at the points where it is undefined.

Example 3.21. Continuity of Trigonometric Functions

Find the limit

$$\lim_{x\to 1} \cos\left(\frac{x^2-1}{x-1}\right).$$

Solution 3.21. Recall from theorem for composition that since the cosine function is continuous everywhere

$$\lim_{x\to 1} \cos(g(x)) = \cos(\lim_{x\to 1} g(x))$$

provided $\lim_{x\to 1} g(x)$ exists. Thus,

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

In Section 3.2.1 we used the numerical evidence in Figure 3.3 to conjecture that

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1. \tag{3.33}$$

However, it is not a simple matter to establish this limit with certainty. The difficulty is that the numerator and denominator both approach zero as $x \to 0$. As discussed in Section 3.3.3, such limits are called indeterminate forms of type $0/0$. Sometimes indeterminate forms of this type can be established by manipulating the ratio algebraically, but in this case no simple algebraic manipulation will work, so we must look for other methods.

The problem with indeterminate forms of type $0/0$ is that there are two conflicting influences at work: as the numerator approaches 0 it drives the magnitude of the ratio toward 0, and as the denominator approaches 0 it drives the magnitude of the ratio toward to $\pm\infty$; depending on the sign of the expression. The limiting behavior of the ratio is determined by the precise way in which these influences offset each other. Later in this lecture we will discuss general methods for attacking indeterminate forms, but for the limit in (3.33), we can use a method called squeezing.

In the method of squeezing one proves that a function f has a limit L at a number c by trapping the function between two other functions, g and h, whose limits at c are known to be L (Figure 3.32). This is the idea behind the following theorem, which we state without proof.

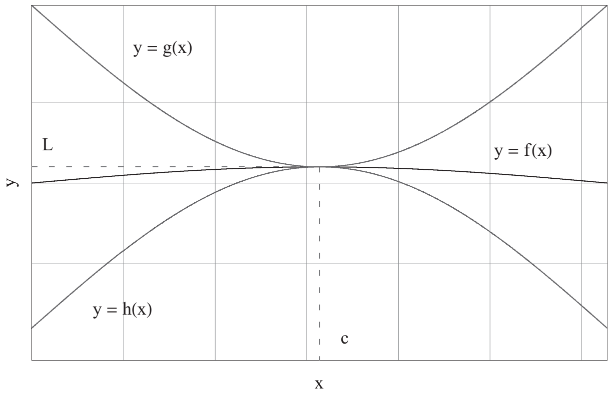


Figure 3.32. Squeezing of functions. Given three functions f, g, and h where g and h are superior or inferior to f.

Theorem 3.17. *Squeezing Theorem*

Let f, g, and h be functions satisfying

$g(x) \leq f(x) \leq h(x)$

for all x in some open interval containing the number c, with the possible exception that the inequalities need not hold at c. If g and h have the same limit as x approaches c, say

$lim_{x \to c}\, g(x) = lim_{x \to c}\, h(x) = L$

then f also has this limit as x approaches c, that is,

$lim_{x \to c}\, f(x) = L.$ ■

The Squeezing Theorem also holds for one-sided limits and limits at $+\infty$ or $-\infty$. How do you think the hypotheses of the theorem would change in those cases?

The usefulness of the Squeezing Theorem will be evident in our proof of the following theorem.

Theorem 3.18. *Special Values*

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

and

$$\lim_{x \to 0} \frac{1 - \cos(x)}{x} = 0. \blacksquare$$

Proof 3.18. In this proof we will interpret x as an angle in radian measure, and we will assume to start that $0 \leq x \leq \pi/2$. It follows from the area formula of a sector of a circle $A = r^2\,\theta/2$ that the area of a sector of radius 1 and central angle x is $x/2$. Moreover, it is suggested by Figure

3.33 that the area of this sector lies between the area of two triangles, one with area $\tan(x)/2$ and one with area $\sin(x)/2$. Thus

$$\frac{\tan(x)}{2} \geq \frac{x}{2} \geq \frac{\sin(x)}{2}.$$

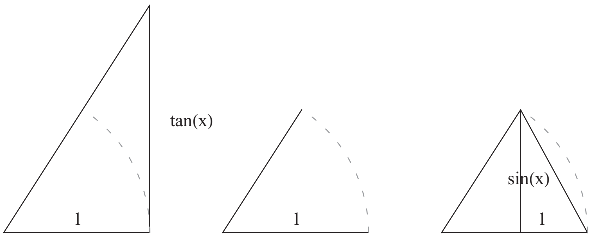


Figure 3.33. The different areas related to the area of the sector.

Multiplying through by $2/\sin(x)$ yields

$$\frac{1}{\cos(x)} \geq \frac{x}{\sin(x)} \geq 1$$

and then taking reciprocals and reversing the inequalities yields

$$\cos(x) \leq \frac{\sin(x)}{x} \leq 1.$$

Moreover, these inequalities also hold for $\frac{-\pi}{2} < x < 0$, since replacing x by $-x$ and using the identities $\sin(-x) = -\sin(x)$ and $\cos(-x) = \cos(x)$ leaves this inequalities unchanged. Finally, since the function $\cos(x)$ and 1 both have limits of 1 as $x \to 0$, it follows from the Squeezing Theorem that $\sin(x)/x$ also has a limit of 1 as $x \to 0$.

For the proof of the second relation we will use the limit of the first relation, the continuity of the sin function, and the trigonometric identity $\sin(x)^2 = 1 - \cos(x)^2$. We obtain

$$\lim_{x\to 0} \frac{1-\cos(x)}{x} = \lim_{x\to 0} \left(\frac{1-\cos(x)}{x} \frac{1+\cos(x)}{1+\cos(x)} \right) =$$

$$\lim_{x\to 0} \frac{\sin(x)^2}{(1+\cos(x))\,x} = \lim_{x\to 0} \frac{\sin(x)}{x} \left(\lim_{x\to x} \frac{\sin(x)}{1+\cos(x)} \right) = (1) \left(\frac{0}{1+1} \right) = 0.$$

QED

3.4.8 Intermediate-Value Theorem

Figure 3.34 shows the graph of a function that is continuous on the closed interval $[a, b]$. The figure suggests that if we draw any horizontal line $y = k$, where k is between $f(a)$ and $f(b)$, then that line will cross the curve $y = f(x)$ at least once over the interval $[a, b]$.

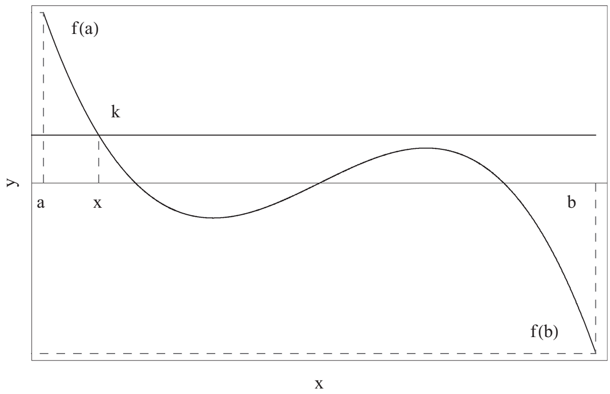


Figure 3.34. Graph of a function with continuous behavior in the interval $[a, b]$.

Stated in numerical terms, if f is continuous on $[a, b]$, then the function f must take on every value k between $f(a)$ and $f(b)$ at least once as x varies from a to b. For example, the polynomial

$$p(x\_) := x^7 - x + 3$$

has a value of 3 at $x = 1$ and a value of 129 at $x = 2$. Thus it follows from the continuity of p that the equation $3 - x + x^7 == k$ has at least one solution in the interval $[1, 2]$ for every value of k between 3 and 129. This idea is stated more precisely in the following theorem.

Theorem 3.19. *Intermediate-Value Theorem*

If f is continuous on a closed interval $[a, b]$ and k is any number between $f(a)$ and $f(b)$, inclusive, then there is at least one number x in the interval $[a, b]$ such that $f(x) = k$. ■

Although this theorem is intuitively obvious, its proof depends on a mathematically precise development of the real number system, which is beyond the scope of this text.

3.4.9 Approximation of Roots

A variety of problems can be reduced to solving an equation $f(x) = 0$ for its roots. Sometimes it is possible to solve for the roots exactly using algebra, but often this is not possible and one must settle for decimal approximations of the roots. One procedure for approximating roots is based on the following consequences of the Intermediate-Value Theorem.

Theorem 3.20. *Root Approximation*

If f is continuous on $[a, b]$, and if $f(a)$ and $f(b)$ are nonzero and have opposite signs, then there is at least one solution of the equation $f(x) = 0$ in the interval (a, b). ■

This result, which is illustrated in Figure 3.35, can be proved as follows.

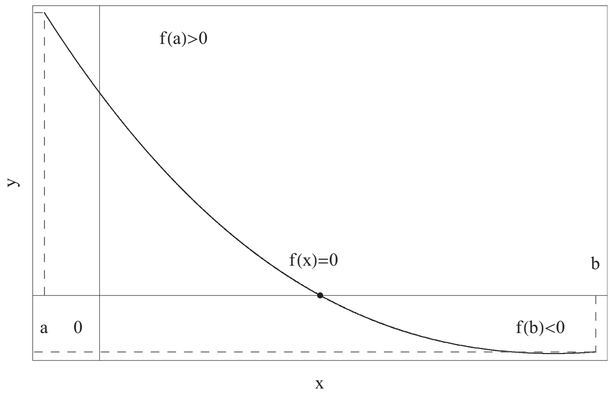


Figure 3.35. Graph of a function $f(x)$ allowing a root in the interval $[a, b]$.

Proof 3.18. Since $f(a)$ and $f(b)$ have opposite signs, 0 is between $f(a)$ and $f(b)$. Thus by the Intermediate-Value Theorem there is at least one number x in the interval $[a, b]$ such that $f(x) = 0$. However, $f(a)$ and $f(b)$ are nonzero, so that x must lie in the interval (a, b), which completes the proof.

QED

Example 3.22. Root finding of a Polynomial

The equation

$$x^5 - x + 8x^2 + 1 = 0$$

can not be solved algebraically because the left hand side has no simple factors. However, if we graph $p(x) = x^5 - x + 8x^2 + 1$ (Figure 3.36), then we are led to the conjecture that there is one real root and that this root lies inside the interval $[-2.1, 2]$. The existence of a root is also confirmed by Theorem 3.20, since $p(-2.1) = -2.46101$ and $p(2) = 63$ have opposite signs. Approximate this root to two decimal-place accuracy.

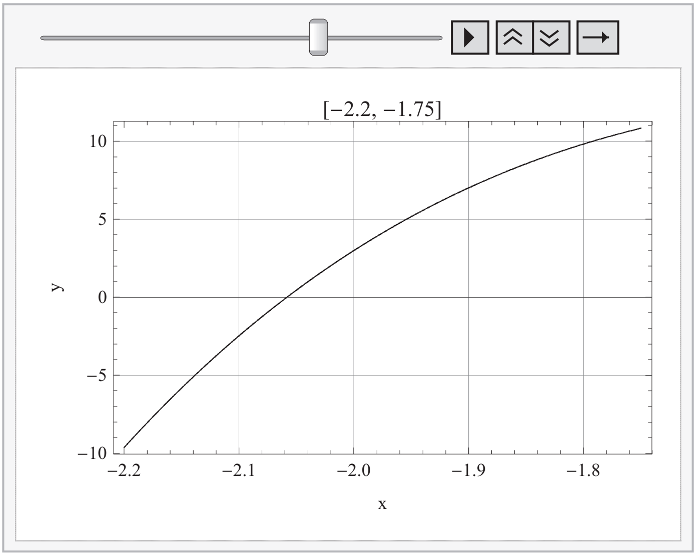


Figure 3.36. Graph of a function $f(x)$ allowing a root in the interval $[a, b]$. Each picture shows a reduction of the interval to locate the root more precisely.

Solution 3.22. The polynomial is defined by

$$p(x\_) := x^5 + 8x^2 - x + 1$$

The following sequence of intervals shrinks the interval length in such a way that the conditions of Theorem 3.20 are satisfied. The intervals are given in curled brackets in the second argument of Map[].

```
Map[{p[#[[1]]], p[#[[2]]]} &, {{-2.1, 2}, {-2.1, -1.8},
    {-2.06, -2.05}, {-2.059, -2.056}, {-2.0585, -2.058}}] //
 TableForm[#, TableHeadings → {{}, {"p(left)", "p(right)"}}] &
```

p(left)	p(right)
-2.46101	63
-2.46101	9.82432
-0.0879704	0.464937
-0.031969	0.135084
-0.00402787	0.0238737

The table shows that the interval is chosen in such a way that the signs of the polynomial $p(x)$ changes. However, the exact value of the root can be determined by the *Mathematica* function FindRoot[].

```
FindRoot[p[x] == 0, {x, -3.1}]
```

```
{x → -2.05843}
```

Stating that the real value of $x = -2.0584$ is the intersection of the polynomial $p(x)$ with the horizontal x-axis.▲

3.4.10 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

3.4.10.1 Test Problems

T1. What does it mean for a function to be continuous? Give examples to illustrate the fact that a function that is not continuous on its entire domain may still be continuous on selected intervals within the domain.

T2. What conditions must be satisfied by a function if it is to be continuous at an interior point of its domain? At an endpoint?

T3. What does it mean for a function to be continuous on an interval?

T4. What are the basic types of discontinuity? Give an example of each. What is a removable discontinuity? Give an example.

T5. It is often said that a function is continuous if you can draw its graph without having to lift your pen from the paper. Why is that?

3.4.10.2 Exercises

E1. Write an equation that expresses the fact that a function f is continuous at the number 4.

E2. If f is continuous on $(-\infty, \infty)$, what can you say about its graph?

E3. If f and g are continuous functions with $f(3) = 5$ and $\lim_{x\to 3} |2 f(x) - g(x)| = 4$, find $g(3)$.

E4. Use the definition of continuity and the properties of limits to show that the functions given are continuous at the given number a.

a. $f(x) = x^2 + \sqrt{8-x}$, $\quad a = 5$,

b. $f(x) = \left(x + 2x^3\right)^4$, $\quad a = -1$,

c. $f(x) = \frac{2x-3x^2}{1+t^5}$, $\quad a = 1$.

E5. Use the definition of continuity and the properties of limits to show that the function is continuous on the given interval.

a. $f(x) = \frac{2x+3}{x-2}$, with $x \in (2, \infty)$,

b. $f(x) = 2\sqrt{3-x}$, with $x \in (-\infty, 3]$.

E6. Explain, why the function is continuous at every number in its domain. State the domain.

a. $f(x) = \frac{x}{x^2+5x+6}$,

b. $f(x) = \sqrt[5]{x}\left(1 + x^7\right)$,

c. $f(x) = e^{-7x}\cos(3x)$,

d. $f(x) = \ln\left(x^5 - 2\right)$,

e. $f(x) = \cos\left(e^{-\sqrt{x^3}}\right)$.

E7. Locate the discontinuities of the function and illustrate by graphing.

a. $f(x) = \frac{1}{1+e^{1/x}}$,

b. $f(x) = \ln\left(\tan\left(\sqrt{x}\right)^2\right)$.

E8. Show that f is continuous on $(-\infty, \infty)$.

a. $f(x) = \begin{cases} x^2 & \text{if } x < 1 \\ \sqrt{x} & \text{if } x \geq 1 \end{cases}$,

$$f(x) = \begin{cases} \sin(x) & \text{if } x < \pi/4 \\ \cos(x) & \text{if } x \geq \pi/4 \end{cases}.$$

E9. Use the Intermediate Value Theorem to show that there is a root of the given equation in the specified interval.

a. $x^4 + x - 3 = 0$ for $x \in (1, 2)$,

b. $\cos(x) = x$ for $x \in (0, 1)$,

c. $\ln(x) = e^{-x}$ for $x \in (1, 2)$,

d. $\sqrt[3]{x} = 1 - x$ for $x \in (0, 1)$

E10 The gravitational force exerted by the earth on a unit mass at a distance r from the center of the planet is

$$F(r) = \begin{cases} \frac{G\,M\,r}{R^3} & \text{if } r < R \\ \frac{G\,M}{r^2} & \text{if } r \geq R \end{cases} \tag{1}$$

where M is the mass of the earth, R is its radius, and G is the gravitational constant. Is F a continuous function of r?

3.5 Derivatives

Many real world phenomena involve changing quantities — the speed of a rocket, the inflation of a currency, the number of bacteria in a culture, the shock intensity of an earthquake, the voltage of an electrical signal, and so forth. In this section we will develop the concept of a derivative, which is the mathematical tool that is used to study rates at which quantities change.

3.5.1 Slope of a Curve

The slope of a curve at a point can be translated into useful information in many applications, so a consideration of the notion of the slop of a curve is warranted.

Consider the function $y = f(x)$ whose graph is shown in Figure 3.37. We focus on the point $P(x_0, f(x_0))$. One has an intuitive notion that the "steepness" of the curve varies at different points. For example, we view the graph of $y = f(x)$ in Figure 3.37 as the cross section of a hill and imagine a hiker walking the hill from left to right.

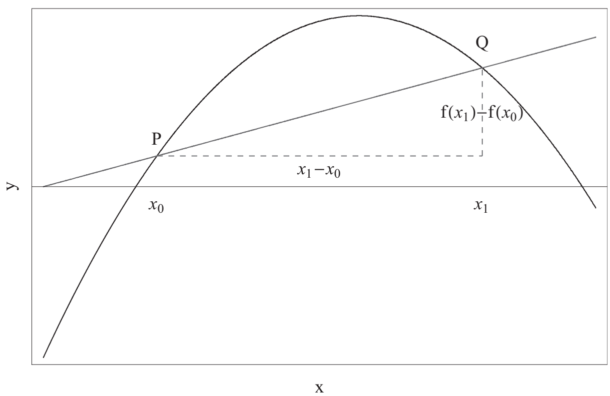


Figure 3.37. Secant line of a function $f(x)$.

The hiker will find the trek fairly arduous at point P, but the climb gets easier as he approaches the summit. Rather than rely on comparative notions of "less steep" or "more steep", we seek a numeric value to attach to each point on the curve that will describe "how steep" the curve is at that point. For straight lines, steepness is the same at every point, and the measure used to describe steepness is the slope of the line. Note that slope not only describes "how steep" a line is, but also whether the line rises or falls. Our goal is to define the slope of a curve $y = f(x)$, even though $f(x)$ isn't linear.

Since we know how to calculate the slope of a line through two points, let us consider a line joining point P with another point $Q(x_1, f(x_1))$ on the curve. By analogy with secants to circles, a line determined by two points on a curve is called a secant line to the curve. The slope of the secant line PQ is given by

$$m_{\text{sc}} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{3.34}$$

As the sampling point Q is chosen closer to P, that is, as x_1 is selected closer to x_0, the slope m_{sc} more nearly approximate what we might reasonably call the "slope" of the curve $y = f(x)$ at the point P. Thus from (3.34), the slop of the curve $y = f(x)$ at $P(x_0, f(x_0))$ should be defined by

$$m = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{3.35}$$

Example 3.23. Slope of a Function

Consider the function

$$\boldsymbol{f(\text{x\_}) := 6\,x - x^2}$$

and the point $P(2, f(2)) = (2, 8)$.

Find the slope of secant lines to the graph of $y = f(x)$ determined by P and points on the graph at $x = 3$ and $x = 1.5$. Also find the slope of the graph of $y = f(x)$ at the point P.

Solution 3.23. The secant line to the graph of f through P and $Q(1.5, f(1.5)) = (1.5, 6.75)$ has slope

$$\boldsymbol{m = \frac{6.75 - 8}{1.5 - 2}}$$

2.5

The secant line to the graph f through P and $Q(3, f(3)) = (3, 9)$ has slope

$$\boldsymbol{m = \frac{9 - 8}{3 - 2}}$$

1

The slope of the graph of f at the point P is

$$m = \lim_{x_1 \to 2} \frac{f(x_1) - f(2)}{x_1 - 2}$$

2

▲

We now have an interpretation of a limit of slopes of secant lines as the slope of the position curve at the instant in question.

3.5.2 Slopes and Rate of Change

The slope can be viewed as rate of change — the rate of change of position with respect to time, or the rate of change of a function's value with respect to its input. Rates of change occur in many applications. For example:

- An economist might be interested in the rate at which production cost changes with the quantity of a product that is manufactured.
- A microbiologist might be interested in the rate at which the number of bacteria in a colony changes with time.
- An engineer might be interested in the rate at which the length of a metal rod changes with temperature.
- A medical researcher might be interested in the rate at which the radius of an artery changes with the concentration of alcohol in the bloodstream.

In general, if x and y are quantities related by an equation $y = f(x)$, we can consider the rate at which y changes with x. We distinguish between an average rate of change, represented by the slope of a secant line to the graph of $y = f(x)$, and an instantaneous rate of change, represented by the slope of the curve at a point.

Definition 3.5. *Average Rate of Change*

If $y = f(x)$, then the average rate of change of y with respect to x over the interval $[x_0, x_1]$ is

$$r = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \blacksquare \tag{3.36}$$

Geometrically, the average rate of change of y with respect to x over the interval $[x_0, x_1]$ is the slope of the secant line to the graph $y = f(x)$ through the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$:

$$r = m_{sc}$$

(see Figure 3.38)

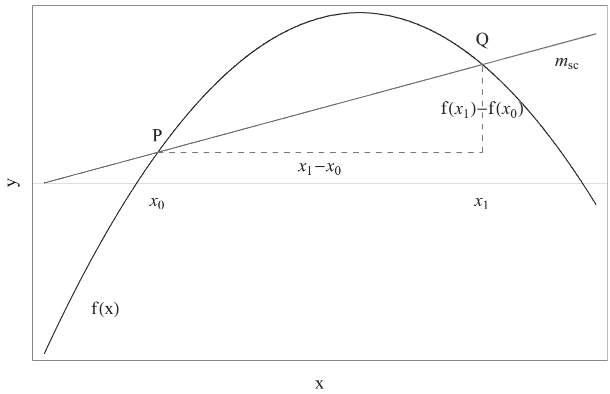


Figure 3.38. m_{sc} is the average rate of change of y with respect to x over the interval $[x_0, x_1]$.

Definition 3.6. *Instantaneous Rate of Change*

If $y = f(x)$, then the instantaneous rate of change of y with respect to x when $x = x_0$ is

$$m = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \blacksquare \tag{3.37}$$

Geometrically the instantaneous rate of change of y with respect to x when $x = x_0$ is the slope of the graph of $y = f(x)$ at the point $(x_0, f(x_0))$ see Figure 3.39.

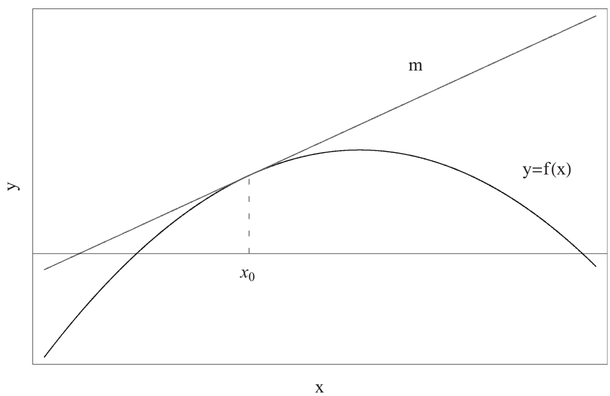


Figure 3.39. Instantaneous rate of change of y with respect to x when $x = x_0$.

Example 3.24. Rate of Change

Find the average and instantaneous rate of change in the interval [3, 5] and at $x = -4$ for the function

$$f(x\_) := x^2 + 1$$

Solution 3.24. We apply formula (3.36) with $x_0 = 3$ and $x_1 = 5$. This yields

$$m = \frac{f(5) - f(3)}{5 - 3}$$

8

Thus on the average, y increases 8 units per unit increase in x over the interval [3, 5]. The instantaneous rate of change follows from formula (3.37) at $x_0 = 4$

$$r = \lim_{x_1 \to -4} \frac{f(x_1) - f(-4)}{x_1 + 4}$$

−8

Thus, for a small change in x from $x = -4$, the value of y will change approximately eight times as much in the opposite direction. That is, because the instantaneous rate of change is negative, the value of y decreases as values of x move through $x = -4$ from left to right.

On the other hand if the x_1 values approach any value $x = x_0$ then the limiting process gives

$$r = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

2 x0

Thus the instantaneous rate of change of y with respect to x at $x = x_0$ is $2\,x_0$. Observe that the result of the numeric calculation can be obtained from this more general result by setting $x_0 = -4$.▲

3.5.3 Derivative

In the preceding subsections we argued that the slope of a graph of $y = f(x)$ at $x = x_0$ should be given by

$$m = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{3.38}$$

The ratio

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{3.39}$$

is called a **difference quotient**. As we saw the difference quotient can be interpreted as the average rate of change of $f(x)$ over the interval $[x_0\, x_1]$, and its limit as $x_1 \to x_0$ is the instantaneous rate of change of $f(x)$ at $x = x_0$.

The geometric problem of finding the slope of a curve, and the somewhat paradoxical notions of instantaneous rate of change, are all resolved by a limit of a difference quotient. The fact that problems in such disparate areas are unified by this expression is celebrated in the definition of the derivative of a function at a value in its domain.

Definition 3.7. *Derivative*

Suppose that x_0 is a number in the domain of a function f. If

$$\lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

exists, then the value of this limit is called the derivative of f at $x = x_0$ and is denoted by $f'(x_0)$. That is

$$f'(x_0) = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

If the limit of the difference quotient exists, $f'(x_0)$ is the slope of the graph of f at the point $P(x_0, f(x_0))$ (or at $x = x_0$). If this limit does not exist, then the slope of the graph of f is undefined at P (or at $x = x_0$). ■

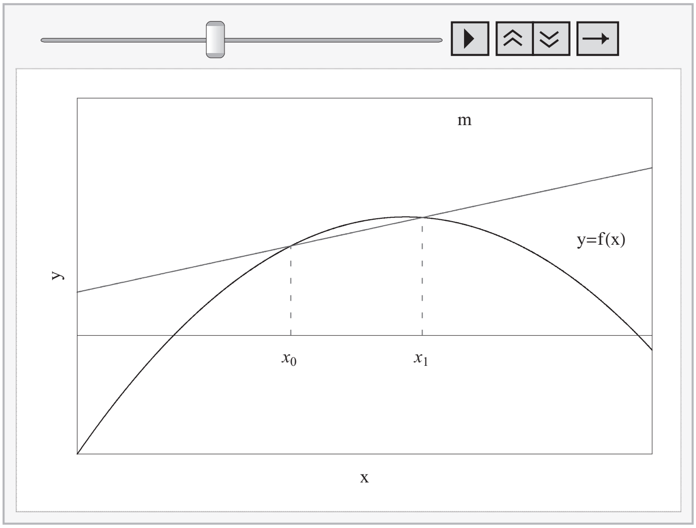


Figure 3.40. Movement of the x_1 value toward x_0 starting from a secant line generating a tangent line.

Now that we have defined the derivative of a function, we can begin to answer a question that fueled much of the early development of calculus. Mathematicians of the seventeenth century were perplexed by the problem of defining a tangent line to a general curve. Of course, in the case of a circle the definition was apparent: a line is tangent to a circle if it meets the circle at a single point. But it was also clear that this simple definition would not suffice in many cases. For example, the y-axis intersects the parabola $y = x^2$ at a single point but does not appear to be tangent to the curve. On the other hand, the line $y = 1$ does seem to be tangent to the graph $y = \sin(x)$ even though it intersects this graph infinitely often (Figure 3.41).

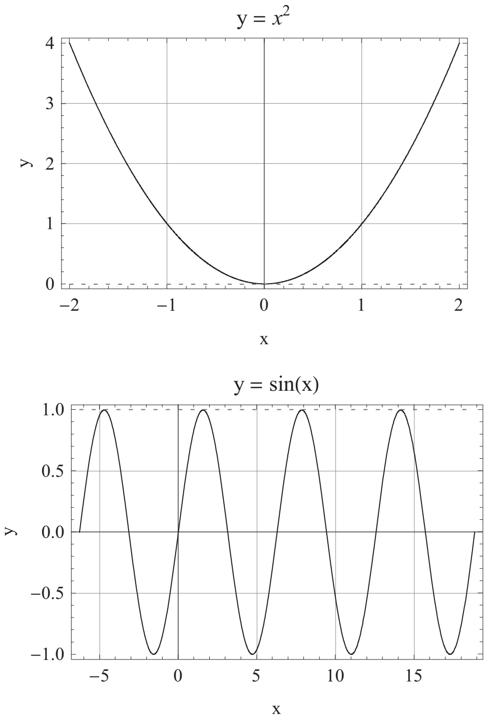


Figure 3.41. Tangent lines on functions (dashed line).

By the end of the first half of the seventeenth century, mathematicians such as Descartes and Fermat had developed a variety of procedures for constructing tangent lines. However, a general definition of a tangent line to a curve was still missing. Roughly speaking, a line should be tangent to the graph of a function $y = f(x)$ at a point $(x_0, f(x_0))$ provided the line has the same direction as the graph at the point. Since the direction of a line is determined by its slope, we would expect a line to be tangent to the graph at $(x_0, f(x_0))$ if the slope of the line is equal to the slope of the graph of f at x_0. Thus, we can now use the derivative to define the tangent line to a curve when the curve is the graph of a function $y = f(x)$.

Definition 3.8. *Tangent Line*

Suppose that x_0 is a number in the domain of a function f. If

$$f'(x_0) = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{3.40}$$

exists, then we define the tangent line to the graph of f at the point $P(x_0, f(x_0))$ to be the line whose equation is

$$y - f(x_0) = f'(x_0)(x - x_0). \tag{3.41}$$

We also call this the tangent line to the graph of f at $x = x_0$. ■

Remark 3.9. Tangent lines to graphs do not have all of the properties of tangent lines to circles. For example, a tangent line to a circle intersects the circle only at the point of tangency, whereas a tangent line to a general graph may intersect the graph at points other than the point of tangency. The most prominent example is a line the tangent to a line is the line itself.

In general the slope of a curve $y = f(x)$ will depend on the point $(x, f(x))$ at which the slope is computed. That is, the slope is itself function of x. To illustrate this let us use (3.40) to compute $f'(x_0)$ at a general x-value x_0 for the curve $y = x^2 + 1$. the computations are

$$\mathbf{f(x\_) := x^2 + 1}$$

$$\mathbf{fp = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}}$$

```
2 x0
```

Now we can use the general formula $f'(x_0) = 2\,x_0$ to compute the slope of the tangent line at any point along the curve $y = x^2 + 1$ simply by substituting the appropriate value for $x = x_0$. Fore example if $x_0 = 2$ $f'(2) = 4$.

To generalize this idea replacing x_0 by x in (3.40), the slope of a graph of $y = f(x)$ at a general point $(x, f(x))$ is given by

$$f'(x) = \lim_{x_1 \to x} \frac{f(x_1) - f(x)}{x_1 - x}. \tag{3.42}$$

The fact that this describes the slope-producing function is so important that there is a common terminology associated with it.

Definition 3.9. *Derivative of f*

The function f' defined by the formula

$$f'(x) = \lim_{w \to x} \frac{f(w) - f(x)}{w - x}. \tag{3.43}$$

is called the derivative of f with respect to x. The domain of f' consist of all x in the domain of f for which the limit exists. ■

Remark 3.10. Despite the presence of the symbol w in the definition, formula (3.43) defines the function f' as a function of the single variable x. To calculate the value of $f'(x)$ at a particular input value x, we fix the value of x and let $w \to x$ in (3.43). The answer to this limit no longer involves the symbol w; w disappears at the step in which the limit is evaluated.

Recalling from the last section that the slope of the graph of $y = f(x)$ can be interpreted as the instantaneous rate of change of y with respect to x, it follows that the derivative of a function f can be interpreted in several ways:

The derivative f' of a function f can be interpreted as a function whose value at x is the slope of the graph of $y = f(x)$ at x, or, alternatively, it can be interpreted as a function whose value at x is the instantaneous rate of change of y with respect to x at x.

Example 3.25. Derivative

Find the derivative with respect to x of $f(x) = x^3 - x$. Graph f and f' together, and discuss the relationship between the two graphs.

Solution 3.25. Later in this chapter we will develop efficient methods for finding derivatives, but for now, we will find the derivative directly from formula (3.40) in the definition of f'. The computation are as follows

$$f(\text{x\_}) := x^3 - x$$

$$\text{fpr} = \lim_{w \to x} \frac{f(w) - f(x)}{w - x}$$

```
-1 + 3 x^2
```

The graph of these two expressions is shown in Figure 3.42. Since $f'(x)$ can be interpreted as the slope of the graph of $y = f(x)$ at x, the derivative $f'(x)$ is positive where the graph of f has positive slope, it is negative where the graph of f has negative slope, and it is zero where the graph of f is horizontal.

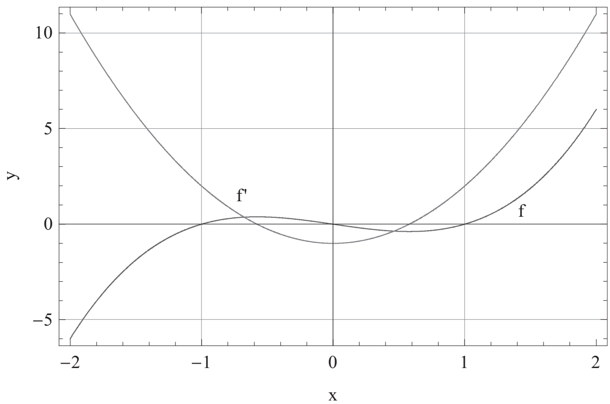


Figure 3.42. Graph of the function $f(x) = x^3 - x$ and its derivative $f'(x) = 3x^2 - 1$.▲

Example 3.26. Derivative of a Line

At each value of x, the tangent line to a line $y = m x + b$ coincides with the line itself and hence all tangent lines have slope m. This suggests geometrically that if $f(x) = m x + b$, then $f'(x) = m$ for all x.

Solution 3.26. This is confirmed by the following computation

$$f(\text{x\_}) := b + m x$$

$$\lim_{w \to x} \frac{f(w) - f(x)}{w - x}$$

m

Note: compare this result with the discussion about lines in Section 2.8.▲

3.5.4 Differentiability

Observe that a function f must be defined at $x = x_0$ in order for the difference quotient

$$\frac{f(w) - f(x_0)}{w - x_0} \tag{3.44}$$

to make sense, since this quotient references a value for $f(x_0)$. Since a value for $f(x_0)$ is required before the limit of this quotient can be considered, values in the domain of the derivative function f' must also be in the domain of f.

Theorem 3.21. *Differentiability*

For a number x_0 in the domain of a function f, we say that f is differentiable at x_0, or that the derivative of f exists at x_0, if

$$\lim_{w \to x_0} \frac{f(w) - f(x_0)}{w - x_0}$$

exists. ■

Thus, the domain of f' consists of all values of x at which f is differentiable. If x_0 is not in the domain of f or if the limit does not exist, then we say that f is not differentiable at x_0, or that the derivative of f does not exist at x_0. If f is differentiable at every value of x in an open interval (a, b), then we say that f is differentiable on (a, b). This definition also applies to infinite open intervals of the form $(a, +\infty)$, $(-\infty, b)$, and $(-\infty, +\infty)$. In the case where f is differentiable on $(-\infty, +\infty)$, we will say that f is differentiable everywhere.

Geometrically, if f is differentiable at a value x_0 for x, then the graph of f has a tangent line at x_0. If f is defined at x_0 but is not differentiable at x_0, then either the graph of f has no well-defined tangent line at x_0 or it has a vertical tangent line at x_0. Informally, the most commonly encountered circumstances of non differentiability occur where the graph of f has

- a corner
- a vertical line, or
- a discontinuity.

Figure 3.43 illustrates each of these situations.

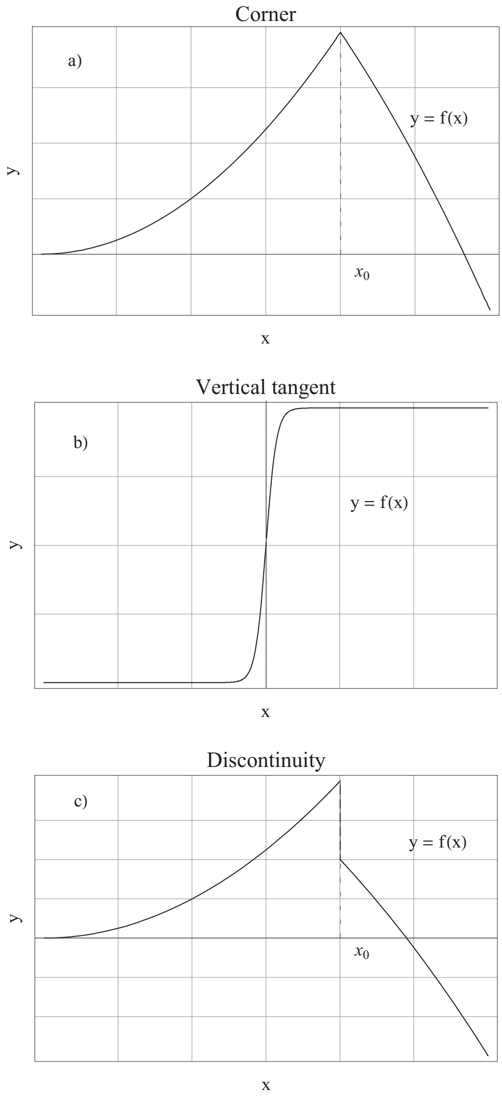


Figure 3.43. Different situations of non differentiability.

It makes sense intuitively that a function is not differentiable where its graph has a corner, since there is no reasonable way to define the graph's slope at a corner. For example, Figure 3.43a and 3.43c shows a typical corner point $P(x_0, f(x_0))$ on the graph of a function f. At this point, the slopes of secant lines joining P and nearby points Q have different limiting values, depending on whether Q is to the left or to the right of P. Hence, the slope of the secant lines do not have a two-sided limit.

A vertical tangent line occurs at a place on a continuous curve where the slopes of secant lines approach $+\infty$ or approach $-\infty$. Since an infinite limit is a special way of saying that a limit does not exist, a function f is not differentiable at a point of vertical tangency (Figure 3.43b).

Example 3.27. Differentiability

The graph of $y = |x|$ has a corner at $x = 0$, which implies that $f(x) = |x|$ is not differentiable at $x = 0$. Show this.

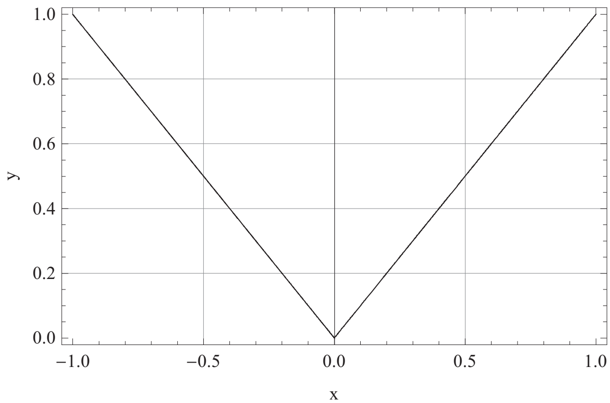


Figure 3.44. Graph of the function $f(x) = |x|$.

Solution 3.27. We define the function $f(x) = |x|$ by

$$f(\text{x\_}) := |x|$$

From formula (3.43) with $x = 0$, the value of $f'(0)$, is given by the limits

$$\lim_{w \to 0^-} \frac{f(w) - f(0)}{w + 0}$$

```
-1
```

and

$$\lim_{w \to 0^+} \frac{f(w) - f(0)}{w + 0}$$

```
1
```

the values of these limits are different and thus the limit does not exist and so does not the derivative.▲

It makes sense intuitively that a function f cannot be differentiable where it has a jump discontinuity, since the value of the function changes precipitously at the jump. The following theorem shows that a function f must be continuous at a value x_0 in order for it to be differentiable there (or stated another way, a function f cannot be differentiable where it is not continuous).

Theorem 3.22. *Differentiability and Continuity*

If f is differentiable at $x = x_0$, then f must also be continuous at x_0. ■

Proof 3.22. We are given that f is differentiable at x_0, so it follows from (3.43) that $f'(x_0)$ exists and is given by

$$f'(x_0) = \lim_{w \to x_0} \frac{f(w) - f(x_0)}{w - x_0}.$$

To show that f is continuous at x_0, we must show that

$$\lim_{w \to x_0} f(w) = f(x_0)$$

or equivalently

$$\lim_{w \to x_0} (f(w) - f(x_0)) = 0.$$

However, this can be proved using (3.40) as follows

$$\lim_{w \to x_0} (f(w) - f(x_0)) =$$

$$\lim_{w \to x_0} \left(\frac{f(w) - f(x_0)}{w - x_0} (w - x_0) \right) = \lim_{w \to x_0} \left(\frac{f(w) - f(x_0)}{w - x_0} \right) \lim_{w \to x_0} (w - x_0)$$

$$= \quad f'(x_0)\, 0 = 0.$$

QED

Remark 3.11. The converse to Theorem 3.22 is false. That is, a function may be continuous at an input value, but not differentiable there. For example, the function $f(x) = |x|$ is continuous at $x = 0$ but not differentiable at $x = 0$. In fact, any function whose graph has a corner and is continuous at a location of the corner will be continuous but not differentiable at the corner.

The relationship between continuity and differentiability was of great historical significance in the development of calculus. In the early nineteenth century mathematicians believed that if a continuous function had many points of non differentiability, these points, like the tips of a saw blade, would have to be separated from each other and joined by smooth curve segments. This misconception was shattered by a series of discoveries beginning in 1834. In that year a Bohemian priest, philosopher, and mathematician named Bernhard Bolzano discovered a procedure for constructing a continuous function that is not differentiable at any point. Later, in 1860, the great German mathematician Karl Weierstrass produced the first formula for such functions. The graphs of such functions are impossible to draw; it is as if the corners are so numerous that any segment of the curve, when suitably enlarged, reveals more corners. The discovery of these pathological functions was important in that it made mathematicians distrustful of their geometric intuition and more relaid at on precise mathematical proof. However, these functions remained only mathematical curiosities until the early 1980s, when applications of them began to emerge. During recent decades, such functions have started to play a fundamental role in the study of geometric objects called fractals. Fractals have revealed an order to natural phenomena that were previously distinguished as random or chaotic.

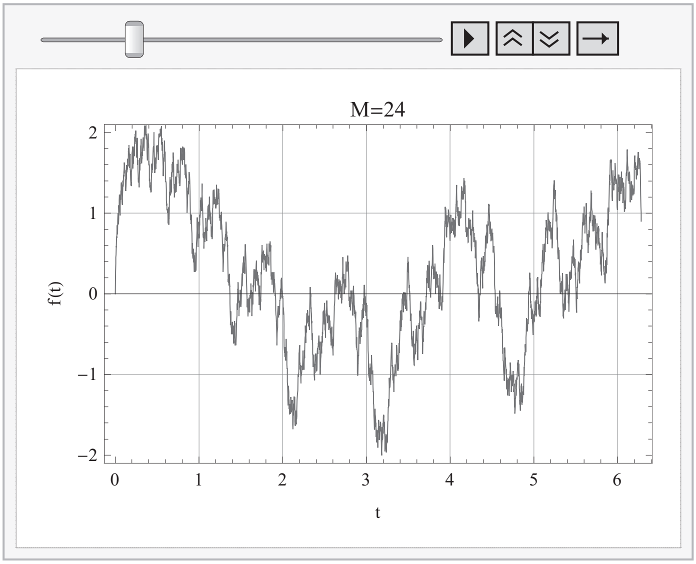


Figure 3.45. Approximation of the Weierstrass function which is nowhere differentiable.

3.5.5 Derivative Notation

The process of finding a derivative is called differentiation. You can think of differentiation as an operation on functions that associates a function f' with a function f. When the independent variable is x, the differentiation operation is often denoted by $\frac{d}{dx} f(x) = \frac{\partial f(x)}{\partial x}$ which is read the derivative of $f(x)$ with respect to x. Thus

$$\frac{d}{dx} f(x) = \frac{\partial f(x)}{\partial x} = f'(x). \tag{3.45}$$

For example, with this notation the derivatives obtained in Example 3.26 and 3.27 can be expressed as

$$\frac{d}{dx} \left(x^3 - x\right) = 3\, x^2 - 1 \tag{3.46}$$

and

$$\frac{d}{dx} (m\, x + b) = m. \tag{3.47}$$

To denote the value of a derivative at a specific value $x = x_0$ with the notation (3.45), we would write

$$\frac{d}{dx} f(x)\,|_{x=x_0} = f'(x_0). \tag{3.48}$$

For example for the derivatives above we find

$$\frac{d}{\mathrm{dx}}\left(x^3 - x\right)|_{x=1} = \left(3\,x^2 - 1\right)|_{x=1} = \left(3 \times 1^2 - 1\right) = 2 \tag{3.49}$$

and

$$\frac{d}{\mathrm{dx}}(m\,x + b)\,|_{x=5} = m\,|_{x=5} = \mathrm{m}. \tag{3.50}$$

When letters other than x and y are used for the independent and dependent variables, then the various notations for the derivative must be adjusted accordingly. For example, if $y = f(u)$, then the derivative with respect to u would be written as

$$\boldsymbol{\frac{\partial f(u)}{\partial u}}$$

$f'(u)$

In particular if

$$\boldsymbol{y = u^3 - u}$$

$u^3 - u$

then the derivative with respect to u is

$$\boldsymbol{\frac{\partial y}{\partial u}}$$

$-1 + 3\,u^2$

3.5.6 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

3.5.6.1 Test Problems

T1. What is the significance of the formula

$$\lim_{h\to 0} \frac{f\,(x+h) - f(x)}{h}\,? \tag{2}$$

Interpret the formula geometrically and physically.

T2. What does it mean for a line to be tangent to a curve C at a point P?

T3. How do you find the tangent to the curve $y = f(x)$ at a point $(x_0,\ y_0)$ on the curve?

T4. How does the slope of the curve $y = f(x)$ at $x = x_0$ relate to the function's rate of change with respect to x at $x = x_0$? To the derivative of f at x_0?

3.5.6.2 Exercises

E1. A curve has equation $y = f(x)$.

a. Write an expression for the slope of the secant line through the points $P(4,\ f(4))$ and $Q(x,\ f(x))$.

b. Write an expression for the slope of the tangent line at P.

E2. Find an equation of the tangent line to the curve at the given point.

$y = \frac{x-2}{x-3}$ at (2, 0),

b. $y = \frac{2x}{(x+2)^2}$ at (1, 2/9),

c. $y = \sqrt{x}$ at (4, 2),

d. $y = 3x^2 - 4x$ at (−2, 20)

E3. Find the slope of the tangent to the curve $y = 3 + 4x^2 - 3x^3$ at the point $x = a$.

a. Find equations of the tangent lines at $x = 1$ and $x = 2$.

b. Graph the curve and both tangents on a common screen.

E4. Find the slope of the tangent to the curve $y = 1/\sqrt{x}$ at the point $x = a$.

a. Find equations of the tangent lines at $x = 1$ and $x = 4$.

b. Graph the curve and both tangents on a common screen.

E5. If a ball is thrown into the air with a velocity of 40 m/s, its height after t seconds is given by $y = 40t - 16t^2$. Find the velocity when $t = 2$.

E6. If $f(x) = 1 - x^4$, find $g'(0)$ and use it to find an equation of the tangent line to the curve $y = 1 - x^4$ at the point (0, 1).

E7. Each limit represents the derivative of some function f at some number a. State such an f and a in each case.

a. $\lim_{x\to1} \frac{x^4+x-2}{x-1}$,

b. $\lim_{x\to0} \frac{\sqrt[4]{16+x}-2}{x}$,

c. $\lim_{x\to0} \frac{(1+x)^{10}-1}{x}$,

d. $\lim_{x\to5} \frac{2^x-32}{x-5}$,

e. $\lim_{x\to0} \frac{\cos(\pi+x)+1}{x}$,

f. $\lim_{x\to\pi/4} \frac{\tan(x)-1}{x-\pi/4}$.

E8. A particle moves along a straight line with equation of motion $s = f(t)$, where s is measured in meters and t in seconds. Find the velocity and the speed when $t = 5$ for:

a. $f(t) = 150 + 50t - 4.5t^2$,

b. $f(t) = t^{-1} - t$.

E9. If a cylindrical tank holds 100 000 liter of water, which can be drained from the bottom of the tank in an hour, then Torricelli's Law gives the volume of water remaining in the tank after t minutes as

$$V(t) = 100\,000\left(1 - \frac{t}{60}\right)^2 \text{ with } 0 \le t \le 60 \tag{1}$$

Find the rate at which the water is flowing out of the tank (the instantaneous rate of change of V with respect to t) as a function of t. What are its units? For times $t = 0$, 10, 20, 30, 40, 50, and 60 min, find the flow rate and the amount of water remaining in the tank. Summarize your findings in a sentence or two. At what time is the flow rate the greatest? The least?

E10 Determine whether $f'(0)$ exists.

a. $f(x) = \begin{cases} x\sin\left(\frac{1}{x}\right) & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$,

b. $f(x) = \begin{cases} x^2\sin\left(\frac{1}{x}\right) & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$.

3.6 Differentiation Rules

The following sections collect rules for derivatives applied to arbitrary functions. These rules are basic in a sense that they can be applied to any kind of function. They are also useful to calculate general expressions.

3.6.1 Derivative of a Constant

The graph of a constant function $f(x) = c$ is a horizontal line $y = c$, and hence the tangent line to this graph has slope 0 at every value of x. Thus we should expect the derivative of a constant function to be 0 for all x.

Theorem 3.23. *Differentiation of a Constant*

The derivative of a constant function is 0; that is, if c is any real number, then

$$\frac{\partial c}{\partial x} = 0. \blacksquare$$

Proof 3.22. Let $f(x) = c$. Then from the definition of the derivative,

$$\frac{\partial c}{\partial x} = f'(x) = \lim_{w\to x} \frac{f(w) - f(x)}{w - x} = \lim_{w\to x} \frac{c - c}{w - x} = \lim_{w\to x} 0 = 0.$$

Example 3.28. Differentiation of a Constant

Let us define the constant function $f(x) = 5$ by

f(x_) := 5

for all x, then $f'(x) = 0$ for all x; that is

$$\frac{\partial f(x)}{\partial x}$$

0

▲

3.6.2 Derivative of a Power

For our next derivative rule, we will need the algebraic identity

$$w^n - x^n = (w - x)\left(w^{n-1} + w^{n-2}\, x + w^{n-3}\, x^2 + \ldots + w\, x^{n-2} + x^{n-1}\right) \tag{3.51}$$

which is valid for any positive integer n. This identity may be verified by expanding the right-hand side of the equation and noting the cancellation of terms. For example, with $n = 4$ we have

Expand$\left[(w - x)\left(w^3 + w^2\, x + w\, x^2 + x^3\right)\right]$

$w^4 - x^4$

Theorem 3.24. *Power Rule*

If n is a positive integer, then

$$\frac{\partial}{\partial x} x^n = n\, x^{n-1}. \blacksquare$$

Proof 3.24. Let $f(x) = x^n$. Then from the definition of the derivative we obtain

$$\begin{aligned}
\left(\frac{\partial x^n}{\partial x}\right) = f'(x) &= \lim_{w\to x} \frac{f(w) - f(x)}{w - x} = \lim_{w\to x} \frac{w^n - x^n}{w - x} \\
&= \lim_{w\to x} \frac{(w - x)\left(w^{n-1} + w^{n-2}\,x + w^{n-3}\,x^2 + \ldots + w\,x^{n-2} + x^{n-1}\right)}{w - x} \\
&= \lim_{w\to x} \left(w^{n-1} + w^{n-2}\,x + w^{n-3}\,x^2 + \ldots + w\,x^{n-2} + x^{n-1}\right) \\
&= x^{n-1} + x^{n-1} + \ldots + x^{n-1} + x^{n-1} \\
&= n\,x^{n-1}
\end{aligned}$$

QED

The derivative of x raised to a positive power is the product of the integer exponent and x raised to the next lower integer power.

Example 3.29. Power Rule

Let

$$\boldsymbol{f}(\mathbf{x\_}) := \boldsymbol{x}^5$$

then the derivative of this function is

$$\frac{\partial f(x)}{\partial x}$$

$$5\,x^4$$

Other examples are

$$\frac{\partial x}{\partial x}$$

$$1$$

or with a larger power

$$\frac{\partial x^{17}}{\partial x}$$

$$17\,x^{16}$$

▲

3.6.3 Derivative of a Constant Times a Function

If a function is multiplied by a constant we should know how the product behaves in a differentiation process. The following theorem tells you that a constant in a product can just be extracted from the differentiation.

Theorem 3.25. *Constants in Functions*

If f is differentiable at x and c is any real number, then $c\, f$ is also differentiable at x and

$$\frac{\partial(c\, f(x))}{\partial x} = c\, \frac{\partial f(x)}{\partial x}.\blacksquare$$

Proof 3.25.

$$\frac{d}{\mathrm{dx}}(c\, f(x)) = \lim_{w\to x} \frac{c\, f(w) - c\, f(x)}{w - x} = \lim_{w\to x} c \lim_{w\to x} \frac{f(w) - f(x)}{w - x} = c\, \frac{d}{d\,x} f(x).$$

QED

A constant factor can be moved through a derivative sign.

Example 3.30. Function Multiplied by a Constant

Given the function

$$\boldsymbol{f(\mathrm{x\_}) := 4\, x^8}$$

THe derivative follows by

$$\boldsymbol{\frac{\partial f(x)}{\partial x}}$$

$$32\, x^7$$

Another example is

$$\boldsymbol{\frac{\partial\left(\pi\, x^{12}\right)}{\partial x}}$$

$$12\, \pi\, \mathrm{x}^{11}$$

and still another one shows

$$\boldsymbol{\frac{\partial \frac{x^{156}}{\pi}}{\partial x}}$$

$$\frac{156\, x^{155}}{\pi}$$

▲

3.6.4 Derivative of Sums and Differences

In addition to products with constants we should know how a sum of functions is treated in a differentiation. The following theorem summarizes this

Theorem 3.26. *Sum of Functions*

If f and g are differentiable at x, then so are $f+g$ and $f-g$ and

$$\frac{\partial(f(x)+g(x))}{\partial x}=\frac{\partial f(x)}{\partial x}+\frac{\partial g(x)}{\partial x}$$

and

$$\frac{\partial(f(x)-g(x))}{\partial x}=\frac{\partial f(x)}{\partial x}-\frac{\partial g(x)}{\partial x}.\blacksquare$$

Proof 3.26.

$$\frac{d}{\mathrm{dx}}(f(x)+g(x)) = \lim_{w\to x}\frac{f(w)+g(w)-(f(x)+g(x))}{w-x}=$$

$$\lim_{w\to x}\frac{f(w)-f(x)}{w-x}+\lim_{w\to x}\frac{g(w)-g(x)}{w-x}=\frac{d}{dx}f(x)+\frac{d}{dx}g(x)$$

The proof of $f-g$ is similar (take it as an exercise).

QED

The derivative of a sum equals the sum of the derivatives, and the derivative of a difference equals the difference of the derivatives.

Example 3.31. Sum and Difference of Functions

The following example show the differentiation of sums and differences

$$\frac{\partial\left(x^7+x^5\right)}{\partial x}$$

$$7\,x^6+5\,x^4$$

A difference of two monomials is

$$\frac{\partial\left(6\,x^{12}-9\,x^5\right)}{\partial x}$$

$$72\,x^{11}-45\,x^4$$

▲

Although Theorem 3.26 was stated for sums and differences of two terms, it can be extended to any mixture of finitely many sums and differences of differentiable functions. For example,

$$\frac{\partial\left(9\,x^5 - 8\,x^4 + 7\,x^3 + 9\right)}{\partial x}$$

$$45\,x^4 - 32\,x^3 + 21\,x^2$$

3.6.5 Derivative of a Product

Derivatives of products are determined due to the product rule of differentiation.

Theorem 3.27. *The Product Rule*

If f and g are differentiable at x, then so is the product f g and

$$\frac{\partial(f(x)\,g(x))}{\partial x} = f(x)\,\frac{\partial g(x)}{\partial x} + g(x)\,\frac{\partial f(x)}{\partial x}.\blacksquare$$

Proof 3.27. The earlier proofs in the previous sections were straightforward applications of the definition of the derivative. However, this proof requires a trick — adding and subtracting the term $f(w)\,g(x)$ to the numerator in the derivative definition as follows

$$\frac{d}{\mathrm{dx}}\,(f(x)\;g(x)) =$$

$$\lim_{w\to x}\frac{f(w)\,g(w) - f(x)\,g(x)}{w-x} = \lim_{w\to x}\frac{f(w)\,g(w) - f(w)\,g(x) + f(w)\,g(x) - f(x)\,g(x)}{w-x} =$$

$$\lim_{w\to x} f(w)\,\frac{g(w)-g(x)}{w-x} + \lim_{w\to x} g(x)\,\frac{f(w)-f(x)}{w-x} =$$

$$\lim_{w\to x} f(w)\,\lim_{w\to x}\frac{g(w)-g(x)}{w-x} + \lim_{w\to x} g(x)\,\lim_{w\to x}\frac{f(w)-f(x)}{w-x} =$$

$$\lim_{w\to x} f(w)\,\frac{d}{d\,x}\,g(x) + \lim_{w\to x} g(x)\,\frac{d}{d\,x}\,f(x) =$$

$$f(x)\,\frac{d}{d\,x}\,g(x) + g(x)\,\frac{d}{d\,x}\,f(x)$$

In the last step $f(w) \to f(x)$ as $w \to x$ because f is continuous at x and $g(x) \to g(x)$ as $w \to x$ because $g(x)$ does not involve w and hence remains unaffected.

QED

The derivative of a product of two functions is the first function times the derivative of the second plus the second function times the derivative of the first.

Example 3.32. Product Rule

Find the derivative of the product

$$\mathbf{p = \left(x^5 - 6\,x^2\right)\left(7\,x^9 + x^3\right)}$$

$$\left(x^5 - 6\,x^2\right)\left(7\,x^9 + x^3\right)$$

Solution 3.32. The first approach to find the result is to use the product rule

$$\mathbf{res1} = \frac{\partial p}{\partial x}$$

$$\left(5\,x^4 - 12\,x\right)\left(7\,x^9 + x^3\right) + \left(x^5 - 6\,x^2\right)\left(63\,x^8 + 3\,x^2\right)$$

The second approach is based on the expansion of the product and then apply the other rules

p1 = Expand[*p*]

$$7\,x^{14} - 42\,x^{11} + x^8 - 6\,x^5$$

The derivative of this polynomial is

$$\mathbf{res2} = \frac{\partial \mathbf{p1}}{\partial x}$$

$$98\,x^{13} - 462\,x^{10} + 8\,x^7 - 30\,x^4$$

If both results are the same the difference of them should vanish. We check this as

Simplify[res1 − res2]

0

The result demonstrates that the two approaches deliver the same result.▲

3.6.6 Derivative of a Quotient

In addition to the product rule there exists a quotient rule allowing to treat quotients of functions.

Theorem 3.28. *The Quotient Rule*

If f and g are differentiable at x and $g(x) \neq 0$, then $f\,/\,g$ is differentiable at x and

$$\frac{\partial\left(\frac{f(x)}{g(x)}\right)}{\partial x} = \frac{g(x)\,\frac{\partial f(x)}{\partial x} - f(x)\,\frac{\partial g(x)}{\partial x}}{g(x)^2}.\blacksquare$$

Proof 3.28.

$$\frac{d}{\mathrm{dx}}\left(\frac{f(x)}{g(x)}\right) = \lim_{w\to x}\frac{\frac{f(w)}{g(w)} - \frac{f(x)}{g(x)}}{w-x} = \lim_{w\to x}\frac{f(w)\,g(x) - f(x)\,g(w)}{(w-x)\,g(x)\,g(w)} =$$

$$\lim_{w\to x}\frac{f(w)\,g(x) - f(x)\,g(x) + f(x)\,g(x) - f(x)\,g(w)}{(w-x)\,g(x)\,g(w)} =$$

$$\lim_{w\to x} \frac{g(x)\,\frac{f(w)-f(x)}{(w-x)} - f(x)\,\frac{g(x)-f(x)}{(w-x)}}{g(x)\,g(w)} =$$

$$\frac{\lim_{w\to x} g(x)\ \lim_{w\to x} \frac{f(w)-f(x)}{(w-x)} - \lim_{w\to x} f(x)\ \lim_{w\to x} \frac{g(x)-f(x)}{(w-x)}}{\lim_{w\to x} g(x)\,g(w)} =$$

$$\frac{\lim_{w\to x} g(x)\ \frac{d}{dx} f(x) - \lim_{w\to x} f(x)\ \frac{d}{dx} g(x)}{\lim_{w\to x} g(x)\,g(w)} =$$

$$\frac{g(x)\ \frac{d}{dx} f(x) - f(x)\ \frac{d}{dx} g(x)}{g(x)\,g(x)} =$$

$$\frac{g(x)\ \frac{d}{dx} f(x) - f(x)\ \frac{d}{dx} g(x)}{g(x)^2}$$

QED

The derivative of a quotient of two functions is the denominator times the derivative of the numerator minus the numerator times the derivative of the denominator, all divided by the denominator squared.

Example 3.33. Quotient Rule

Let

$$\boldsymbol{f(x) := \frac{x^4 - 4}{x^3 + 3}}$$

Solution 3.33. The derivative of the ratio of the two functions $f_1 = x^4 - 4$ and $f_2 = x^3 + 3$ is derived by using the quotient rule. Check the result by your own calculation with paper and pencil.

$$\boldsymbol{g = \text{Simplify}\left[\frac{\partial f(x)}{\partial x}\right]}$$

$$\frac{x^2\left(x^4 + 12\,x + 12\right)}{\left(x^3 + 3\right)^2}$$

The graph of the two functions is given in the following plot.

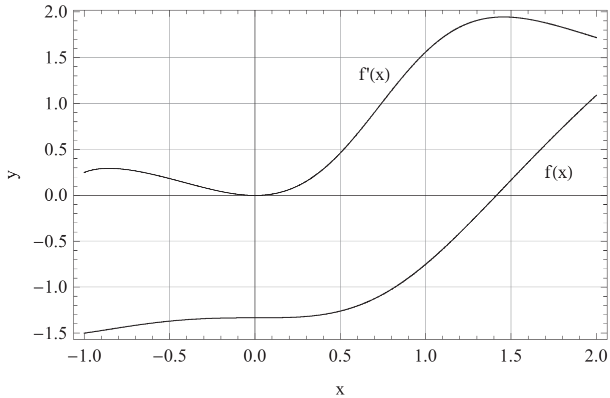


Figure 3.46. The function $f(x) = \frac{x^4-4}{x^3+3}$ and its derivative f' as a function of x.

3.6.7 The Chain Rule

In this section we will derive a formula that express the derivative of a composition $f \circ g$ in terms of the derivatives of f and g. This formula will enable us to differentiate complicated functions using known derivatives of simpler functions.

If we know the derivatives of f and g how can we use this information to find the derivative of the composition $f \circ g$?

The key to solving this problem is to introduce dependent variables

$$y = (f \circ g)(x) = f(g(x)) \qquad \text{and} \qquad u = g(x) \tag{3.52}$$

so that $y = f(u)$. We are interested in using the known derivatives

$$\frac{dy}{du} = f'(u) \qquad \text{and} \qquad \frac{du}{dx} = g'(x) \tag{3.53}$$

to find the unknown derivative

$$\frac{dy}{dx} = \frac{d}{dx}(f(g(x))). \tag{3.54}$$

Stated another way, we are interested in using the known rate of change dy/du and du/dx to find the unknown rate of change dy/dx. But intuition suggests that rates of change multiply. For example, if y changes at 4 times the rate of change of u and u changes at 2 times the rate of change of x, then y changes at $4 \times 2 = 8$ times the rate of change of x. This suggests that

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}. \tag{3.55}$$

These ideas are formalized in the following theorem.

Theorem 3.29. *The Chain Rule*

If g is differentiable at x and f is differentiable at g(x), then the composition $f \circ g$ is differentiable at x. Moreover,

$$(f \circ g)'(x) = f'(g(x))\, g'(x).$$

Alternatively, if

$$y = f(g(x)) \qquad \text{and} \qquad u = g(x)$$

then $y = f(u)$ and

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}. \blacksquare$$

The proof of this result is omitted. We take this theorem just to apply it.

Example 3.34. Chain Rule

Find the derivative $h'(x)$ if

$$\boldsymbol{h(x\_) := 4\sqrt{x^2-7}}$$

Solution 3.34. We first find functions f and g such that $f \circ g = h$. Observe that if $g(x) = x^2 - 7$ and $f(x) = 4\sqrt{x}$, then

$$(f \circ g)(x) = f(g(x)) = 4\sqrt{g(x)} = 4\sqrt{x^2-7} = h(x)$$

Also,

$$\boldsymbol{\frac{\partial\left(4\sqrt{u}\right)}{\partial u}}$$

$$\frac{2}{\sqrt{u}}$$

and

$$\boldsymbol{\frac{\partial\left(x^2-7\right)}{\partial x}}$$

$$2\,x$$

Using the chain rule,

$$h'(x) = f'(g(x))\, g'(x) = \frac{2}{\sqrt{g(x)}}\, 2\,x = \frac{4\,x}{\sqrt{x^2-7}}$$

which agrees with the result

$$\frac{\partial h(x)}{\partial x}$$

$$\frac{4\,x}{\sqrt{x^2-7}}$$

▲

3.6.8 Derivatives of Logarithmic and Exponential Functions

In this section we will obtain derivative formulas for logarithmic and exponential functions, and we will discuss the general relationship between the derivative of a one-to-one function and its inverse function.

The natural logarithm plays a special role in calculus that can be motivated by differentiating $\log_b(x)$, where b is an arbitrary base. For this purpose, recall that $\log_b(x)$ is continuous for $x > 0$. We will also need the limit

$$\lim_{v\to 0}\,(1+v)^{1/v} = e. \tag{3.56}$$

where e is the exponential constant whose numeric value is given by

N[e, 30]

2.71828182845904523536028747135

Using the definition of a derivative, we obtain

$$
\begin{aligned}
\frac{d}{dx}\log_b(x) &= \lim_{w\to x}\frac{\log_b(w)-\log_b(x)}{w-x}\\
&= \lim_{w\to x}\left(\frac{1}{w-x}\log_b\left(\frac{w}{x}\right)\right)\\
&= \lim_{w\to x}\left(\frac{1}{w-x}\log_b\left(\frac{x+(w-x)}{x}\right)\right)\\
&= \lim_{w\to x}\left(\frac{1}{w-x}\log_b\left(1+\frac{w-x}{x}\right)\right)\\
&= \lim_{w\to x}\left(\frac{1}{x}\frac{x}{w-x}\log_b\left(1+\frac{w-x}{x}\right)\right)\\
&= \lim_{v\to 0}\left(\frac{1}{x}\frac{1}{v}\log_b(1+v)\right)\\
&= \frac{1}{x}\lim_{v\to 0}\left(\frac{1}{v}\log_b(1+v)\right)\\
&= \frac{1}{x}\lim_{v\to 0}\log_b(1+v)^{1/v}\\
&= \frac{1}{x}\log_b\left(\lim_{v\to 0}(1+v)^{1/v}\right)\\
&= \frac{1}{x}\log_b e
\end{aligned}
\tag{3.57}
$$

Thus

$$\frac{d}{dx}\log_b(x) = \frac{1}{x}\log_b e \qquad , \; x > 0. \tag{3.58}$$

For logarithms there exists a conversion formula converting the logarithm of base b to the natural logarithm ln by $\log_b e = 1/\ln(b)$, so we can rewrite the derivative formula as

$$\frac{d}{dx}\log_b(x) = \frac{1}{x\ln(b)}, \qquad x > 0. \tag{3.59}$$

In the special case where the basis $b = e$, we have $\ln(e) = 1$, so this formula becomes

$$\frac{d}{dx}\ln(x) = \frac{1}{x}, \qquad x > 0. \tag{3.60}$$

Thus among all possible bases, the base $b = e$ produces the simplest formula for the derivative of $\log_b(x)$. This is one of the reasons why the natural logarithm function is preferred over other logarithms in science and engineering.

Example 3.35. Logarithms I

The following Figure shows the natural logarithm

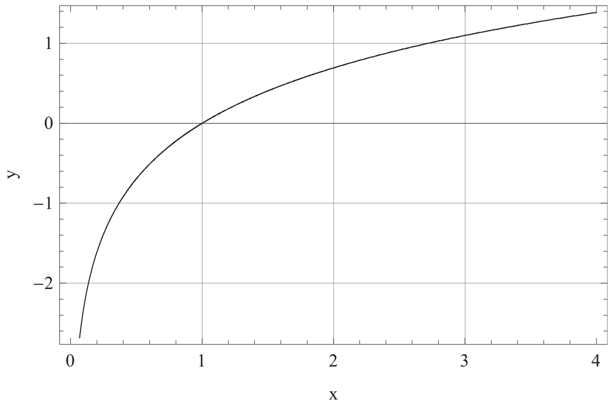


Figure 3.47. The logarithm ln(x) for positive values of x.

Find the derivative at $x = 1$, $x = 2$, and $x = 4$. Is there any horizontal tangent line to the logarithm?

Solution 3.35. In *Mathematica* the natural logarithm ln(x) is denoted by the function Log[]. If we have a different basis b than e the notation is Log[b,x]. Taking into account this notation the natural logarithm is differentiated by

$$\textbf{res} = \frac{\partial \log(x)}{\partial x}$$

$$\frac{1}{x}$$

The derivatives at the locations $x = 1$, $x = 2$, and $x = 4$ are found by inserting the specified values

(res /. #1 &) /@ {x → 1, x → 2, x → 4}

$$\left\{1, \frac{1}{2}, \frac{1}{4}\right\}$$

For very large arguments of x the derivative of ln(x) becomes smaller and smaller but does not vanish.▲

If u is a differentiable function of x, and if $u(x) > 0$, then applying the chain rule to (3.59) and (3.60) produces the following generalized derivative formulas

$$\frac{d}{dx}\log_b(u(x)) = \frac{1}{u(x)\ln(b)}\frac{d}{dx}u(x) \tag{3.61}$$

and

$$\frac{d}{dx}\ln(u(x)) = \frac{1}{u(x)}\frac{d}{dx}u(x). \tag{3.62}$$

Example 3.36. Logarithms II

Find the derivative of

$$f(x\_) := \log(x^2 + 1)$$

Solution 3.36. The derivative follows directly by

$$\frac{\partial f(x)}{\partial x}$$

$$\frac{2x}{1 + x^2}$$

or using formula (3.62) $\frac{1}{x^2+1} \frac{d}{dx} (x^2 + 1)$ showing the composition rule.▲

When possible the properties of logarithms should be used to convert products, quotients, and exponents into sums, differences, and constant multiples before differentiating a function involving logarithms.

Example 3.37. Logarithms III

Another more interesting example is the following to get the derivative of the function

$$f(x\_) := \log\left(\frac{x^2 \log(x)}{\sqrt{x+1}}\right)$$

Solution 3.37. First apply the rules for logarithms and expand the function to

inter = PowerExpand[$f(x)$]

$$2\log(x) - \frac{1}{2}\log(x+1) + \log(\log(x))$$

Then differentiate term by term

$$\frac{\partial \text{inter}}{\partial x}$$

$$-\frac{1}{2(x+1)} + \frac{2}{x} + \frac{1}{x\log(x)}$$

which is much more simpler to derive than using the chain rule in the original function.▲

By equation (3.59) we know that

$$\frac{d}{dx}\log_b(x) = \frac{1}{x\ln(b)} \tag{3.63}$$

is a nonzero function, so that the inverse function for $\log_b(x)$ is differentiable on $(-\infty, +\infty)$.

To obtain a derivative formula for the exponential function with base b, we rewrite $y = b^x$ as

$$x = \log_b(y) \tag{3.64}$$

and differentiate with respect to x, where $y = y(x)$ to obtain

$$1 = \frac{1}{y \ln(b)} \frac{d}{d\,x} y \tag{3.65}$$

Solving for dy / dx and replacing y by b^x we have

$$\frac{d}{d\,x} y = y \ln(b) = b^x \ln(b). \tag{3.66}$$

Thus we have shown that

$$\frac{d}{d\,x} b^x = b^x \ln(b). \tag{3.67}$$

In the special case where $b = e$, we have $\ln(e) = 1$ so that (3.67) becomes

$$\frac{d}{d\,x} e^x = e^x. \tag{3.68}$$

Moreover, if u is a differentiable function of x, then it follows from (3.67) and (3.68) that

$$\frac{d}{d\,x} b^{u(x)} = b^{u(x)} \ln(b) \frac{d\,u(x)}{d\,x} \qquad \text{and} \qquad \frac{d}{d\,x} e^{u(x)} = e^{u(x)} \frac{d\,u(x)}{d\,x}. \tag{3.69}$$

Example 3.38. Exponentials

Use relation (3.69) to derive the derivative of $f(x) = 2^{\ln(x)}$, $g(x) = e^{-2\,x}$, and $h(x) = e^{x^3}$.

Solution 3.38. The first function f delivers

$$\frac{\partial 2^{\log(x)}}{\partial x}$$

$$\frac{\log(2)\, 2^{\log(x)}}{x}$$

The second function g gives

$$\frac{\partial e^{-2\,x}}{\partial x}$$

$$-2\, e^{-2\,x}$$

Function $h(x)$ is differentiated by

$$\frac{\partial e^{x^3}}{\partial x}$$

$$3\, e^{x^3}\, x^2$$

▲

3.6.9 Derivatives of Trigonometric Functions

The main objective of this section is to obtain formulas for the derivatives of trigonometric functions.

For the purpose of finding derivatives of the trigonometric functions sin(x), cos(x), tan(x), cot(x), sec(x), and csc(x), we will assume that x is measured in radians. We will also need the following limits, which were stated in Theorem 3.18

$$\lim_{h\to 0} \frac{\sin(h)}{h} = 1 \qquad \text{and} \qquad \lim_{h\to 0} \frac{(1-\cos(h))}{h} = 0. \tag{3.70}$$

We begin with the problem of differentiation of sin(x). Introducing an alternative form for limits

$$f'(x) = \lim_{w\to x} \frac{f(w)-f(x)}{w-x} = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \tag{3.71}$$

where we replaced w by $x+h$ meaning that the location w is h units apart from x.

For the definition of the derivative, we have

$$\begin{aligned} \frac{d}{dx}\sin(x) &= \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h} \\ &= \lim_{h\to 0} \left(\sin(x)\frac{\cos(h)-1}{h} + \cos(x)\frac{\sin(h)}{h}\right) \\ &= \lim_{h\to 0} \left(\cos(x)\frac{\sin(h)}{h} - \sin(x)\left(\frac{1-\cos(h)}{h}\right)\right) \end{aligned} \tag{3.72}$$

Since sin(x) and cos(x) do not involve h, they remain unchanged as $h \to 0$; thus

$$\lim_{h\to 0}\sin(x) = \sin(x) \qquad \text{and} \qquad \lim_{h\to 0}\cos(x) = \cos(x). \tag{3.73}$$

Consequently,

$$\begin{aligned} \frac{d}{dx}\sin(x) &= \cos(x)\lim_{h\to 0}\frac{\sin(h)}{h} - \sin(x)\lim_{h\to 0}\left(\frac{1-\cos(h)}{h}\right) \\ &= \cos(x)(1) - \sin(x)(0) \\ &= \cos(x) \end{aligned} \tag{3.74}$$

Thus, we have shown that

$$\frac{d}{dx}\sin(x) = \cos(x). \tag{3.75}$$

The derivative of cos(x) can be obtained similarly, resulting in the formula

$$\frac{d}{dx}\cos(x) = -\sin(x). \tag{3.76}$$

The derivatives of the remaining trigonometric functions are

$$\frac{d}{dx}\tan(x) = \sec(x)^2 \tag{3.77}$$

$$\frac{d}{dx}\sec(x) = \sec(x)\tan(x) \tag{3.78}$$

$$\frac{d}{dx}\cot(x) = -\csc(x)^2 \tag{3.79}$$

$$\frac{d}{dx}\csc(x) = -\csc(x)\cot(x). \tag{3.80}$$

These formulas can all be obtained from (3.59) and (3.60) using the relationships

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \qquad \cot(x) = \frac{\cos(x)}{\sin(x)}, \tag{3.81}$$

$$\sec(x) = \frac{1}{\cos(x)}, \qquad \csc(x) = \frac{1}{\sin(x)}. \tag{3.82}$$

For example,

$$\begin{aligned}\frac{d}{dx}\tan(x) &= \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right) = \frac{\cos(x)\,\frac{d}{dx}\sin(x) - \sin(x)\,\frac{d}{dx}\cos(x)}{\cos(x)^2} \\ &= \frac{\cos(x)\cos(x) - \sin(x)\,(-\sin(x))}{\cos(x)^2} \\ &= \frac{\cos(x)^2 + \sin(x)^2}{\cos(x)^2} \\ &= \frac{1}{\cos(x)^2} = \sec(x)^2\end{aligned} \tag{3.83}$$

Remark 3.12. The derivative formulas for the trigonometric functions should be memorized. We emphasize again that in all of the derivative formulas for the trigonometric functions, x is measured in radians.

Example 3.39. Trigonometric Functions I

Find the derivative $f'(x)$ if

$$\boldsymbol{f(\text{x\_}) := x^2 \tan(x)}$$

Solution 3.39. Using the product rule we obtain

$$\boldsymbol{\frac{\partial f(x)}{\partial x}}$$

$$x^2 \sec^2(x) + 2\,x\tan(x)$$

▲

Example 3.40. Trigonometric Functions II

Find the derivative of the function

$$f(x\_) := \frac{\sin(x)}{\cos(x) + 1}$$

Solution 3.40. Using the quotient rule and the rules for trigonometric functions we gain

$$\text{res} = \frac{\partial f(x)}{\partial x}$$

$$\frac{\cos(x)}{\cos(x) + 1} + \frac{\sin^2(x)}{(\cos(x) + 1)^2}$$

This result can be simplified by using the identities for trigonometric relations

Simplify[res]

$$\frac{1}{\cos(x) + 1}$$

▲

3.6.10 Higher Derivatives

If the derivative f' of a function f is itself differentiable, then the derivative of f' is denoted by f'' and is called the second derivative of f. As long as we have differentiability, we can continue the process of differentiating derivatives to obtain third, fourth, and even higher derivatives of f. The successive derivatives of f are denoted by

$$f', \ f'' = (f')', \ f''' = (f'')', \ f^{(4)} = (f''')', \ f^{(5)} = \left(f^{(4)}\right)', \ \ldots \tag{3.84}$$

These are called the first derivative, the second derivative, the third derivative, and so forth. Beyond the third derivative, it is too clumsy to continue using primes, so we switch from primes to integers in parentheses to denote the order of the derivative. In this notation it is easy to denote a derivative of arbitrary order by writing

$$f^{(n)} = \frac{d^n f}{dx^n} \qquad \text{which is the } n^{\text{th}} \text{ derivative of } f. \tag{3.85}$$

Example 3.41. Higher Derivatives of Polynomials

Given a polynomial of fifth order

$$f(x\_) := x^5 - 5x^3 + 7x^2 + 12$$

Determine the higher order derivatives.

Solution 3.41. The first order derivative is given by

$$\frac{\partial f(x)}{\partial x}$$

$$5x^4 - 15x^2 + 14x$$

The second order derivative is

$$\frac{\partial^2 f(x)}{\partial x \partial x}$$

$$20\,x^3 - 30\,x + 14$$

The third order derivative reads

$$\frac{\partial^3 f(x)}{\partial x \partial x \partial x}$$

$$60\,x^2 - 30$$

The fourth order derivative is

$$\frac{\partial^4 f(x)}{\partial x^4}$$

$$120\,x$$

The fifth is

$$\frac{\partial^5 f(x)}{\partial x^5}$$

$$120$$

and the sixth order derivative vanishes because the fifth order derivative was constant.

$$\frac{\partial^6 f(x)}{\partial x^6}$$

$$0$$

All higher order derivatives greater than or equal to 6 will vanish, too.▲

Successive derivatives can also be denoted as follows:

$$f'(x) = \frac{d\,f(x)}{d\,x} \tag{3.86}$$

$$f''(x) = \frac{d}{d\,x}\left(\frac{d\,f(x)}{d\,x}\right) = \frac{d^2\,f(x)}{d\,x^2} \tag{3.87}$$

$$f'''(x) = \frac{d}{d\,x}\left(\frac{d^2\,f(x)}{d\,x^2}\right) = \frac{d^3\,f(x)}{d\,x^3} \ldots \tag{3.88}$$

In general we write

$$f^{(n)}(x) = \frac{d^n\,f(x)}{d\,x^n} \tag{3.89}$$

which is read the n^{th} derivative of f with respect to x.

When a dependent variable is involved, say $y = f(x)$, then successive derivatives can be denoted by writing

$$\frac{dy}{dx}, \ \frac{d^2 y}{dx^2}, \ \frac{d^3 y}{dx^3}, \ \ldots, \ \frac{d^n y}{dx^n}, \ \ldots \tag{3.90}$$

or more briefly

$$y', \ y'', \ y''', \ y^{(4)}, \ y^{(5)}, \ \ldots, y^{(n)}, \ \ldots \tag{3.91}$$

3.6.11 L'Hôpital's Rule

In this section we will discuss a general method for using derivatives to find limits. This method will enable us to establish limits with certainty that earlier in the text we were only able to conjecture using numerical or graphical evidence. The method that we will discuss in this section is an extremely powerful tool that is used internally by many computer programs to calculate limits of various types.

In the earlier section we discussed limits that can be determined by inspection or by some appropriate algebraic manipulation. Two special exceptions to this were the limits in Theorem 3.18,

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1 \qquad \text{and} \qquad \lim_{x \to 0} \left(\frac{1 - \cos(x)}{x} \right) = 0. \tag{3.92}$$

The first of these relations was shown by the Squeezing Theorem 3.17 and some careful manipulation of inequalities, and the second equation then followed using the identity $\sin^2(x) + \cos^2(x) = 1$. However, the two equations are actually special cases of derivatives, as can be seen by

$$\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{\sin(x) - \sin(0)}{x - 0} = \frac{d}{dx} \sin(x) \,|_{x=0} = \cos(0) = 1 \tag{3.93}$$

and

$$\begin{gathered} \lim_{x \to 0} \frac{1 - \cos(x)}{x} = \\ \lim_{x \to 0} -\left(\frac{\cos(x) - 1}{x} \right) = -\lim_{x \to 0} \left(\frac{\cos(x) - \cos(0)}{x - 0} \right) = -\left(\frac{d}{dx} \cos(x) \,|_{x=0} \right) = \sin(0) = 0. \end{gathered} \tag{3.94}$$

What makes the limits in (3.92) bothersome is the fact that the numerator and the denominator both approaches 0 as $x \to 0$. Such limits are called indeterminate forms of type $0/0$. As illustrated above, the definition of a derivative provides an important class of examples of indeterminate forms of type $0/0$. Our goal here is to develop a general method, based on the derivative, for evaluating indeterminate forms.

Consider the limit

$$\lim_{x \to 0} \frac{e^{2x} - 1}{\sin(x)}. \tag{3.95}$$

Unlike (3.92), the limit in (3.95) is not easily seen as the evaluation of the derivative of a function at $x = 0$. However (3.95) can be expressed as a ratio of two derivatives

$$\lim_{x\to 0} \frac{e^{2x}-1}{\sin(x)} = \lim_{x\to 0}\left(\frac{\left(e^{2x}-e^{2(0)}\right)/(x-0)}{(\sin(x)-\sin(0))/(x-0)}\right) = \frac{\frac{d}{dx}\left(e^{2x}\right)|_{x=0}}{\frac{d}{dx}(\sin(x))\,|_{x=0}} = \frac{2\,e^{0}}{\cos(0)} = 2. \tag{3.96}$$

The method in (3.96) can be stated more generally. Suppose that f and g are differentiable functions at $x = a$ and that

$$\lim_{x\to a} \frac{f(x)}{g(x)} \tag{3.97}$$

is an indeterminate form of type $0/0$, that is,

$$\lim_{x\to a} f(x) = 0 \qquad \text{and} \qquad \lim_{x\to a} g(x) = 0. \tag{3.98}$$

In particular, the differentiability of f and g at $x = a$ implies that f and g are continuous at $x = a$, and hence from (3.98)

$$f(a) = \lim_{x\to a} f(x) = 0 \qquad \text{and } g(a) = \lim_{x\to a} g(x) = 0. \tag{3.99}$$

Furthermore, since f and g are differentiable at $x = a$,

$$\lim_{x\to a} \frac{f(x)}{x-a} = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = f'(a) \tag{3.100}$$

and

$$\lim_{x\to a} \frac{g(x)}{x-a} = \lim_{x\to a} \frac{g(x)-g(a)}{x-a} = g'(a). \tag{3.101}$$

If $g'(a) \neq 0$ then the indeterminate form in (3.97) can be evaluated as the ratio of derivative values:

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f(x)/(x-a)}{g(x)/(x-a)} = \frac{\lim_{x\to a} \frac{f(x)-f(a)}{x-a}}{\lim_{x\to a} \frac{g(x)-g(a)}{x-a}} = \frac{f'(a)}{g'(a)}. \tag{3.102}$$

If $f'(x)$ and $g'(x)$ are continuous at $x = a$, the result (3.102) is a special case of L'Hôpital's rule, which converts an indeterminate form of type $0/0$ into a new limit involving derivatives. Moreover, L'Hôpital's rule is also true for limits at $-\infty$ and at $+\infty$. We state the following result without proof.

Theorem 3.30. *L'Hôpital's Rule*

Suppose that f and g are differentiable functions on an open interval containing $x = a$, except possibly at $x = a$, and that

$$\lim_{x\to a} f(x) = 0 \text{ and } \lim_{x\to a} g(x) = 0.$$

If $\lim_{x\to a} f'(x)/g'(x)$ has a finite limit, or if this limit is $+\infty$ or $-\infty$, then

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

Moreover, this statement is also true in the case of a limit as $x \to a^-$, $x \to a^+$, $x \to -\infty$, or as $x \to +\infty$. ■

Remark 3.13. Note that in L'Hôpital's rule the numerator and denominator are differentiated separately, which is not the same as differentiating $f(x)/g(x)$. L'Hôpital's rule is also applicable to indeterminate forms of ∞/∞.

In the following examples, we will apply L'Hôpital's rule.

Example 3.42. L'Hôpital's Rule

In each part confirm that the limit is an indeterminate form of type $0/0$, and evaluate L'Hôpital's rule for the function.

$$f(\text{x\_}) := \frac{x^2 - 4}{x - 2}$$

Solution 3.42. Let us examine the numerator and denominator separately

$$\lim_{x\to 2} \text{Numerator}[f(x)]$$

0

$$\lim_{x\to 2} \text{Denominator}[f(x)]$$

0

L'Hôpital's rule now delivers by differentiation numerator and denominator separately

$$\frac{\lim_{x\to 2} \frac{\partial \text{Numerator}[f(x)]}{\partial x}}{\lim_{x\to 2} \frac{\partial \text{Denominator}[f(x)]}{\partial x}}$$

4

Another example is given by the function

$$g(\text{x\_}) := \frac{x^{-4/3}}{\sin\left(\frac{1}{x}\right)}$$

$$\lim_{x\to\infty} \text{Numerator}[g(x)]$$

∞

$$\lim_{x\to\infty} \text{Denominator}[g(x)]$$

∞

L'Hôpital's rule delivers for $g(x)$

$$\frac{\lim_{x\to\infty} \frac{\partial \text{Numerator}[g(x)]}{\partial x}}{\lim_{x\to\infty} \frac{\partial \text{Denominator}[g(x)]}{\partial x}}$$

0

▲

3.6.12 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

3.6.12.1 Test Problems

T1. What role does the derivative play in defining slopes, tangents, and rates of change?

T2. What is the derivative of a function f? How is its domain related to the domain of f? Give examples.

T3. What does it mean for a function to be differentiable on an open interval? On a closed interval?

T4. Describe geometrically when a function typically does not have a derivative at a point.

T5. What rules do you know for calculating derivatives? Give some examples.

T6. What is implicit differentiation? When do you need it? Give examples.

T7. What is the rule for calculating the derivative of a composite of two differentiable functions? How is such a derivative evaluated? Give examples.

3.6.12.2 Exercises

E1. Sketch the graph of the function $f(x) = e^x$, playing particular attention to how the graph crosses the y-axis. What fact allows you to do this?

- **a.** What types of functions are $f(x) = e^x$ and $g(x) = x^e$?
- **b.** Which of the two functions in part a) grows more rapidly when x is large?

E2. Differentiate the functions.

- **a.** $f(x) = \sqrt{40}$,
- **b.** $f(x) = \frac{3}{5} x^9$,
- **c.** $f(x) = 187.342 \ldots$,
- **d.** $f(x) = 2x^7 - 3x^2 + x - 12$,
- **e.** $f(x) = 3 - \frac{1}{2} x$,
- **f.** $f(x) = \frac{1}{2}(x^5 + 8)$,
- **g.** $f(x) = x^{-3/2}$,
- **h.** $f(x) = \frac{4}{x^2}$,
- **i.** $f(x) = c\,x^{-8}$,
- **j.** $f(x) = \sqrt{x} - 2e^{-x}$,
- **k.** $f(x) = e^{-x} + \frac{1}{e^x}$,
- **l.** $f(x) = \sqrt{x} + \frac{1}{\sqrt{x}}$,
- **m.** $f(x) = (x + 5)(x - 6)$,
- **n.** $f(x) = \left(e^x + \frac{1}{x^3}\right)^2$,

E3. Find the first and second derivatives of the function.

- **a.** $f(x) = x^4 - 4x^3 + 2x^2 - 15x$,
- **b.** $f(x) = \sqrt{x} + \sqrt[3]{x}$.

E4. Find the first and second derivatives of the function. Check to see that your answers are reasonable by comparing the graphs of f, f', and f''.

- **a.** $f(x) = 2x - 5x^{5/6}$,
- **b.** $f(x) = e^x - x^5$.

E5. The equation of motion of a particle is $s = t^3 - 3t$, where s is in meters and t is in seconds. Find

a. the velocity and acceleration as function of t,

b. the acceleration after 2 s, and

c. the acceleration when the velocity is 0.

E6. Find equations of the tangent line and normal line to the curve at the given point.

a. $y = x^5 + 3\, e^x$ at $x = 0$,

b. $y = (1 + 2\, x)^3$ at $x = 1$.

E7. Find $f'(x)$. Compare the graphs of $f(x)$ and $f'(x)$ and use them to explain why your answer is reasonable.

a. $f(x) = e^x - 4\, x$,

b. $f(x) = x - \frac{1}{x}$,

c. $f(x) = 5\, x^5 - 3\, x^3 + x$,

d. $f(x) = 2\, x^{12} - 3\, x^3 + 5$.

E8. Find the n^{th} derivative of each function by calculating the first few derivatives and observing the pattern that occurs.

a. $f(x) = x^n$,

b. $f(x) = \frac{1}{x}$.

E9. For what values of a and b is the line $2\, x + y = b$ tangent to the parabola $y = a\, x^2$ when $x = 2$?

E10 Draw a diagram showing two perpendicular lines that intersect on the y-axis and are both tangent to the parabola $y = x^2$. Where do these lines intersect?

4

Extreme Value Problems

4.1 Introduction

In this chapter we will study various applications of the derivative. For example, we will use methods of calculus to analyze functions and their graphs. In the progress we will show how calculus and graphing of functions can provide most of the important information about the behavior of a function. Another application of the derivative will be in the solution of optimization problems. For example, if cost is the main consideration, we might be interested in finding the least expensive way to perform the task. Mathematically, optimization problems can be reduced to finding the largest or smallest value of a function on some interval, and determining where the largest or smallest value occurs. Using the derivative, we will develop the mathematical tools necessary for solving such problems.

4.2 Maxima and Minima

In this section we will discuss methods for finding the high and low points on the graph of a function. The idea we develop here will have important applications.

4.2.1 Relative Maxima and Minima

If we imagine the graph of a function f to be a two-dimensional mountain range with hills and valleys, then the tops of the hills are called relative maxima, and the bottoms of the valleys are called relative minima (see Figure 4.1).

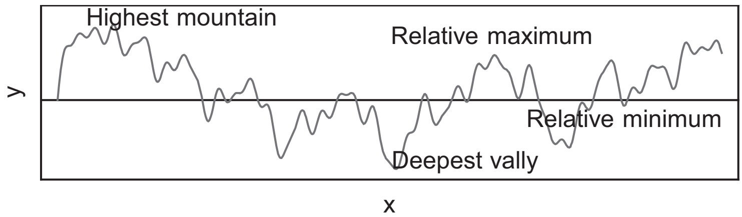


Figure 4.1. Landscape of a two-dimensional mountain.

The relative maxima are the high points in their immediate vicinity, and the relative minima are the low points. Note that a relative maximum need not be the highest point in the entire mountain range, and a relative minimum need not be the lowest point — they are just high and low points relative to the nearby terrain. The ideas are captured in the following definition

Definition 4.1. *Maxima and Minima*

A function f is said to have a relative maximum at x_0 if there is an open interval containing x_0 on which $f(x_0)$ is the largest value, that is $f(x_0) \geq f(x)$ for all x in the interval. Similarly, f is said to have a relative minimum at x_0 if there is an open interval containing x_0 on which $f(x_0)$ is the smallest value, that is $f(x_0) \leq f(x)$ for all x in the interval. If f has either a relative maximum or a relative minimum at x_0, then f is said to have a relative extremum at x_0. ■

Example 4.1. Relative Extrema

Locate the relative extrema of the four functions graphed in the following Figure

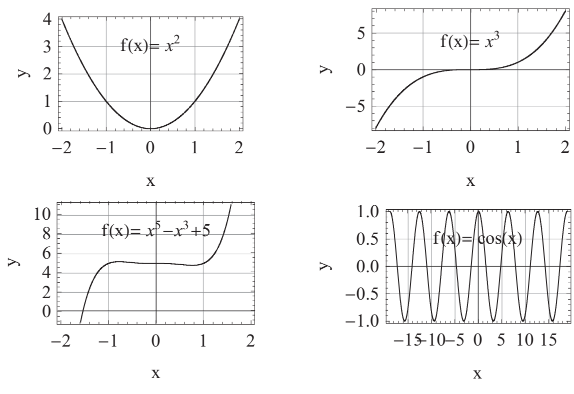


Figure 4.2. Different function showing relative maxima and minima.

Solution 4.1. The first function $f(x) = x^2$ has a relative minimum but no relative maximum. The function $f(x) = x^3$ does not have an extremum. The function $f(x) = x^5 - x^3 + 5$ has a relative maximum at $x = -\sqrt{3/5}$ and a relative minimum at $x = \sqrt{3/5}$. The function $f(x) = \cos(x)$ has relative maxima at all even multiples of π and relative minima at all odd multiples of π.▲

Points at which relative extrema occur can be viewed as the transition points that separate the regions where a graph is increasing from those where it is decreasing. As suggested in Figure 4.3, the relative extrema of a continuous function f occur at points where the graph of f either has a horizontal tangent line or is not differentiable. This is the content of the following theorem.

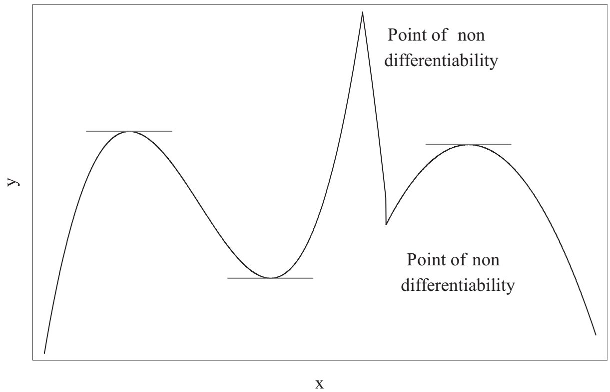


Figure 4.3. Differentiability, horizontal lines and non-differentiability of a function.

Theorem 4.1. *Extremum Theorem*

Suppose that f is a function defined on an open interval containing the number x_0. If f has a relative extremum at $x = x_0$, then either $f'(x_0) = 0$ or f is not differentiable at x_0. ■

Proof 4.1. Assume that f has a relative extreme value at x_0. There are two possibilities either f is differentiable at x_0 or it is not. If it is not, then we are done. If f is differentiable at x_0, then we must show that $f'(x_0) = 0$. It cannot be the case that $f'(x_0) > 0$, for then f would not have a relative extreme value at x_0. For the same reason, it cannot be the case that $f'(x_0) < 0$. We conclude that if f has a relative extreme value at x_0 and if f is differentiable at x_0, then $f'(x_0) = 0$.

4.2.2 Critical Numbers

Values in the domain of f at which either $f'(x) = 0$ or f is not differentiable are called critical numbers of f. Thus Theorem 4.1 can be rephrased as follows:

- *If a function is defined on an open interval, its relative extrema on the interval, if any, occur at critical numbers.*

Sometimes we will want to distinguish critical numbers at which $f'(x) = 0$ from those at which f is not differentiable. We will call a point on the graph of f at which $f'(x) = 0$ a stationary point of f.

It is important not to put too much into Theorem 4.1 — the theorem asserts that the set of critical numbers is a complete set of candidates for locations of relative extrema, but it does not say that a critical number must yield a relative extremum. That is, there may be critical numbers at which a relative extremum does not occur. For example, for the eight critical numbers shown in Figure 4.4

relative extrema occur at each x_0 marked in the upper two rows, but not any x_0 marked in the bottom two rows.

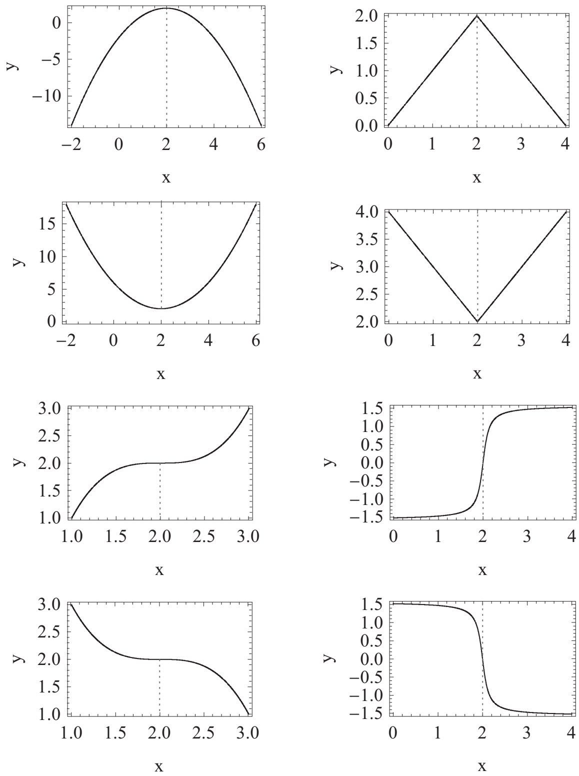


Figure 4.4. Different kind of critical numbers.

4.2.3 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

4.2.3.1 Test Problems

T1. Explain the difference between an absolute minimum and a local minimum.

T2. Suppose f is a continuous function defined on a closed interval $[a, b]$.

- **a.** What theorem guarantees the existence of an absolute maximum value and an absolute minimum value for f?
- **b.** What steps would you take to find those maximum and minimum values?

T3. What can be said about the extreme values of a function that is continuous on a closed interval?

T4. What is the Second Derivative Test for Local Extreme Values? Give examples of how it is applied.

4.2.3.2 Exercises

E1. Sketch the graph of a function that has two local maxima, one local minimum, and no absolute minimum.

E2. Sketch the graph of a function that has three local minima, two local maxima, and seven critical numbers.

E3. Sketch the graph of a function on $[-1, 2]$ that has an absolute maximum but no absolute minimum.

Sketch the graph of a function on [-1,2] that is discontinuous but has both an absolute maximum and an absolute minimum.

E5. Sketch the graph of a function f that is continuous on [1, 5] and has the given properties.

a. Absolute maximum at 5, absolute minimum at 2, local maximum at 3, local minima at 2 and 4.

b. Absolute minimum at 2, absolute maximum at 3, local minimum at 4.

c. f has no local maximum or minimum, but 2 and 4 are critical numbers.

E6. Find the critical numbers of the function.

a. $f(x) = 5x^2 + 4x$,

b. $f(x) = x^3 + 3x^2 - 24x$,

c. $f(x) = |3x - 4|$,

d. $f(x) = \frac{x-1}{x^2+4}$,

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

f. $f(x) = 3x^4 + 4x^3 - 6x^2$,

g. $f(x) = x^3 + x^2 + x$.

E7. A formula for the *derivative* of a function f is given. How many critical numbers does f have?

a. $f'(x) = 6e^{-|x|/10}\sin(x) - 1$,

b. $f'(x) = \frac{100\cos(x)^2}{10+x^2} - 1$.

E8. Between 0°C and 30°C, the volume V of 1 kg of water at a temperature T is given approximately by the formula

$$V = 999.87 - 0.06426\,T + 0.0085043\,T^2 - 0.0000679\,T^3 \tag{1}$$

Find the temperature at which water has its maximum density.

4.3 Rolle's Theorem and the Mean Value Theorem

Let us start with the statement, if the derivative of a function is everywhere zero, then the function is constant.

We will begin with a special case of the Mean Value Theorem, called Rolle's Theorem, in honor of the mathematician Michel Rolle. This theorem states the geometrically obvious fact that if the graph of a differentiable function intersects the x-axis at two places, a and b, then somewhere between a and b there must be at least one place where the tangent line is horizontal. The precise statement of the theorem is as follows:

Theorem 4.2. *Rolle's Theorem*

Let f be differentiable on (a, b) and continuous on $[a, b]$. If $f(a) = f(b) = 0$, then there is at least one number x_0 in (a, b) such that $f'(x_0) = 0$. ■

We will not give a proof of this theorem but show its validity by a simple example.

Example 4.2. Rolle's Theorem

The function $f(x) = \sin(x)$ has roots at $x = 0$ and $x = 2\pi$. Verify the hypothesis and conclusion of Rolle's Theorem for $f(x) = \sin(x)$ on $[0, 2\pi]$.

Solution 4.2. Since f is continuous and differentiable everywhere, it is differentiable on $(0, 2\pi)$ and continuous on $[0, 2\pi]$. Thus Rolle's Theorem guarantees that there is at least one number x_0 in the interval $(0, 2\pi)$ such that $f'(x_0) = 0$. Since $f'(x) = \cos(x)$, we can find x_0 by solving the equation by the *Mathematica* function Reduce[]

Reduce[cos(x) == 0, x]

$$c_1 \in \mathbb{Z} \bigwedge \left(x = 2\pi c_1 - \frac{\pi}{2} \bigvee x = 2\pi c_1 + \frac{\pi}{2} \right)$$

which are obviously the two solutions we are looking for with $c_1 = 0$ and $c_1 = 1$ belonging to the specified interval. The solutions are either $x_0 = \pi/2$ or $x_0 = 3\pi/2$ (see Figure 4.5).

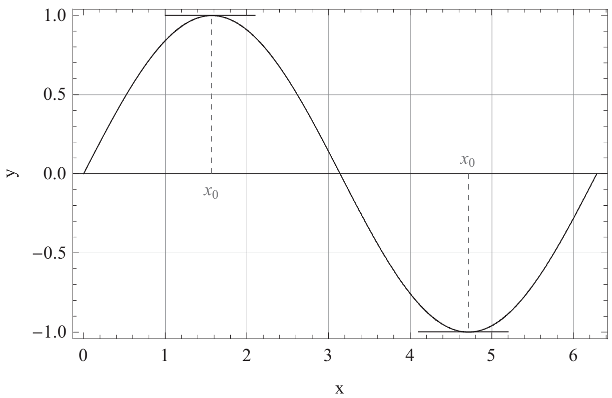


Figure 4.5. Example for Rolle's theorem demonstrating that there are vanishing derivatives for a specified range.▲

The hypotheses in Roll's Theorem are critical—if f fails to be differentiable at even one place in the interval, then the conclusion may not hold. For example, the function $f(x) = |x| - 1$ has roots at $x = \pm 1$, yet there is no horizontal tangent line to the graph of f over the interval $(-1, 1)$ (see Figure 4.6).

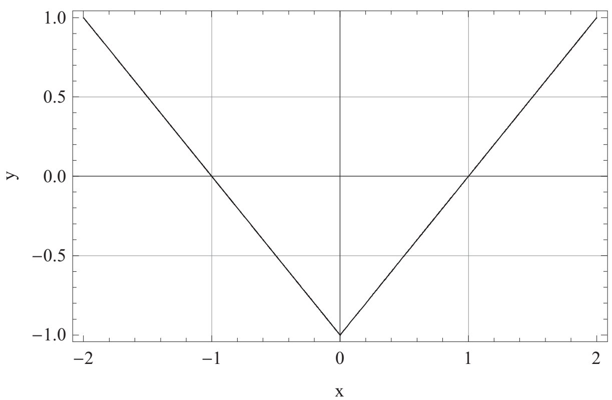


Figure 4.6. Rolle's Theorem is not satisfied for $f(x) = |x| - 1$ because there is no tangent line at $x = 0$.

The mean value theorem is a technical result whose applications are more important than the theorem itself. The mean value theorem is, like the intermediate value and extreme value theorems, an existence theorem. It asserts the existence of a point in an interval where a function has a particular behavior, but it does not tell you how to find the point.

Theorem 4.3. *Mean Value Theorem*

If f is continuous on $[a, b]$ and is differentiable on (a, b), then there is a point x_0 in (a, b) at which

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}. \blacksquare \tag{4.1}$$

In physical terms, the mean value theorem says that the average velocity of a moving object during an interval of time is equal to the instantaneous velocity at some moment in the interval. Geometrically, the theorem says that a secant line drawn through two points on a smooth graph is parallel to the tangent line at some intermediate point on the curve. There may be more than one such point, as in Figure 4.5. Consideration of these physical and geometric interpretations should make the theorem believable.

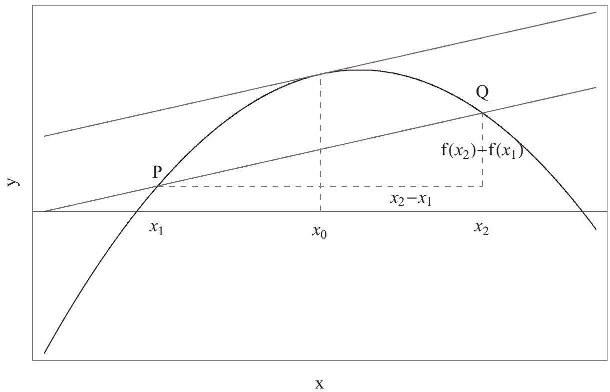


Figure 4.7. The slope of the secant line $(f(x_2) - f(x_1))/(x_2 - x_1)$ is equal to the slope of the tangent line of the curve at a certain point x_0.

We will not prove the mean value theorem. However, we will use it as a tool to make some estimations. For now, we will concentrate on some applications. These tell us that if we know something about $f'(x)$ for all x in $[a, b]$, then we can conclude something about the relation between values of $f(x)$ at different points in $[a, b]$.

Remark 4.1. Consequences of the Mean Value Theorem

Suppose that f is differentiable on (a, b) and continuous on $[a, b]$.

1. In addition let us assume that there are two numbers A and B such that $A \le f'(x) \le B$ for all x in (a, b). Then for any two distinct points x_1 and x_2 in $[a, b]$, $A \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le B$.

2. If $f'(x) = 0$ on (a, b), then f is constant on $[a, b]$.

3. Let $F(x)$ and $G(x)$ be functions such that $F'(x) = G'(x)$ for all x in an open interval (a, b). Then there is a constant C such that $F(x) = G(x) + C$ for all x in (a, b).

The first consequence holds since $[f(x_2) - f(x_1)]/(x_2 - x_1) = f'(x_0)$ for some x_0, between x_2, and x_1, by the mean value theorem applied to f on the interval with endpoints x_2 and x_1. Since x_0, is in (a, b), $f'(x_0)$ lies between A and B, so does the difference quotient. In particular, if $f'(x_0) = 0$, we can choose $A = B = 0$, which implies that $f(x_1) = f(x_2)$. Hence f is constant. To obtain the third consequence, observe that $F(x) - G(x)$ has zero derivative, so it is a constant. This consequence of the mean value theorem is a fact about anti derivatives which we used in Section 5.1.3.

Example 4.3. Mean Value Theorem I

The velocity of a train is kept between 40 and 50 kilometers per hour during a trip of 200 kilometers. What can you say about the duration of the trip?

Solution 4.3. Before presenting a formal solution using the mean value theorem, let us use common sense. If the velocity is at least 40 kilometers per hour, the trip takes at most $200/40 = 5$ hours. If the velocity is at most 50 kilometers per hour, the trip takes at least $200/50 = 4$ hours. Thus, the trip takes between 4 and 5 hours.

To use the mean value theorem, let $f(t)$ be the position of the train at time t; let a and b be the beginning and ending times of the trip. By consequence 1 with $A = 40$ and $B = 50$, we have $40 \leq (f(b) - f(a))/(b - a) \leq 50$. But $f(b) - f(a) = 200$, so $40 \leq 200/(b - a) \leq 50$ which is $1/5 \leq 1/(b - a) \leq 1/4$, and $5 \geq b - a \geq 4$. Hence the trip takes somewhere between 4 and 5 hours, as we found above.▲

Example 4.4. Mean Value Theorem II

You are driving on a straight highway on which the speed limit is 55 km$/h$. At 8.05am a police car clocks your velocity at 50 km$/h$ and at 8.10 am a second police car posted 5 km down the road clocks your velocity at 55 km$/h$. Explain why the police have the right to charge you with a speeding violation.

Solution 4.4. You traveled 5 km in 5 min (= 1/12 h), so your average velocity was 60 km$/h$. However, the Mean Value Theorem guaranties the police that your instantaneous velocity was 60 km$/h$ at least once over the 5 km section of highway.▲

4.3.1 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

4.3.1.1 Test Problems

T1. Explain the practical meaning of Roll's theorem.

T2. What is the Mean Value Theorem useful for?

T3. Explain the geometric meaning of Roll's Theorem and the Mean Value Theorem.

4.3.1.2 Exercises

E1. The function

$$f(x) = \begin{cases} x & 0 \le x < 1 \\ 0 & x = 1 \end{cases} \tag{1}$$

is zero at $x = 0$ and $x = 1$ and differentiable on $(0, 1)$, but its derivative on $(0, 1)$ is never zero. How can this be? Doesn't Rolle's Theorem say the derivative has to be zero somewhere in $(0,1)$? Give reasons for your answer.

E2. Find the value or values of x_0 that satisfy the equation of the Mean Value Theorem for the following functions:

a. $f(x) = x^2 + 2x - 1$ with $x \in [0, 1]$,

b. $f(x) = \sqrt{x-1}$ with $x \in [1, 3]$,

c. $f(x) = x^{2/3}$ with $x \in [0, 1]$,

d. $f(x) = x + \frac{1}{x}$ with $x \in [1/2, 2]$.

E3. Use Rolle's Theorem to prove that between every two zeros of $x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ there lies a zero of $n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \ldots + a_1$.

E4. Which of the functions listed below satisfy the hypotheses of the Mean Value Theorem on the given interval, and which do not? Give reasons for your answers.

a. $f(x) = x^{2/3}$ with $x \in [-1, 8]$,

b. $f(x) = x^{4/5}$ with $x \in [0, 1]$,

c. $f(x) = \begin{cases} \frac{\sin(x)}{x} & -\pi \le x < 0 \\ 0 & x = 0 \end{cases}$,

d. $f(x) = \sqrt{x(1-x)}$ with $x \in [0, 1]$.

E5. For what values of a, m and b does the function

$$f(x) = \begin{cases} 3 & x = 0 \\ -x^2 + 3x + a & 0 < x < 1 \\ m x + b & 1 \le x \le 2 \end{cases} \tag{2}$$

satisfy the hypotheses of the Mean Value Theorem on the interval $[0, 2]$?

E6. Find the function with the given derivative whose graph passes through the point P.

a. $f'(x) = 2x - 1$, $P(0, 0)$,

b. $f'(x) = 8 - \csc(x)^2$, $P(\pi/4, 0)$,

c. $f'(x) = \frac{1}{x^2} + 2x$, $P(-1, 1)$,

d. $f'(x) = \sec(x)\tan(x) - 1$, $P(0, 0)$.

E7. Give the velocity and initial position of a body moving along a coordinate line. Find the body's position at time t.

a. $v = 9.8t + 5$, $s(0) = 10$,

b. $v = 32t - 2$, $s(1/2) = 4$,

c. $v = \frac{2}{\pi}\cos\left(\frac{2t}{\pi}\right)$, $s(\pi^2) = 1$,

d. $v = \sin(\pi t)$, $s(0) = 0$.

E8. A trucker handed in a ticket at a toll booth showing that in 2 hours he had covered 159 km on a toll road with speed limit 65 km/h. The trucker was cited for speeding. Why?

E9. A runner at the Cairo Pyramid Run ran the 8 km in half an hours. Show that at least twice the runner was running at exactly 15 km/h.

E10 Show that for any number a and b, the inequality $|\sin(b) - \sin(a)| \le |b - a|$ is true.

4.4 First and Second Derivative Test

To develop an effective method for finding relative extrema of a function f, we need some criteria that will enable us to distinguish between the critical numbers where relative extrema occur and those where they do not. One such criterion can be motivated by examining the sign of the first derivative of f on each side of the eight critical numbers in Figure 4.4.

4.4.1 First Derivative Test

The findings in Figure 4.4 can be used to classify the properties of extrema in the following way:

- At the two relative maxima in the top row, f' is positive to the left of x_0 and negative to the right.
- At the two relative minima in the second row, f' is negative to the left of x_0 and positive to the right.
- At the first two critical numbers in the row before the bottom row f' is positive on both sides of x_0.
- At the last two critical numbers in the bottom row, f' is negative on both sides of x_0.

These observations suggest that a function f will have relative extrema at those critical numbers where f' changes sign. Moreover, if the sign changes from negative to positive, then a relative minimum occurs. This is the contents of the following theorem.

Theorem 4.4. *First Derivative Test*

Suppose f is continuous at a critical number x_0.

a) If $f'(x) > 0$ on an open interval extending left from x_0 an $f'(x) < 0$ on an open interval extending right from x_0, the f has a relative maximum at x_0.

b) If $f'(x) < 0$ on an open interval extending left from x_0 an $f'(x) > 0$ on an open interval extending right from x_0, the f has a relative minimum at x_0.

c) If $f'(x)$ has the same sign (either $f'(x) > 0$ or $f'(x) < 0$) on an open interval extending left from x_0 and on an open interval extending right from x_0, then f does not have a relative extremum at x_0. ■

Example 4.5. First Derivative Test I

Locate the relative maxima and minima of

$$f(x\_) := 3x^3 - 15x^2$$

Solution 4.5. The function f is defined and continuous for all positive real values of x and its derivative is

der = $\frac{\partial f(x)}{\partial x}$

$9\,x^2 - 30\,x$

This quadratic polynomial defines the critical values. We find those values by solving the equation

$$\frac{\partial f(x)}{\partial x} = 0 \tag{4.2}$$

The solutions are given by the two values

sol = Flatten[Solve[der == 0, *x*]]

$\left\{x \to 0,\, x \to \frac{10}{3}\right\}$

$x = 0$ and $x = 10/3$. To apply the first derivative test, we examine the sign of $f'(x)$ on intervals extending to the left and right of the critical numbers. Since the sign of the derivative changes at $x = 0$ from negative to positive if we sweep the interval from larger to smaller values, there is a relative maximum at $x = 0$. The calculations are for the upper border of the interval

der /. *x* → *x* + 0.1 /. sol⟦1⟧

-2.91

for the lower border we find

der /. *x* → *x* − 0.1 /. sol⟦1⟧

3.09

This shows that the slope changes from a positive to a negative sign, representing a relative maximum.

For the second critical number at $x = 10/3$ the same procedure yields that the derivative $f'(x)$ changes sign from a positive value to a negative value if we approach the critical value from above. This means that at $x = 10/3$ there exists a relative minimum of f. The calculation at the upper boundary of the interval yields

der /. *x* → *x* + 0.1 /. sol⟦2⟧

3.09

On the lower boundary the derivative delivers

der /. *x* → *x* − 0.1 /. sol⟦2⟧

-2.91

The graph of the function shows that at $x = 0$ there exists a relative maximum and at $x = 10/3$ we find a relative minimum.

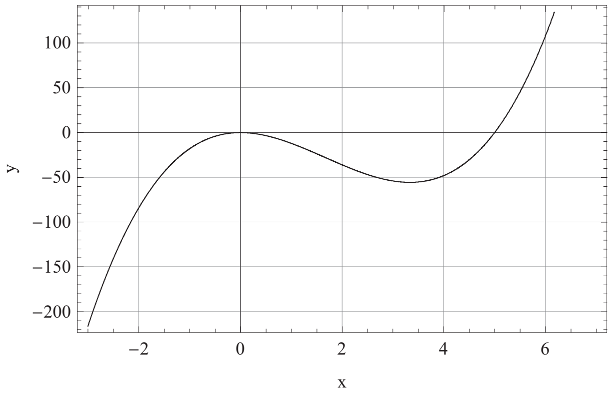


Figure 4.8. Graph of the function $f(x) = 3\,x^3 - 15\,x^2$.▲

Example 4.6. First Derivative Test II

Locate the relative extrema of

$$\boldsymbol{f(\mathrm{x\_}) := x^3 - 3\,x^2 + 3\,x - 1}$$

if any.

Solution 4.6. Since f is differentiable everywhere, the only possible critical number correspond to stationary points. Differentiating f yields

$$\textbf{der = Factor}\left[\frac{\partial f(x)}{\partial x}\right]$$

$$3\,(x-1)^2$$

Solving $f'(x) = 0$ yields

sol = Solve[der == 0, x]

$\{\{x \to 1\}, \{x \to 1\}\}$

However, the derivative $f'(x) = 3\,(x-1)^2 \geq 0$ for all x, so $f'(x)$ does not change sign at $x = 1$; consequently, f does not have a relative extremum at $x = 1$. Thus f has no relative extrema.

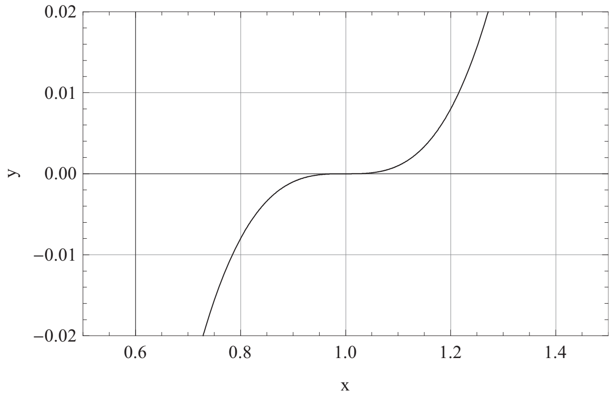


Figure 4.9. Graph of the function $f(x) = x^3 - 3x^2 + 3x - 1$.▲

4.4.2 Second Derivative Test

There is another test for relative extrema that is often easier to apply than the first derivative test. It is based on the geometric observation that a function f has a relative maximum at a stationary point if the graph of f is concave down on an open interval containing the point, and it has a relative minimum if it is concave up (Figure 4.10)

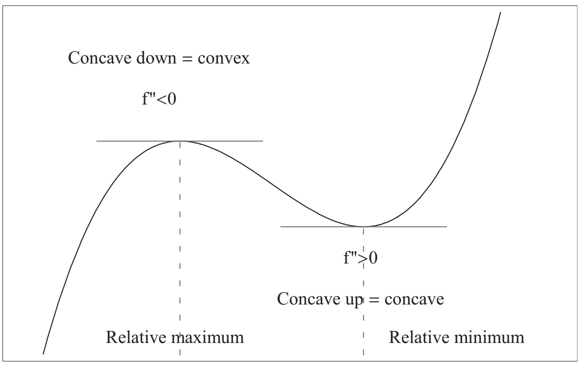


Figure 4.10. Behavior of second order derivatives.

Theorem 4.5. *Second Derivative Test*

Suppose that f is twice differentiable at x_0.

a) If $f'(x_0) = 0$ and $f''(x_0) > 0$, then f has a relative minimum at x_0.

b) If $f'(x_0) = 0$ and $f''(x_0) < 0$, then f has a relative maximum at x_0.

c) If $f'(x_0) = 0$ and $f''(x_0) = 0$, then the test is inconclusive, that is, f may have a relative *maximum, a relative minimum, or neither at* x_0. ■

Example 4.7. Second Derivative I

Locate the relative maxima and minima of

$$f(\text{x\_}) := x^4 - 2x^2$$

and confirm that your results are consistent with the graph of f.

Solution 4.7. First we calculate the derivative

$$\text{der} = \frac{\partial f(x)}{\partial x}$$

$$4x^3 - 4x$$

The solution with respect to x gives us

sol = Solve[der == 0, *x*]

$\{\{x \to -1\}, \{x \to 0\}, \{x \to 1\}\}$

The second derivative we need for testing is given by

$$\text{der2} = \frac{\partial^2 f(x)}{\partial x \, \partial x}$$

$$-4 + 12\,x^2$$

Inserting the found values for the critical numbers into the second derivative yields either a criteria for a maximum or minimum or neither. The first solution at $x = 1$ provides

der2 /. sol⟦1⟧

8

a positive second derivative; that is at $x = 1$ there is a relative minimum. The second critical value at $x = 0$ yields

der2 /. sol⟦2⟧

-4

a negative second derivative; that is there is a relative maximum. The last critical value gives

der2 /. sol⟦3⟧

8

a positive second derivative; that is there exists a relative minimum at that position

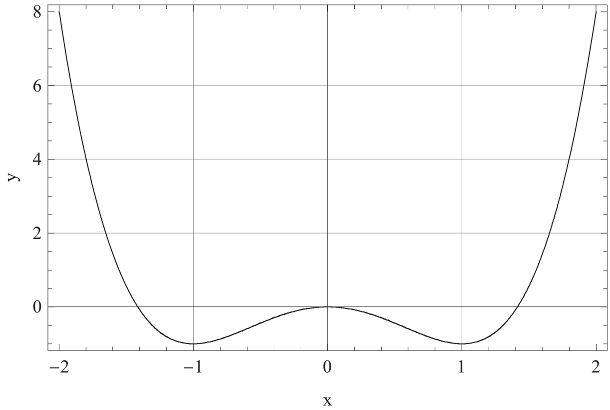


Figure 4.11. Graph of the function $f(x) = x^4 - 2x^2$.

The graph of the function shows that two relative minima at $x = 1$ and $x = -1$ and one relative maxima at $x = 0$ exists.▲

Example 4.8. Second Derivative Test II

Generate the graph of the function

$$f(\text{x\_}) := \frac{\log(x)}{x}$$

and locate the position of all relative extrema.

Note that the domain of $f(x)$ is $(0, +\infty)$. Figure 4.12 shows a graph of $f(x) = \ln(x)/x$. This figure suggests that the graph has a relative maximum.

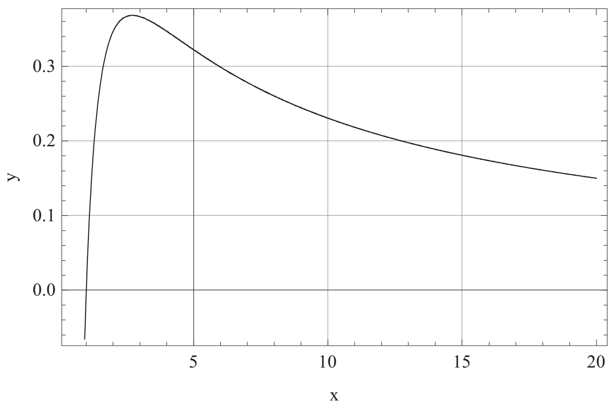


Figure 4.12. Graph of the function $f(x) = \ln(x)/x$.

Solution 4.8. The first derivative of the function is given by

der = $\dfrac{\partial f(x)}{\partial x}$

$$\frac{1}{x^2} - \frac{\ln(x)}{x^2}$$

Solving for $f'(x) = 0$ yields the stationary point

sol = Flatten[Solve[der == 0, *x*]]

$\{x \to e\}$

Inserting this point into the second derivative yields

der2 = $\dfrac{\partial^2 f(x)}{\partial x \partial x}$ **/. sol**

$$-\frac{1}{e^3}$$

a negative value so there is a relative maximum at $x = e$.▲

4.4.3 Inflection Points

Points where a graph changes from concave to convex, ore vice versa, are of special interest, so there is some terminology associated with them.

Definition 4.2. *Inflection Point*

If f is continuous on an open interval containing a value x_0, and if f changes the direction of its concavity at the point $(x_0, f(x_0))$, then we say that f has an inflection point at x_0 and we call the point $(x_0, f(x_0))$ on the graph of f an inflection point of f.■

This definition represents the observation that the inflection point of a curve $y = f(x)$ mark the points where the slope of the tangent lines change from increasing to decreasing, or vice versa. Thus, in the case where f is differentiable, $f'(x)$ will have a relative maximum or relative minimum at any inflection point of f (see Figure 4.13).

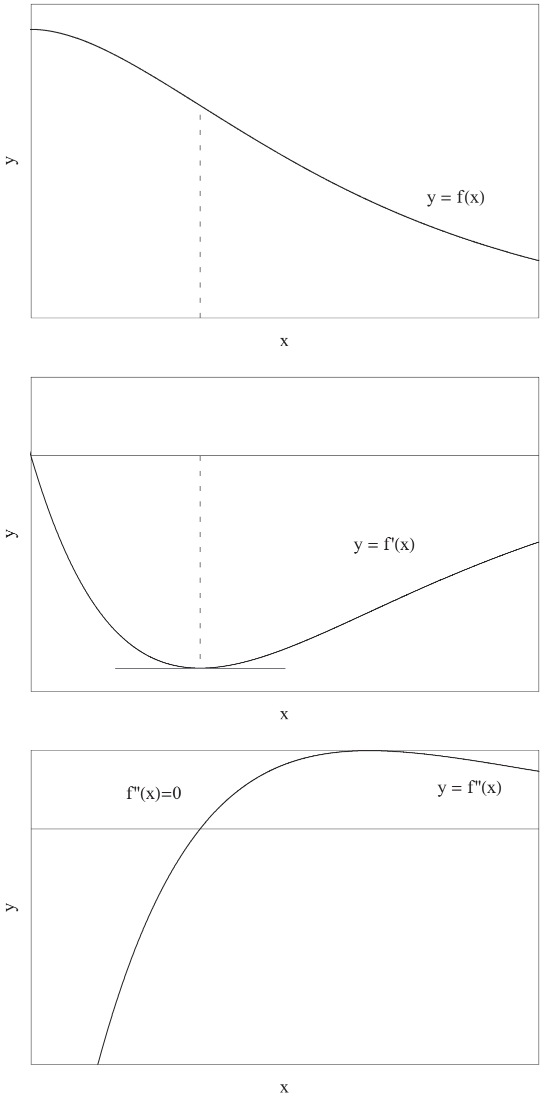


Figure 4.13. Behavior of functions at an inflection point.

For a differentiable function $y = f(x)$, the rate of change of y with respect to x will have a relative extremum at any inflection point of f. That is, an inflection point defines a place on the graph of $y = f(x)$ where the graph is steepest or where the graph is least steep in the vicinity of the point.

Example 4.9. Inflection Point

Show that the function

$$f(\mathrm{x\_}) := x\,e^{-x}$$

has an inflection point.

Solution 4.9. The first derivative of the function is

$$\mathbf{der1} = \mathbf{Factor}\left[\frac{\partial f(x)}{\partial x}\right]$$

$$-e^{-x}(x-1)$$

The relative extrema follow from $f'(x) = 0$. The solution for this example shows that

sol = Flatten[Solve[der == 0, *x*]]

$\{x \to 1\}$

$x = 1$ is a critical number where a relative maximum occurs. This is checked by the second derivative test

$$\mathbf{der2} = \mathbf{Factor}\left[\frac{\partial^2 f(x)}{\partial x \partial x}\right]$$

$$e^{-x}(x-2)$$

Inserting the critical value into the second derivative shows that the value of f'' is negative

der2 /. sol

$(e-2)\,e^{-e}$

which results to a relative maximum. The inflection point follows from the vanishing of the second derivative. If we solve the equation $f''(x) = 0$ we find

sol2 = Flatten[Solve[der2 == 0, *x*]]

$\{x \to 2\}$

Checking the second derivative in the neighborhood of this point shows that the second derivative is positive for values greater than $x = 2$

der2 /. *x* → *x* + .1 /. sol2

0.0122456

and is negative for values smaller than $x = 2$.

der2 /. *x* → *x* − .1 /. sol2

−0.0149569

The graphical representation of the curve is given in the following figure

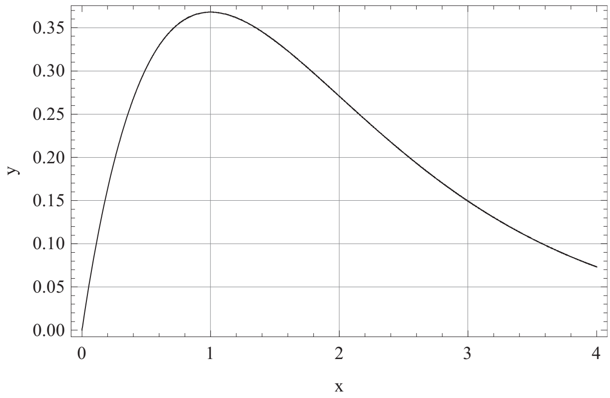


Figure 4.14. Graph of the function $f(x) = x\,e^{-x}$.▲

4.4.4 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

4.4.4.1 Test Problems

T1. What is the First Derivative Test for Local Extreme Values? Give examples of how it is applied.

T2. What is an inflection point? Give an example. What physical significance do inflection points sometimes have?

T3. What do the derivatives of a function tell you about the shape of its graph?

T4. How do you find the absolute extrema of a continuous function on a closed interval? Give examples.

T5. What does it mean for a function to have a local extreme value on its domain? An absolute extreme value? How are local and absolute extreme values related, if at all? Give examples.

4.4.4.2 Exercises

E1. Answer the following questions about the functions whose derivatives are given: i) What are the critical points of f? ii) On what intervals is f increasing or decreasing? iii) At what points, if any, does f assume local maximum and
minimum values?

a. $f'(x) = (x-1)^2\,(x+2)$,

b. $f'(x) = (x-1)\,(x+2)\,(x-3)$,

c. $f'(x) = x(x-1)$,

d. $f'(x) = x^{-1/2}(x-3)$,

e. $f'(x) = (x-1)^2\,(x+2)^2$,

f. $f'(x) = (x-1)\,(x+2)$,

E2. Find the local extrema of each function on the given interval. In addition graph the function and its derivative together. Comment on the behavior of f in relation to the signs and values of f'.

a. $f(x) = \frac{x}{2} - 2\sin\left(\frac{x}{2}\right)$, $\quad 0 \le x \le 2\pi$,

b. $f(x) = -2\cos(x) - \cos(x)^2$, $\quad -\pi \le x \le \pi$,

c. $f(x) = \sec(x)^2 - 2\tan(x)$, $\quad \frac{-\pi}{2} < x < \frac{\pi}{2}$,

d. $f(x) = \csc(x)^2 - 2\cot(x)$, $\quad 0 < x < \pi$.

E3. For the listed functions below find the intervals on which the function is increasing and decreasing. Then identify the function's local extreme values, if any, saying where they are taken on. Which, if any, of the extreme values are absolute? Support your findings with a graph on a computer.

a. $f(x) = -x^2 - 3x + 3$,
b. $f(x) = x^3 - 18x$,
c. $f(x) = 6x - x^3$,
d. $f(x) = (x + 7)^3$,
e. $f(x) = 3x^2 - 4x^3$,
f. $f(x) = 3x^3 + 16x$,
g. $f(x) = \frac{x^3}{3x^2+1}$,
h. $f(x) = x^2\sqrt{5 - x}$,
i. $f(x) = \frac{x^2-3}{x-2}, \quad x \neq 2$,
j. $f(x) = x^{2/3}(x^2 - 4)$.

E4. Prove that the function

$$f(x) = x^{101} + x^{51} + x + 1 \tag{1}$$

has neither a local maximum nor a local minimum.

E5. An object with weight W is dragged along a horizontal plane by a force acting along a rope attached to the object. If the rope makes an angle θ with the plane, then the magnitude of the force is

$$F = \frac{\mu W}{\mu \sin(\theta) + \cos(\theta)} \tag{2}$$

where μ is a positive constant called the coefficient of friction and where $0 \leq \theta \leq \pi/2$. Show that F is minimized when $\tan(\theta) = \mu$.

E6. A cubic function is a polynomial of degree 3; that is, it has the form $f(x) = a x^3 + b x^2 + c x + d$, where $a \neq 0$.

a. Show that a cubic function can have two, one or no critical numbers. Give examples and sketches to illustrate the three possibilities.
b. How many local extreme values can a cubic function have?

E7. As x moves from left to right through the point $c = 2$ is the graph of $f(x) = x^3 - 3x + 2$ rising, or is it falling? Give reasons for your answer.

E8. Use the sign pattern for the derivative

$$\frac{\mathrm{df}}{\mathrm{dx}} = 6(x-1)(x-2)^2(x-3)^3(x-4)^4 \tag{3}$$

to identify the points where f has local maximum and minimum values.

E9. Suppose that f is continuous on $[a, b]$ and that c is an interior point of the interval. Show that if $f'(x) \leq 0$ on $[a, b]$ and $f'(x) \geq 0$ on $(x, b]$, then $f(x)$ is never less than $f(c)$ on $[a, b]$.

E10 Assume that the brakes of an automobile produce a constant deceleration of k m/s. a) Determine what k must be to bring an automobile traveling 60 km/h to rest in a distance of 100 m from the point where the brakes are applied. b) With the same k, how far would a car traveling 30 km/h travel before being brought to a stop?

4.5 Applications in Optimization

In this section we will show how the methods discussed in the last section can be used to solve various applied optimization problems.

4.5.1 Classes of Optimization Problems

The applied optimization problems that we will consider in this section fall into the following two categories:

- Problems that reduce to maximizing or minimizing a continuous function over a finite closed interval.
- Problems that reduce to maximizing or minimizing a continuous function over a finite interval or a finite interval that is not closed.

For problems of the first type we know that the solution can be obtained by examining the values of the function at the critical numbers and at the endpoints. However, for problems of the second type there may or may not be a solution. If the function is continuous and has exactly one relative extremum of the appropriate type in an interval.

4.5.2 Standard Applications

In his *On a Method for the Evaluation of Maxima and Minima*, the seventeenth century French mathematician Pierre de Fermat solved an optimization problem very similar to the one posed in our first example. Fermat's work on such optimization problems prompted the French mathematician Laplace to proclaim Fermat the "true inventor of the differential calculus". Although this honor must still reside with Newton and Leibniz, it is the case that Fermat developed procedures that anticipated parts of differential calculus.

Example 4.10. Geometric Problem

A garden is to be laid out in a rectangular area and protected by a chicken wire fence. What is the largest possible area of the garden if only 100 running meters of chicken wire is available for the fence?

Solution 4.10. Let

- x be the length of the rectangle in meter
- y is the width of the rectangle in meter
- A is the area of the rectangle in square meter.

The area of the rectangle is then given by

$$\textbf{Area} = \boldsymbol{x\,y}$$

$$x\,y$$

Since the perimeter of the rectangular area is 100 meter, the variables x and y are related by the equation

$$\textbf{perimeter} = \mathbf{2}\,\boldsymbol{x} + \mathbf{2}\,\boldsymbol{y} = \mathbf{100}$$

$$2\,x + 2\,y = 100$$

Solving for one of the variables say y of the perimeter relation and substituting this result into the area relation, we find

sol = Flatten[Solve[perimeter, *y*]]

$\{y \to 50 - x\}$

and the substitution delivers

area = Area /. sol

$(50 - x)\, x$

Because x represents a length it cannot be negative, and because the two sides of length x cannot have a combined length exceeding the total perimeter of 100 meter, the variable x must satisfy

$$0 \leq x \leq 50$$

Thus we have reduced the problem to that of finding the value of x in [0, 50], for which A is a maximum. Since the area is represented by a polynomial in x it is continuous on [0, 50], and so the maximum must occur at an endpoint of this interval or at a critical number. From the area relation we get

$$\mathbf{der} = \frac{\partial\, \mathbf{area}}{\partial x}$$

$50 - 2\, x$

Setting this relation equal to zero and solving with respect to x, we obtain

sol = Flatten[Solve[der == 0, *x*]]

$\{x \to 25\}$

The second derivative test shows that

$$\frac{\partial^2\, \mathbf{area}}{\partial x\, \partial x}$$

-2

represents a maximum. Substituting the found value into the area relation, we find that the maximal value of the area is given by

area /. sol

625

The relation for the area is shown in Figure 4.15

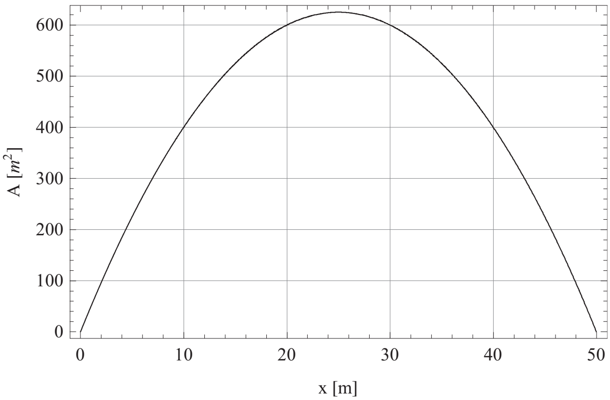


Figure 4.15. Change of the area by changing the length of the rectangle *x*.▲

Example 4.11. Equipment Costs

Figure 4.16 shows an offshore oil well located at a point *W* that is 5 km from the closest point A on a straight shoreline. Oil is to be piped from *W* to the shore point *B* that is 8 km from *A*. The pipes are laid under water in a straight line from *W* to some shore point *P* between *A* and *B* and then on to *B* via pipe along the shoreline. If the cost of the construction is \$1,000,000.00 per km under water and \$500,000 per km over land, where should the point *P* be located to minimize the cost of laying the pipe?

Solution 4.11. Let

- x be the distance between *A* and *P*
- C the cost in million dollars for the entire pipeline

From Figure 4.16 the length of pipe under water is the distance between *W* and *P*. By the Theorem of Pythagoras, the length is

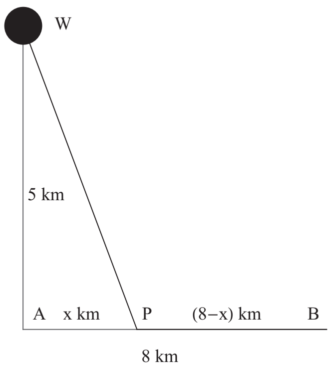


Figure 4.16. Geometry for an oil pipeline from *W* to *B* off shore and on shore.

$$\textbf{PipeLengthWP} = \sqrt{x^2 + 5^2}$$

$$\sqrt{x^2 + 25}$$

Also from Figure 4.16 the length of the pipe over land is the distance between *P* and *B*, which is

$$\textbf{PipeLengthPB} = 8 - x$$

$$8 - x$$

From these two relations it follows that the cost in million of $ is given by

$$\textbf{Cost} = \frac{\textbf{PipeLengthPB}}{\textbf{2}} + \textbf{1 PipeLengthWP}$$

$$\sqrt{x^2 + 25} + \frac{8 - x}{2}$$

Because the distance between *A* and *B* is 8 km, the distance *x* between *A* and *P* must satisfy

$0 \leq x \leq 8.$

We have thus reduced our problem of finding the value of *x* in the interval [0, 8] for which the cost function must be a minimum. Since the cost function is continuous on the closed interval [0, 8], we can use the methods developed in the preceding section to find the minimum. From the cost function the first order derivative follows by

$$\textbf{der1} = \frac{\partial \textbf{Cost}}{\partial x}$$

$$\frac{x}{\sqrt{x^2 + 25}} - \frac{1}{2}$$

Setting the first derivative equal to zero and solving with respect to *x*, we find

sol = Flatten[Solve[der1 == 0, *x*]]

$$\left\{x \to \frac{5}{\sqrt{3}}\right\}$$

Since this number is contained in the interval [0, 8], the minimum must occur at this value. The second order derivative test confirms this

$$\frac{\partial^2 \textbf{Cost}}{\partial x \partial x} \textbf{ /. sol}$$

$$\frac{3\sqrt{3}}{40}$$

the result is positive and thus the cost function has a minimum. The following graph shows the cost function as a function of the distance *x*

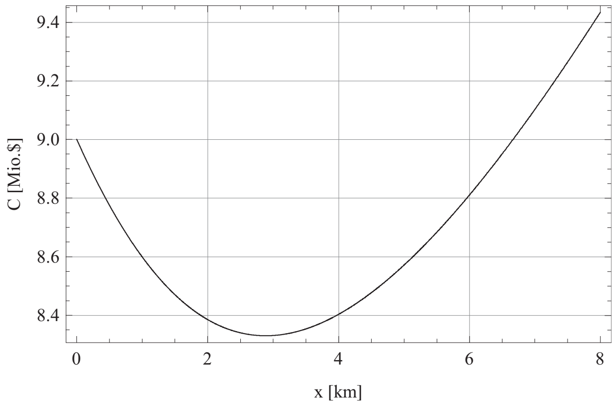


Figure 4.17. Cost function for an oil pipeline from *W* to *B* off shore and on shore.

The total costs for this project are found by inserting the found value for x into the cost function; that is

N[Cost /. sol]

8.33013

The costs are given in Mio. \$.▲

Example 4.12. Motion of a Projectile

The motion of a projectile was and is an example of importance because in a baseball or golf play, we need to know where the ball touches down if we give it a strike. Ancient people needed also to know where the stones or bullets go if they are thrown by a bow. Applications for military purposes are evident. The most important point for this kind of problem is under which condition the projectile is thrown to the far end point possible.

In this example let us consider the motion of a projectile or a ball in the atmosphere. In our considerations, we neglect air resistance. Furthermore, we consider only kinematics; we also demand that the projectile follows a parabolic orbit with a vertical symmetry axis and with constant horizontal velocity. The motion of the ball takes place in a three-dimensional space; thus, the velocity and the location of the ball is a certain vector with three components, respectively. These components are independent of each other and, thus, can be considered on its own. Thus we are able to separate each direction of the motion from the others. If we assume that the projectile is moving in a plane, we only need two coordinates to describe the motion. Let us further assume that the ball is thrown with a finite velocity $\vec{v}_0$ inclined by an angle α with respect to the horizontal direction. To simplify things, let us first define the origin as the starting point of the ball. Later, we will generalize this to the situation where the starting point does not coincide with the origin.

Solution 4.12. The track of the ball is defined by the parametric representation in time t by

$$\textbf{track} = \left\{t\,\mathrm{vx} + \mathrm{x0},\ \frac{1}{2}(-g)\,t^2 + \mathrm{vy}\,t + \mathrm{y0}\right\};$$

where x_0 and y_0 are the starting point and v_x and v_y are the velocities in x- and y-direction, respectively. The assumption that the origin is the starting point of the track causes the vanishing of x_0 and y_0. This condition is specifically initiated by the following *Mathematica* line where the initial condition is set.

cond1 = Flatten[Solve[Thread[{0, 0} == track /. $t \to$ 0, List], {x0, y0}]]

$\{x0 \to 0, y0 \to 0\}$

Inserting these initial conditions into the track, we find a simplified representation of the track in component form by

tracS = track /. cond1

$$\left\{t\,\text{vx},\ t\,\text{vy} - \frac{g\,t^2}{2}\right\}$$

The assumption that the ball is thrown in a certain direction with inclination α to the horizon and initial velocity v allows us to determine the parameters vx and vy in the track representation. To introduce α we assume that the magnitude of the velocity can be divided into an x and y component by $v_x = v\cos(\alpha)$ and $v_y = v\sin(\alpha)$. These two components are equal to the first order derivative of our track at the initial state. If we solve these two equations with respect to the velocity components we find

$$\textbf{cond2} = \textbf{Flatten}\left[\textbf{Solve}\left[\textbf{Thread}\left[\{v\cos(\alpha), v\sin(\alpha)\} == \frac{\partial\,\textbf{tracS}}{\partial t}\ /.\ t \to 0, \textbf{List}\right], \{\textbf{vx}, \textbf{vy}\}\right]\right]$$

$\{\text{vx} \to v\cos(\alpha), \text{vy} \to v\sin(\alpha)\}$

Again inserting the results into the track coordinates, we end up with the final representation of the path by

tracS1 = tracS /. cond2

$$\left\{t\,v\cos(\alpha),\ t\,v\sin(\alpha) - \frac{g\,t^2}{2}\right\}$$

This representation contains two parameters v and α, the magnitude of the velocity and the inclination angle, respectively. Choosing this parameters allows us to plot the track of the ball as shown in Figure 4.18

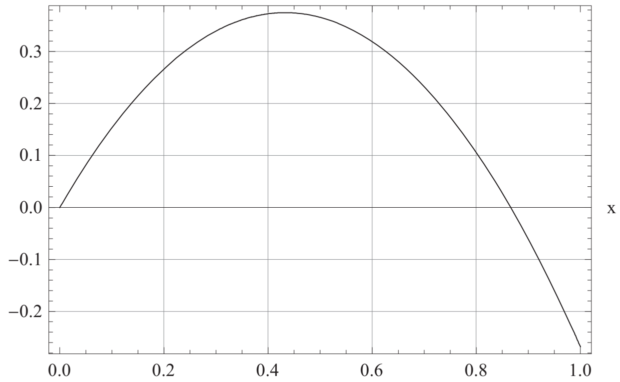


Figure 4.18. Track of a ball thrown into the air.

The motion of the ball for different times can be represented by the known coordinates inserting different times. This is done in the following line

$$\textbf{Ball} = \textbf{Table}\Big[\Big\{\textbf{RGBColor[0, 0, 0.62501]}, \textbf{PointSize[.08]}, \textbf{Point}\Big[\textbf{tracS1} \;/.\; \Big\{v \to 1, \alpha \to \frac{\pi}{3}, g \to 1\Big\}\Big]\Big\}, \{t, 0, 1.7, .1\}\Big];$$

Combining both sets of graphs allows us to display the movement of the ball. The illustration shows that the ball moves first upward and then downward, hitting the ground at a finite distance.

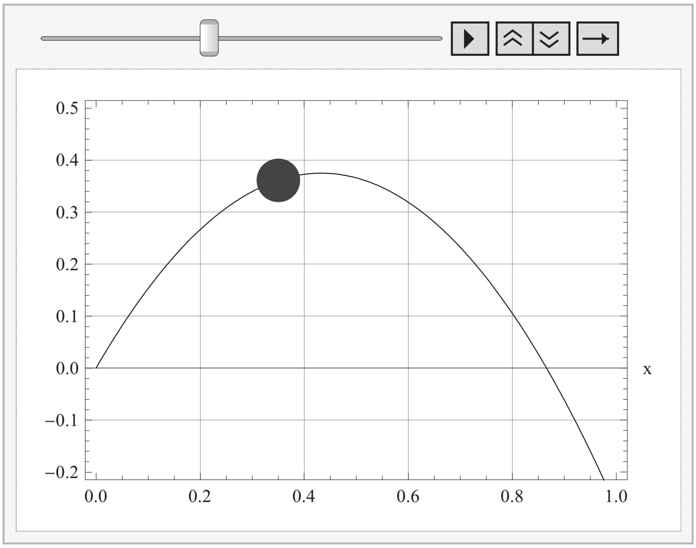


Figure 4.19. Animation of the ball motion thrown from a ground position.

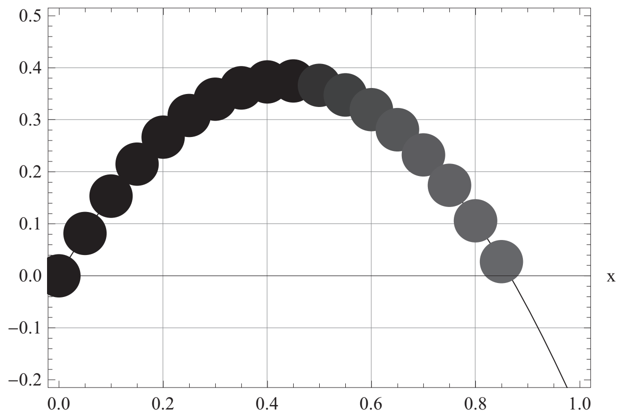


Figure 4.20. The complete sequence of motion steps displayed in a single graph.

This sequence of pictures was generated by plotting the parametric representation of the motion. The general equation of the path $y(x)$ can be obtained be eliminating the variable t in the track representation:

Thread[{*x*, *y*} == tracS1, List]

$$\left\{x = t\, v \cos(\alpha),\ y = t\, v \sin(\alpha) - \frac{g\, t^2}{2}\right\}$$

generalTrack = Flatten[Simplify[Solve[Eliminate[Thread[{*x*, *y*} == tracS1, List], *t*], *y*]]]

$$\left\{y \to x \tan(\alpha) - \frac{g\, x^2 \sec^2(\alpha)}{2\, v^2}\right\}$$

This relation is of the form of a parabola passing through the point (0, 0). The following figure shows the path of the ball:

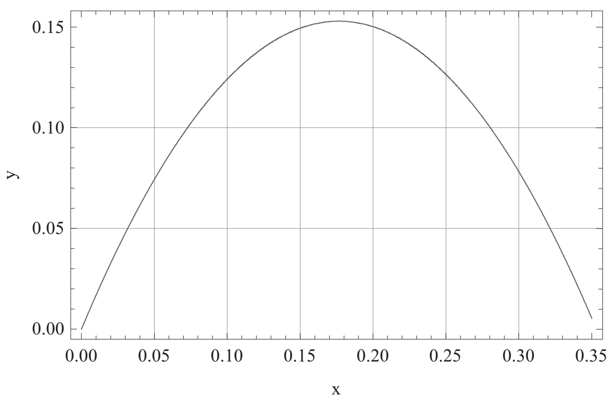


Figure 4.21. Path of the ball.

The range of the flight can be determined from the condition that the y elevation vanishes. This condition serves to determine x from the relation

ranges = Simplify[Solve[generalTrack /. {Rule → Equal, y → 0}, x]]

$$\left\{\{x \to 0\}, \left\{x \to \frac{v^2 \sin(2\,\alpha)}{g}\right\}\right\}$$

The solution of the quadratic equation in x delivers two solutions. Because we are looking for positive ranges, we select the non vanishing second solution:

range = ranges⟦2⟧

$$\left\{x \to \frac{v^2 \sin(2\,\alpha)}{g}\right\}$$

To optimize the distance for x, we consider this expression as a function of α and ask for the extreme values of its dependence. This means we calculate the first order derivative of x with respect to α.

$$\textbf{eq1 = Simplify}\left[\frac{\partial(x\ /.\ \textbf{range})}{\partial\alpha}\right]$$

$$\frac{2\,v^2 \cos(2\,\alpha)}{g}$$

Setting the first order derivative equal to zero and solving with respect to α gives

s1 = Solve[eq1 == 0, α]

$$\left\{\left\{\alpha \to -\frac{\pi}{4}\right\}, \left\{\alpha \to \frac{\pi}{4}\right\}\right\}$$

The second derivative test gives the second order derivative as

$$\textbf{s2 = Simplify}\left[\frac{\partial^2(x\ /.\ \textbf{range})}{\partial\alpha\,\partial\alpha}\right]$$

$$-\frac{4\,v^2 \sin(2\,\alpha)}{g}$$

Inserting the found angles in this expression delivers the results

PowerExpand[Simplify[(s2 /. α → #1 &) /@ α1]]

$\alpha 1$

so that the angles $\pi/4$ and $-\pi/4$ result to a maximal throwing distance. The result gained is shown in the following animation. The optimal ball track is plotted first while increasing or decreasing the angle α results into a path which is shorter than the optimal one.

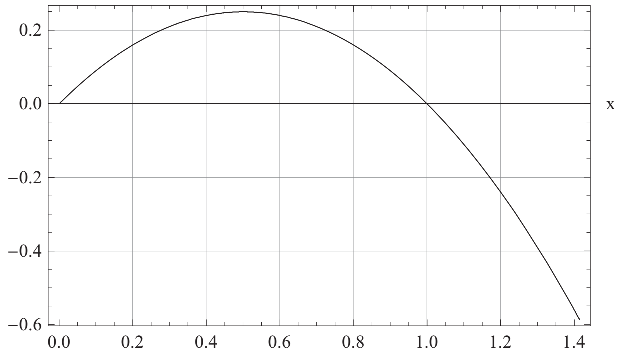


Figure 4.22. Optimal track of the ball.

Then three different path of the ball are generated with different angles α by the following lines

$$\textbf{Ball} = \textbf{Table}\left[\left\{\textbf{RGBColor}[0, 0, 0.62501], \textbf{PointSize}[.08], \textbf{Point}\left[\textbf{tracS1} \,/.\, \left\{v \to 1, \alpha \to \frac{\pi}{4}, g \to 1\right\}\right]\right\}, \{t, 0, 1.8, .1\}\right];$$

$$\textbf{Ball} = \textbf{Join}\left[\textbf{Ball}, \textbf{Table}\left[\left\{\textbf{RGBColor}[0.501961, 0, 0.25098], \textbf{PointSize}[.08], \textbf{Point}\left[\textbf{tracS1} \,/.\, \left\{v \to 1, \alpha \to \frac{\pi}{3}, g \to 1\right\}\right]\right\}, \{t, 0, 1.8, .1\}\right]\right];$$

$$\textbf{Ball} = \textbf{Join}\left[\textbf{Ball}, \textbf{Table}\left[\left\{\textbf{RGBColor}[0, 0, 1], \textbf{PointSize}[.08], \textbf{Point}\left[\textbf{tracS1} \,/.\, \left\{v \to 1, \alpha \to \frac{\pi}{7}, g \to 1\right\}\right]\right\}, \{t, 0, 1.8, .1\}\right]\right];$$

Combining all sets of data allows us to display the movement of the ball. In figure 4.23 a configuration with a suboptimal solution is shown. The optimal solution is represented by the solid parabola.

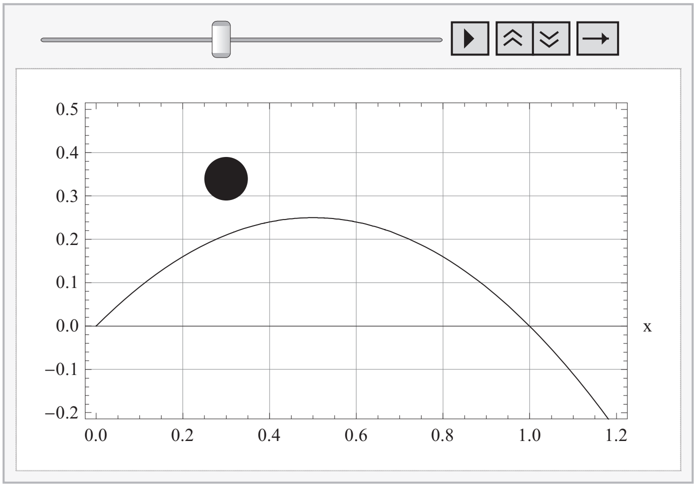


Figure 4.23. The optimal track of a ball compared with two sub optimal paths. The sub optimal path uses slightly different initial conditions in the throwing angle (see text above).▲

Example 4.13. Colors of a Rainbow

Rainbows are created when raindrops scatter sunlight. They have fascinated mankind since ancient times and have inspired attempts at scientific explanation since the time of Aristoteles. In this application we use the ideas of Descartes and Newton to explain the shape and location of a rainbow. We assume for our consideration that the raindrops are approximated by spheres.

Figure 4.24 shows a ray of sunlight entering a spherical raindrop at A. Some of the light is reflected but the line AB shows the path of the part that enters the drop. Notice that the light is refracted toward the normal line OA and in fact Snel's Law says that $\sin(\alpha) = k \sin(\beta)$, where α is the angle of incidence, β is the angle of refraction, and $k \approx 4/3$ is the index of refraction for water. At point B some of the light passes through the drop and is refracted into the air but the line BC shows the part that is reflected. Note that the angle of incidence equals the angle of reflection. When the ray reaches C, part of it is reflected, but for the time being we are more interested in the part that leaves the raindrop at C. The angle of deviation $D(\alpha)$ is the amount of clockwise rotation that the ray has undergone during this three-stage process.

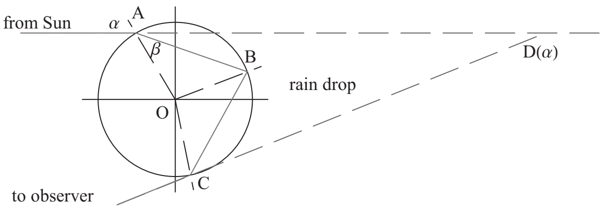


Figure 4.24. Path of a ray from the Sun to the observer. The points A and C are refraction points while B is a reflection point for the ray.

This relation can be read off from Figure 4.24 and is given by

$$D(\alpha) = (\alpha - \beta) + (\pi - 2\,\beta) + (\alpha - \beta) = 2\,\alpha - 4\,\beta + \pi. \tag{4.3}$$

Solution 4.13. Taking into account the relation between the angles in the refraction process $\sin(\alpha) = k\sin(\beta)$ we can rewrite the angle of deviation as

$$\boldsymbol{d(\alpha\_) := 2\,\alpha - 4\sin^{-1}\left(\frac{\sin(\alpha)}{k}\right) + \pi}$$

Since we are looking for the minimal deviation angle we have to determine the first order derivative of the deviation angle. This relation is given by

$$\textbf{Simplify}\left[\frac{\partial\, d(\alpha)}{\partial\alpha}\right]$$

$$2 - \frac{4\cos(\alpha)}{k\sqrt{1 - \frac{\sin^2(\alpha)}{k^2}}}$$

The minimal value for α is gained if we set this expression equal to zero and solve for the angle α

$$\alpha\textbf{m} = \textbf{Simplify}\left[\alpha \text{ /. } \textbf{Solve}\left[\frac{\partial\, d(\alpha)}{\partial\alpha} = 0, \alpha\right]\right]$$

$$\left\{-\cos^{-1}\left(-\frac{\sqrt{k^2-1}}{\sqrt{3}}\right), \cos^{-1}\left(-\frac{\sqrt{k^2-1}}{\sqrt{3}}\right), -\cos^{-1}\left(\frac{\sqrt{k^2-1}}{\sqrt{3}}\right), \cos^{-1}\left(\frac{\sqrt{k^2-1}}{\sqrt{3}}\right)\right\}$$

Since the equation is a quartic equation in α we find four solutions. In addition the result shows that the minimal deviation angle depends crucially on the refraction index k. If we insert the numeric value for the refraction index we get

$$\alpha\textbf{mk} = \alpha\textbf{m} \text{ /. } k \to \frac{\textbf{4.}}{\textbf{3}}$$

$$\{-2.10502, 2.10502, -1.03657, 1.03657\}$$

This corresponds to an angle measured in degree

$$\frac{180\,\alpha\mathbf{mk}}{\pi}$$

{−120.609, 120.609, −59.3911, 59.3911}

To select the minimal value we use the second order derivative test by inserting the found values into the second order derivative

$$\frac{\partial^2 d(\alpha)}{\partial\alpha\,\partial\alpha} \text{ /. Solve}\left[\frac{\partial d(\alpha)}{\partial\alpha} = 0, \alpha\right] \text{ /. } k \to \frac{4.}{3}$$

{−2.53546, 2.53546, −2.53546, 2.53546}

The second and fourth value of our result is greater than zero so the second and fourth solution will result to a minimum. This minimal value is given at $\alpha_m \approx 59.4\,°$. The corresponding angle of deviation is determined by using the fourth value of our result and inserting it into the formula

$$h = d(\alpha\mathbf{mk}[\![4]\!]) \text{ /. } k \to \frac{4}{3}$$

2.40804

which corresponds to an angle of

$$\frac{180\,h}{\pi}$$

137.97

in degrees. The result so far gained means that many rays with $\alpha \approx 138\,°$ become deviated by approximately the same amount. It is the concentration of rays coming from near the direction of minimum deviation to create the brightness of the primary rainbow.

4.5.3 Optimal Designs

This section collects some example demonstrating the design of simple geometric objects like cans, rain gutters, and cups. Optimal designs are important because they will save costs for material and production. In addition the design is based on an approach which is systematic and reproducible.

Example 4.14. Can Design

A cylindrical can is to be made to hold 1 L of oil. Find the dimensions that will minimize the cost of the metal to manufacture the can.

Solution 4.14. A standard can is characterized by two geometric lengths the radius r of the can and the height h of the can. In order to minimize the cost of the metal, we minimize the total surface area of the cylinder (top, bottom, and sides). Taking into account that we have two circles (top and bottom) with area $A_{t/b} = \pi\,r^2$ each and a rectangle of area $A_s = 2\,\pi\,r\,h$ we have the total area of the can given by

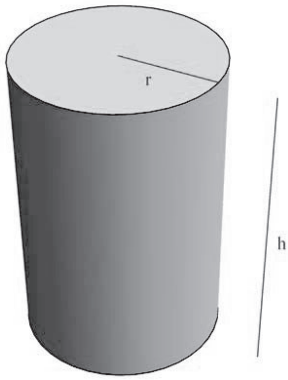


Figure 4.25. Open surface of a can of radius *r* and height *h*.

$$A = A_t + A_b + A_s = 2\,\pi\,r^2 + 2\,\pi\,r\,h.$$

To eliminate h from this expression we make use of the fact that the can has a finite volume V which is given as

$$V = \pi\,r^2\,h$$

This means that the height of the can be expressed as

$$h = \frac{V}{\pi\,r^2}$$

which reduces the formula for the area to

$$A = 2\,\pi\,r^2 + 2\,\pi\,r\,\frac{V}{\pi\,r^2} = 2\,\pi\,r^2 + \frac{2\,V}{r}$$

This formula is shown in Figure 4.26 as a function of r.

$$\boldsymbol{A(\mathrm{r\_}) := 2\,\pi\,r^2 + \frac{2\,V}{r}}$$

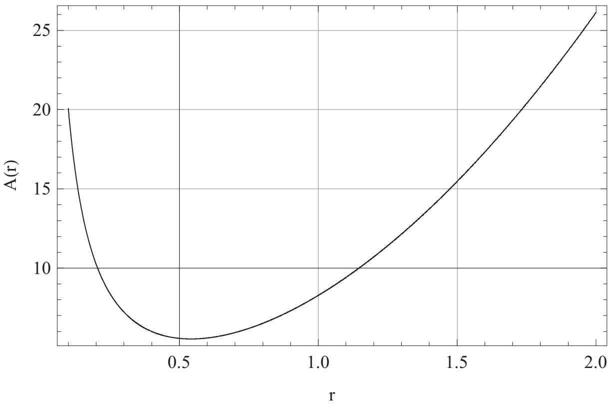


Figure 4.26. Surface area of a can as a function of the radius r if the volume V is known here we used $V = 1$.

To find the critical numbers, we differentiate the are formula for the surface area with respect to r.

$$\textbf{derivative} = \frac{\partial A(r)}{\partial r} = 0$$

$$4\pi r - \frac{2V}{r^2} = 0$$

and equate this relation equal to zero. The solution of this relation withe respect to r will give us the critical numbers for r

sols = Solve[derivative, *r*]

$$\left\{\left\{r \to -\sqrt[3]{-\frac{1}{2\pi}}\sqrt[3]{V}\right\}, \left\{r \to \frac{\sqrt[3]{V}}{\sqrt[3]{2\pi}}\right\}, \left\{r \to \frac{(-1)^{2/3}\sqrt[3]{V}}{\sqrt[3]{2\pi}}\right\}\right\}$$

Since the original rational function consists of a third order polynomial in the numerator we get three different solutions of which only the real solution is useful for practical applications. Thus the potential candidate for the minimal surface is given by

cand1 = sols⟦2⟧

$$\left\{r \to \frac{\sqrt[3]{V}}{\sqrt[3]{2\pi}}\right\}$$

If we look at this solution, we realize that the radius is directly related to the cubic square root of the volume. To show that the derived value will generate a minimal surface we apply the second derivative test to the surface area of the can

$$\frac{\partial^2 A(r)}{\partial r \partial r} \textbf{ /. cand1}$$

$$12\pi$$

which results to a value greater than zero and thus the value $r = V^{1/3} / (2\pi)^{1/3}$ represents the radius for a minimal surface of the can. The value of the height corresponding to this value is derived if we insert the found result into the relation for h (compare the formula above), we find

$$\frac{V}{\pi r^2} \text{ /. cand1}$$

$$\frac{2^{2/3} \sqrt[3]{V}}{\sqrt[3]{\pi}}$$

which corresponds to twice the radius of the minimal surface. So the ratio of $h/r = 2$ for the minimum of the surface. In other words the height of the can should be equal to the diameter of the can to have a minimal surface.▲

Example 4.15. Rain Gutter Design

A rain gutter is to be constructed from a metal sheet of with L which is divided into three segments with equal length a. By bending up one-third of the sheet on each side an angle θ the gutter is formed. How should θ be chosen so that the gutter will carry the maximum amount of water?

Solution 4.15. To maximize the volume of the gutter we have to maximize the cross section of the gutter. To do this we divide the section into two triangles and one rectangle see Figure 4.27.

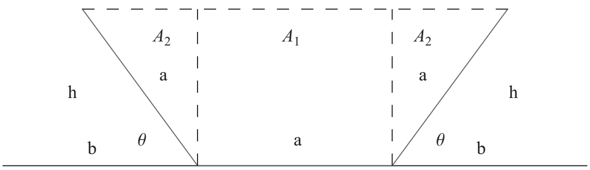


Figure 4.27. Section of the gutter. a is the length of the segment, θ the bending angle and h the height of the gutter.

The total cross-section of the gutter is given by

$$A = A_1 + 2 A_2$$

where the areas are given respectively by

$$\mathbf{A1 = a\, a \sin(\theta)}$$

$$a^2 \sin(\theta)$$

and

$$\mathbf{A2 = \frac{1}{2} a\, a \sin(\theta) \cos(\theta)}$$

$$\frac{1}{2} a^2 \sin(\theta) \cos(\theta)$$

The total cross section of the gutter thus is

A = A1 + 2 A2

$a^2 \sin(\theta) + a^2 \sin(\theta)\cos(\theta)$

The maximal value is gained if we search for the critical numbers of θ. This means we have to solve the equation

$$\mathbf{eq} = \frac{\partial A}{\partial \theta} = 0$$

$-a^2 \sin^2(\theta) + a^2 \cos^2(\theta) + a^2 \cos(\theta) = 0$

with respect to θ

sol1 = Solve[eq, θ]

$$\left\{\{\theta \to -\pi\}, \left\{\theta \to -\frac{\pi}{3}\right\}, \left\{\theta \to \frac{\pi}{3}\right\}, \{\theta \to \pi\}\right\}$$

To check which of the four solutions are related to the maximum we use the second derivative test. Which gives

$$\mathbf{der2} = \frac{\partial^2 A}{\partial\theta\partial\theta}$$

$-a^2 \sin(\theta) - 4\, a^2 \sin(\theta)\cos(\theta)$

Inserting the different bending angles derived we find

(der2 /. #1 &) /@ sol1

$$\left\{0, \frac{3\sqrt{3}\ a^2}{2}, \frac{1}{2}(-3)\sqrt{3}\ a^2, 0\right\}$$

The result shows that the angle $\theta = \pi/3$ will give the maximal cross section of the gutter. This corresponds to an angle in degree of $\theta = 60\,°$.▲

Example 4.16. Conical Cup Design

A cone-shaped drinking cup is made from a circular piece of metal of radius R by cutting out a sector and joining the edges CA and CB. Find the maximum capacity of such a cup.

Solution 4.16. The geometric cut of the metal sheet is shown in Figure 4.28

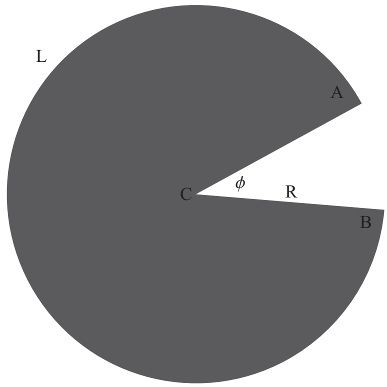


Figure 4.28. Metal sheet for a conical cup wit radius R and opening angle ϕ.

The total length of the outer rim of the sheet is given by

$$\boldsymbol{L = 2\,\pi\,R - R\,\phi}$$

$$2\,\pi\,R - R\,\phi$$

where ϕ is the opening angle of the removed segment. The length of the outer line is equal to the upper rim of the cone see Figure 4.30. This relation is given in the following equation.

$$\mathbf{eq1 = 2\,\pi\,r == L}$$

$$2\,\pi\,r = 2\,\pi\,R - R\,\phi$$

The solution with respect to r gives the opening radius of the cone

sol0 = Flatten[Solve[eq1, *r*]]

$$\left\{r \to \frac{2\,\pi\,R - R\,\phi}{2\,\pi}\right\}$$

In addition this opening radius is related to the radius R of the original disk segment. If we insert the expressions for r we gain a relation connecting the opening angle θ of the cone with the cutting angle of the metal sheet.

$$\mathbf{eq2 = r == R\sin(\theta)\ /.\ sol0}$$

$$\frac{2\,\pi\,R - R\,\phi}{2\,\pi} = R\sin(\theta)$$

The solution with respect to θ delivers

sol01 = Flatten[Solve[eq2, θ]]

$$\left\{\theta \to \sin^{-1}\left(\frac{2\,\pi - \phi}{2\,\pi}\right)\right\}$$

The height of the cone is related to the disk radius R and the opening angle of the cone by

sol02 = Flatten[Solve[*h* == *R* cos(*θ*), *h*]]

$\{h \to R\cos(\theta)\}$

To determine the volume of the cone we can use the relation of the radius of the cone at height z above the origin given by

$u(z) = z\tan(\theta)$

$z\tan(\theta)$

The volume is evaluated by the integral. This formula will be discussed in more detail in a later section. The important result is the expression for the volume V itself.

$$V = \text{Simplify}\left[\pi \int_0^h u(z)^2\, dz \;/.\; \text{sol02} \;/.\; \text{sol01}\right]$$

$$\frac{R^3 \sqrt{(4\pi - \phi)\phi}\,(\phi - 2\pi)^2}{24\pi^2}$$

This formula is equivalent to the well known formula for a cone given by

$$\text{V1} = \text{Simplify}\left[\frac{1}{3}\pi h r^2 \;/.\; \text{sol02} \;/.\; \text{sol0} \;/.\; \text{sol01}\right]$$

$$\frac{R^3 \sqrt{(4\pi - \phi)\phi}\,(\phi - 2\pi)^2}{24\pi^2}$$

where we inserted the relations for the radii r and R and the angle θ. If we plot the volume as a function of ϕ, we observe that there are several critical numbers for ϕ see Figure 4.29.

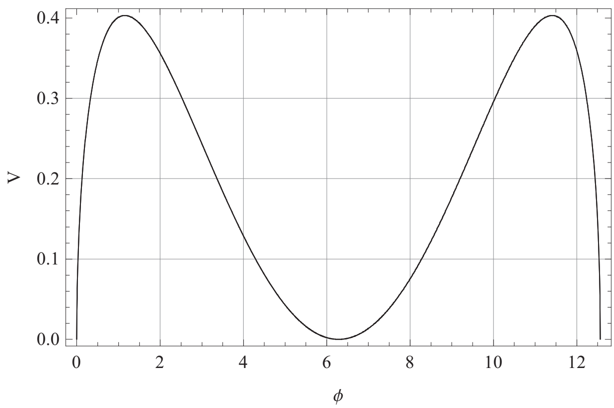


Figure 4.29. Behavior of the volume with variation of ϕ. The critical ϕ values greater than 2π are unimportant for practical applications.The volume V is plotted for $R = 1$.

To find the critical numbers related to the maxima and minima of the volume we have to calculate the first order derivative of V with respect to ϕ.

$$\textbf{Simplify}\left[\frac{\partial V}{\partial \phi}\right]$$

$$\frac{R^3\left(-3\,\phi^3+18\,\pi\,\phi^2-28\,\pi^2\,\phi+8\,\pi^3\right)}{24\,\pi^2\,\sqrt{(4\,\pi-\phi)\,\phi}}$$

The solution of this expression set equal to zero delivers 3 values for the critical numbers.

$$\textbf{sol1} = \textbf{Solve}\left[\frac{\partial V}{\partial \phi} = 0, \phi\right]$$

$$\left\{\{\phi \to 2\,\pi\}, \left\{\phi \to \frac{2}{3}\left(3-\sqrt{6}\right)\pi\right\}, \left\{\phi \to \frac{2}{3}\left(3+\sqrt{6}\right)\pi\right\}\right\}$$

The volumes related to this critical numbers follow by inserting the ϕ values into the volume

$$\textbf{Simplify}[V \textbf{ /. sol1}]$$

$$\left\{0, \frac{2\,\pi\,R^3}{9\,\sqrt{3}}, \frac{2\,\pi\,R^3}{9\,\sqrt{3}}\right\}$$

It is obvious that the first value will result to a volume which is unimportant. The second and third value generate a finite value for the volume. However, the third value for ϕ is greater than 2π and thus is not relevant for practical applications. The same result is gained by the second derivative test by inserting the values for ϕ into the second derivative of V with respect to ϕ

$$\textbf{Simplify}\left[\frac{\partial^2 V}{\partial \phi\, \partial \phi} \textbf{ /. sol1}\right]$$

$$\left\{\frac{R^3}{6\,\pi}, -\frac{R^3}{\sqrt{3}\,\pi}, -\frac{R^3}{\sqrt{3}\,\pi}\right\}$$

The result shows that the second solution for ϕ will deliver a maximum of the volume. The related angles in degree are given in the following list

$$N\left[\frac{180\,\phi}{\pi} \textbf{ /. sol1}\right]$$

{360., 66.0612, 653.939}

To estimate the opening angle of the cone, we can insert the values for ϕ into the relation with θ. This shows us again, that only the second angle is of practical use in the design

$$\textbf{sol2} = N[\textbf{sol01 /. sol1}]$$

$$\{\{\theta \to 0.\}, \{\theta \to 0.955317\}, \{\theta \to -0.955317\}\}$$

The opening angles in degree are given as

$$N\left[\frac{180\,\theta}{\pi} \textbf{ /. sol2}\right]$$

{0., 54.7356, −54.7356}

The opening radius of the cone is generated in a similar way by inserting the ϕ-values into the relation for r

sol0 /. sol1

$$\left\{\{r \to 0\}, \left\{r \to \frac{2\,\pi\,R - \frac{2}{3}\left(3 - \sqrt{6}\right)\pi\,R}{2\,\pi}\right\}, \left\{r \to \frac{2\,\pi\,R - \frac{2}{3}\left(3 + \sqrt{6}\right)\pi\,R}{2\,\pi}\right\}\right\}$$

The height of the cone follows in the same way by

sol02 /. sol01 /. sol1

$$\left\{\{h \to R\}, \left\{h \to \sqrt{1 - \frac{\left(2\,\pi - \frac{2}{3}\left(3 - \sqrt{6}\right)\pi\right)^2}{4\,\pi^2}}\;R\right\}, \left\{h \to \sqrt{1 - \frac{\left(2\,\pi - \frac{2}{3}\left(3 + \sqrt{6}\right)\pi\right)^2}{4\,\pi^2}}\;R\right\}\right\}$$

Using the second solution this optimal cone can be graphically given as shown in Figure 4.30

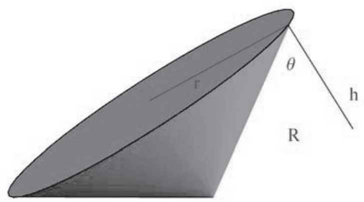


Figure 4.30. Cone resulting from a metal sheet of radius R and opening angle ϕ. The opening cone angle θ, the cone radius r and the cone height h are functions of R and ϕ.▲

4.5.4 Newton's Method

In Section 2.9 we showed how to approximate the roots of a polynomial $p(x) = 0$ by using the Factoring Theorem. In this section we will study a technique, called Newton's Method, that is usually more efficient than the mentioned method. Newton's Method is the technique used by many commercial and scientific computer programs for finding roots.

In algebra we learn that the solution of a first-degree equation $a\,x + b = 0$ is given by the formula $x = -b / a$, and the solutions of a second-degree equation

$$a\,x^2 + b\,x + c = 0$$

are give by the quadratic formula. Closed formulas also exist for the solutions of all third- and fourth-degree equations. In the beginning of the 19th century the French mathematician Galois proofed that it is impossible to construct a similar formula for the solutions of a general fifth-degree equation or higher. This proof is based on the symmetry or group properties of the equation. Thus, for a specific fifth-degree polynomial such as

$$x^5 + 18\,x^4 - 5\,x^3 - 2\,x^2 + 12\,x + 17 = 0$$

it may be difficult or impossible to find exact values for all solutions. Similar difficulties occur for non-polynomial equations such as

$$x^3 + \tan(x) = 0.$$

For such equations the solutions are generally approximated in some way, often by the method invented by Newton.

Suppose that we are trying to find a root r of the equation $f(x) = 0$, and suppose that by some method we are able to obtain an initial rough estimate, x_1, of r, say by generating the graph of $y = f(x)$ examining the intersection with the x-axis. If now $f(x_1) = 0$ we are done then $r = x_1$ (see Figure 4.31). However, if $f(x_1) \neq 0$, then we simplify the problem, that of finding a root to a related linear equation. The linear approximation to $y = f(x)$ in $x = x_1$ is given by the tangent line to the graph of f at x_1, so it might be reasonable to expect that the intersection of the tangent line provides an improved approximation to the root r. If we call this intercept x_2, we can treat x_2 in the same way we did x_1. So if $f(x_2) = 0$, the $r = x_2$. If $f(x_2) \neq 0$, then construct the tangent line to the graph of f at x_2, and take x_3 to be the x-intercept of this tangent line. Continuing in this way, we can generate a succession of values $x_1,\ x_2,\ x_3,\ x_4,\ \ldots$ that will usually approach r. This procedure for approximating r is called Newton's Method.

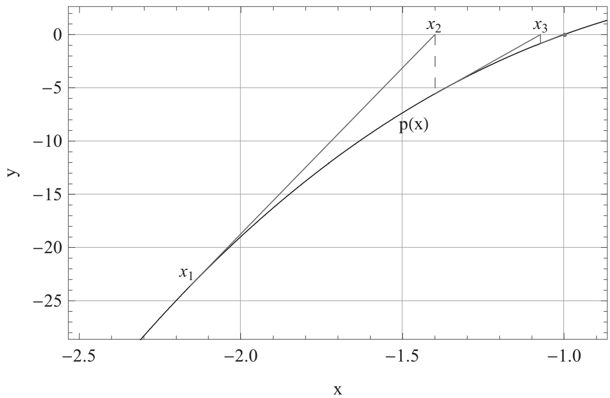


Figure 4.31. Iteration process of Newton's Method.

To implement Newton's Method analytically, we must derive a formula that will tell us how to calculate each improved approximation from the preceding approximation. For this purpose, we note that the point-slope form of the tangent line to $y = f(x)$ at the initial approximation x_1 is

$$y - f(x_1) = f'(x_1)\,(x - x_1). \tag{4.4}$$

If $f'(x_1) \neq 0$, then this line is not parallel to the x-axis and consequently it crosses the x-axis at some point $(x_2,\ 0)$. Substituting the coordinates of this point in the formula above yields

$$0 - f(x_1) = f'(x_1)(x_2 - x_1). \tag{4.5}$$

Solving for x_2 we obtain

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}. \tag{4.6}$$

The next approximation can be obtained more easily. If we view x_2 as the starting approximation and x_3 the new approximation, we can simply apply the given formula with x_2 in place of x_1 and x_3 in place of x_2. This yields

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}. \tag{4.7}$$

provided $f'(x_2) \neq 0$. In general, if x_n is the nth approximation, then it is evident from the two steps given above that the improved approximation x_{n+1} is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \text{ with } n = 1, 2, 3, \ldots. \tag{4.8}$$

This formula is realized in *Mathematica* by the following line

newtonsMethod(f_, x1_) := $x - \frac{f}{\frac{\partial f}{\partial x}}$ /. $x \to$ x1

The replacement $x \to x_1$in the function $f = f(x)$ is necessary because we are dealing with numerical values in the calculation. The iteration of this function can be carried out by a special *Mathematica* function called Nest[] or NestList[]. These function generate a nested expression of the newtonsMethod[] function and deliver the approximation of the root. For example if we are going to determine one of the roots of the polynomial

$$p(x) = x^3 - 4x^2 + 5 == 0$$

defined as

p(x_) := $x^3 - 4x^2 + 5$

whose graph is given in Figure 4.32

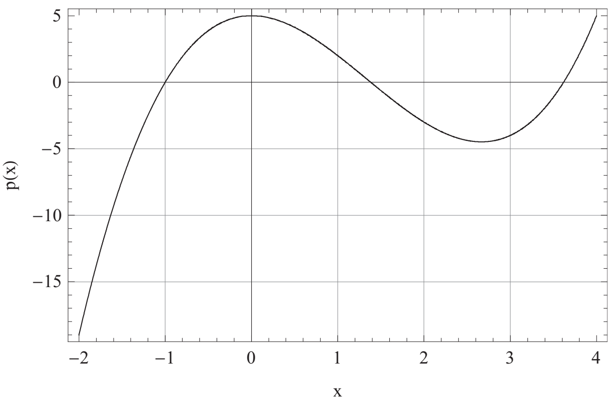


Figure 4.32. Graph of the polynomial $p(x) = x^3 - 4x^2 + 5$.

If we apply Newton's Method to this polynomial we get for an initial value $x_1 = 1.155$ the following list of 7 approximations

res = NestList[newtonsMethod($p(x)$, #1) &, 0.155, 7]

{0.155, 4.357, 3.82396, 3.64124, 3.61838, 3.61803, 3.61803, 3.61803}

The result is a list of approximations of the root starting with the initial value.

Example 4.17. Newton's Method I

Use Newton's Method to find $\sqrt[7]{2}$.

Solution 4.17. Observe that finding $\sqrt[7]{2}$ is equivalent to finding the positive root of the equation

$$x^7 - 2 = 0$$

so we take

p(x_) := $x^7 - 2$

and apply Newton's Method to this formula with an appropriate initial value $x_1 = 0.85$. The iteration of the method delivers

NestList[newtonsMethod($p(x)$, #1) &, 0.85, 17]

{0.85, 1.48613, 1.30035, 1.17368, 1.11532, 1.10442, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409, 1.10409}

This means that $\sqrt[7]{2} = 1.10409\ldots$.▲

Example 4.18. Newton's Method II

Find the solution of the equation $\cos(x) = x$.

Solution 4.18. We rewrite the equation as

$\mathbf{p(x\_) := \cos(x) - x}$

and apply Newton's Method to this expression to get

res = NestList[newtonsMethod($p(x)$, #1) &, 0.155, 7]

{0.155, 0.876609, 0.742689, 0.739088, 0.739085, 0.739085, 0.739085, 0.739085}

The symbolic expression for this iteration can be found by replacing the numerical value by a general symbol as in the next line shown

Simplify[NestList[newtonsMethod($p(x)$, #1) &, x_1, 2]]

$$\left\{x_1, \frac{x_1 \sin(x_1) + \cos(x_1)}{\sin(x_1) + 1}, x_1 + \frac{\cos(x_1) - x_1}{\sin(x_1) + 1} - \frac{x_1 \sin(x_1) + \cos(x_1) - (\sin(x_1) + 1) \cos\left(\frac{x_1 \sin(x_1)+\cos(x_1)}{\sin(x_1)+1}\right)}{(\sin(x_1) + 1)\left(\sin\left(\frac{x_1 \sin(x_1)+\cos(x_1)}{\sin(x_1)+1}\right) + 1\right)}\right\}$$

4.5.5 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

4.5.5.1 Test Problems

T1. Describe the main steps of an optimal design process.

T2. List the necessary mathematical steps for an optimal design.

4.5.5.2 Exercises

E1. What is the smallest perimeter possible for a rectangle whose area is 16 m^2 and what are its dimensions?

E2. Your iron works has contracted to design and build a 500 m^3 square-based, open-top, rectangular steel holding tank for a paper company. The tank is to be made by welding thin stainless steel plates together along their edges. As the production engineer, your job is to find dimensions for the base and height that will make the tank weigh as little as possible.

a. What dimensions do you tell the shop to use?

b. Briefly describe how you took weight into account.

E3. Find the volume of the largest right circular cone that can be inscribed in a sphere of radius 3 m.

E4. A rectangle is to be inscribed under the arch of the curve $y = 4\cos(x/2)$ from $x = -\pi$ to $x = \pi$. What are the dimensions of the rectangle with largest area, and what is the largest area?

E5. Find the dimensions of a right circular cylinder of maximum volume that can be inscribed in a sphere of radius 10 cm. What is the maximum volume?

E6. A window is in the form of a rectangle surmounted by a semicircle. The rectangle is of clear glass, whereas the semicircle is of tinted glass that transmits only half as much light per unit area as clear glass does. The total perimeter is fixed. Find the proportions of the window that will admit the most light. Neglect the thickness of the frame.

E7. Suppose that $c(x) = x^3 - 20\,x^2 + 20\,000\,x$ is the cost of manufacturing x items. Find a production level that will minimize the average cost of making x items.

E8. A right triangle whose hypotenuse is $\sqrt{2}$ m long is revolved about one of its legs to generate a right circular cone. Find the radius, height, and volume of the cone of greatest volume that can be made this way.

E9. A silo is to be constructed in the form of a cylinder surmounted by a hemisphere. The cost of construction per square unit of surface area is twice as great for the hemisphere as it is for the cylindrical sidewall and the bottom. Determine the dimensions to be used if the volume is fixed and the cost of construction is to be kept to a minimum. Neglect the thickness of the silo and waste in construction.

E10 A rectangular plot of farmland will be bounded on one side by a river and on the other three sides by a singlestrand electric fence. With 800 m of wire at your disposal, what is the largest area you can enclose, and what are its dimensions?

E11 Let $f(x) = 3x - x^3$. Show that the equation $f(x) = -4$ has a solution in the interval [2, 3] and use Newton's method to find it.

E12 Let $f(x) = x^4 - x^3$. Show that the equation $f(x) = 75$ has a solution in the interval [3, 4] and use Newton's method to find it.

5

Integration

5.1 Introduction

Traditionally, that part of calculus concerned with finding tangent lines and rate of change is called differential calculus and that part with finding integrals is called integral calculus. However, we will see in this chapter that the two branches are closely related that the distinction between differential and integral calculus is often hard to discern.

In this chapter we will begin with an overview of the problem of finding areas—we will discuss what the term *area* means, and we will outline two approaches to defining and calculating areas. Following this overview, we will discuss the Fundamental Theorem of Calculus, which is the theorem that relates the problem of finding tangent lines and areas, and we will discuss techniques for calculating areas.

5.1.1 Finding Areas

The main goal of this chapter is to study the following major problem of calculus:

- *Given a function f that is continuous and non negative on an interval* $[a, b]$, *find the area between the graph of f and the interval* $[a, b]$ *on the x-axis.*

In this introductory section we will consider the problem of calculating areas of plane regions with curvilinear boundaries. All of the results in this section will be reexamined in more detail in later sections, so our purpose here is simply to introduce the fundamental concepts, compare Figure 5.1.

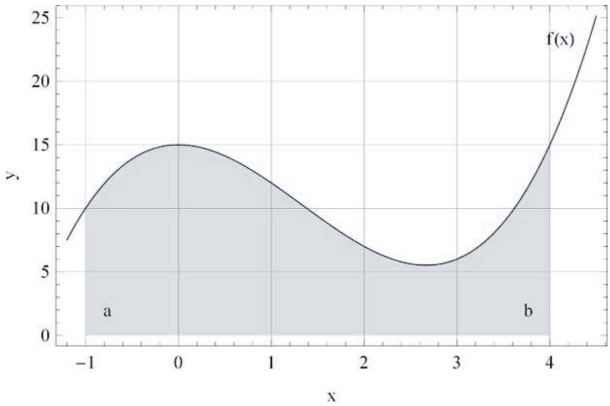


Figure 5.1. Area under a graph $y = f(x)$ based on the interval $[a, b]$.

From a strictly logical point of view, we should first provide a precise definition of the term *area* before discussing methods for calculating areas. However, in this section we will treat the concept of area intuitively, postponing a more formal definition to later sections.

Formulas for the areas of simple plane geometric objects like squares, rectangles, triangles, etc. were well known in many early civilizations. On the other hand, obtaining formulas for regions with curvilinear boundaries caused problems for early mathematicians. The first real progress on such problems was made by the Greek mathematician Archimedes who obtained the area of region bounded by arcs of circles, parabolas, spirals, and various other curves by ingenious use of a procedure later known as the *method of exhaustion*.

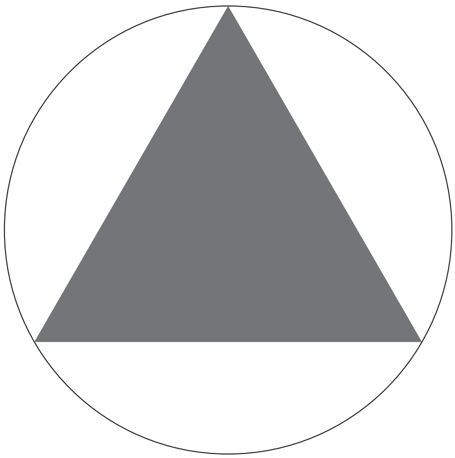


Figure 5.2. Change of the number of corners used to approximate the are of a circle starting with $n = 3$.

That method, when applied to a circle of radius r, consists of inscribing a succession of regular polygons in the circle and allowing the number of sides n to increase indefinitely (see Figure 5.3). As n increases the polygons tend to exhaust the region inside the circle, and the areas of those polygons become better and better approximations to the exact area of the circle.

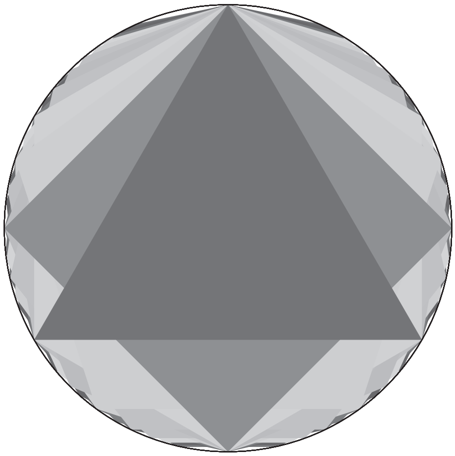


Figure 5.3. Different orders of approximating the are of a circle by polygons.

To see how this works numerically, let $A(n)$ denote the are of a regular n-sided polygon inscribed in a circle of radius 1. The following table shows the values of $A(n) = n\cos\left(\frac{\pi}{n}\right)\sin\left(\frac{\pi}{n}\right)$ for large values of n.

n	A(n)
100.	3.13953
200.	3.14108
300.	3.14136
400.	3.14146
500.	3.14151
600.	3.14154
700.	3.14155
800.	3.14156
900.	3.14157
1000.	3.14157

Note that for large values of n the area $A(n)$ appears to be close to π, as one would expect. This suggests that for a circle of radius 1, the method of exhaustion is equivalent to an equation of the form

$$\lim_{n\to\infty} A(n) = \pi. \tag{5.1}$$

However, Greek mathematicians were very suspicious of the concept of infinity and intentionally avoided explanations that referred to the limiting behavior of some quantity. As a consequence, obtaining exact answers by the classical method of exhaustion was a cumbersome procedure. In our discussion of the area problem we will consider a more modern version of the method of exhaustion that explicitly incorporates the notion of a limit. Because our approach uses a collection of rectangles to exhaust an area, we will refer to it as the rectangle method.

5.1.2 Rectangle Method

There are two basic methods for finding the area of the region having the form shown in Figure 5.4 — the rectangle method and the anti-derivative method. The idea behind the rectangle method is as follows:

Divide the interval $[a, b]$ into n equal sub intervals, and over each sub interval construct a rectangle that extends from the x-axis to any point on the curve $y = f(x)$ that is above the sub interval; the particular point does matter—it can be above the center, above the endpoint, or above any other point in the sub interval.

For each n, the total area of the rectangles can be viewed as an approximation to the exact area under the curve over the interval $[a, b]$. Moreover, it is evident intuitively that as n increases these approximations will get better and better and will approach the exact area as a limit.

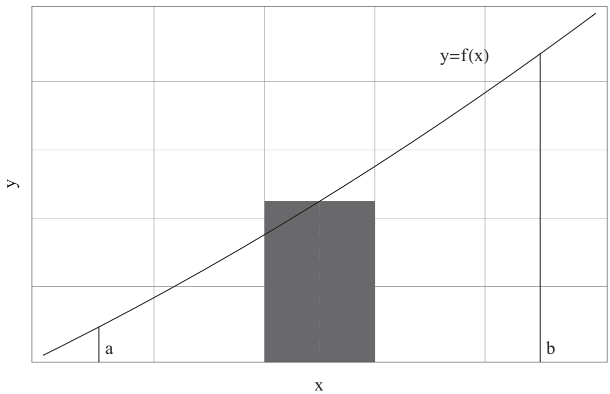


Figure 5.4. Approximation of the area below a function $f(x)$ by a rectangle.

Later, this procedure will serve both as a mathematical definition and a method of computation— we will define the area under $y = f(x)$ over the interval $[a, b]$ as the limit of the area of the approximating rectangles, and we will use the method itself to approximate this area (see Figure 5.5).

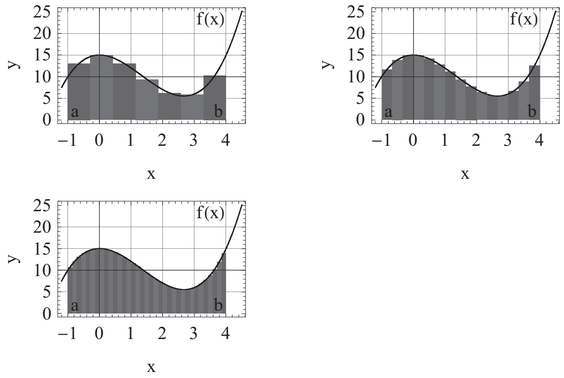


Figure 5.5. Approximation of the area under a graph $y = f(x)$ based on the interval $[a, b]$ by successively increasing the number of rectangles.

To illustrate this idea, we will use the rectangle method to approximate the area under the curve $y = x^2$ over the interval $[0, 1]$. We begin by dividing the interval $[0, 1]$ into n equal sub intervals, from which it follows that each sub interval has length $1/n$; the endpoints of the sub interval occur at

$$0, \ \frac{1}{n}, \ \frac{2}{n}, \ \frac{3}{n}, \ \ldots, \ \frac{n-1}{n}, \ 1. \tag{5.2}$$

We want to construct a rectangle over each of these sub intervals whose height is the value of the function $f(x) = x^2$ at some number in the sub interval. To be specific, let us use the right endpoint, in which case the heights of our rectangles will be

$$\left(\frac{1}{n}\right)^2, \ \left(\frac{2}{n}\right)^2, \ \left(\frac{3}{n}\right)^2, \ \ldots, \ 1. \tag{5.3}$$

and since each rectangle has a base of with $1/n$, the total area A_n of the n rectangles will be

$$A_n = \left(\left(\frac{1}{n}\right)^2 + \left(\frac{2}{n}\right)^2 + \left(\frac{3}{n}\right)^2 + \ldots + 1^2\right)\left(\frac{1}{n}\right). \tag{5.4}$$

In general using sigma notation we write

$$A_n = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 \tag{5.5}$$

represented in *Mathematica* notation is

$$\boldsymbol{A(\mathbf{n\_}) := \frac{\sum_{k=1}^{n} \left(\frac{k}{n}\right)^2}{n}}$$

This notation tells us to form the sum of the terms that result when we substitute successive integers for k in the expression $(k/n)^2$, starting with $k = 1$ and ending with $k = n$. Each value of a positive integer n then determines a value of the sum.

For example if $n = 4$, then the total area of the four approximating rectangles will be

$$\boldsymbol{A(4)}$$

$$\frac{15}{32}$$

The following table shows some approximations for larger n

n	A(n)
4.	0.46875
10.	0.385
100.	0.33835
1000.	0.333834
10 000.	0.333383
100 000.	0.333338

5.1.3 Anti-derivative Method

Despite the intuitive appeal of the rectangle method, the limits involved can be evaluated directly only in certain special cases. For this reason, work on the area problem remained at a rudimentary level until the later half of the seventeenth century. Two results that were to prove to be a major breakthrough in the area problem were discovered by the mathematicians Isaac Barrow, Isaac Newton, and Gottfried Leibniz in Germany. These results appeared without fanfare, as Proposition 11 of Lecture X and Proposition 19 of Lecture XI in Isaac Barrow's Lectiones geometricae. Each of the two results can be used to solve the area problem.

The solution based on the ideas by Barrow was preferred by Isaac Newton and provides us with a paradoxical effective indirect approach to the area problem. According to this line of argument, to find the area under the curve in Figure 5.1, one should first consider the seemingly harder problem of finding the area $A(x)$ between the graph f and the interval $[a, x]$, where x denotes an arbitrary number in $[a, b]$. If one can discover a formula for the area function $A(x)$, then the area under the curve from a to b can be obtained simply by substituting $x = b$ into the formula.

This may seem to be a surprising approach to the area problem. After all, why should the problem of determining the area $A(x)$ for every x in the interval $[a, b]$ be more tractable than the problem of computing a single value $A(b)$? However, the basis for this approach is the observation that although the area function $A(x)$ may be difficult to compute, its derivative $A'(x)$ is easy to find. To illustrate, let us consider some examples of area functions $A(x)$ that can be computed from simple geometry.

Example 5.1. Anti derivative

For each of the following functions f $\{f(x) = 2,\ f(x) = x + 1\}$ find the area $A(x)$ between the graph of f and the interval $[a, x] = [-1, x]$, and find the derivative $A'(x)$ of this area function.

Solution 5.1. For the first case the area is given by a rectangle of height 2 and length $x - (-1)$ so that the area function is

$$A(x) = 2\,(x - (-1)) = 2\,(x + 1) = 2\,x + 2.$$

For this area function the derivative is

$$A'(x) = 2 = f(x).$$

For the second case where the height is given by $x + 1$ and the width is also x+1 corresponding to a isosceles right triangle the are function is

$$A(x) = \frac{1}{2}(x + 1)(x + 1) = \frac{x^2}{2} + x + \frac{1}{2}.$$

The derivative of this area function is

$$A'(x) = x + 1 = f(x).\blacktriangle$$

Note that in every case in Example 5.1 the relation

$$A'(x) = f(x) \tag{5.6}$$

holds. That is, the derivative of the area function $A(x)$ is the function whose graph forms the upper boundary of the region. To find the area function $A'(x)$, we can look for a function whose derivative is $f(x)$. this is called an anti differentiation problem because we are trying to find $A(x)$ by undoing a differentiation.

To see how this anti derivative method applies to a specific example, let us return to the problem of finding the area between the graph of $f(x) = x^2$ and the interval [0, 1]. If we let $A(x)$ denote the area between the graph f and the interval $[0, x]$, then (5.6) tells us that $A'(x) = f(x) = x^2$. By simple guesswork, we see that one function whose derivative is $f(x) = x^2$ is $x^3 / 3$. Therefore $A(x) = x^3 / 3$ and the area between the graph of f and the interval [0, 1] is $A(1) = 1/3$. Note that this conclusion agrees with our numerical estimation given above.

Although the anti derivative method provides us with a convenient solution to the area problem, it appears to have little to do with the rectangle method. It would be nice to have a solution that more clearly elucidates the connection between the operation of summing areas of rectangles on the one hand and the operation of anti differentiation on the other. Fortunately, the solution to the area problem based on Barrow's work reveals just this connection. In addition, it allows us to formulate in modern language the approach to the area problem preferred by Leibniz.

5.1.4 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

5.1.4.1 Test Problems

T1. How can you sometimes estimate quantities like distance traveled, area, and average value with finite sums? Why might you want to do so?

T2. What is the relation between integrals and area? Describe some other interpretations of definite integrals.

5.1.4.2 Exercises

E1. Using rectangles whose height is given by the value of the function at the midpoint of the rectangle's base (the midpoint rule) estimate the area under the graphs of the following functions, using first two and then four rectangles.

$f(x) = x^2$ with $x \in [0, 1]$,

b. $f(x) = x^3$ with $x \in [0, 1]$,

c. $f(x) = \sqrt{x}$ with $x \in [0, 2]$,

d. $f(x) = \frac{1}{x}$ with $x \in [1, 3]$,

e. $f(x) = 3 + x^2$ with $x \in [-3, 3]$.

E2. Inscribe a regular n-sided polygon inside a circle of radius 1 and compute the area of the polygon for the following values of $n = \{4, 8, 16\}$. Compare the areas found with the area of the circle.

E3. Inscribe a regular n-sided polygon inside a circle of radius 1 and compute the area of one of the n congruent triangles formed by drawing radii to the vertices of the polygon. Compute the limit of the area of the inscribed polygon as $n \to \infty$. Repeat the computations for a circle of radius r.

E4. For the following functions use a Computer Algebra System (CAS) to perform the following steps.

a. Plot the functions over the given interval.

b. Subdivide the interval into 200, and 1000 sub intervals of equal length and evaluate the function at the midpoint of each sub interval

c. Compute the average value of the function values generated in part b).

$f(x) = \frac{\sin(x)}{x}$ with $x \in [-2, 2]$,

$f(x) = x \sin(x)$ with $x \in [0, \pi]$, and

$f(x) = x \sin\left(\frac{1}{x}\right)$ with $x \in \left[\frac{\pi}{2}, \pi\right]$.

5.2 Integration

We see that the rectangle method and the use of anti differentiation provide us with quite different approaches to the area problem. The rectangle method is a frontal assault on the problem, whereas anti differentiation is more in the front of a sneak attack. In this section we will carefully study both approaches to the problem.

Here we will discuss integral calculus as a tool

- to compute areas
- to compute volumes of 3-dimensional (and higher-dimensional) solids
- to compute average values of variable functions
- to solve differential equations

5.2.1 Indefinite Integrals

In the last section we saw the potential for anti differentiation to play an important role in finding exact areas. In this section we will develop some fundamental results about anti differentiation that will ultimately lead us to systematic procedures for solving many anti derivative problems.

Definition 5.1. *Indefinite Integrals*

Let $f(x)$ be a function defined on an interval I. If $F(x)$ is a function also defined on I such that

$$F'(x) = f(x)$$

then $F(x)$ is called an indefinite integral (or a primitive function or an anti derivative) of $f(x)$. $F(x)$ is often denoted by

$$\int f(x)\,dx. \blacksquare$$

For example, the function $F(x) = x^3 / 3$ is an anti derivative of $f(x) = x^2$ in the interval $(-\infty, +\infty)$ because for each x in this interval

$$F'(x) = \frac{d}{\mathrm{dx}}\left(\frac{1}{3}x^3\right) = x^2 = f(x). \tag{5.7}$$

However, $F(x) = x^3 / 3$ is not the only anti derivative of f on this interval. If we add any constant C to $x^3 / 3$, then the function $G(x) = x^3 / 3 + C$ is also an anti derivative of f on $(-\infty, +\infty)$, since

$$G'(x) = \frac{d}{\mathrm{dx}}\left(\frac{1}{3}x^3 + C\right) = x^2 + 0 = f(x). \tag{5.8}$$

In general, once any single anti derivative is known, other anti derivatives can be obtained by adding constants to the known anti derivative. Thus

$$\frac{1}{3}x^3,\ \frac{1}{3}x^3 + 6,\ \frac{1}{3}x^3 + \sqrt{5},\ \frac{1}{3}x^3 + \sin(5) \tag{5.9}$$

are all anti derivatives of $f(x) = x^2$.

It is reasonable to ask if there are anti derivatives of a function f that cannot be obtained by adding some constant to a known anti derivative F. The answer is no—once a single anti derivative of f on an interval I is known, all other anti derivatives on that interval are obtainable by adding constants to the known anti derivative. The following Theorem summarizes this.

Theorem 5.1. *Anti derivatives*

If $F(x)$ is an anti derivative of $f(x)$ on an interval I then for any constant C the function $F(x) + C$ is also an anti derivative on that interval. Moreover, each anti derivative of $f(x)$ on the interval I can be expressed in the form $F(x) + C$ by choosing the constant C appropriately. $\blacksquare$

Proof 5.1. If $F(x) + C$ is an anti derivative, then $(F(x) + C)' = F'(x) + C' = f(x) + 0$. The anti derivative is defined up to addition of an arbitrary constant.

The process of finding anti derivatives is called anti differentiation or integration. Thus, if

$$\frac{d}{\mathrm{dx}} F(x) = f(x) \tag{5.10}$$

then integrating the function $f(x)$ produces an anti derivative of the form $F(x) + C$. To emphasize this process (5.10) is recast using integral notation

$$\int f(x)\,dx = F(x) + C \tag{5.11}$$

where C is understood to represent an arbitrary constant. It is important to note that (5.10) and

(5.11) are just different notations to express the same fact. For example,

$$\int x^2\, dx = \frac{1}{3}x^3 + C \text{ is equivalent to } \frac{d}{\mathrm{dx}}\left(\frac{1}{3}x^3 + C\right) = x^2. \tag{5.12}$$

Note that if we differentiate an anti derivative of $f(x)$, we obtain $f(x)$ back again. Thus,

$$\frac{d}{\mathrm{dx}}\left[\int f(x)\, dx\right] = f(x). \tag{5.13}$$

QED

The expression $\int f(x)\,\mathrm{dx}$ is called an indefinite integral. The adjective indefinite emphasizes that the result of anti differentiation is a generic function, described only up to a constant added to the result. The elongated S that appears on the left hand side of (5.11) is called an integral sign, the function $f(x)$ is called the integrand, and the constant C is called the constant of integration. This notation was introduced by Leibniz. In his early papers Leibniz used the notation *omn. (an abbreviation of the Latin word omnes)* to denote integration. Then on October 29, 1675 he wrote, "It will be useful to write $\int$ for omn, thus $\int \ell$ for omn ℓ ...". Two or three weeks later he refined the notation further and wrote $\int[]\,\mathrm{dx}$ rather than $\int$ alone. This notation is so useful and so powerful that its development by Leibniz must be regarded as a major milestone in the history of mathematics and science.

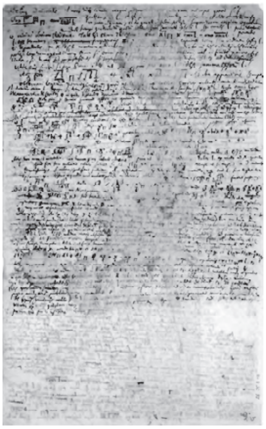


Figure 5.6. Leibniz's paper from October 29, 1675. Part of a letter in Leibniz's own hand describing some of his calculus work. By courtesy of Gottfried Wilhelm Leibniz Bibliothek. Niedersächsische Landesbibliothek.

The differential symbol, dx, in the differentiation and antiderifferentiation operations

$$\frac{d}{dx}[] \quad \text{and} \quad \int [\,]\, dx \tag{5.14}$$

serves to identify the independent variable. If an independent variable other than x is used, say t, then the notation must be adjusted appropriately. Thus,

$$\frac{d}{dt}[F(t)] = f(t) \quad \text{and} \quad \int [f(t)]\, dt = F[t] + C \tag{5.15}$$

are equivalent statements. Note that Leibniz used the same notation for operator arguments enclosed in square brackets. This notation is reused in *Mathematica* notation today.

Example 5.2. Indefinite Integrals

Given the power function x^n, $n \neq -1$. These functions have the anti derivative $\frac{1}{n+1}\ x^{n+1}$ which can be calculated using *Mathematica* by

$$\int \boldsymbol{x^n\, dx}$$

$$\frac{x^{n+1}}{n+1}$$

The exponential function e^x has an anti derivative e^x.

$$\int \boldsymbol{e^x\, dx}$$

$$e^x$$

The reciprocal function $\frac{1}{x}$ has anti derivative $\ln|x|$

$$\int \boldsymbol{\frac{1}{x}\, dx}$$

$$\ln(x)$$

The trigonometric functions have anti derivatives $\int \sin(x)\, dx = -\cos(x)$, $\int \cos(x)\, dx = \sin(x)$

$$\int \boldsymbol{\sin(x)\, dx}$$

$$-\cos(x)$$

and

$$\int \boldsymbol{\cos(x)\, dx}$$

$$\sin(x)$$

The rational $f(x) = 1 \big/ \left(1 + x^2\right)$ has the anti derivative $\int \frac{1}{1+x^2}\, dx = \arctan(x)$

$$\int \boldsymbol{\frac{1}{x^2+1}\, dx}$$

$$\tan^{-1}(x)$$

▲

For simplicity, the dx is sometimes absorbed into the integrand. For example,

$$\int 1\, dx \qquad \text{can be written as} \qquad \int dx \tag{5.16}$$

$$\int \frac{1}{x^2}\, dx \qquad \text{can be written as} \qquad \int \frac{dx}{x^2}. \tag{5.17}$$

The integral sign and differential serve as delimiters, flanking the integrand on the left and right, respectively. In particular, we do not write $\int dx\, f(x)$ when we intend $\int f(x)\, dx$.

It is not obvious but true that every continuous function has an anti derivative. This will be discussed later.

Usually it is difficult to compute the anti derivative of a given function. "To compute" means to express the results by means of other well known functions. In many cases this is impossible. The following examples show this for special functions

$$\int \sin(x^2)\, dx, \;\; \text{or} \;\; \int e^{x^2}\, dx, \;\; \text{or} \;\; \int \frac{1}{\sqrt{1-x^4}}\, dx \tag{5.18}$$

cannot be expressed by elementary functions. To demonstrate this we calculate the related integrals with the help of *Mathematica*

$$\textbf{fresnelS} = \int \sin(x^2)\, dx$$

$$\sqrt{\frac{\pi}{2}}\, S\left(\sqrt{\frac{2}{\pi}}\, x\right)$$

In this case the Fresnel integral S results which cannot be represented by elementary functions. However, the graph of Fresnel integrals S are shown in Figure 5.7

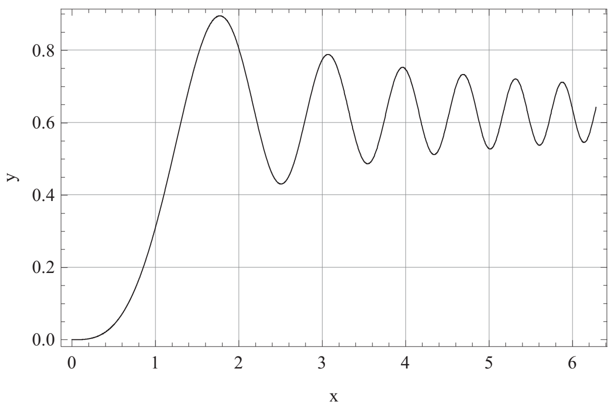


Figure 5.7. The Fresnel integral S as a function of x.

For the second integral we find

$$\textbf{errE} = \int e^{x^2}\, dx$$

$$\frac{1}{2}\sqrt{\pi}\ \text{erfi}(x)$$

including the error function which also is not an elementary function. Its graphical representation shows a behavior like exponential functions.

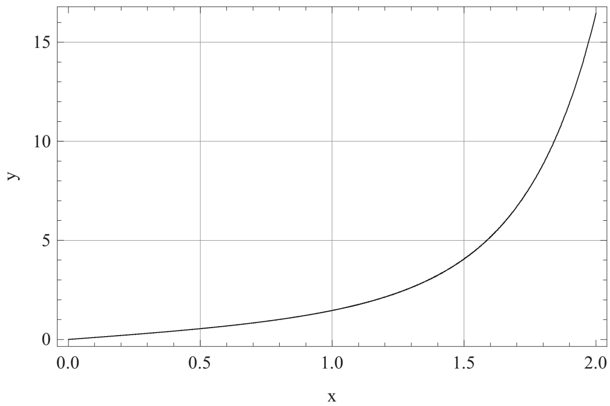


Figure 5.8. The error function.

The last integral results into a Jacobi elliptic function *F*

$$\textbf{jacbi} = \int \frac{1}{\sqrt{1 - x^4}}\, dx$$

$$F\left(\sin^{-1}(x)\,\middle|\,-1\right)$$

The graph of the Jacobi elliptic function is shown in the following figure.

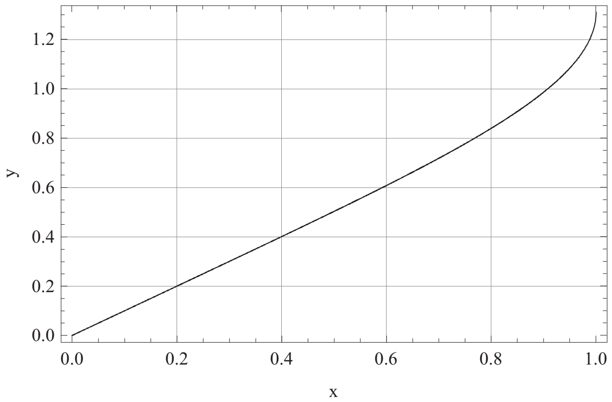


Figure 5.9. The Jacobi elliptic function *F*.

Remark 5.1. By definition the following relation is always true $\int f'(x)\,dx = f(x)$.

5.2.2 Computation Methods for Integrals

To compute the indefinite integral is the inverse of the computation of the derivative ("anti-derivative"). Therefore, all rules for computing derivatives imply corresponding rules for integration.

5.2.2.1 Properties of the Indefinite Integral

Our first properties of anti derivatives follow directly from the simple constant factor, sum, and difference rules for derivatives.

Theorem 5.2. *Algebraic Properties of Indefinite Integrals*

Suppose that $F(x)$ and $G(x)$ are anti derivatives of $f(x)$ and $g(x)$, respectively, and that c is a constant. Then:

a) A constant factor can be moved through an integral sign; that is

$$\int c\, f(x)\,dx = c\int f(x)\,dx = c\,F(x) + C$$

b) An anti derivative of a sum is the sum of the anti derivatives; that is,

$$\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx = F(x) + G(x) + C.$$

c) An anti derivative of a difference is the difference of the anti derivatives; that is,

$$\int [f(x) - g(x)]\,dx = \int f(x)\,dx - \int g(x)\,dx = F(x) - G(x) + C.\blacksquare$$

Proof 5.2. In general, to establish the validity of an equation of the form

$$\int h(x)\,\mathrm{dx} = H(x) + C$$

one must show that

$$\frac{d}{\mathrm{dx}} H(x) = h(x).$$

We are given that $F(x)$ and $G(x)$ are anti derivatives of $f(x)$ and $g(x)$, respectively, so we know that

$$\frac{d}{\mathrm{dx}} F(x) = f(x) \qquad \text{and} \qquad \frac{d}{\mathrm{dx}} G(x) = g(x).$$

Thus

$$\frac{d}{\mathrm{dx}}[c\,F(x)] = c\,\frac{d}{\mathrm{dx}}[F(x)] = c\,f(x)$$

$$\frac{d}{\mathrm{dx}}[F(x) + G(x)] = \frac{d}{\mathrm{dx}}[F(x)] + \frac{d}{\mathrm{dx}}[G(x)] = f(x) + g(x)$$

$$\frac{d}{\mathrm{dx}}[F(x) - G(x)] = \frac{d}{\mathrm{dx}}[F(x)] - \frac{d}{\mathrm{dx}}[G(x)] = f(x) - g(x)$$

which proves the three statements of the theorem.

Another way to prove these statements is the use of the algebraic properties of derivatives.

The first one corresponds to Linearity:

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

hence

$$\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$$

The second relation is related to linearity and how to handle constants in these expressions:

$$[c\,f(x)]' = c\,f'(x) \text{ where } c \in \mathbb{R} \text{ is } a \text{ constant.}$$

Hence

$$\int c\,f(x)\,\mathrm{dx} = c\int f(x)\,\mathrm{dx}\,.$$

QED

The conclusion of these relations together with the formula for monomials

$$\int x^n\,\mathrm{dx} = \frac{1}{n+1}\,x^{n+1} \tag{5.19}$$

allow us to integrate arbitrary polynomials represented by

$$p(x) = a_0 + a_1\,x + \ldots + a_n\,x^n. \tag{5.20}$$

Other basic formulas relating derivatives and anti derivative are collected in the following table.

Table 5.1. Basic differentials and anti derivatives

Differentiation Formula	*Integration Formula*
$\frac{d}{dx}[x] = 1$	$\int dx = x + C$
$\frac{d}{dx}\left[\frac{x^{r+1}}{r+1}\right] = x^r$	$\int x^r\, dx = \frac{x^{r+1}}{r+1} + C$
$\frac{d}{dx}[\sin(x)] = \cos(x)$	$\int \cos(x)\, dx = \sin(x) + C$
$\frac{d}{dx}[-\cos(x)] = \sin(x)$	$\int \sin(x)\, dx = -\cos(x) + C$
$\frac{d}{dx}[\tan(x)] = \sec(x)^2$	$\int \sec(x)^2\, dx = \tan(x) + C$
$\frac{d}{dx}[e^x] = e^x$	$\int e^x\, dx = e^x + C$
$\frac{d}{dx}\left[\frac{b^x}{\ln(b)}\right] = b^x$	$\int b^x\, dx = \frac{b^x}{\ln(b)} + C$
$\frac{d}{dx}[\ln(x)] = \frac{1}{x}$	$\int \frac{1}{x}\, dx = \ln(x) + C$
$\frac{d}{dx}[\arctan(x)] = \frac{1}{1+x^2}$	$\int \frac{1}{1+x^2}\, dx = \arctan(x) + C$
$\frac{d}{dx}[\arcsin(x)] = \frac{1}{\sqrt{1-x^2}}$	$\int \frac{1}{\sqrt{1-x^2}}\, dx = \arcsin(x) + C$

Differentiation and integration of basic functions.▲

5.2.2.2 Products of Functions

In this section we will discuss an integration technique that is essentially an anti derivative formulation of the formula for differentiating a product of two functions.

In this section we will examine a method of integration that is based on the product rule for differentiation. To motivate the general formula, we will consider the problem of evaluating

$$\int x\cos(x)\, dx. \tag{5.21}$$

Our approach to this problem will be by means of a two-step process. The first step is to choose a function whose derivative is the sum of two functions, one of which is $f(x) = x\cos(x)$. For example, the function $g(x) = x\sin(x)$ has this property, since by the product rule

$$\frac{\partial(x\sin(x))}{\partial x}$$

$$\sin(x) + x\cos(x)$$

The second step in evaluating $\int x\cos(x)\, dx$ is to subtract from our chosen function an anti derivative for the extra function that is produced by the product rule. What results will then be an anti derivative for $x\cos(x)$. For example, from the anti derivative of $\sin(x)$, we conclude that

$$\mathbf{x\,\sin(x) - (-\cos(x))}$$

$$x\sin(x) + \cos(x)$$

is an anti derivative of $x\cos(x)$. Indeed, this conclusion is easily verified since

$$\frac{\partial(\mathbf{x\,\sin(x) + \cos(x)})}{\partial \mathbf{x}}$$

$$x\cos(x)$$

It follows that

$$\int x\cos(x)\,\mathrm{dx} = x\sin(x) + \cos(x) + C. \tag{5.22}$$

This two-step process is an illustration of a method of integration known as integration by parts. More generally, suppose that we wish to evaluate an integral of the form $\int f(x)\,g(x)\,\mathrm{dx}$. If $G(x)$ is an anti derivative of $g(x)$, then the product rule for derivatives, the function

$$h(x) = f(x)\,G(x) \tag{5.23}$$

of the two functions satisfies the equation

$$h'(x) = \frac{d}{\mathrm{dx}}[f(x)\,G(x)] = f'(x)\,G(x) + f(x)\,g(x). \tag{5.24}$$

Consequently, if we subtract an anti derivative for $f'(x)\,G(x)$ from the function $f(x)\,G(x)$, the result will be an anti derivative for $f(x)\,g(x)$. We may express this conclusion symbolically by writing

$$\int f(x)\,g(x)\,\mathrm{dx} = f(x)\,G(x) - \int f'(x)\,G(x)\,\mathrm{dx} \tag{5.25}$$

which is one version of the integration by parts formula. By using this formula, we can sometimes reduce a difficult integration problem to an easier one.

This formula is known as *partial integration* or *integration by parts*. The point here is, that it may be easier to compute $\int f(x)\,g'(x)\,dx$ than to compute $\int f'(x)\,g(x)\,dx$.

In practice, it is usual to write (5.25) by letting

$$u = f(x), \qquad \mathrm{du} = f'(x)\,\mathrm{dx} \tag{5.26}$$

$$v = G(x), \qquad \mathrm{dv} = G'(x) = g(x)\,\mathrm{dx}. \tag{5.27}$$

This yields the following alternative form for (5.25)

$$\int u\,\mathrm{dv} = u\,v - \int v\,\mathrm{du}. \tag{5.28}$$

To illustrate the use of formula (5.28) we will reevaluate the example with $\int x\cos(x)\,\mathrm{dx}$. The first step is to make a choice of u and dv. We will let $u = x$ and $\mathrm{dv} = \cos(x)\,\mathrm{dx}$ from which it follows that $\mathrm{du} = \mathrm{dx}$ and $v = \sin(x)$. Then from formula (5.28) we get

$$\int x\cos(x)\,\mathrm{dx} = \\ x\sin(x) - \int \sin(x)\,\mathrm{dx} \;=\; x\sin(x) - (-\cos(x)) + C \;=\; x\sin(x) + \cos(x) + C. \tag{5.29}$$

Remark 5.2. In the calculation of $v = \sin(x)$ from $\mathrm{dv} = \cos(x)\,\mathrm{dx}$, we omitted a constant of integration. Had we included a constant of integration and written $v = \sin(x) + C_1$, the constant C_1 would have eventually canceled out. This is always the case in integration by parts, and it is common to omit consideration of a constant of integration when going from dv to v. However, in certain cases a clever choice of a constant of integration can simplify the computation of $\int v\,\mathrm{du}$.

To use integration by parts successfully, the choice of u and dv must be made so that the new integral is easier than the original. For example, if we decided above to let

$$u = \cos(x), \qquad \mathrm{dv} = x\,\mathrm{dx}, \qquad \mathrm{du} = -\sin(x)\,\mathrm{dx}, \qquad v = \frac{x^2}{2}$$

then we would have obtained

$$\int x\cos(x)\,\mathrm{dx} = \frac{x^2}{2}\cos(x) - \int \frac{x^2}{2}(-\sin(x))\,\mathrm{dx} = \frac{x^2}{2}\cos(x) + \frac{1}{2}\int x^2\sin(x)\,\mathrm{dx}.$$

For this choice of u and dv, the new integral is actually more complicated than the original. In general there are no hard and fast rules for choosing u and dv; it is mainly a matter of experience that comes from lots of practice.

Example 5.3. Integration by Parts I

Evaluate $\int x\,e^x\,\mathrm{dx}$.

Solution 5.3. In this case the integrand is the product of the algebraic function x with the exponential function e^x. According the integration by parts rule we set

$$u = x \qquad \text{and} \qquad \mathrm{dv} = e^x\,\mathrm{dx}$$

so that

$$\mathrm{du} = \mathrm{dx} \qquad \text{and} \qquad v = e^x.$$

Thus, from (5.28) we gain

$$\int x\,e^x\,\mathrm{dx} = \int u\,\mathrm{dv} = u\,v - \int v\,\mathrm{du} = x\,e^x - \int e^x\,\mathrm{dx} = x\,e^x - e^x + C.\blacktriangle$$

Example 5.4. Integration by Parts II

Evaluate $\int x\sin(x)\,\mathrm{dx}$.

Solution 5.4. Put $u = x$, $\mathrm{dv} = \sin(x)\,\mathrm{dx}$; i.e. $v = -\cos(x)$

Then integration by parts yields

$$\int u\,\mathrm{dv} =$$

$$u\,v - \int v\,\mathrm{du} = \int x\sin(x)\,\mathrm{dx} = -x\cos(x) - \int -\cos(x)\,\mathrm{dx} = -x\cos(x) + \sin(x) + C.$$

(Check this by computing the derivative of the right hand side!) The integration by using a CAS is

$$\int \boldsymbol{x\sin(x)\,dx}$$

$\sin(x) - x\cos(x)$

which agrees with the paper and pencil result.▲

Example 5.5. Integration by Parts III

Evaluate $\int \ln(x)\,\mathrm{dx}$ for $x > 0$.

Solution 5.5. Put (a little trick)

$v = x$ and $\mathrm{dv} = 1\,\mathrm{dx}$ $u = \ln(x)$ and $\mathrm{du} = \frac{1}{x}\,\mathrm{dx}$ then we gain

$$\int u\,\mathrm{dv} = u\,v - \int v\,\mathrm{du} = \int \ln(x)\,\mathrm{dx} = x\ln(x) - \int x\,\frac{1}{x}\,\mathrm{dx} = x\ln(x) - x.$$

You can check this by differentiation of the final result. On the other hand *Mathematica* delivers

$$\int \boldsymbol{\ln(x)\,dx}$$

$x\ln(x) - x$

which is the same as the pedestrian result.▲

Example 5.6. Integration by Parts IV

Evaluate $\int \sin(x)^2\,\mathrm{dx}$.

Solution 5.6. Put $u' = \sin(x)$, $u = -\cos(x)$ and $v = \sin(x)$

$$\int \sin(x)\sin(x)\,\mathrm{dx} = -\cos(x)\sin(x) - \int -\cos(x)\cos(x)\,\mathrm{dx}$$

Now it seems, nothing simpler has been found. But remember $\cos^2 + \sin^2 = 1$, $\cos^2 = 1 - \sin^2$. Therefore we can continue as follows

$$\int \sin(x)^2\,dx = -\cos x \sin x + \int \left(1 - \left(\sin(x)^2\right)\right) dx$$

$$2\int \sin(x)^2\,dx = \int 1\,\mathrm{dx} - \cos(x)\,\sin(x)$$

$$\int \sin(x)^2\,dx = \frac{1}{2}(x - \cos(x)\,\sin(x))$$

You can check this by your own computation. The same result follows with *Mathematica* by

TrigExpand$\left[\int \sin^2(x)\,dx\right]$

$$\frac{x}{2} - \frac{1}{2}\sin(x)\cos(x)$$

▲

5.2.2.3 Integration by Substitution

In this section we will study a technique, called substitution, that can often be used to transform complicated integration problems into simpler ones.

The method of substitution can be motivated by examining the chain rule from the viewpoint of antiderifferentiation. For this purpose, suppose that F is an anti derivative of f and that g is an differentiable function. The chain rule implies that the derivative of $F(g(x))$ can be expressed as

$$\frac{d}{\mathrm{dx}}[F(g(x))] = F'(g(x))\,g'(x) \tag{5.30}$$

which we can write in integral form as

$$\int F'(g(x))\,g'(x)\,\mathrm{dx} = F(g(x)) + C \tag{5.31}$$

or since F is an anti derivative of f

$$\int f(g(x))\,g'(x)\,\mathrm{dx} = F(g(x)) + C \tag{5.32}$$

For our purposes it will be useful to let $u = g(x)$ and to write $\mathrm{du}/\mathrm{dx} = g'(x)$ in the differential form $\mathrm{du} = g'(x)\,\mathrm{dx}$. With this notation (5.31) can be expressed as

$$\int f(u)\,\mathrm{du} = F(u) + C \tag{5.33}$$

The process of evaluating an integral of form (5.32) by converting it into form (5.33) with the substitution

$$u = g(x) \qquad \text{and} \qquad \mathrm{du} = g'(x)\,\mathrm{dx} \tag{5.34}$$

is called the method of u-substitution. Here our emphasis is not on the interpretation of the expression $\mathrm{du} = g'(x)\,\mathrm{dx}$ as a function of dx. Instead, the differential notation serves primarily as a useful bookkeeping device for the method of u-substitution. The following example shows how the method works.

Example 5.7. Integration by Substitution

Evaluate $\int 2\,x \cos(x^2)\,dx$

Solution 5.7. If we let $u = x^2$, then $du/dx = 2\,x$, which implies that $du = 2\,x\,dx$. Thus, the given integral can be written as

$$\int 2\,x \cos(x^2)\,dx = \int \cos(u)\,du = \sin(u) + C = \sin(x^2) + C.$$

The result corresponds to the symbolic integration with *Mathematica*

$$\int \mathbf{2\,x \cos(x^2)\,d x}$$

$\sin(x^2)$

▲

It is important to realize that in the method of u-substitution you have control over the choice of u, but once you make that choice you have no control over the resulting expression for du. Thus in the last example we chose $u = x^2$ but $du = 2\,x\,dx$ was computed. Fortunately, our choice of u that was easy to evaluate. However, in general, the method of u-substitution will fail if the chosen u and the computed du cannot be used to produce an integrand in which no expressions involving x remain, or if you cannot evaluate the resulting integral. Thus for example, the substitution $u = x^2$, $du = 2\,x\,dx$ will not work for the integral

$$\int 2\,x \sin(x^4)\,dx$$

because this substitution results in the integral

$$\int \sin(u^2)\,du$$

which still cannot be evaluated in terms of familiar functions.

In general, there are no hard and fast rules for choosing u, and in some problems no choice of u will work. In such cases other methods need to be used, some of which will be discussed later. Making appropriate choices for u will come with experience, but you may find the following guidelines, combined with a mastery of the basic integrals helpful.

- **Step 1**: Look for some composition $f(g(x))$ within the integrand for which the substitution
  $$u = g(x), \qquad du = g'(x)\,dx$$
  produces an integral that expressed entirely in terms of u and du. This may or may not be possible.
- **Step 2**: If you are successful in Step 1, then try to evaluate the resulting integral in terms of u. Again, this may or may not be possible.
- **Step 3**: If you are successful in Step 2, then replace u by $g(x)$ to express your final answer in terms of x.

Easy to Recognize Substitutions

The easiest substitutions occur when the integrand is the derivative of a known function, except for a constant added to or subtracted from the independent variable.

Example 5.8. Substitutions I

Evaluate the integral $\int \sin(x+9)\,dx$

Solution 5.8. The integral can be evaluated by applying the substitution $u = x + 9$ and $du = dx$

$$\int \sin(x+9)\,dx = \int \sin(u)\,du = -\cos(u) + C = -\cos(x+9) + C. \blacktriangle$$

Another easy u-substitution occurs when the integrand is the derivative of a known function, except for a constant that multiplies or divides the independent variable. The following example illustrates two ways to evaluate such integrals.

Example 5.9. Substitutions II

Evaluate the integral $\int \cos(5\,x)\,dx$.

Solution 5.9. Using the substitution $u = 5\,x$ and $du = 5\,dx$, we find

$$\int \cos(5\,x)\,dx = \int \cos(u)\,\frac{1}{5}\,du = \frac{1}{5}\int \cos(u)\,du = \frac{1}{5}\sin(u) + C = \frac{1}{5}\sin(5\,x) + C. \blacktriangle$$

More generally, if the integrand is a composition of the form $f(a\,x+b)$, where $f(x)$ is an easy to integrate function, then the substitution $u = a\,x + b$, $du = a\,dx$ will work.

Less Apparent Substitutions

The method of substitution is relatively straightforward, provided the integrand contains an easily recognized composition $f(g(x))$ and the remainder of the integrand is a constant multiple of $g'(x)$. If this is not the case, the method may still apply but can require more computation.

Example 5.10. Substitutions III

Evaluate $\int x^2 \sqrt{x-1}\,dx$.

Solution 5.10. The composition $\sqrt{x-1}$ suggests the substitution $u = x - 1$ so that $du = dx$. From the first relation we derive

$$x^2 = (u+1)^2 = u^2 + 2\,u + 1$$

so that

$$\int x^2 \sqrt{x-1}\,dx =$$

$$\int \left(u^2 + 2\,u + 1\right) \sqrt{u}\ \mathrm{du} = \int \left(u^{5/2} + 2\,u^{3/2} + u^{1/2}\right) \mathrm{du}\ = \frac{2}{7}\,u^{7/2} + \frac{4}{5}\,u^{5/2} + \frac{2}{3}\,u^{3/2} + C$$

$$= \frac{2}{7}\,(x-1)^{7/2} + \frac{4}{5}\,(x-1)^{5/2} + \frac{2}{3}\,(x-1)^{3/2} + C.\blacktriangle$$

Example 5.11. Substitutions IV

Evaluate $\int \cos(x)^3\ \mathrm{dx}$.

Solution 5.11. The only compositions in the integrand that suggest themselves are

$$\cos(x)^3 = (\cos(x))^3 \qquad \text{and} \qquad \cos(x)^2 = (\cos(x))^2.$$

However, neither the substitution $u = \cos(x)$ nor the substitution $u = \cos(x)^2$ work (verify!). In this case, an appropriate substitution is not suggested by the composition contained in the integrand. On the other hand, note from Equation (5.31) that the derivative $g'(x)$ appears as a factor in the integrand. This suggests that we write

$$\int \cos(x)^3\ \mathrm{dx}\ = \int \cos(x)^2 \cos(x)\,\mathrm{dx}$$

and solve the equation $\mathrm{du} = \cos(x)\,\mathrm{dx}$ for $u = \sin(x)$. Since $\sin(x)^2 + \cos(x)^2 = 1$, we then have

$$\int \cos(x)^3\ \mathrm{dx} = \int \cos(x)^2 \cos(x)\,\mathrm{dx} =$$

$$\int \left(1 - \sin(x)^2\right) \cos(x)\,\mathrm{dx} = \int \left(1 - u^2\right) \mathrm{du}\ = u - \frac{1}{3}\,u^3 + C = \sin(x) - \frac{1}{3}\,\sin(x)^3 + C.\blacktriangle$$

Example 5.12. Substitutions V

Evaluate $\int \frac{\mathrm{dx}}{a^2+x^2}$ where $a \neq 0$ is a constant.

Solution 5.12. Some simple algebra and an appropriate u-substitution will allow us to use one of the standard integration formulas given in Table 5.1.

$$\int \frac{\mathrm{dx}}{a^2 + x^2} = \frac{1}{a} \int \frac{\mathrm{dx}/a}{1 + (x/a)^2} = \frac{1}{a} \int \frac{\mathrm{du}}{1 + u^2} = \frac{1}{a}\,\arctan(x) + C = \frac{1}{a}\,\arctan\left(\frac{x}{a}\right) + C.\blacktriangle$$

5.2.2.4 Integration of Rational Functions by Partial Fractions

Recall that a rational function is a ratio of two polynomials. In this section we will give a general method for integrating rational functions that is based on the idea of decomposing a rational function into a sum of simple rational functions that can be integrated by the methods in earlier sections.

In algebra one learns to combine two or more fractions into a single fraction by finding a common denominator. For example,

$$\frac{2}{x-4}+\frac{3}{x+1}=\frac{2\,(x+1)+3\,(x-4)}{(x-4)\,(x+1)}=\frac{5\,x-10}{x^2-3\,x-4}. \tag{5.35}$$

However, for purposes of integration, the left side of (5.35) is preferable to the right side since each of the terms is easy to integrate:

$$\int\frac{5\,x-10}{x^2-3\,x-4}\,\mathrm{dx} = \int\frac{2}{x-4}\,\mathrm{dx}+\int\frac{3}{x+1}\,\mathrm{dx}=2\ln|x-4|+3\ln|x+1|+C. \tag{5.36}$$

Thus, it is desirable to have some method that will enable us to obtain the left side of (5.35), staring with the right side. To illustrate how this can be done, we begin by noting that on the left side the numerator are constants and the denominators are factors of the denominators on the right side. Thus, to find the left side of (5.35), starting from the right side, we could factor the denominator of the right side and look for constants A and B such that

$$\frac{5\,x-10}{x^2-3\,x-4}=\frac{A}{x-4}+\frac{B}{x+1}. \tag{5.37}$$

One way to find the constants A and B is to multiply (5.37) through by $(x-4)\,(x+1)$ to clear fractions. This yields

$$5\,x-10=A\,(x+1)+B\,(x-4) \tag{5.38}$$

This relationship holds for all x, so it holds in particular if $x=4$ or $x=-1$. Substituting $x=4$ in (5.38) makes the second term on the right drop out and yields the equation $10=5\,A$ or $A=2$; and substituting $x=-1$ in (5.38) makes the first term on the right drop out and yields the equation $-15=5\,B$ or $B=3$. Substituting these values in (5.37) we obtain

$$\frac{5\,x-10}{x^2-3\,x-4}=\frac{2}{x-4}+\frac{3}{x+1}. \tag{5.39}$$

which agree with (5.35).

A second method for finding the constants A and B is to multiply out the right side of (5.38) and collect like powers of x to obtain

$$5\,x-10=A\,(x+1)+B\,(x-4)=(A+B)\,x+A-4\,B. \tag{5.40}$$

Since the polynomials on the two sides are identical, their corresponding coefficients must be the same. Equating the corresponding coefficients on the two sides yields the following system of equations in the unknowns A and B:

$$A+B=5 \tag{5.41}$$

$$A-4\,B=-10. \tag{5.42}$$

Solving this system yields $A=2$ and $B=3$ as before.

The terms on the right side of (5.39) are called partial fractions of the expression on the left side because they each constitute part of the expression. To find those partial fractions we first had to make a guess about their form, and then we had to find the unknown constants. Our next objective is to extend this idea to general rational functions. For this purpose, suppose that $P(x)\,/\,Q(x)$ is a proper rational function, by which we mean that the degree of the numerator is less than the degree of the denominator. There is a Theorem in advanced algebra which states that every proper rational function can be expressed as a sum

$$\frac{P(x)}{Q(x)} = F_1(x) + F_2(x) + \ldots + F_n(x) \tag{5.43}$$

where $F_1(x)$, $F_2(x)$,…, $F_n(x)$ are rational functions of the form

$$\frac{A}{(a\,x+b)^k} \quad \text{or} \quad \frac{A\,x+B}{\left(a\,x^2+b\,x+c\right)^k} \tag{5.44}$$

in which the denominators are factors of $Q(x)$. The sum is called the partial fraction decomposition of $P(x)/Q(x)$, and the terms are called partial fractions. As in our opening example, there are two parts to finding a partial fraction decomposition and finding the unknown constants.

The first step in finding the form of the partial fraction decomposition of a proper rational function $P(x)/Q(x)$ is to factor $Q(x)$ completely into linear and irreducible quadratic factors, and then collect all repeated factors so that $Q(x)$ is expressed as a product of distinct factors of the form

$$(a\,x+b)^m \quad \text{and} \quad \left(a\,x^2+b\,x+c\right)^m. \tag{5.45}$$

From these factors we can determine the form of the partial fraction decomposition using two rules that we will now discuss.

If all of the factors of $Q(x)$ are linear, then the partial fraction decomposition of $P(x)/Q(x)$ can be determined by using the following rule:

Theorem 5.3. *Linear Factor Rule*

For each factor of the form $(a\,x+b)^m$, the partial fraction decomposition contains the following sum of m partial fractions:

$$\frac{A_1}{a\,x+b} + \frac{A_2}{(a\,x+b)^2} + \ldots + \frac{A_m}{(a\,x+b)^m}$$

where A_1, A_2, ..., A_m are constants to be determined. In the case where $m = 1$, only the first term in the sum appears. ■

Example 5.13. Partial Fractions I

Evaluate $\int \frac{dx}{x^2+x-2}$.

Solution 5.13. The integrand is a proper rational function that can be written as

$$\frac{1}{x^2+x-2} = \frac{1}{(x-1)\,(x+2)}. \tag{5.46}$$

The factor $x-1$ and $x+2$ are both linear and appear to the first power, so each contributes one term to the partial fraction decomposition by the linear factor rule. Thus, the decomposition has the form

$$\frac{1}{(x-1)(x+2)} = \frac{A}{x-1} + \frac{B}{x+2} \tag{5.47}$$

where A and B are constants to be determined. Multiplying this expression through by $(x-1)(x+2)$ yields

$$1 = A(x+2) + B(x-1). \tag{5.48}$$

As discussed earlier, there are two methods for finding A and B: we can substitute values of x that are chosen to make terms on the right drop out, or we can multiply out on the right and equate corresponding coefficients on the two sides to obtain a system of equations that can be solved for A and B. We will use the first approach.

Setting $x = 1$ makes the second term in (5.48) drop out and yields $1 = 3A$ or $A = 1/3$; and setting $x = -2$ makes the second term in (5.48) drop out and yields $1 = -3B$ or $B = -1/3$. Substituting these values in (5.47) yields the partial fraction decomposition

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

The integration can now be completed as follows

$$\int \frac{dx}{x^2+x-2} = \frac{1}{3}\int \frac{dx}{x-1} - \frac{1}{3}\int \frac{dx}{x+2} = \frac{1}{3}\ln|x-1| - \frac{1}{3}\ln|x+2| + C = \frac{1}{3}\ln\left|\frac{x-1}{x+2}\right| + C. \blacktriangle$$

If the factors of $Q(x)$ are linear and none are repeated, as in the last example, then the recommended method for finding the constants in the constants in the partial fraction decomposition is to substitute appropriate values of x to make terms drop out. However, if some of the linear factors are repeated, then it will not be possible to find all of the constants in this way. In this case the recommended procedure is to find as many constants as possible by substitution and then find the rest by equating coefficients. This is illustrated in the next example.

Example 5.14. Partial Fractions II

Evaluate $\int \frac{2x+4}{x^3-2x^2}\,dx$.

Solution 5.14. The integrand can be rewritten as

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

Although x^2 is a quadratic factor; it is not irreducible since $x^2 = x\,x$. Thus, by the linear factor rule, x^2 introduces two terms (since $m = 2$) of the form

$$\frac{A}{x} + \frac{B}{x^2}$$

and the factor $x-2$ introduces one term (since $m = 1$) of the form

$$\frac{C}{x-2}$$

so the partial fraction decomposition is

$$\frac{2x+4}{x^3-2x^2}=\frac{A}{x}+\frac{B}{x^2}+\frac{C}{x-2}.$$

Multiplying by $x^2(x-2)$ yields

$$2x+4= A\,x\,(x-2)+B\,(x-2)+C\,x^2$$

Expansion of both sides will result to

$$2x+4 = x^2(A+C)+x(B-2A)-2B$$

The resulting determining system for the unknowns is gained by comparing powers of x which gives

$$4=-2B$$

$$2= B-2A$$

$$0= A+C$$

so that $A=-C$, $B=-2$, and $A=(B-2)/2=-2$, thus $C=2$. The constants A, B, and C inserted into the partial fraction will give

$$\frac{2x+4}{x^3-2x^2}=\frac{-2}{x}+\frac{-2}{x^2}+\frac{2}{x-2}$$

Thus

$$\int\frac{2x+4}{x^3-2x^2}\,\mathrm{dx}=$$

$$-2\int\frac{\mathrm{dx}}{x}-2\int\frac{\mathrm{dx}}{x^2}+2\int\frac{\mathrm{dx}}{x-2}=-2\ln\left|x\right|+\frac{2}{x}+2\ln\left|x-2\right|+C=2\ln\left|\frac{x-2}{x}\right|+\frac{2}{x}+C.\blacktriangle$$

There are some cases in which the method of partial fractions is inappropriate. For example, it would be inappropriate to use partial fractions to perform the integration

$$\int\frac{3x^2+2}{x^3+2x-8}\,\mathrm{dx} = \int\frac{u'}{u}\,\mathrm{dx} = \int\frac{\mathrm{du}}{u}=\ln|u|+C=\ln\left|x^3+2x-8\right|+C \tag{5.49}$$

since the substitution $u=x^3+2x-8$ is more direct. Similarly, the integration

$$\int\frac{2x-1}{x^2+1}\,\mathrm{dx}= \int\frac{2x}{x^2+1}\,\mathrm{dx}-\int\frac{\mathrm{dx}}{x^2+1}=\ln\left|x^2+1\right|-\arctan(x)+C \tag{5.50}$$

requires only a little algebra since the integrand is already in partial fraction form.

5.2.2.5 Trigonometric Substitution

In this section we will discuss a method for evaluating integrals containing radicals by making substitutions involving trigonometric functions. We will also show how integrals containing quadratic polynomials can sometimes be evaluated by completing the square.

To start, we will be concerned with integrals that contain expressions of the form

$$\sqrt{a^2 - x^2}\,,\ \sqrt{a^2 + x^2}\,,\ \sqrt{x^2 - a^2}$$

in which a is a positive constant. The basic idea for evaluating such integrals is to make a substitution for x that will eliminate the radical. For example, to eliminate the radical in the expression $\sqrt{a^2 - x^2}$, we can make the substitution

$$x = a\sin(\theta) \qquad \text{with } -\pi/2 \le \theta \le \pi/2$$

which yields

$$\sqrt{a^2 - x^2} = \sqrt{a^2 - a^2\sin(\theta)^2} = a\sqrt{1 - \sin(\theta)^2} = a\sqrt{\cos(\theta)^2} = a\,|\cos(\theta)| = a\cos(\theta)$$

The restriction on θ serves two purposes —it enables us to replace $|\cos(\theta)|$ by $\cos(\theta)$ to simplify the calculations, and it also ensures that the substitutions can be rewritten as $\theta = \arcsin(x/a)$, if needed.

Example 5.15. Trig Substitution

Evaluate $\int \frac{dx}{x^2\sqrt{4-x^2}}$.

Solution 5.15. To eliminate the radical, we make the substitution

$$x = 2\sin(\theta) \qquad \text{with } dx = 2\cos(\theta)\,d\theta.$$

This yields

$$\int \frac{dx}{x^2\sqrt{4-x^2}} = \int \frac{2\cos(\theta)\,d\theta}{(2\sin(\theta))^2\sqrt{4-4\sin^2(\theta)}} =$$

$$\int \frac{2\cos(\theta)\,d\theta}{(2\sin(\theta))^2\,2\cos(\theta)} = \frac{1}{4}\int \frac{d\theta}{\sin^2(\theta)} = \frac{1}{4}\int \csc^2(\theta)\,d\theta = -\frac{1}{4}\cot(\theta) + C.$$

At this point we have completed the integration; however, because the original integral was expressed in terms of x, it is desirable to express $\cot(\theta)$ in terms of x as well. This can be done using trigonometric identities, but the expression can also be obtained by writing the substitution $x = 2\sin(\theta)$ as $\sin(\theta) = x/2$ which inserted into $\cot(\theta)$ gives

$$\cot(\theta) = \frac{\cos(\theta)}{\sin(\theta)} = \frac{\sqrt{1-\sin^2(\theta)}}{\sin(\theta)} = \frac{2\sqrt{1-\left(\frac{x}{2}\right)^2}}{x} = \frac{\sqrt{4-x^2}}{x}.$$

Substituting this into the integral yields

$$\int \frac{dx}{x^2\sqrt{4-x^2}} = -\frac{1}{4}\cot(\theta) + C = -\frac{1}{4}\frac{\sqrt{4-x^2}}{x} + C.\blacktriangle$$

Integrals that involve quadratic expressions $a\,x^2 + b\,x + c$, where $a \neq 0$ and $b \neq 0$, can often be evaluated by first completing the square, then making an appropriate substitution. The following example illustrates this idea.

Example 5.16. Quadratic completing and Substitution

Evaluate $\int \frac{x}{x^2-4x+8}$ dx.

Solution 5.16. Completing the square yields

$$x^2 - 4x + 8 = \left(x^2 - 4x + 4\right) + 8 - 4 = (x-2)^2 + 4.$$

Thus, the substitution

$$u = x - 2 \text{ and du} = \text{dx}$$

yields

$$\int \frac{x}{x^2-4x+8}\,\text{dx} = \int \frac{x}{(x-2)^2+4}\,\text{dx} = \int \frac{u+2}{u^2+4}\,\text{du} = \int \frac{u}{u^2+4}\,\text{du} + 2\int \frac{\text{du}}{u^2+4} =$$

$$\frac{1}{2}\int \frac{2u}{u^2+4}\,\text{du} + 2\int \frac{\text{du}}{u^2+4} = \frac{1}{2}\ln\left(u^2+4\right) + 2\left(\frac{1}{2}\right)\arctan\left(\frac{u}{2}\right) + C =$$

$$\frac{1}{2}\ln\left((x-2)^2+4\right) + \arctan\left(\frac{x-2}{2}\right) + C.\blacktriangle$$

5.2.3 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

5.2.3.1 Test Problems

T1. What is sigma notation? What advantage does it offer? Give examples.

T2. How are indefinite integrals defined and what do we need to evaluate them?

T3. What kind of algebraic properties do indefinite integrals have? State a few of them.

T4. Which techniques do we know to evaluate indefinite integrals?

5.2.3.2 Exercises

E1. Verify by differentiation that the following formulas are correct:

a. $\int \frac{x}{\sqrt{x^2+1}}\,dx = \sqrt{x^2+1} + C,$

b. $\int \cos(x)^3\,dx = \sin(x) - \frac{1}{3}\sin(x)^3 + C,$

c. $\int x\cos(x)\,dx = x\sin(x) + \cos(x) + C,$

d. $\int \frac{x}{\sqrt{a+bx}}\,dx = \frac{2}{3b^2}(bx - 2a)\sqrt{a+bx} + C.$

E2. Evaluate the indefinite integrals in the following list by using the given substitutions to reduce the integrals to standard form.

a. $\int \sin(3x)\,dx$ using $u = 3x$,

b. $\int x\sin(2x)^2\,dx$ using $u = 2x^2$,

c. $\int \sec(2x)\tan(2x)\,dx$ using $u = 2x$,

d. $\int \left(1-\cos\left(\frac{x}{2}\right)\right)^2 \sin\left(\frac{x}{2}\right)dx$ using $u = 1 - \cos\left(\frac{x}{2}\right)$,

$\int 12\left(x^4+4x^2+1\right)^2\left(x^3+2x\right)dx$ using $u = x^4+2x^2+1$,

f. $\int \sqrt{x}\,\sin\left(x^{3/2}-1\right)^2 dx$ using $u = x^{3/2}-1$,

g. $\int \frac{1}{x^2}\cos\left(\frac{1}{x}\right)^2 dx$ using $u = \frac{1}{x}$,

h. $\int \frac{dx}{\sqrt{2x+7}}$ using $u = 2x+7$ or $u = \sqrt{2x+7}$

E3. Find the general indefinite integral for the following expressions:

a. $\int\left(x^2+\frac{1}{x^2}\right)dx$,

b. $\int\left(\sqrt{x^3}+\sqrt[3]{x^2}\right)dx$,

c. $\int x\left(x^2+3\right)^2 dx$,

d. $\int(1-x)\left(3+x^2\right)dx$,

e. $\int \sin(x)+\sinh(x)\,dx$,

f. $\int\left(x^4+x^3-6x+1\right)dx$.

E4. Evaluate the following integrals:

a. $\int 3x\sqrt{8-3x^2}\;dx$,

b. $\int \frac{\left(1+\sqrt{x}\right)^3}{\sqrt{x}}dx$,

c. $\int \frac{1}{\sqrt{x}\left(1+\sqrt{x}\right)^2}dx$,

d. $\int \sin\left(\frac{x}{2}\right)^5\cos\left(\frac{x}{2}\right)dx$,

e. $\int \frac{4x}{\sqrt{2x^2+1}}dx$,

f. $\int \sin(3x-5)\,dx$,

g. $\int x^2\left(\frac{x^3}{12}-1\right)^5 dx$.

E5. Use a graph to estimate the x-intercepts of the curve $u = x + x^2 - x^4$. Then use this information to estimate the area of the region that lies under the curve and above the x-axis.

E6. Find the general indefinite integral. Illustrate by graphing several members of the family on the same screen.

a. $\int\left(\sin(x)+\frac{1}{2}x\right)dx$,

b. $\int\left(e^{-x}-3x^2\right)dx$.

E7. Evaluate the following integrals:

a. $\int 2\sqrt{\cos(x)}\;\sin(x)\,dx$,

b. $\int(2x+1+2\cos(2x+1))\,dx$,

c. $\int\left(x-\frac{2}{x}\right)\left(x+\frac{2}{x}\right)dx$,

d. $\int \frac{(x+1)^2-1}{x^4}dx$,

e. $\int \sqrt{x}\,\sin\left(2x^{3/2}\right)dx$.

E8. Integrate the following rational functions:

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

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

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

d. $f(x) = \frac{2x^2-3}{x^4-1}$,

$f(x) = \frac{2x-3}{x^4-1}$,

E9. Use trigonometric substitutions to solve the following integrals:

a. $\int \sqrt{1-x^2}\, dx$,

b. $\int \frac{1}{\sqrt{1-x^2}}\, dx$,

c. $\int \frac{1+x^2}{\sqrt{1-x^2}}\, dx$

5.3 Definite Integrals

In this section we will introduce the concept of a "definite integral" which will link the concept of area to other important concepts such as length, volume, density, probability, and work.

As in earlier sections, let us begin by assuming that f is non negative and continuous on an interval $[a, b]$, in which case the area A under the graph of f over the interval $[a, b]$ is represented by the definite integral

$$A = \int_a^b f(x)\, dx. \tag{5.51}$$

This relation is shown in Figure 5.10.

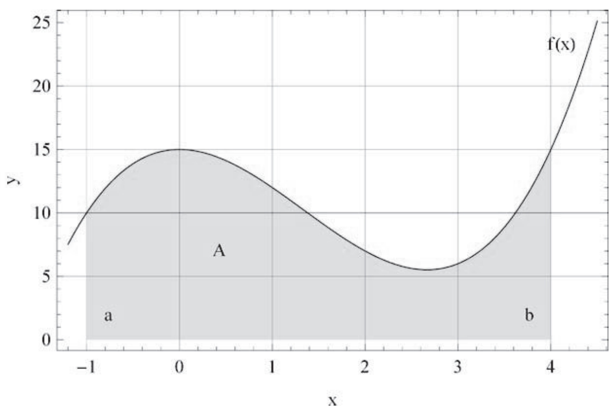


Figure 5.10. Area under a graph $y = f(x)$ based on the interval $[a, b]$.

In section 5.1.2 we introduced an approximation of the area of a curve by using a rectangular approximation. We assumed that for each positive number n, the interval $[a, b]$ was subdivided into n sub intervals of equal length to create bases for the approximating rectangles. For some functions it may be more convenient to use rectangles with different widths; however, if we are to exhaust an area with rectangles of different widths, then it is important that successive subdivisions be constructed in such a way that the widths of the rectangles approach zero as n increases. Thus, we must preclude the kind of situation that occur in which the right half of the interval is never subdivided. If this kind of subdivision were allowed, the error in the approximation would not approach zero as n increases.

A partition of the interval $[a, b]$ is a collection of numbers

$$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b \tag{5.52}$$

that divided $[a, b]$ into n sub intervals of length

$$\Delta x_1 = x_1 - x_0,\ \Delta x_2 = x_2 - x_1,\ \Delta x_3 = x_3 - x_2,\ \ldots \Delta x_n = x_n - x_{n-1}. \tag{5.53}$$

The partition is said to be regular provided the sub intervals all have the same length

$$\Delta x_k = \Delta x = \frac{b-a}{n}. \tag{5.54}$$

For a regular partition, the widths of the approximating rectangles approach zero as n is made large. Since this need not be the case for a general partition, we need some way to measure the size of these widths. One approach is to let max Δx_k denote the largest of the sub interval widths. The magnitude $\max(\Delta x_k)$ is called the mesh size of the partition. For example Figure 5.11 shows a partition of the interval $[0, 6]$ into four sub intervals with a mesh size of 2.

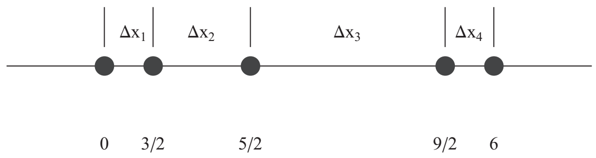


Figure 5.11. A partition of an interval [0,6] with $\max(\Delta x_k) = \Delta x_3 = \frac{9}{2} - \frac{5}{2} = 2$.

If we allow non-equal sub interval widths, we can approximate the area under a given function $f(x)$ by the following sum

$$\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k. \tag{5.55}$$

We need to replace the expression $n \to \infty$ in the approximation in such a way that guarantees that the lengths of all sub intervals approach zero. We will use the expression $\max(\Delta x_k) \to 0$ for this purpose. Based on our intuitive concept of area, we would then expect the net signed area A between the graph of f and the interval $[a, b]$ to satisfy the equation

$$A = \lim_{\max(\Delta x_k) \to 0} \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k. \tag{5.56}$$

The limit that appears in this expression is one of the fundamental concepts of integral calculus and forms the basis for the following.

Definition 5.2. *Integrability*

A function f is said to be integrable on a finite closed interval $[a, b]$ if the limit

$$\lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} f(x_k^*)\, \Delta x_k$$

exists and does not depend on the choice of partitions or on the choice of the numbers x_k^* in the sub intervals. When this is the case we denote the limit by the symbol

$$\int_a^b f(x)\, dx \;=\; \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} f(x_k^*)\, \Delta x_k$$

which is called the definite integral of f from a to b. The numbers a and b are called the lower limit of integration and the upper limit of integration, respectively, and $f(x)$ is called the integrand. ■

The notation used for the definite integral deserves some comment. Historically, the expression "$f(x)\,dx$" was interpreted by the infinitesimal area of a rectangle with height $f(x)$ and infinitesimal width dx. By summing these infinitesimal areas, the entire area under the curve was obtained. The integral symbol $\int$ is an elongated S that was used to indicate this summation. For us, the integral symbol $\int$ and the symbol dx can serve as reminders that the definite integral is actually a limit of a summation as $\Delta x_k \to 0$.

The sum that appears in Definition 5.2 is called a Riemann sum, and the definite integral is sometimes called a Riemann integral in honor of the German mathematician Bernhard Riemann who formulated many of the basic concepts of integral calculus.

Example 5.17. Definite Integral

Use Definition 5.2 to show that if $f(x) = C$ is a constant function, then

$$\int_a^b f(x)\, dx \;=\; C(b-a).$$

Solution 5.17. Since $f(x) = C$ is constant, it follows that no matter how the values of x_k^* are chosen,

$$\sum_{k=1}^{n} f(x_k^*)\, \Delta x_k = \sum_{k=1}^{n} C\, \Delta x_k = C \sum_{k=1}^{n} \Delta x_k = C(b-a).$$

The following Figure 5.12 represents the geometric situation.

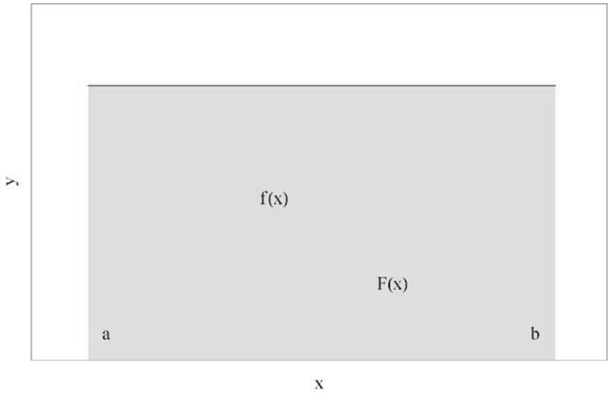


Figure 5.12. Function $f(x) = C$ on an interval $[a, b]$ and the definite integral $\int_a^b f(x)\,dx = C(b - a)$.

Since every Riemann sum has the same value $C(b-a)$, it follows that

$$\lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \lim_{\max(\Delta x_k)\to 0} C(b-a) \;=\; C(b-a).\blacktriangle$$

Note that in Definition 5.2, we do not assume that the function f is necessarily continuous on the interval $[a, b]$.

Example 5.18. Discontinuous Function

Define a function f on the interval $[0, 1]$ by $f(x) = 1$ if $0 < x \leq 1$ and $f(0) = 0$. Show that

$$\int_0^1 f(x)\,dx = 1.$$

Solution 5.18. We first note that sine

$$\lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} 1 = 1 \neq 0 = f(0)$$

f is not continuous on the interval $[0, 1]$ and any choice of the x_k^* corresponding to this partition. Then either $x_1^* = 0$ or it does not. If not, then

$$\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \sum_{k=1}^{n} \Delta x_k = 1.$$

On the other hand, if $x_1^* = 0$, then $f(x_1^*) = f(0) = 0$ and

$$\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \sum_{k=2}^{n} \Delta x_k = -\Delta x_1 + \sum_{k=1}^{n} \Delta x_k = 1 - \Delta x_1.$$

In either case we see that the difference between the Riemann sum

$$\sum_{k=1}^{n} f(x_k^*)\,\Delta x_k$$

and 1 is at most Δx_1. Since Δx_1 approaches zero as $\max(\Delta x_k) \to 0$, it follows that

$$\int_0^1 f(x)\,dx = 1. \blacktriangle$$

Although Example 5.18 shows that a function does not have to be continuous on an interval to be integrable on that interval, we will be interested primarily in the definite integrals of continuous functions. Our earlier discussion of areas suggests that a function that is continuous on an interval should also be integrable on that interval. This is the content of the next theorem, which we state without proof.

Theorem 5.4. *Continuity and Integrability*

If a function f is continuous on an interval $[a, b]$, then f is integrable on $[a, b]$. ■

We can use this theorem to clarify the connection between the definite integral and the area calculated by the limit. Suppose that f is a continuous function on the interval $[a, b]$. Recall that we defined the area A between the graph of f and the interval $[a, b]$ to be given by the limit

$$A = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\,\Delta x. \tag{5.57}$$

On the other hand it follows from Theorem 5.4 and Definition 5.2 that we can use regular partitions of $[a, b]$ to compute the definite integral of f over $[a, b]$ as the limit

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\,\Delta x. \tag{5.58}$$

Since the two limits are the same, we conclude that

$$A = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\,\Delta x = \int_a^b f(x)\,dx = \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k. \tag{5.59}$$

In other words, the definite integral of a continuous function f from a to b may always be interpreted as the net signed area between the graph of f and the interval $[a, b]$. Of course, if f is non negative, this is simply the area beneath the graph of f and above the interval $[a, b]$.

5.3.1 Properties of the Definite Integrals

It is assumed in Definition 5.2 that $[a, b]$ is a finite closed interval with $a < b$, and hence the upper limit of integration in the definite integral is greater than the lower limit of integration. However, it will be convenient to extend this definition to allow for cases in which the upper and lower limits of integration are equal or the lower limit of integration is greater than the upper limit of integration. For this purpose we make the following special definitions.

Definition 5.3. *Integrability and Limits*

If a is the domain of f, we define

$$\int_a^a f(x)\,\mathrm{dx} = 0.$$

If f is integrable on $[a, b]$*, then we define*

$$\int_a^b f(x)\,\mathrm{dx} = -\int_b^a f(x)\,\mathrm{dx}.\blacksquare$$

Remark 5.3. The first relation of this definition is consistent with the intuitive idea that the area between a point on the x-axis and a curve $y = f(x)$ should be zero. The second relation of this definition is simply a useful convention; it states that interchanging the limits of integration reverses the sign of the integral.

Example 5.19. Integrability and Limits

The following examples show the influence of the integration limits on the integral value. The first example uses equal limits which means that we are integrating on a point

$$\int_1^1 x^2\, dx$$

0

Integration on a point shows that the integral value vanishes. Another example with different limits interchanged shows

$$\int_1^0 \sqrt{1-x^2}\; dx$$

$$-\frac{\pi}{4}$$

and limits interchanged

$$\int_0^1 \sqrt{1-x^2}\; dx$$

$$\frac{\pi}{4}$$

The two integrals deliver the same value $\pi/4$ but differ in their sign because the integral limits were interchanged.▲

Because definite integrals are defined as limits, they inherit many of the properties of limits. For example, we know that constants can be moved through limit signs and that the limit of a sum or difference is the sum or difference of the limits. Thus, you should not be surprised by the following theorem, which we state without formal proof.

Theorem 5.5. *Algebraic Properties of Definite Integrals*

If f and g are integrable on $[a, b]$ and if c is a constant, then $c\,f$, $f+g$, $f-g$ are integrable on $[a, b]$ and

$$a)\ \int_a^b c\,f(x)\,\mathrm{dx} = c\int_a^b f(x)\,\mathrm{dx}$$

$$b)\ \int_a^b [f(x)+g(x)]\,\mathrm{dx} = \int_a^b f(x)\,\mathrm{dx} + \int_a^b g(x)\,\mathrm{dx}$$

$$c)\ \int_a^b [f(x)-g(x)]\,\mathrm{dx} = \int_a^b f(x)\,\mathrm{dx} - \int_a^b g(x)\,\mathrm{dx}\ \blacksquare$$

Part b) of this theorem can be expressed to more than two functions. Some properties of definite integrals can be motivated by interpreting the integral as an area. For example, if f is continuous and non negative on the interval $[a, b]$, and if c is a point between a and b, then the area under $y = f(x)$ over the interval $[a, b]$ can be split into two pats and expressed as the area under the graph from a to c plus the area under the graph from c to b, that is,

$$\int_a^b f(x)\,\mathrm{dx} = \int_a^c f(x)\,\mathrm{dx} + \int_c^b f(x)\,\mathrm{dx}. \tag{5.60}$$

This is a special case of the following theorem about definite integrals, which we state without proof.

Theorem 5.6. *Additivity of Areas*

If f is integrable on a closed interval containing the three numbers a, b, and c, then

$$\int_a^b f(x)\,\mathrm{dx} = \int_a^c f(x)\,\mathrm{dx} + \int_c^b f(x)\,\mathrm{dx}$$

no matter how the numbers are ordered. ■

The following theorem, which we state without proof, can also be motivated by interpreting definite integrals as areas.

Theorem 5.7. *Inequality Relations*

a) If f is integrable on $[a, b]$ and $f(x) \geq 0$ for all x in $[a, b]$, then

$$\int_a^b f(x)\,\mathrm{dx} \geq 0.$$

b) If f and g are integrable on $[a, b]$ and $f(x) \geq g(x)$ for all x in $[a, b]$, then

$$\int_a^b f(x)\,\mathrm{dx} \;\geq\; \int_a^b g(x)\,\mathrm{dx}.\blacksquare$$

Geometrically, part a) of this theorem states the obvious fact that if f is non-negative on $[a, b]$, then the area between the graph of f and the interval $[a, b]$ is also non negative. Part b) has its simplest integration when f and g are non negative on $[a, b]$, in which case the theorem states that if the graph of f does not go below the graph of g, then the area under the graph of f is at least as large as the area under the graph of g.

Remark 5.4. Part b) of this theorem states that one can integrate both sides of the inequality $f(x) \geq g(x)$ without altering the sense of the inequality. We also note that in the case where $b > a$, both parts of the theorem remain true if $\geq$ is replaced by $\leq$, $>$, or $<$ throughout.

Example 5.20. Algebraic Properties

Evaluate the integral

$$\int_0^1 \left(5 - 3\sqrt{1 - x^2}\right) \mathrm{dx}.$$

Solution 5.20. From parts a) and c) of Theorem 5.5 we can write

$$\int_0^1 \left(5 - 3\sqrt{1 - x^2}\right) \mathrm{dx} = \int_0^1 5\,\mathrm{dx} - \int_0^1 3\sqrt{1 - x^2}\,\mathrm{dx} = \int_0^1 5\,\mathrm{dx} - 3\int_0^1 \sqrt{1 - x^2}\,\mathrm{dx}.$$

The first integral can be integrated as the area of a rectangle of height 5 and base 1, so its value is 5, and from Example 5.19 the value of the second integral is $\pi/4$. Thus,

$$\int_0^1 \left(5 - 3\sqrt{1 - x^2}\right) \mathrm{dx} = 5 - 3\left(\frac{\pi}{4}\right) = 5 - \frac{3\pi}{4}.\blacktriangle$$

5.3.2 Discontinuities and Integrability

The problem of determining when functions with discontinuities are integrable is quite complex and beyond the scope of this text. However, there are a few basic results about integrability that are important to know. Let us give first a basic definition to fix some terminology.

Definition 5.4. *Integrability and Limits*

A function f that is defined on an interval I is said to be bounded on I if there is a positive number M such that

$$-M \;\leq\; f(x) \;\leq\; M$$

for all x in the interval I. Geometrically, this means that the graph of f over the interval I lies between the lines $y = -M$ and $y = M$. $\blacksquare$

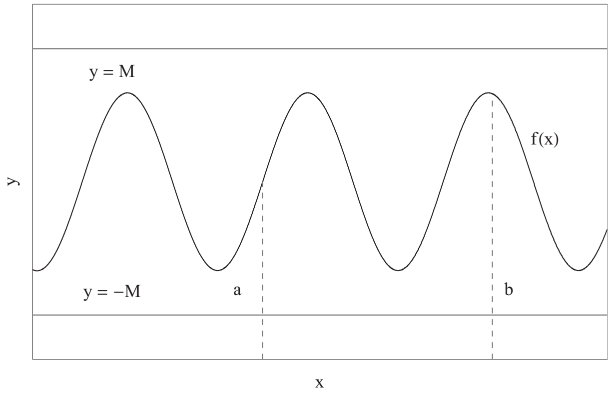


Figure 5.13. Bounded function on an interval $[a, b]$.

For example, a continuous function f is bounded on every finite closed interval because the Extreme-Value Theorem which states that if a function f is continuous on a finite closed interval $[a, b]$, then f has both an absolute maximum and an absolute minimum on $[a, b]$, which implies that f has an absolute maximum and an absolute minimum on the interval; hence, its graph will lie between the line $y = -M$ and $y = M$, provided we make M large enough. In contrast, a function that has a vertical asymptote inside of an interval is not bounded on the interval because its graph over the interval cannot be made to lie between the lines $y = -M$ and $y = M$, no matter how large we make the value of M.

The following theorem, which is stated without proof, provides some facts about integrability for functions with discontinuities.

Theorem 5.8. *Integration of Discontinuous Functions*

Let f be a function that is defined on the finite closed interval $[a, b]$.

a) If f has finitely many discontinuities in $[a, b]$ but is bounded on $[a, b]$, then f is integrable on $[a, b]$.

b) If f is not bounded on $[a, b]$, then f is not integrable on $[a, b]$. ■

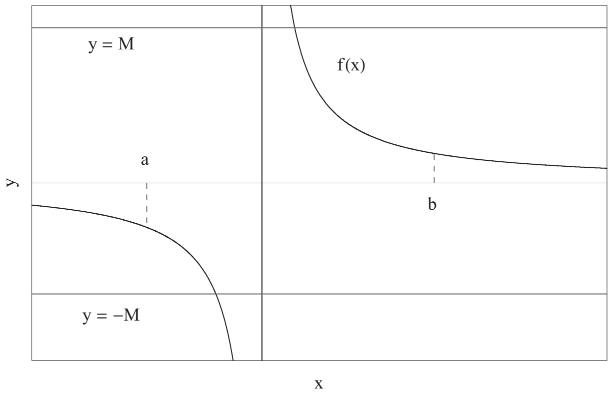


Figure 5.14. Not bounded function on an interval $[a, b]$.

5.3.3 The Fundamental Theorem of Calculus

In this section we will establish two basic relationships between definite and indefinite integrals that together constitute a result called Fundamental Theorem of Calculus. One part of this theorem will relate the rectangle and anti derivative methods for calculating areas, and the second part will provide a powerful method for evaluating definite integrals using anti derivatives.

5.3.3.1 The Fundamental Theorem of Calculus

As in an earlier section, let us begin by assuming that f is non negative and continuous on the interval $[a, b]$, in which case the area A under the graph of f over the interval $[a, b]$ is represented by the definite integral

$$A = \int_a^b f(x)\,\mathrm{dx}. \tag{5.61}$$

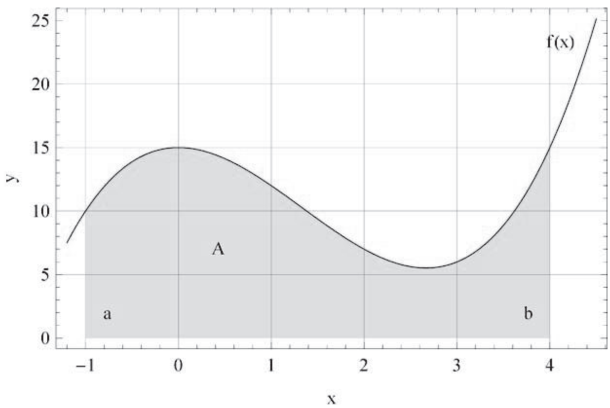


Figure 5.15. Area under a graph $y = f(x)$ based on the interval $[a, b]$.

Recall that our discussion of the anti derivative method suggested that if $A(x)$ is the area under the graph of f from a to x then

$$A'(x) = f(x) \tag{5.62}$$

$$A(a) = 0 \tag{5.63}$$

$$A(b) = A. \tag{5.64}$$

The formula $A'(x) = f(x)$ states that $A(x)$ is an anti derivative of $f(x)$, which implies that every other anti derivative of $f(x)$ on $[a, b]$ can be obtained by adding a constant to $A(x)$.

Accordingly, let

$$F(x) = A(x) + C \tag{5.65}$$

be any anti derivative of $f(x)$, and consider what happens when we subtract $F(a)$ from $F(b)$:

$$F(b) - F(a) = [A(b) + C] - [A(a) + C] = A(b) - A(a) = A - 0 = A. \tag{5.66}$$

Hence equation (5.61) can be expressed as

$$\int_a^b f(x)\,\mathrm{dx} = F(b) - F(a). \tag{5.67}$$

In words this equation states:

The definite integral can be evaluated by finding any anti derivative of the integrand and then subtracting the value of its anti derivative at the lower limit of integration from its value at the upper limit of integration.

Although our evidence for this result assumed that f is non negative on $[a, b]$, this assumption is not essential.

Theorem 5.9. *Fundamental Theorem of Calculus, Part I*

If f is continuous on $[a, b]$ and F is any anti derivative of f on $[a, b]$, then

$$\int_a^b f(x)\,\mathrm{dx} = F(b) - F(a).\blacksquare$$

Proof 5.9. Let $x_1, x_2, \ldots, x_{n-1}$ be any number in $[a, b]$ such that

$$a < x_1 < x_2 < \ldots < x_{n-1} < b.$$

The values dividing $[a, b]$ into n sub intervals

$$[a, x_1],\ [x_1, x_2],\ \ldots [x_{n-1}, b]$$

whose lengths, as usual, we denote by

$$\Delta x_1,\ \Delta x_2,\ \ldots,\ \Delta x_n.$$

By hypothesis, $F'(x) = f(x)$ for all x in $[a, b]$, so F satisfies the fact that the mean derivative in a closed interval can be expressed by the difference quotient. Hence, we can find numbers x_1^*, x_2^*, x_3^*, …, x_n^* in the respective sub interval such that

$$F(x_1) - F(a) = F'(x_1^*)(x_1 - a) = f(x_1^*)\,\Delta x_1$$

$$F(x_2) - F(x_1) = F'(x_2^*)(x_2 - x_1) = f(x_2^*)\,\Delta x_2$$

$$F(x_3) - F(x_2) = F'(x_3^*)(x_3 - x_2) = f(x_3^*)\,\Delta x_3$$

$$\vdots$$

$$F(b) - F(x_{n-1}) = F'(x_n^*)(b - x_{n-1}) = f(x_n^*)\,\Delta x_n$$

Adding the preceding equations yields

$$F(b) - F(a) = \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k. \tag{5.68}$$

Let us now increase n in such a way that $\max(\Delta x_k) \to 0$. Since f is assumed to be continuous, the right side of (5.68) approaches $\int_a^b f(x)\,dx$ by Theorem 5.4 and Definition 5.2. However, the left side of (5.68) is independent of n; that is the left hand side of (5.68) remains constant as n increases. Thus

$$F(b) - F(a) = \lim_{\max(\Delta x_k) \to 0} \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \int_a^b f(x)\,dx.$$

QED

It is standard to denote the difference $F(b) - F(a)$ as

$$F(x)\,|_a^b = F(b) - F(a) \qquad \text{or} \qquad [F(x)]_a^b = F(b) - F(a). \tag{5.69}$$

For example using the first of these notations we can write

$$\int_a^b f(x)\,dx = F(x)\,|_a^b\,. \tag{5.70}$$

Example 5.21. Definite Integrals Notation

Evaluate $\int_1^2 x\,dx$.

Solution 5.21. The function $F(x) = \frac{1}{2}x^2$ is an anti derivative of $f(x) = x$; thus we can write

$$\int_1^2 x\,dx = \frac{1}{2}x^2\,|_1^2 = \frac{1}{2}(2)^2 - \frac{1}{2}(1)^2 = 2 - \frac{1}{2} = \frac{3}{2}.\blacktriangle$$

Example 5.22. Definite Integrals Notation

Find the area under the curve $y = \cos(x)$ over the interval $[0, \pi/2]$ and $[0, \pi]$.

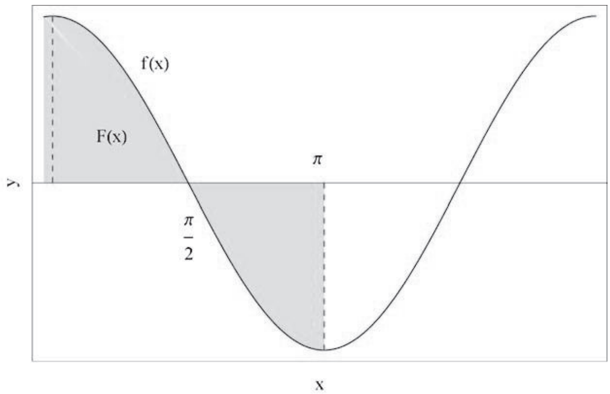


Figure 5.16. Function $f(x) = \cos(x)$ on a interval $[0, \pi]$.

Before you start to evaluate the area make a conjecture of the integral value.

Solution 5.22. The given integral can be interpreted as the signed area between the graph of $y = \cos(x)$ and the interval $[0, \pi]$. The graph in Figure 5.16 suggests that over the interval $[0, \pi]$ the portion of area above the x-axis is the same as the portion of area below the x-axis, so we conjecture that the signed area is zero; this implies that the value of the integral is zero. This is confirmed by the computation

$$\int_0^{\pi} \cos(x)\, dx$$

0

which is carried out step by step as

$$\int_0^{\pi} \cos(x)\, \mathrm{dx} = \sin(x)\,\Big|_0^{\pi} = \sin(\pi) - \sin(0) = 0.$$

The integral on the interval $[0, \pi/2]$ equals

$$\int_0^{\frac{\pi}{2}} \cos(x)\, dx$$

1

which is step by step

$$\int_0^{\pi/2} \cos(x)\, \mathrm{dx} = \sin(x)\,\Big|_0^{\pi/2} = \sin\left(\frac{\pi}{2}\right) - \sin(0) = 1 - 0 = 1. \blacktriangle$$

Observe that in the preceding example we did not include a constant of integration in the anti derivatives. In general, when applying the Fundamental Theorem of Calculus there is no need to include a constant of integration because it will drop out anyway. To see that this is so, let F be any anti derivative of the integrand on $[a, b]$, and let C be any constant; then

$$\int_a^b f(x)\,\mathrm{dx} = F(x) + C\,|_a^b = F(b) + C - (F(a) + C) = F(b) - F(a). \tag{5.71}$$

Thus, for purposes of evaluating a definite integral we can omit the constant of integration in

$$\int_a^b f(x)\,\mathrm{dx} = F(x) + C\,|_a^b \tag{5.72}$$

and express (5.70) as

$$\int_a^b f(x)\,\mathrm{dx} = \left[\int f(x)\,\mathrm{dx}\right]_a^b \tag{5.73}$$

which relates the definite and indefinite integrals.

Example 5.23. Definite Integrals and Indefinite Integrals

Evaluate the integral $\int_1^9 \sqrt{x}\,\mathrm{dx}$

Solution 5.23. The integral can be solved by solving the indefinite integral and substituting the boundaries into the resulting expression

$$\int_1^9 \sqrt{x}\,\mathrm{dx} = \left[\int \sqrt{x}\,\mathrm{dx}\right]_1^9 = \left[\frac{2}{3}\,x^{3/2}\right]_1^9 = \frac{2}{3}\,(27-1) = \frac{52}{3}.$$

or step by step using *Mathematica*

$$\textbf{integ} = \int \sqrt{x}\,dx$$

$$\frac{2\,x^{3/2}}{3}$$

(integ /. x → 9) − (integ /. x → 1)

$$\frac{52}{3}$$

▲

5.3.3.2 The Mean-Value Theorem for Integrals

To reach our goal of showing that continuous functions have anti derivatives, we will need to develop a basic property of definite integrals, known as the Mean-Value Theorem for Integrals. In the next section we will use this theorem to extend the familiar idea of average value so that it applies to continuous functions, but here we will need it as a tool for developing other results.

Let f be a continuous non negative function on $[a, b]$, and let m and M be the minimum and maximum values of $f(x)$ on this interval. Consider the rectangles of heights m and M over the interval $[a, b]$. It is clear geometrically from Figure 5.17 that the area

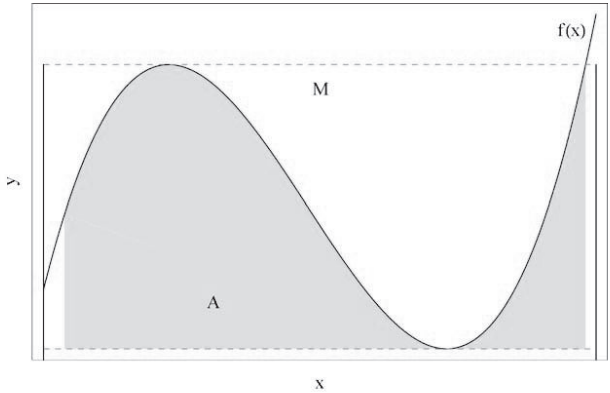


Figure 5.17. Area under a graph $y = f(x)$ showing a maximum M and a minimum m based on the interval $[a, b]$.

$$A = \int_a^b f(x)\,dx \tag{5.74}$$

under $y = f(x)$ is at least as large as the area of the rectangle of height m and no larger than the area of the rectangle of height M. It seems reasonable, therefore, that there is a rectangle over the interval $[a, b]$ of some appropriate height $f(x^*)$ between m and M whose area is precisely A; that is

$$\int_a^b f(x)\,dx = f(x^*)\,(b-a). \tag{5.75}$$

This is a special case of the following result.

Theorem 5.10. *Mean-Value Theorem for Integrals*

If f is continuous on a closed interval $[a, b]$, then there is at least one number x^* in $[a, b]$ such that

$$\int_a^b f(x)\,dx = f(x^*)\,(b-a).\blacksquare$$

Proof 5.10. By the Extremal-Value Theorem if a function f is continuous on a finite closed interval $[a, b]$, then f has both an absolute maximum and an absolute minimum on $[a, b]$; so f assumes a maximum M and a minimum value m on $[a, b]$. Thus, for all x in $[a, b]$, we find

$$m \le f(x) \le M$$

and from Theorem 5.7

$$\int_a^b m\,dx \le \int_a^b f(x)\,dx \le \int_a^b M\,dx \tag{5.76}$$

or

$$m\,(b-a) \;\leq\; \int_a^b f(x)\,\mathrm{dx} \;\leq\; M\,(b-a) \tag{5.77}$$

or

$$m \;\leq\; \frac{1}{b-a}\int_a^b f(x)\,\mathrm{dx} \;\leq\; M. \tag{5.78}$$

That implies that

$$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{dx} \tag{5.79}$$

is a number between m and M, and since $f(x)$ assumes the values m and M on $[a, b]$, it follows from the Intermediate Value Theorem stated above that $f(x)$ must assume the value (5.79) at some x^* in $[a, b]$; that is

$$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{dx} \;=\; f(x^*) \qquad \text{or} \qquad \int_a^b f(x)\,\mathrm{dx} \;=\; f(x^*)\,(b-a).$$

QED

Example 5.24. Mean-Value Theorem

Since $f(x) = x^2$ is continuous on the interval [1, 4], the Mean-Value Theorem for Integrals guarantees that there is a number x^* in [1, 4] such that the Theorem is satisfied.

Solution 5.24. Consider the integral

$$\int_1^4 x^2\,\mathrm{dx} \;=\; f(x^*)\,(4-1) \;=\; (x^*)^2\,(4-1) = 3\,(x^*)^2.$$

But

$$\int_1^4 x^2\,dx$$

21

so that

eq = 3 (x*)^2 == 21

$3\,(x^*)^2 = 21$

and the solution for x^* is

s1 = Solve[eq, x*]

$\{\{x^* \to -\sqrt{7}\}, \{x^* \to \sqrt{7}\}\}$

which numerically is

***N*[s1]**

$\{\{x^* \to -2.64575\}, \{x^* \to 2.64575\}\}$

is a number taken from the interval [1, 4] whose existence is guaranteed by the Mean-Value Theorem for integrals.▲

5.3.3.3 Part 2 of the Fundamental Theorem of Calculus

In Section 5.2 we suggested that f is continuous and non negative on $[a, b]$, and if $A(x)$ is the area under the graph of $y = f(x)$ over the interval $[a, x]$ then $A'(x) = f(x)$. But $A(x)$ can be expressed as the definite integral

$$A(x) = \int_a^x f(t)\,\mathrm{dt} \tag{5.80}$$

where we have used t rather than x as the variable of integration to avoid confusion with the x that appears as the upper limit of integration. Thus the relationship $A'(x) = f(x)$ can be expressed as

$$\frac{d}{\mathrm{dx}}\left[\int_a^x f(t)\,\mathrm{dt}\right] = f(x). \tag{5.81}$$

This is a special case of the following more general result, which applies even if f has negative values.

Theorem 5.11. *Fundamental Theorem of Calculus Part 2*

If f is continuous on an interval I, then f has an anti derivative on I. In particular, if a is any number in I, then the function F defined by

$$F(x) = \int_a^x f(t)\,\mathrm{dt}$$

is an anti derivative of f on I; that is $F'(x) = f(x)$ for each x in I, or in an alternative notation

$$\frac{d}{\mathrm{dx}}\left[\int_a^x f(t)\,\mathrm{dt}\right] = f(x).■$$

Proof 5.11. We will show first that $F(x)$ is defined at each x in the interval I. If $x > a$ and x is in the interval I, then Theorem 5.4 applied to the interval $[a, x]$ and the continuity of f on I ensure that $F(x)$ is defined; and if x is in the interval I and $x \le a$, then Definition 5.3 combined with Theorem 5.4 ensures that $F(x)$ is defined. Thus, $F(x)$ is defined for all x in I.

Next we will show that $F'(x) = f(x)$ for each x in the interval I. If x is not an endpoint of I, then it follows from the definition of a derivative that

$$\begin{aligned} F'(x) = \lim_{w\to x} \frac{F(w) - F(x)}{w - x} &= \lim_{w\to x}\left(\frac{1}{w-x}\left[\int_a^w f(t)\,\mathrm{dt} - \int_a^x f(t)\,\mathrm{dt}\right]\right) = \\ \lim_{w\to x}\left(\frac{1}{w-x}\left[\int_a^w f(t)\,\mathrm{dt} + \int_x^a f(t)\,\mathrm{dt}\right]\right) &= \lim_{w\to x}\left(\frac{1}{w-x}\int_x^w f(t)\,\mathrm{dt}\right) \end{aligned} \tag{5.82}$$

Applying the Mean-Value Theorem of Integrals to $\int_x^w f(t)\,\mathrm{dt}$, we obtain

$$\frac{1}{w-x}\int_x^w f(t)\,\mathrm{dt} = \frac{1}{w-x}[f(t^*)\,(w-x)] = f(t^*) \tag{5.83}$$

where t^* is some number between x and w. Because t^* is between x and w, it follows that $t^* \to x$ as $w \to x$. Thus $f(t^*) \to (x)$ as $w \to x$, since f is assumed continuous at x. Therefore, it follows from (5.82) and (5.83) that

$$F'(x) = \lim_{w\to x}\left(\frac{1}{w-x}\int_x^w f(t)\,\mathrm{dt}\right) = \lim_{w\to x} f(x^*) = f(x).$$

QED

If x is an endpoint in the interval I, then the two-sided limits in the proof must be replaced by the appropriate one-sided limits, but otherwise the arguments are identical.

In words the last formula in Theorem 5.11 states:

If a definite integral has a variable upper limit of integration, a constant lower limit of integration, and a continuous integrand, then the derivative of the integral with respect to its upper limit is equal to the integrand evaluated at the upper limit.

Example 5.25. FTC Part 2

Find

$$\frac{d}{\mathrm{dx}}\left[\int_1^x t^3\,\mathrm{dt}\right]$$

by applying Part 2 of the Fundamental Theorem of Calculus, and then confirm the result by performing the integration and then differentiating.

Solution 5.25. The integrand is a continuous function, so that

$$\frac{d}{\mathrm{dx}}\left[\int_1^x t^3\,\mathrm{dt}\right] = x^3$$

Alternatively, evaluating the integral and then differentiating yields

$$\mathbf{int} = \int_1^x t^3\, dt$$

$$\frac{x^4}{4} - \frac{1}{4}$$

$$\frac{\partial\,\mathbf{int}}{\partial x}$$

$$x^3$$

so the two methods for differentiating the integral agree.▲

5.3.3.4 Differentiation and Integration are Inverse Processes

The two parts of the Fundamental Theorem of Calculus (FTC), when taken together, tell us that differentiation and integration are inverse processes in the sense that each undoes the effect of the other. To see why this is so, note that Part 1 of the FTC implies that

$$\int_a^x f'(t)\,\mathrm{dt} = f(x) - f(a) \tag{5.84}$$

which tells us that if the value of $f(a)$ is known, then the function f can be recovered from the derivative f' by integrating. Conversely, Part 2 of the FTC states that

$$\frac{d}{\mathrm{dx}}\left[\int_a^x f(t)\,\mathrm{dt}\right] = f(x) \tag{5.85}$$

which tells us that the function f can be recovered from its integral by differentiating. Thus, differentiation and integration can be viewed as inverse processes.

It is common to treat parts 1 and 2 of the FTC as a single theorem, and refer to it simply as the Fundamental Theorem of Calculus. This theorem ranks as one of the greatest discoveries in the history of science, and its formulation by Newton and Leibniz is generally regarded to be the discovery of calculus.

5.3.4 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

5.3.4.1 Test Problems

T1. What is a Riemann sum? Why might you want to consider such a sum?

T2. What is the relation between definite integrals and area? Describe some other interpretations of definite integrals.

T3. What is the Fundamental Theorem of Calculus? Why is it so important? Illustrate each part of the theorem with an example.

T4. Describe the rules for working with definite integrals. Give examples.

T5. What is the average value of an integrable function over a closed interval? Must the function assume its average value? Explain.

5.3.4.2 Exercises

E1. Evaluate the Riemann sum for $f(x) = \frac{1}{2} - 3\,x$ with $2 \le x \le 9$ with six sub intervals, taking the sample points to be left endpoints. Explain, with the aid of a diagram, what the Riemann sum represents.

E2. Use a calculator or computer to make a table of values of right Riemann sums R_n for the integral $\int_0^\pi \sin(x)\,dx$ with $n = 5$, 10, 50, and 100. What value do these numbers appear to be approaching?

E3. Express the integral as a limit of sums. Then evaluate, using a computer algebra system to find both the sum and the limit.

a. $\int_0^\pi \sin(3\,x)\,dx$,

b. $\int_1^7 x^8\,dx$.

E4. Write as a single integral in the form $\int_a^b f(x)\,dx$:

$$\int_{-2}^{5} \sin(x)^4 \cos(x)^5\,dx + \int_{5}^{12} \sin(x)^4 \cos(x)^5\,dx - \int_{-2}^{-4} \sin(x)^4 \cos(x)^5\,dx \tag{1}$$

E5. Use the form of the definition of the definite integral to evaluate the integral.

$\int_{-2}^{6}(1+2x)\,dx$,

b. $\int_0^4 (1+2x^2)\,dx$,

c. $\int_0^4 (2-x^3)\,dx$,

d. $\int_1^4 x^4\,dx$.

E6. Use properties of integrals to prove the inequalities:

a. $\int_1^3 \sqrt{x^4+1}\,dx \geq \frac{26}{3}$,

b. $\int_0^{\pi/2} x\sin(x)\,dx \leq \frac{\pi^2}{8}$.

E7. Use areas to evaluate the following integrals:

a. $\int_a^b 3x\,dx$ with $0 < a < b$,

b. $\int_0^b 5x\,dx$ with $b > 0$,

c. $\int_0^b \frac{x}{3}\,dx$ with $b > 0$,

d. $\int_a^b x\,dx$ with $0 < a < b$.

E8. Suppose f has absolute minimum value m and absolute maximum value M. Between what two values must $\int_0^2 f(x)\,dx$ lie? Which property of integrals allows you to make your conclusion?

E9. Use the Max-Min Inequality to find upper and lower bounds for the value of

$$\int_0^1 \frac{1}{1+x^2}\,dx. \tag{2}$$

E10 If your CAS can draw rectangles associated with Riemann sums, use it to draw rectangles associated with Riemann sums that converge to the integrals listed below. Use $n = 6$, 10, 20, and 50 sub intervals of equal length in each case and compare your found results with the given values.

a. $\int_0^1 (1-x)\,dx = \frac{1}{2}$,

b. $\int_0^1 (x^2+1)\,dx = \frac{4}{3}$,

c. $\int_0^{\pi/4} \sec(x)^2\,\mathrm{dx} = 1$.

5.4 Improper Integrals

Up to now we have focused on definite integrals with continuous integrands and finite intervals of integration. In this section we will extend the concept of a definite integral to include infinite intervals of integration and integrands that become infinite within the interval of integration.

It is assumed in the definition of the definite integrals

$$\int_a^b f(x)\,\mathrm{dx} \tag{5.86}$$

that $[a, b]$ is a finite interval and that the limit that defines the integral exists; that is, the function f is integrable. We discussed that continuous functions are integrable, as are bounded functions with finitely many points of discontinuity. We also observed that functions that are not bounded on the interval of integration are not integrable. Thus, for example, a function with a vertical asymptote within the interval of integration would not be integrable.

Our main objective in this section is to extend the concept of a definite integral to allow for infinite intervals of integration and integrands with vertical asymptotes within the interval of integration.

We will call the vertical asymptotes infinite discontinuities, and we will call integrals with infinite intervals of integration or infinite discontinuities within the interval of integration **improper integrals**. Here are some examples:

Improper integrals with infinite intervals of integration:

$$\int_1^\infty \frac{dx}{x^2},\ \int_{-\infty}^0 e^x\, dx,\ \int_{-\infty}^\infty \frac{dx}{1+x^2} \tag{5.87}$$

Improper integrals with infinite discontinuities in the interval of integration:

$$\int_{-3}^3 \frac{dx}{x^2},\ \int_1^2 \frac{dx}{1-x},\ \int_0^\pi \tan(x)\, dx \tag{5.88}$$

Improper integrals with infinite discontinuities and infinite intervals of integration:

$$\int_0^\infty \frac{dx}{\sqrt{x}},\ \int_{-\infty}^\infty \frac{dx}{x^2-9},\ \int_1^\infty \sec(x)\, dx \tag{5.89}$$

5.4.1 Improper Integrals

To motivate a reasonable definition for improper integrals of the form

$$\int_a^\infty f(x)\, dx \tag{5.90}$$

let us begin with the case where f is continuous and non negative on $[a, \infty)$, so we can think of the integral as the area under the curve $y = f(x)$ over the interval $[a, \infty)$ (see Figure 5.18).

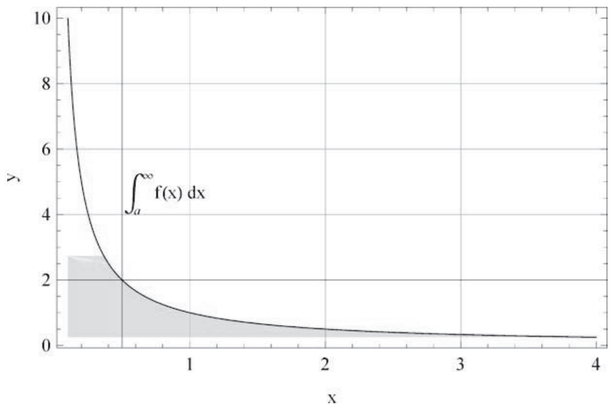


Figure 5.18. Area under a graph $y = 1/x$ based on the interval $[a, \infty)$.

At first, you might be inclined to argue that this area is infinite because the region has infinite extent. However, such an argument would be based on vague intuition rather than precise mathematical logic, since the concept of area has only been defined over intervals of finite extent. Thus, before we can make any reasonable statement about the area of the region in Figure 5.18, we need to begin by defining what we mean by the area of this region. For that purpose, it will help to focus on a specific example.

Suppose we are interested in the area A of the region that lies below the curve $y = 1/x^2$ and above the interval $[1, \infty)$ on the x-axis. Instead of trying to find the entire area at once, let us begin by calculating the portion of the area that lies above the interval $[1, \ell]$, where $\ell > 1$ is arbitrary. That area is

$$\int_1^\ell \frac{dx}{x^2} = -\frac{1}{x}\Big|_1^\ell = 1 - \frac{1}{\ell} \tag{5.91}$$

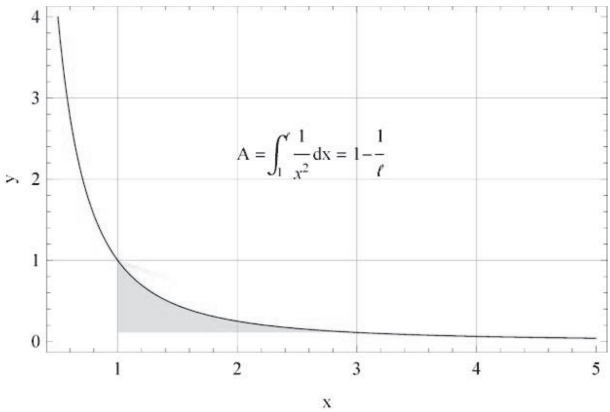


Figure 5.19. Area under a graph $y = 1/x^2$ based on the interval $[1, \ell]$.

If we now allow ℓ to increase so that $\ell \to \infty$, then the portion of the area over the interval $[1, \ell]$ will begin to fill out the area over the entire interval $[1, \infty)$ (see the animation in Figure 5.20).

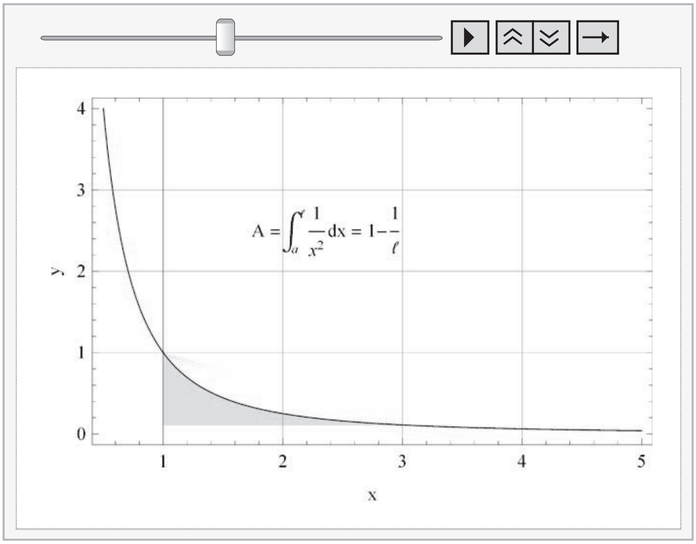


Figure 5.20. Area under a graph $y = 1/x^2$ based on the interval $[1, \ell]$.

Hence we can reasonably define the area A under $y = 1/x^2$ over the interval $[1, \infty)$ to be

$$A = \int_1^\infty \frac{dx}{x^2} = \lim_{\ell\to\infty} \int_1^\ell \frac{dx}{x^2} = \lim_{\ell\to\infty}\left(1 - \frac{1}{\ell}\right) = 1. \tag{5.92}$$

Thus, the area has a finite value of 1 and is not infinite as we first conjectured.

With the preceding discussion as our guide, we make the following definition which is applicable to functions with both positive and negative values:

Definition 5.5. *Improper Integrals*

The improper integral of a function f over the interval $[a, \infty)$ is defined as

$$\int_a^\infty f(x)\,dx = \lim_{\ell\to\infty} \int_a^\ell f(x)\,dx$$

In the case where the limit exists the proper integral is said to converge, and the limit is defined to be the value of the integral. In the case where the limit does not exist, the improper integral is said to diverge, and it is not assigned a value. ■

If f is non negative on $[a, \infty)$ and the improper integral converges, then the value of the integral is regarded to be the area under the graph of f over the interval $[a, \infty)$; and if the integral diverges, then the area under the graph of f over the interval $[a, \infty)$ is regarded to be infinite.

Example 5.26. Improper Integrals

Evaluate $\int_1^\infty \frac{dx}{x^3}$ and $\int_1^\infty \frac{dx}{x}$.

Solution 5.26. Following the definition, we replace the infinite upper limit by a finite upper limit ℓ, and then take the limit of the resulting integral. This yields

$$\int_1^\infty \frac{dx}{x^3} = \lim_{\ell\to\infty} \int_1^\ell \frac{dx}{x^3} = \lim_{\ell\to\infty} \left[-\frac{1}{2\,x^2}\right]_1^\ell = \lim_{\ell\to\infty} \left(\frac{1}{2} - \frac{1}{2\,\ell^2}\right) = \frac{1}{2}.$$

The second integral is determined by

$$\int_1^\infty \frac{dx}{x} = \lim_{\ell\to\infty} \int_1^\ell \frac{dx}{x} = \lim_{\ell\to\infty} [\ln(x)]_1^\ell = \lim_{\ell\to\infty} (\ln(\ell)) = \infty.$$

In this case the integral diverges and hence has no value.▲

Because the functions $1/x^3$, $1/x^2$, and $1/x$ are non negative over the interval $[1, \infty)$, it follows from the calculations carried out so far that over this interval the area under $y = 1/x^3$ is $1/2$, the area under $y = 1/x^2$ is 1, and that the area under $y = 1/x$ is infinite. However, on the surface the graphs of the three functions seem very much alike and there is nothing to suggest why one of the area should be infinite and the other two finite. One explanation is that $1/x^3$ and $1/x^2$ approach zero more rapidly under the curves $y = 1/x^3$ and $y = 1/x^2$ than under $y = 1/x$ as $\ell \to \infty$, and the difference is just enough that the first two areas are finite and the third is infinite.

Example 5.27. Improper Integrals I

For what values of p does the integral $\int_1^\infty \frac{dx}{x^p}$ converge?

Solution 5.27. We know from the preceding example that the integral diverges if $p = 1$, so let us assume that $p \neq 1$. In this case we have

$$\int_1^\infty \frac{dx}{x^p} = \lim_{\ell\to\infty} \int_1^\ell x^{-p}\, dx = \lim_{\ell\to\infty} \frac{x^{1-p}}{1-p} \Big|_1^\ell = \lim_{\ell\to\infty} \left[\frac{\ell^{1-p}}{1-p} - \frac{1}{1-p}\right].$$

If $p > 1$, then the exponent $1 - p$ is negative and $\ell^{1-p} \to 0$ as $\ell \to \infty$; and if $p < 1$, then the exponent $1 - p$ is positive and $\ell^{1-p} \to \infty$ as $\ell \to \infty$. Thus, the integral converges if $p > 1$ and diverges otherwise. In the convergent case the value of the integral is

$$\int_1^\infty \frac{dx}{x^p} = \left[0 - \frac{1}{1-p}\right] = \frac{1}{p-1} \qquad \text{for } p > 1.\blacktriangle$$

The following theorem summarizes this result:

Theorem 5.12. *Convergence of Power Functions*

$$\int_1^\infty \frac{dx}{x^p} = \frac{1}{p-1} \qquad \text{if} \qquad p > 1$$

if $p \leq 1$ then the integral diverges.■

Example 5.28. Improper Integrals II

Evaluate $\int_0^\infty (1-x)\, e^{-x}\, \text{dx}$.

Solution 5.28. Integrating by parts with $u = 1 - x$ and $\text{dv} = e^{-x}\, \text{dx}$ yields

$$\int (1-x)\, e^{-x}\, \text{dx} = -e^{-x}(1-x) - \int e^{-x}\, \text{dx} = -e^{-x} + x\, e^{-x} + e^{-x} + C = x\, e^{-x} + C.$$

Thus

$$\int_0^\infty (1-x)\, e^{-x}\, \text{dx} = \lim_{\ell\to\infty} \left[x\, e^{-x}\right]_0^\ell = \lim_{\ell\to\infty} \frac{\ell}{e^\ell}.$$

This limit is an indeterminate form of type ∞/∞, so we will apply L'Hospital's rule by differentiat ing the numerator and denominator with respect to ℓ. This yields

$\int_0^\infty (1-x)\, e^{-x}\, \text{dx} = \lim_{\ell\to\infty} \frac{1}{e^\ell} = 0.$

An explanation of why this integral is zero can be obtained by interpreting the integral as the net area between the graph of $y = (1-x)\, e^{-x}$ and the interval $[0, \infty)$.

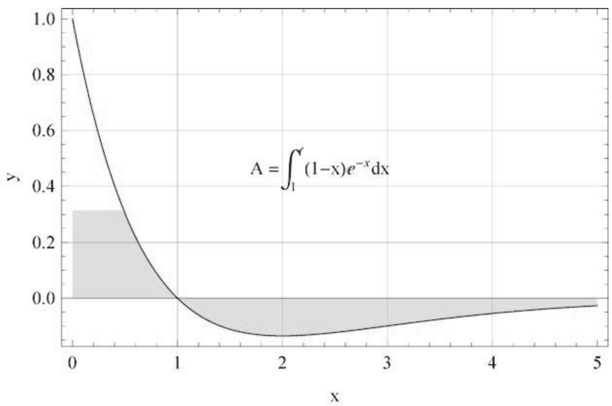


Figure 5.21. Area under a graph $y = (1-x)\, e^{-x}$ based on the interval $[0, \ell]$ as $\ell \to \infty$.▲

To handle the other situations when ∞ or $-\infty$ occurs as one of the boundaries or both we define:

Definition 5.6. *Improper Integrals*

The improper integral of f over the interval $(-\infty, b]$ is defined as

$$\int_{-\infty}^b f(x)\, dx = \lim_{\ell\to-\infty} \int_\ell^b f(x)\, dx. \tag{5.93}$$

The integral is said to converge if the limit exists and diverge if it does not. The improper integral of f over the interval $(-\infty, \infty)$ is defined as

$$\int_{-\infty}^\infty f(x)\, dx = \int_{-\infty}^c f(x)\, dx + \int_c^\infty f(x)\, dx \tag{5.94}$$

where c is any real number. The improper integral is said to converge if both terms converge and diverge if either term diverges. ■

Remark 5.5. In this definition, if f is non negative on the interval of integration, then the improper integral is regarded to be the area under the graph of f over that interval; the area has a finite value if the integral converges and is infinite if it diverges. We also note that in (5.94) it is usual to choose $c = 0$, but the choice does not matter; it can be proved that neither the convergence nor the value of the integral depends on the choice of c.

Example 5.29. Improper Integrals

Evaluate $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\, dx$.

Solution 5.29. We will evaluate the integral by choosing $c = 0$ in (5.94). With this value for c we obtain

$$\int_{-\infty}^{0} \frac{1}{1+x^2}\, dx + \int_{0}^{\infty} \frac{1}{1+x^2}\, dx = \lim_{\ell\to-\infty}\left[\int_{\ell}^{0} \frac{1}{1+x^2}\, dx\right] + \lim_{\ell\to\infty}\left[\int_{0}^{\ell} \frac{1}{1+x^2}\, dx\right] =$$
$$\lim_{\ell\to-\infty}\left[\arctan(x)\right]_{\ell}^{0} + \lim_{\ell\to\infty}\left[\arctan(x)\right]_{0}^{\ell} =$$
$$\lim_{\ell\to-\infty}\left[-\arctan(\ell)\right] + \lim_{\ell\to\infty}\left[\arctan(\ell)\right] = \frac{\pi}{2} + \frac{\pi}{2} = \pi$$

Since the integrand is non negative on the interval $(-\infty, \infty)$, the integral represents the area of the region shown in Figure 5.22.

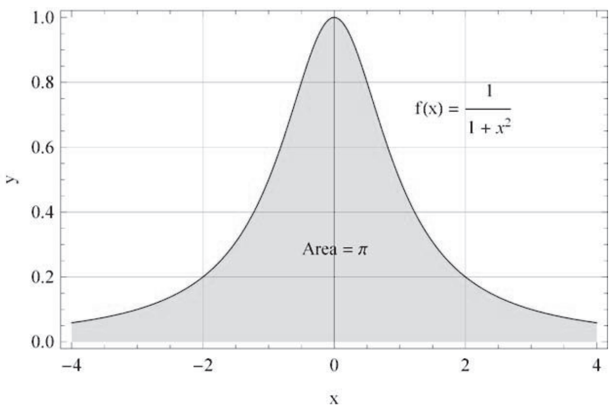


Figure 5.22. Area of the improper integral $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\, dx$.▲

Next we will consider improper integrals whose integrands have infinite discontinuities. We will start with the case where the interval of integration is a finite interval $[a, b]$ and the infinite discontinuity occurs at the right-hand endpoint.

To motivate an appropriate definition for such an integral let us consider the case where f is non negative on $[a, b]$, so we can interpret the improper integral $\int_a^b f(x)\,dx$ as the area of the region in Figure 5.23.

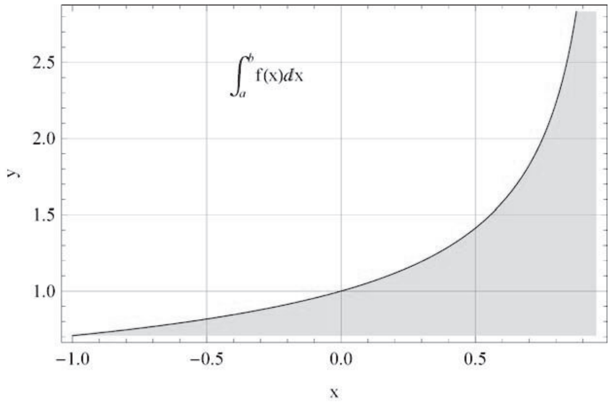


Figure 5.23. Area of the improper integral $\int_a^b f(x)\,dx$ where the integrand becomes indefinite at the right boundary.▲

The problem of finding the area of this region is complicated by the fact that it extends indefinitely in the positive y-direction. However, instead of trying to find the entire area at once, we can proceed indirectly by calculating the portion of the area over the interval $[a, \ell]$ and then letting ℓ approach b to fill out the area of the entire region. Motivated by this idea, we make the following definition:

Definition 5.7. *Improper Integral with Discontinuity in the Interval* $[a.b]$

If f is continuous on the interval $[a, b]$, except for an infinite discontinuity at b, then the improper integral of f over the interval $[a, b]$ is defined as

$$\int_a^b f(x)\,dx = \lim_{\ell \to b^-} \int_a^\ell f(x)\,dx. \tag{5.95}$$

In the case where the limit exits, the improper integral is said to converge, and the limit is defined to be the value of the integral. In the case where the limit dos not exist, the improper integral is said to diverge, and it is not assigned a value.■

Example 5.30. Improper Integral with Singularity I

Evaluate $\int_0^1 \frac{1}{\sqrt{1-x}}\,dx$.

Solution 5.30. The integral is improper because the integrand approaches $+\infty$ as x approaches the upper limit 1 from the left. From (5.95)

$$\int_0^1 \frac{1}{\sqrt{1-x}}\,dx =$$

$$\lim_{\ell\to 1^-}\int_0^\ell \frac{1}{\sqrt{1-x}}\,dx = \lim_{\ell\to 1^-}\left[-2\sqrt{1-x}\,\right]_0^\ell = \lim_{\ell\to 1^-}\left[-2\sqrt{1-\ell}\,+2\right] = 2.$$

The same result is obtained by *Mathematica*

$$\int_0^1 \frac{1}{\sqrt{1-x}}\,dx$$

2

▲

Improper integrals with an infinite discontinuity at the left-hand endpoint or inside the interval of integration are defined as follows.

Definition 5.8. *Improper Integral with Discontinuity in the Interval* $[a.b]$

If f is continuous on the interval $[a, b]$, except for an infinite discontinuity at a, then the improper integral of f over the interval $[a, b]$ is defined as

$$\int_a^b f(x)\,dx = \lim_{k\to a^+}\int_k^a f(x)\,dx.$$

The integral is said to converge if the limit exists and diverge if it does not. If f is continuous on the interval $[a, b]$, except for a infinite discontinuity at a number c in (a, b), then the improper integral of f over the interval $[a, b]$ is defined as

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

The improper integral is said to converge if both terms converge and diverge if either term diverges (Figure 5.24).■

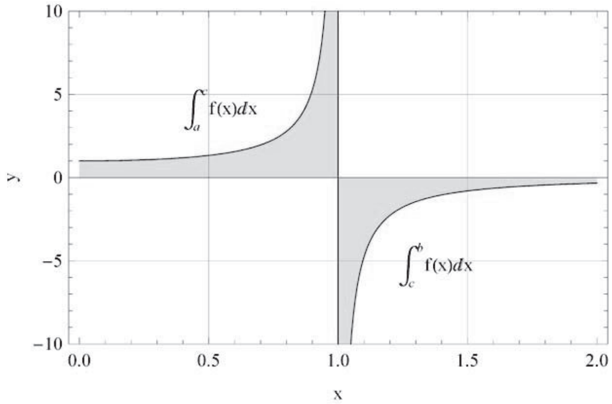


Figure 5.24. Area of the improper integral $\int_a^b f(x)\,dx$ where the integrand becomes indefinite at a point c inside the interval $[a, b]$.▲

Example 5.31. Improper Integral with Singularity II

Evaluate $\int_1^4 \frac{1}{(x-2)^{2/3}}\,dx$.

Solution 5.31. The integral is improper because the integrand approaches $+\infty$ at $x = 2$, which is inside the interval of integration. From Definition 5.8 we obtain

$$\int_1^4 \frac{1}{(x-2)^{2/3}}\,dx = \int_1^2 \frac{1}{(x-2)^{2/3}}\,dx + \int_2^4 \frac{1}{(x-2)^{2/3}}\,dx.$$

But

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

$$\lim_{\ell\to 2^-} \int_1^\ell \frac{1}{(x-2)^{2/3}}\,dx = \lim_{\ell\to 2^-}\left[3\,(\ell-2)^{1/3} - 3\,(1-2)^{1/3}\right] = -3\,(-1)^{1/3} = 3$$

$$\int_2^4 \frac{1}{(x-2)^{2/3}}\,dx = \lim_{k\to 2^+} \int_k^4 \frac{1}{(x-2)^{2/3}}\,dx = \lim_{k\to 2^+}\left[3\,(4-2)^{1/3} - 3\,(k-2)^{1/3}\right] = 3\,(2)^{1/2}$$

Thus we find

$$\int_1^4 \frac{1}{(x-2)^{2/3}}\,dx = \int_1^2 \frac{1}{(x-2)^{2/3}}\,dx + \int_2^4 \frac{1}{(x-2)^{2/3}}\,dx = 3 + 3\,(2)^{1/3}.▲$$

Warning: It is sometimes tempting to apply the Fundamental Theorem of Calculus directly to an improper integral without taking the appropriate limits. To illustrate what can go wrong with this procedure, suppose we ignore the fact that the integral

$$\int_0^2 \frac{1}{(x-1)^2}\, dx$$

is improper and write the wrong relation

$$\int_0^2 \frac{1}{(x-1)^2}\, dx = \left[-\frac{1}{x-1}\right]_0^1 = -1-(1) = -2.$$

This result is clearly nonsense because the integrand is never negative and consequently the integral cannot be negative! To evaluate the improper integral correctly we should write

$$\int_0^2 \frac{1}{(x-1)^2}\, dx = \int_0^1 \frac{1}{(x-1)^2}\, dx + \int_1^2 \frac{1}{(x-1)^2}\, dx$$

But

$$\int_0^1 \frac{1}{(x-1)^2}\, dx = \lim_{\ell\to 1^-} \int_0^\ell \frac{1}{(x-1)^2}\, dx = \lim_{\ell\to 1^-} \left[-\frac{1}{\ell-1} - 1\right] = +\infty$$

so that the total integral diverges.

5.4.2 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

5.4.2.1 Test Problems

T1. How are improper integrals defined? Give examples.

T2. What are the problems in evaluating improper integrals?

T3. Are horizontal asymptotes a problem in improper integrals?

5.4.2.2 Exercises

E1. Explain why each of the following integrals is improper.

- **a.** $\int_0^{\pi/2} \sec(x)\, dx$,
- **b.** $\int_{-\infty}^0 \frac{1}{x^2+2}\, dx$,
- **c.** $\int_0^\infty x^4\, e^{-x^2}\, dx$,
- **d.** $\int_0^3 \frac{x}{x^2-4x+7}\, dx$.

E2. Which of the following integrals are improper? Why?

- **a.** $\int_1^2 \frac{1}{2x-1}\, dx$,
- **b.** $\int_1^2 \ln(x-1)\, dx$,
- **c.** $\int_0^1 \frac{1}{2x-1}\, dx$,
- **d.** $\int_0^3 \frac{\sin(x)}{1+x^2}\, dx$.

E3. Find the area under the curve $f(x) = 1/x^3$ from $x = 1$ to $x = s$ and evaluate it for $s = 10$, 100, and 1000. Then find the total area under this curve for $x \geq 1$.

E4. Determine whether each integral is convergent or divergent. Evaluate those that are convergent.

- **a.** $\int_0^\infty \frac{x}{(x^2+2)^2}\, dx$,

$\int_4^\infty e^{-x/2}\,dx$,

c. $\int_1^\infty \frac{e^{-\sqrt{x}}}{\sqrt{x}}\,dx$,

d. $\int_{-\infty}^\infty \cos(\pi\,x)\,dx$,

e. $\int_{-\infty}^\infty x\,e^{-x^2}\,dx$,

f. $\int_1^\infty \frac{1}{x\ln(x)^2}\,dx$,

g. $\int_0^1 \frac{6}{x^6}\,dx$,

h. $\int_0^\infty \frac{e^x}{e^{2x}+3}\,dx$.

E5. Determine whether each integral is convergent or divergent. Evaluate those that are convergent.

a. $\int_0^\infty \frac{\arctan(x)}{2+e^x}\,dx$,

b. $\int_0^\pi \frac{\sin(x)^2}{\sqrt{x}}\,dx$,

c. $\int_0^1 \frac{\sec(x)^2}{x\sqrt{x}}\,dx$,

d. $\int_1^\infty \frac{2+e^{-x}}{x}\,dx$,

e. $\int_0^\infty \frac{x}{x^4+1}\,dx$.

E6. Find the value of the constant C for which the integral

$$\int_0^\infty \left(\frac{1}{\sqrt{x^2+4}} - \frac{C}{x+2}\right) dx \tag{1}$$

converges. Evaluate the integral for this value of C.

E7. Show that $\int_0^\infty x^2\,e^{-x^2}\,dx = \frac{1}{2}\int_0^\infty e^{-x^2}\,dx$.

E8. Determine how large the number has to be so that

$$\int_a^\infty \frac{1}{x^2+1}\,dx < 0.001 \tag{2}$$

E9. If $f(x)$ is continuous for $x \geq 0$, the *Laplace transform* of f is the function defined by

$$F(s) = \int_0^\infty f(x)\,e^{-s\,x}\,dx \tag{3}$$

and the domain of F is the set consisting of all numbers s for which the integral converges. Find the Laplace transforms of the following functions.

a. $f(x) = 1$,

b. $f(x) = x$,

c. $f(x) = e^x$,

d. $f(x) = \sin(x)$.

E10 The *average speed* of molecules in an ideal gas is

$$\overline{v} = \frac{4}{\sqrt{\pi}}\left(\frac{M}{2\,R\,T}\right)^{3/2}\int_0^\infty v^3\,e^{-M\,v^2/(2\,R\,T)}\,dv \tag{4}$$

where M is the molecular weight of the gas, R is the gas constant, T is the gas temperature, and v is the molecular speed. Show that

$$\overline{v} = \sqrt{\frac{8\,R\,T}{\pi\,M}}\,. \tag{5}$$

5.5 Applications of Definite Integrals

In the last chapter we introduced the definite integral as the limit of a Riemann sum in the context of finding areas. However, Riemann sums and definite integrals have applications that extend far beyond the area problem. In this section we will show how Riemann sums and definite integrals arise in such problems as finding the volume and surface area of a solid, and finding the length of a plane curve.

Although these problems are drivers, the required calculations can all be approached by the same procedure that we used to find areas—breaking the required calculation into small parts making an approximation that is good because that part is small, adding the approximations from the part to produce a Riemann sum that approximates the entire quantity to be calculated, and then taking the limit of the Riemann sums to produce an exact result.

5.5.1 Area Between two Curves

In the last section we showed how to find the area between a curve $y = f(x)$ and an interval on the x-axis. Here we will show how to find the area between two curves.

Before we consider the problem of finding the area between two curves it will be helpful to review the basic principle that underlies the calculation of area as a definite integral. Recall that if f is continuous and non negative on $[a, b]$, then the definite integral for the area A under $y = f(x)$ over the interval $[a, b]$ is obtained in three steps

1. Assuming that the width of the kth strip is Δx_k, approximate the area of that strip by the area of a rectangle of width Δx_k and height $f(x_k^*)$, where x_k^* is a number in the kth sub interval.
2. Add the approximate areas of the strips to approximate the entire area A by the Riemann sum:

$$A \approx \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k \tag{5.96}$$

3. Take the limit of the Riemann sums as the number of sub intervals increases and their widths approaches zero. This causes the error in the approximations to approach zero and produces the following definite integral for the exact area A:

$$A = \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} f(x_k^*)\,\Delta x_k = \int_a^b f(x)\,dx. \tag{5.97}$$

Observe the effect that the limit process has on the various parts of the Riemann sum:

- The quantity x_k^* in the Riemann sum becomes the variable x in the definite integral.
- The interval width Δx_k in the Riemann sum becomes the dx in the definite integral.
- The interval $[a, b]$ is implicit in the Riemann sum as the aggregate of the sub intervals with widths $\Delta x_1, \ldots, \Delta x_n$ but $[a, b]$ is explicitly represented by the upper and lower limits of

integration in the definite integral.

We will now consider the following extension of the area problem.

Theorem 5.13. *First Area Problem*

Suppose that f and g are continuous functions on an interval $[a, b]$ and

$$f(x) \geq g(x) \qquad \text{for} \qquad a \leq x \leq b. \tag{5.98}$$

This means that the curve $y = f(x)$ lies above the curve $y = g(x)$ and that the two can touch but not cross. Find the area A of the region bounded above by $y = f(x)$, below by $y = g(x)$, and on the sides by the lines $x = a$ and $x = b$. ■

To solve this problem we divide the interval $[a, b]$ in n sub intervals, which has the effect of subdividing the region into n strips. If we assume that the width of the kth strip is Δx_k, then the area of the strip can be approximated by the area of a rectangle of width Δx_k and height $f(x_k^*) - g(x_k^*)$, where x_k^* is a number in the kth sub interval. Adding these approximations yields the following Riemann sum that approximates the area A:

$$A \approx \sum_{k=1}^{n} \left[f(x_k^*) - g(x_k^*)\right] \Delta x_k \tag{5.99}$$

Taking the limit as n increases and the widths of the sub interval approach zero yields the following definite integral for the area A between the curves:

$$A = \lim_{\max(\Delta x_k) \to 0} \sum_{k=1}^{n} \left[f(x_k^*) - g(x_k^*)\right] \Delta x_k = \int_a^b [f(x) - g(x)]\, dx. \tag{5.100}$$

In summary, we have the following result:

Theorem 5.14. *Area Formula*

If f and g are continuous functions on the interval $[a, b]$, and if $f(x) \geq g(x)$ for all x in $[a, b]$, then the area of the region bounded above by $y = f(x)$, below by $y = g(x)$, on the left by the line $x = a$, and on the right by the line $x = b$ is

$$A = \int_a^b [f(x) - g(x)]\, dx. \blacksquare \tag{5.101}$$

In the case where f and g are non negative on the interval $[a, b]$, the formula

$$A = \int_a^b [f(x) - g(x)]\, dx = \int_a^b f(x)\, dx - \int_a^b g(x)\, dx \tag{5.102}$$

states that the area A between the curves can be obtained by subtracting the area under $y = g(x)$ from the area under $y = f(x)$ (see Figure 5.25)

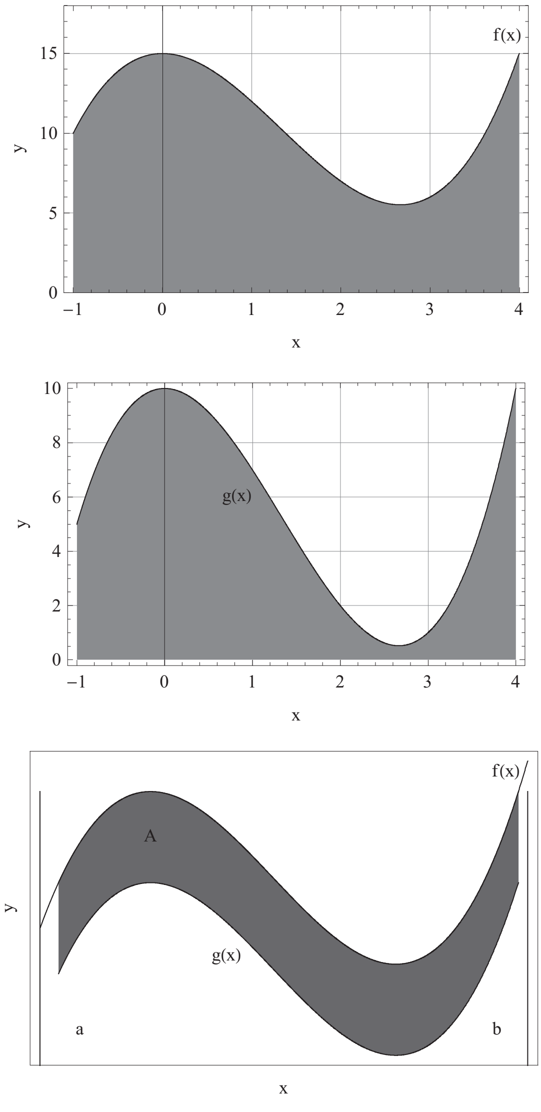


Figure 5.25. Area under a graph $y = f(x) - g(x)$ on the interval $[a, b]$.

When the region is complicated, it may require some careful thought to determine the integrand and limits of integration in (5.101). Here is a systematic procedure that you can follow to set up this formula.

- **Step 1**. Sketch the region and then draw a vertical line segment through the region at an arbitrary point x on the x-axis, connecting the top and bottom boundaries.
- **Step 2**. The y-coordinate of the endpoint of the line segment sketched in Step 1 will be $f(x)$,

the bottom one $g(x)$, and the length of the line segment will be $f(x) - g(x)$. This is the integrand in (5.101).

- **Step 3**. To determine the limits of integration, imagine moving the line segment left and then right. The leftmost position at which the line segment intersects the region is $x = a$ and the rightmost is $x = b$.

Remark 5.6. It is not necessary to make an extremely accurate sketch in Step 1; the only purpose of the sketch is to determine which curve is the upper boundary and which is the lower boundary.

There is a useful way of thinking about this procedure: If you view the vertical line segment as the cross section of the region at the point x, then Formula (5.101) states that the area between the curves is obtained by integrating the length of the cross section over the interval from a to b.

Example 5.32. Area Calculations I

Find the area of the region bounded above by $y = x + 6$, bounded below by $y = x^2$ and bounded on the sides by the lines $x = 0$ and $x = 2$.

Solution 5.32. The region and a cross section are shown in Figure 5.26. The cross section extends from $g(x) = x^2$ on the bottom to $f(x) = x + 6$ on the top. If the cross section is moved through the region, then its leftmost position will be $x = 0$ and its rightmost position will be $x = 2$.

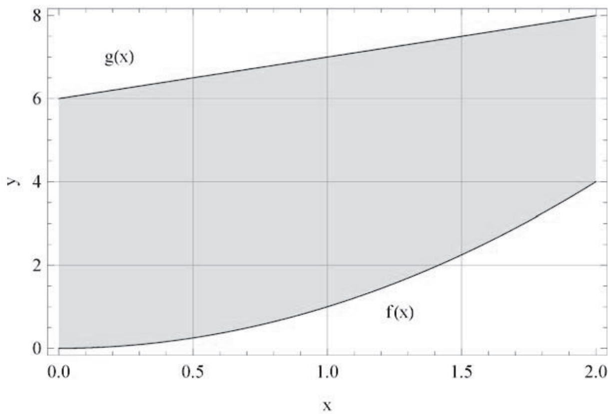


Figure 5.26. Area under a graph $f(x) = x^2$, $g(x) = x + 6$ on the interval $[0, 2]$.

Thus, from (5.101)

$$A = \int_0^2 \left(-x^2 + x + 6\right) dx$$

$$\frac{34}{3}$$

▲

It is possible that the upper and lower boundaries of a region may intersect at one or both endpoints, in which case the sides of the region will be points, rather than vertical line segments. When that occurs you will have to determine the points of intersection to obtain the limits of integration.

Example 5.33. Area Calculations II

Find the area of the region that is enclosed between the curves $y = x^2$ and $y = x + 6$.

Solution 5.33. A sketch of the region (Figure 5.27) shows that the lower boundary is $y = x^2$ and the upper boundary is $y = x + 6$.

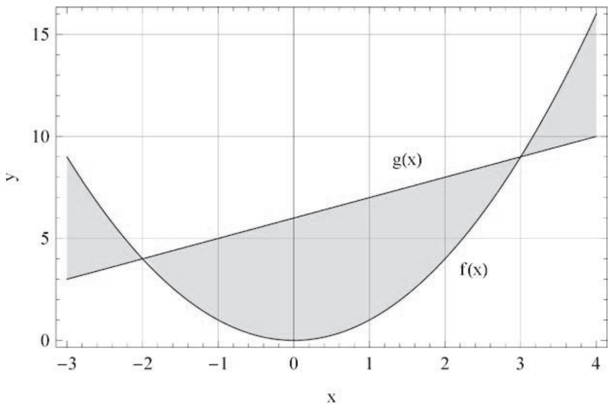


Figure 5.27. Area under a graph $f(x) = x^2$, $g(x) = x + 6$ on the interval $[-2, 3]$.

At the endpoints of the region, the upper and lower boundaries have the same y-coordinate; thus, to find the endpoints we equate

$$y = x^2 \qquad \text{and} \qquad y = x + 6. \tag{5.103}$$

This yields

eq1 = x^2 == x + 6

$x^2 = x + 6$

Solving this equation with respect to x gives us

sol1 = Solve[eq1, *x*]

$\{\{x \to -2\}, \{x \to 3\}\}$

Although the y-coordinates of the endpoints are not essential to our solution, they may be obtained from (5.103) by substituting $x = -2$ and $x = 3$ in either equation. This yields

x^2 /. sol1

$\{4, 9\}$

so the upper and lower boundaries intersect at $(-2, 4)$ and $(3, 9)$.

From (5.101) with $f(x) = x + 6$, $g(x) = x^2$, $a = -2$, and $b = 3$, we obtain the area

$$A = \int_{-2}^{3} \left(-x^2 + x + 6\right) dx$$

$$\frac{125}{6}$$

▲

5.5.2 Volumes by Slicing

In the last section we showed that the area of a plane region bounded by two curves can be obtained by integrating the length of a general cross section over an appropriate interval. In this section we will see that the same basic principle can be used to find volumes of certain three-dimensional solids.

5.5.2.1 Basic Ideas

Recall hat the underlying principle for finding the area of a plane region is to divide the region into thin strips, approximate the area of each strip by the area of a rectangle, add the approximations to form a Riemann sum, and take the limit of the Riemann sums to produce an integral for the area. Under appropriate conditions, the same strategy can be used to find the volume of a solid. The idea is to divide the solid into thin slabs, approximate the volume of each slab, add the approximations to form a Riemann sum, and take the limit of the Riemann sum to produce an integral for the volume.

What makes this method work is the fact that a thin slab has cross sections that do not vary much in size or shape, which, as we will see, makes its volume easy to approximate. Moreover, the thinner the slab, the less variation in its cross section and better the approximation. Thus, once we approximate the volumes of the slab, we can set up a Riemann sum whose limit is the volume of the entire solid. We will give the details shortly, but first we need to discuss how to find the volume of a solid whose cross sections do not vary in size and shape.

One of the simplest examples of a solid with congruent cross sections is a right circular cylinder of radius r, since all cross sections taken perpendicular to the central axis are circular regions of radius r. The volume V of a right circular cylinder of radius r and height h can be expressed in terms of the height and the area of a cross section as

$$V = \pi r^2 h = \text{[area of cross section]} \times \text{[height]}. \tag{5.104}$$

This is a special case of a more general volume formula that applies to solids called right cylinders (see Figure 5.28). A right cylinder is a solid that is generated when a plane region is translated along a line or axis that is perpendicular to the region. The distance h that the region is translated is called the height or sometimes the width of the cylinder, and each cross-section is a duplicate of the translated region.

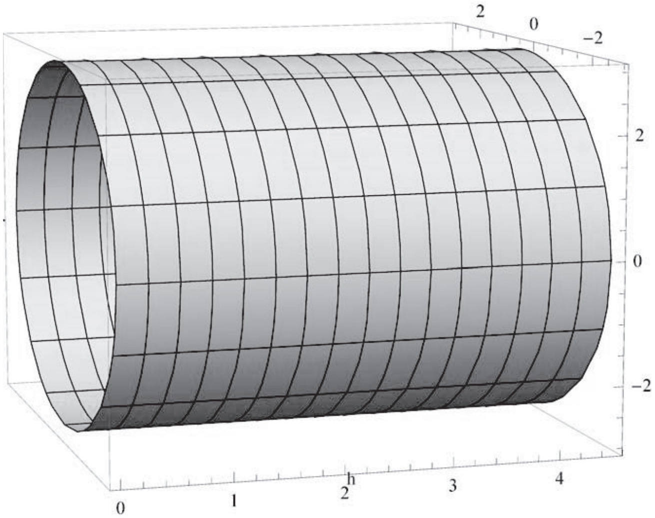


Figure 5.28. Cylinder with radius *r* and height *h*.

We will assume that the volume V of a right cylinder with cross-sectional area A and height h is given by

$$V = A\,h = \text{[area of cross section]} \times \text{[height]}. \tag{5.105}$$

Note that this is consistent with (5.104) for the volume of a right circular cylinder. We now have all of the tools required to solve the following problem.

Theorem 5.15. *Volume Problem*

Let S be a solid that extends along the x-axis and is bounded on the left and right, respectively, by the planes that are perpendicular to the x-axis at x = a and x = b. Find the volume V of the solid, assuming that its cross-sectional area A(x) is known at each x in the interval [a, b]. ■

To solve those problem we divide the interval $[a, b]$ into n sub intervals, which has the effect of dividing the solid into n slabs.

If we assume that the width of the kth slab is Δx_k, then the volume of the slab can be approximated by the volume of a right cylinder of width Δx_k and cross-sectional area $A(x_k^*)$, where x_k^* is a number in the kth sub interval. Adding these approximations yields the following Riemann sum that approximates the volume V:

$$V \approx \sum_{k=1}^{n} A(x_k^*)\,\Delta x_k. \tag{5.106}$$

Taking the limit as n increases and the widths of the sub intervals approaches zero yields the definite integral

$$V = \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} A(x_k^*)\,\Delta x_k = \int_a^b A(x)\,dx. \tag{5.107}$$

In summary, we have the following result:

Theorem 5.16. *Volume Formula I*

Let S be a solid bounded by two parallel planes perpendicular to the x-axis at x = a and x = b. If, for each x in [a, b], the cross-sectional area of S perpendicular to the x-axis is A(x), then the volume of the solid is

$$V = \int_a^b A(x)\,dx \tag{5.108}$$

provided A(x) is integrable. ■

There is a similar result for cross sections perpendicular to the y-axis.

Theorem 5.17. *Volume Formula II*

Let S be a solid bounded by two parallel planes perpendicular to the y-axis at y = c and y = d. If, for each y in [c, d], the cross-sectional area of S perpendicular to the y-axis is A(y), then the volume of the solid is

$$V = \int_c^d A(y)\,dy \tag{5.109}$$

provided A(y) is integrable. ■

Example 5.34. Volume Calculations I

Derive the formula for the volume of a right pyramid whose altitude is h and whose base is a square with sides of length a.

Solution 5.34. As illustrated in Figure 5.29 we introduce a regular coordinate system in which the y-axis passes through the apex and is perpendicular to the base, and the x-axis passes through the base and is parallel to a side of the base.

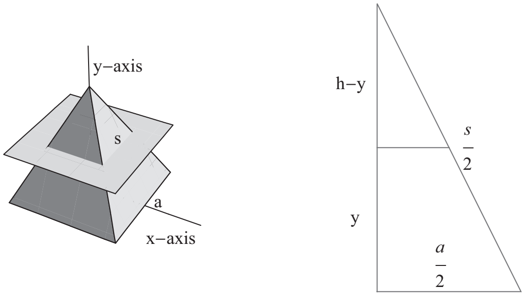


Figure 5.29. Cross section of a pyramid of height h and a baseline of length a.

At any y in the interval $[0, h]$ on the y-axis, the cross section perpendicular to the y-axis is a square. If s denotes the length of a side of this square, then by similar triangles

$$\frac{\frac{1}{2}s}{\frac{1}{2}a} = \frac{h-y}{h} \qquad \text{or} \qquad s = \frac{a}{h}(h-y). \tag{5.110}$$

Thus, the area $A(y)$ of the cross section at y is

$$A(y) = s^2 = \frac{a^2}{h^2}(h-y)^2 \tag{5.111}$$

and the volume is

$$\begin{aligned} V = \int_0^h A(y)\,dy = \int_0^h \frac{a^2}{h^2}(h-y)^2\,dy = \frac{a^2}{h^2}\int_0^h (h-y)^2\,dy = \frac{a^2}{h^2}\left[-\frac{1}{3}(h-y)^3\right]_0^h = \\ \frac{a^2}{h^2}\left[0 + \frac{1}{3}h^3\right] = \frac{1}{3}a^2 h. \end{aligned} \tag{5.112}$$

That is, the volume is $1/3$ of the area of the base times the altitude.▲

5.5.2.2 Volumes by Disks

We will be interested in the following general problem:

Theorem 5.18. *Revolution Problem*

Let f be continuous and non negative on $[a, b]$, and let R be the region that is bounded above $y = f(x)$, below by the x-axis, and on the sides by the lines $x = a$ and $x = b$. Find the volume of the solid of revolution that is generated by revolving the region R about the x-axis.■

We can solve this problem by slicing (see Figure 5.30). For this purpose, observe that the cross section of the solid taken perpendicular to the x-axis at the point x is a circular disk of radius $f(x)$. The area of this region is

$$A(x) = \pi\,[f(x)]^2. \tag{5.113}$$

Thus, the volume of the solid is

$$V = \int_a^b \pi\,[f(x)]^2\,dx. \tag{5.114}$$

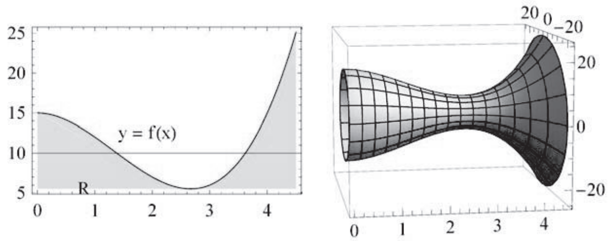


Figure 5.30. Cross section of a body generated by cubic polynomial which is rotated around the *x*-axis. The resulting solid body is shown in the bottom of the figure.

Because the cross sections are disc shaped, the application of this formula is called the method of disks.

Example 5.35. Volume by Disks I

Find the volume of the solid that is obtained when the region under the curve $y = \sqrt{x}$ over the interval [1, 4] is revolved about the x-axis.

Solution 5.35. Applying the volume formula, we find

$$V = \int_a^b \pi[f(x)]^2\,dx = \pi\int_a^b \left(\sqrt{x}\right)^2 dx = \pi\int_a^b x\,dx = \frac{\pi}{2}\,x^2\,\Big|_{a=1}^{b=4} = \pi\left(8 - \frac{1}{2}\right) = \frac{15\,\pi}{2}.\blacktriangle$$

Example 5.36. Volume by Disks II

Derive the formula of the volume of a sphere of radius r.

Solution 5.36. A sphere of radius r can be generated by revolving the upper semicircular disk enclosed between the x-axis and

$$x^2 + y^2 = r^2$$

about the x-axis. Since the upper half of this circle is the graph of $y = f(x) = \sqrt{r^2 - x^2}$, it follows that the volume of the sphere is

$$V = \int_a^b \pi[f(x)]^2\,dx = \int_{-r}^{r} \pi\left[\sqrt{r^2 - x^2}\,\right]^2 dx = \pi\int_{-r}^{r}\left(r^2 - x^2\right)dx = \pi\left[r^2\,x - \frac{x^3}{3}\right]_{-r}^{r} = \frac{4}{3}\,\pi\,r^3.\blacktriangle$$

Not all solids of revolution have solid interiors; some have holes or channels that create interior surfaces. Thus we will be interested in problems of the following type.

Theorem 5.19. *Slicing Problem*

Let f and g be continuous and non negative on $[a, b]$, and suppose that $f(x) \geq g(x)$ for all x in the interval $[a, b]$. Let R be the region that is bounded above by $y = f(x)$, below by $y = g(x)$, and on both sides by the lines $x = a$ and $x = b$. Find the volume of the solid of revolution that is generated by revolving the region R about the x-axis. ■

We can solve this problem by slicing (see Figure 5.31). For this purpose, observe that the cross section of the solid taken perpendicular to the x-axis at the point x is the annular region with inner radius $g(x)$ and outer radius $f(x)$; hence its area is

$$A(x) = \pi\left([f(x)]^2 - [g(x)]^2\right).$$

Thus the volume of the solid is

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

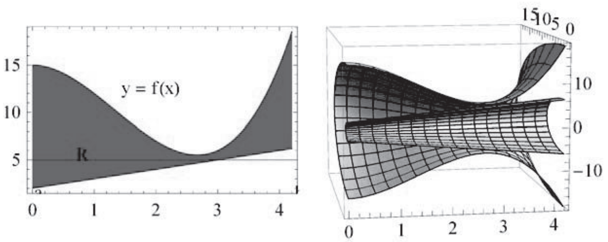


Figure 5.31. Solid cross section of a body with a inner and outer surface. The top figure shows the solid cross section. The bottom figure displays the outer and inner surface of the solid body. Note that the two surfaces are shown in a section through the body. The outer surface is represented by a cubic polynomial while the inner surface is a linear polynomial.

Because the cross section are annular shaped, the application of this formula is called the method of rings.

Example 5.37. Volume by Disks III

Find the volume of the solid generated when the region between the graphs of the equations $f(x) = \frac{1}{2} + x^2$ and $g(x) = x$ over the interval $[0, 2]$ is revolved about the x-axis.

Solution 5.37. The volume is

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

$$\int_0^2 \pi\left(\left[\frac{1}{2} + x^2\right]^2 - x^2\right) dx = \int_0^2 \pi\left(\frac{1}{4} + x^4\right) dx = \pi\left[\frac{x}{4} + \frac{x^5}{5}\right]_0^2 = \frac{69}{10}\pi. \blacktriangle$$

The methods of disks and rings have analogs for regions that are revolved about the y-axis. Using the method of slicing you should have no trouble deducing the following formulas for volumes of the solids.

$$V = \int_c^d \pi [u(y)]^2 \, dy \qquad \text{for solids}$$

and

$$V \int_c^d \pi\left([w(y)]^2 - [v(y)]^2\right) dy \qquad \text{for annular rings.}$$

Example 5.38. Volume by Disks IV

Find the volume of the solid generated when the region enclosed by $y = \sqrt{x}$, $y = 2$, and $x = 0$ is revolved about the y-axis.

Solution 5.38. The cross section taken perpendicular to the y-axis are disks, so we will apply the formula from above. But first we must rewrite $y = \sqrt{x}$ as $x = y^2$. Thus, with $u(y) = y^2$ the volume is

$$V = \int_c^d \pi [u(y)]^2 \, dy = \int_0^2 \pi \, y^4 \, dy = \pi \frac{y^5}{5} \Big|_0^2 = \frac{32 \pi}{5} . \blacktriangle$$

5.5.3 Length of a Plane Curve

In this section we will consider the problem of finding the length of a plane curve.

Although formulas for length of circular arcs appear in early historical records, very little was known about the lengths of more general curves until the mid-seventeenth century. About that time formulas were discovered for a few specific curves such as the length of an arch of a cycloid. However, such basic problems as finding the length of an ellipse defied the mathematicians of that period, and almost no progress was made on the general problem of finding lengths of curves until the advent of calculus in the next century.

5.5.3.1 Regular Curves

Our first objective in this section is to define what we mean by the length, also called the arc length, of a plane curve $y = f(x)$ over an interval $[a, b]$. Once that is done we will be able to focus on computational matters. To avoid some complications that would otherwise occur, we will impose the requirement that f' be continuous on $[a, b]$, in which case we will say that $y = f(x)$ is a smooth curve on $[a, b]$.

We will be concerned with the following problem:

Theorem 5.20. *Arc length problem*

Suppose that $y = f(x)$ is a smooth curve on the interval $[a, b]$. Define and find a formula for the arc length L of the curve $y = f(x)$ over the interval $[a, b]$. ■

The basic idea for defining arc length is to break up the curve into small segments, approximate the curve segments by line segments, add the lengths of the line segments to form a Riemann sum that approximates the arc length L, and take the limit of the Riemann sums to obtain an integral for L. To implement this idea, divide the interval $[a, b]$ into n sub intervals by inserting numbers x_1, x_2, …, x_{n-1} between $a = x_0$ and $b = x_n$. Let P_0, P_1, …, P_n be the points on the curve with x-coordinates $a = x_0, x_1, x_2, \ldots, b = x_n$ and join these points on the curve with straight line segments (see Figure 5.32). These line segments form a polygonal path that we can regard as an approximation to the curve $y = f(x)$.

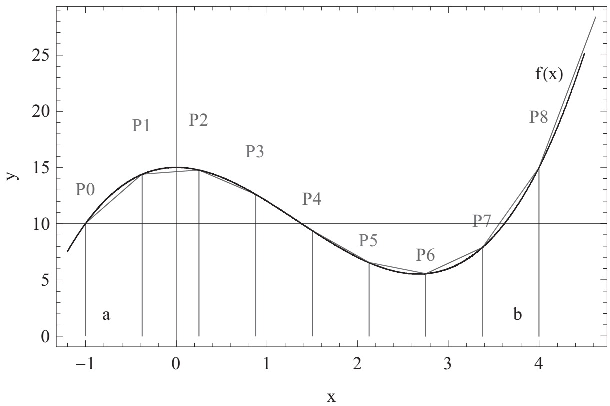


Figure 5.32. Approximation of the length of a curve by line segments connecting the points P_k with P_{k+1}.

The length L_k of the kth line segment in the polygonal path is

$$L_k = \sqrt{(\Delta x_k)^2 + \left(\Delta y_k\right)^2} = \sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2}\,. \tag{5.115}$$

If we now add the lengths of these line segments, we obtain the following approximation to the length L of the curve

$$L \approx \sum_{k=1}^{n} L_k = \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2}\,. \tag{5.116}$$

To put this in the form of a Riemann sum we will apply the Mean-Value Theorem. This theorem implies that there is a number x_k^* between x_{k-1} and x_k such that

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} = f'(x_k^*) \qquad \text{or} \qquad f(x_k) - f(x_{k-1}) = f'(x_k^*)\,\Delta x_k \tag{5.117}$$

and hence we can rewrite the formula for L_k as

$$L \approx \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + (f'(x_k^*)\,\Delta x_k)^2} = \sum_{k=1}^{n} \Delta x_k \sqrt{1 + (f'(x_k^*))^2}\,. \tag{5.118}$$

Thus, taking the limit as n increases and the widths of the sub intervals approach zero yields the following integral that defines the arc length L:

$$L = \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} \Delta x_k \sqrt{1 + (f'(x_k^*))^2} = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx. \tag{5.119}$$

In summary, we have the following definition:

Definition 5.9. *Arc Length*

If $y = f(x)$ is a smooth curve on the interval $[a, b]$, then the arc length L of this curve over $[a, b]$ is defined as

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx. \blacksquare \tag{5.120}$$

This result provides both a definition and a formula for computing arc lengths. Where convenient (5.120) can also be expressed as

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{5.121}$$

Moreover, for a curve expressed in the form $x = g(y)$, where g' is continuous on $[c, d]$, the arc length L from $y = c$ to $y = d$ can be expressed as

$$L = \int_c^d \sqrt{1 + [g'(y)]^2}\, dy = \int_c^d \sqrt{1 + \left(\frac{dg}{dy}\right)^2}\, dy. \tag{5.122}$$

Example 5.39. Arc Length

Find the arc length of the curve $y = x^{3/2}$ from $(1, 1)$ to $\left(2, 2\sqrt{2}\right)$.

Solution 5.39. Since the derivative of this function is

$$\mathbf{der} = \frac{\partial x^{3/2}}{\partial x}$$

$$\frac{3\sqrt{x}}{2}$$

and since the curve extends from $x = 1$ to $x = 2$, it follows that the length is given by

$$\boldsymbol{L} = \int_1^2 \sqrt{\mathbf{der}^2 + 1}\, dx$$

$$\frac{1}{27}\left(22\sqrt{22} - 13\sqrt{13}\right)$$

which is approximately

N[L]

2.08581

Apply to the same problem the other formula using the integration with respect to y as an exercise. The result should be the same. Why?▲

5.5.3.2 Parametric Curves

The following theorem provides a formula for finding the arc length of a curve from parametric equations for the curve. Its derivation is similar to that of formula (5.120) and will be omitted.

Theorem 5.21. *Arc Length for Parametric Curves*

If no segment of the curve represented by the parametric equations

$$x = x(t), \qquad y = y(t) \qquad \text{with } a \leq t \leq b$$

is traced more than once as t increases from a to b, and if dx / dt and dy / dt are continuous functions for $a \leq t \leq b$, then the arc length L of the curve is given by

$$L = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \blacksquare \tag{5.123}$$

Remark 5.7. Note that formulas (5.121) and (5.122) are special cases of (5.123). For example, formula (5.121) can be obtained from (5.122) by writing $y = f(x)$ parametrically as $x = t$, $y = f(t)$; similarly, formula (5.122) can be obtained from (5.123) by writing $x = g(y)$ parametrically as $x = g(t)$, $y = t$. We leave the details as exercises.

Example 5.40. Arc Length of a Parametric Curve

Use (5.123) to find the circumference of a circle of radius a from the parametric equations

$$x = r\cos(t), \qquad y = r\sin(t) \qquad \text{with } 0 \leq t \leq 2\pi.$$

Solution 5.40. The arc length is calculated by applying formula (5.123) as

$$\boldsymbol{L} = \int_0^{2\pi} \sqrt{\left(\frac{\partial(r\sin(t))}{\partial t}\right)^2 + \left(\frac{\partial(r\cos(t))}{\partial t}\right)^2}\, dt$$

$$2\pi\sqrt{r^2}$$

which is equivalent to $L = 2\pi r$.▲

As a rule the integrals that arise in calculating arc length tend to be impossible to evaluate in terms of elementary functions, so it will often be necessary to approximate the integral using numeric methods. The examples given are somehow exceptions of the generic case of arc length calculations.

5.5.4 Area of a Surface of Revolution

In this section we will consider the problem of finding the area of a surface that is generated by revolving a plane curve about a line.

A surface of revolution is a surface that is generated by revolving a plane curve about an axis that lies in the same plane as the curve. For example, the surface of a sphere can be generated by revolving a semicircle about its diameter, and the lateral surface of a right circular cylinder can be generated be revolving a line segment about an axis that is parallel to it.

In this section we will be concerned with the following problem:

Theorem 5.22. *Surface Area Problem*

Suppose that f is a smooth, non negative function on $[a, b]$ and that a surface of revolution is generated by revolving the portion of the curve $y = f(x)$ between $x = a$ and $x = b$ about the x-axis. Define what is meant by the area S of a surface, and find a formula for computing it.■

To motivate an appropriate definition for the area S of a surface of revolution, we will decompose the surface into small sections whose area can be approximated by elementary formulas, add the approximations of the areas of the sections to form a Riemann sum that approximates S, and then take the limit of the Riemann sums to obtain an integral for the exact value of S.

To implement this idea, divide the interval $[a, b]$ into n sub intervals by inserting numbers x_1, x_2, …, x_{n-1} between $a = x_0$ and $b = x_n$. The corresponding points on the graph of f define a polygonal path that approximates the curve $y = f(x)$ over the interval $[a, b]$. When this polygonal path is revolved about the x-axis, it generates a surface consisting of n parts, each of which is a frustum of a right circular cone. Thus, the area of each part of the approximating surface can be obtained from the formula

$$S = \pi(r_1 + r_2)\,\ell$$

for the lateral area S of a frustum of slant height ℓ and base radius r_1 and r_2 (Figure 5.33). As suggested by Figure 5.33, the kth frustum has radii $f(x_{k-1})$ and $f(x_k)$ and height Δx_k.

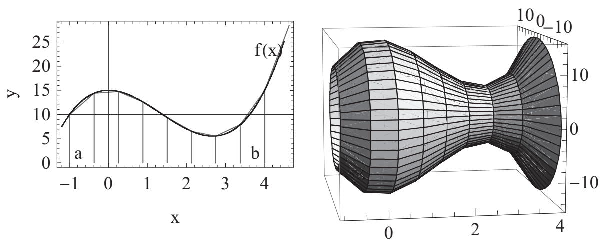


Figure 5.33. Approximation of the surface by frustums of slant height ℓ. Top the generating surface line is approximated by line elements. Bottom the surface is shown for the approximating frustums.

Its slant height is the length L_k of the kth line segment in the polygonal path, which from formula (5.115) is

$$L_k = \sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2}\,. \tag{5.124}$$

Thus the lateral area S_k of the kth frustum is

$$S_k = \pi\,[f(x_{k-1}) + f(x_k)]\,\sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2}\,. \tag{5.125}$$

If we add these areas, we obtain the following approximation to the area S of the entire surface:

$$S \approx \sum_{k=1}^{n} \pi\,[f(x_{k-1}) + f(x_k)]\,\sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2}\,. \tag{5.126}$$

To put this in the form of a Riemann sum, we will apply the Mean-Value Theorem. This theorem implies that there is a number x_k^* between x_{k-1} and x_k such that

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} = f'(x_k^*) \qquad \text{or} \qquad f(x_k) - f(x_{k-1}) = f'(x_k^*)\,\Delta x_k \tag{5.127}$$

and hence we can rewrite (5.126) as

$$S \approx \sum_{k=1}^{n} \pi\,[f(x_{k-1}) + f(x_k)]\,\Delta x_k\,\sqrt{1 + (f'(x_k^*))^2}\,. \tag{5.128}$$

However, this is not yet a Riemann sum because it involves the variables x_{k-1} and x_k. To eliminate these variables from the expression, observe that the average value of the numbers $f(x_{k-1})$ and $f(x_k)$ lies between these numbers, so the continuity of f and the Intermediate-Value Theorem imply that there is a number x_k^{**} between x_{k-1} and x_k such that

$$\frac{1}{2}[f(x_{k-1}) + f(x_k)] = f(x_k^{**}). \tag{5.129}$$

Thus, (5.128) can be expressed as

$$S \approx \sum_{k=1}^{n} 2\,\pi\,f(x_k^{**})\,\Delta x_k\,\sqrt{1 + (f'(x_k^*))^2}\,. \tag{5.130}$$

Although this expression is close to a Riemann sum in form, it is not a true Riemann sum because it involves x_k^* and x_k^{**}, rather than x_k^* alone. However, it is proved in advanced calculus courses that this has no effect on the limit because of the continuity of f. Thus, we can assume that $x_k^{**} = x_k^*$ when taking the limit, and this suggests that S can be defined as

$$S \approx \lim_{\max(\Delta x_k)\to 0} \sum_{k=1}^{n} 2\,\pi\,f(x_k^{**})\,\Delta x_k\,\sqrt{1 + (f'(x_k^*))^2} = \int_a^b 2\,\pi\,f(x)\,\sqrt{1 + (f'(x))^2}\,dx. \tag{5.131}$$

In summary, we have the following definition:

Theorem 5.23. *Area of a Surface of Revolution*

If f is a smooth non negative function on $[a, b]$, then the surface area S of the surface of revolution that is generated by revolving the portion of the curve $y = f(x)$ between $x = a$ and $x = b$

about the x-axis is defined as

$$S = \int_a^b 2\pi f(x) \sqrt{1 + (f'(x))^2}\, dx. \blacksquare$$

This result provides both a definition and a formula for computing surface areas. Where convenient, this formula can also be expressed as

$$S = \int_a^b 2\pi f(x) \sqrt{1 + (f'(x))^2}\, dx = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{5.132}$$

Moreover, if g is non negative and $x = g(y)$ is a smooth curve on the interval $[c, d]$, then the area of the surface that is generated by revolving the portion of a curve $x = g(y)$ between $y = c$ and $y = d$ about the y-axis can be expressed as

$$S = \int_c^d 2\pi g(y) \sqrt{1 + (g'(y))^2}\, dy = \int_c^d 2\pi x \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy. \tag{5.133}$$

Example 5.41. Surface of Revolution

Find the area of the surface that is generated by revolving the portion of the curve $y = x^3$ between $x = 0$ and $x = 1$ about the x-axis.

Solution 5.41. Since $y = x^3$, we have $dy/dx = 3x^2$, and hence the surface area S is

$$S = \int_a^b 2\pi f(x) \sqrt{1 + (f'(x))^2}\, dx = \int_0^1 2\pi x^3 \sqrt{1 + (3x^2)^2}\, dx = \int_0^1 2\pi x^3 \sqrt{1 + 9x^4}\, dx =$$

$$\frac{2\pi}{36} \int_1^{10} u^{1/2}\, du = \frac{2\pi}{36} \frac{2}{3} u^{3/2} \big|_1^{10} = \frac{\pi}{27} \left(10^{3/2} - 1\right) = 3.56312. \blacktriangle$$

5.5.5 Numerical Integration

Our usual procedure for evaluating a definite integral is to find an anti derivative of the integrand and apply the Fundamental Theorem of Calculus. However, if an anti derivative of the integral cannot be found, then we must settle for a numerical approximation of the integral. In the last sections we discussed procedures for approximating areas using Riemann sums. In this section we will adapt those ideas to approximate general definite integrals, and we will discuss some new approximation methods that often provide more accuracy with less computation.

5.5.5.1 Approximations by Riemann Sums

Recall from Section 5.2.3 that the definite integral of a continuous function f over an interval $[a, b]$ may be computed as

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\, \Delta x \tag{5.134}$$

where the sum that appears on the right side is called a Riemann sum. In this formula, the interval $[a, b]$ is divided into n sub intervals of width $\Delta x = (b-a)/n$, and x_k^* denotes an arbitrary point in the kth sub interval. It follows that as n increases the Riemann sum will eventually be a good approximation to the integral, which we denote by writing

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{n} f(x_k^*)\,\Delta x \tag{5.135}$$

or equivalently

$$\int_a^b f(x)\,dx \approx \Delta x\left[f(x_1^*) + f(x_2^*) + \ldots + f(x_n^*)\right]. \tag{5.136}$$

In this section we will denote the values of f at the endpoints of the sub intervals by

$$y_0 = f(a),\ \ y_1 = f(x_1),\ \ y_2 = f(x_2),\ \ldots,\ \ y_{n-1} = f(x_{n-1}),\ \ y_b = f(b) \tag{5.137}$$

and will denote the values of f at the midpoints of the sub intervals by

$$y_{m_1},\ \ y_{m_2},\ \ \ldots,\ \ y_{m_n} \tag{5.138}$$

(see Figure 5.34).

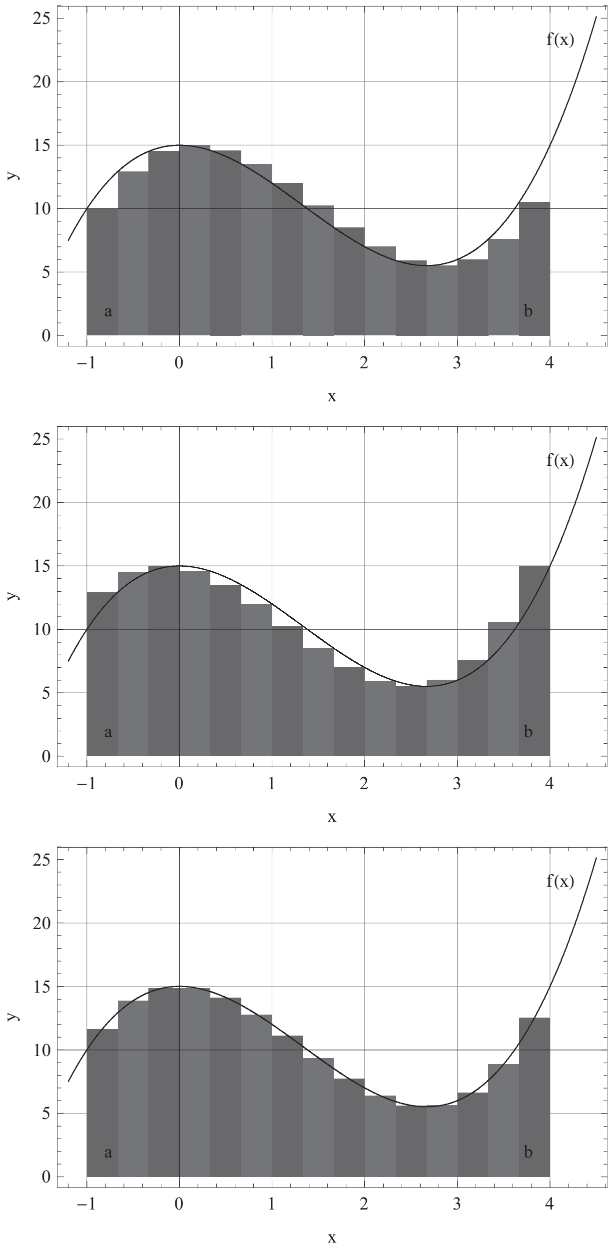
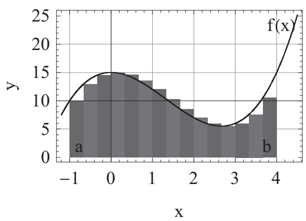
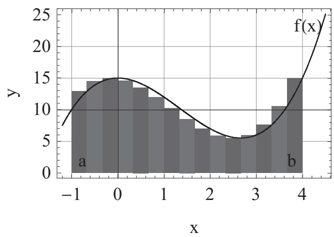
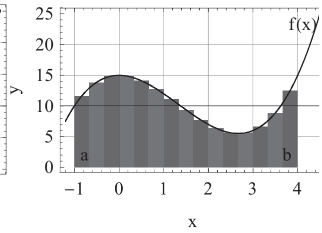


Figure 5.34. Different approximation methods to choose the value x_k^* as left-, right-, and midpoint in the interval $[x_{k-1}, x_k]$.

With this notation the left endpoint, right endpoint, and midpoint approximations can be expressed as shown in Table 5.2.

Table 5.2. Formulas to approximate a definite integral by left, right, and midpoint approximations

Left endpoint approximation	*Right endpoint approximation*	*Midpoint approximation*
$\int_a^b f(x)\,dx = \frac{(b-a)}{n}[y_0 + y_1 + \ldots + y_{n-1}]$	$\int_a^b f(x)\,dx = \frac{(b-a)}{n}[y_1 + y_2 + \ldots + y_n]$	$\int_a^b f(x)\,dx = \frac{(b-a)}{n}[y_{m_1} + y_{m_2} + \ldots + y_{m_n}]$
		

Ways of integral approximation.

5.5.5.2 Trapezoidal Approximation

The left-hand and right-hand endpoint approximations are rarely used in approximations; however, if we take the average of the left-hand and right-hand endpoint approximations, we obtain a result, called the trapezoidal approximation, which is commonly used

$$\int_a^b f(x)\,dx = \frac{(b-a)}{2\,n}[y_0 + 2\,y_1 + 2\,y_2 + \ldots + 2\,y_{n-1} + y_n] \tag{5.139}$$

The name trapezoidal approximation can be explained by considering the case in which $f(x) \geq 0$ on $[a, b]$, so that $\int_a^b f(x)\,dx$ represents the area under $f(x)$ over $[a, b]$. Geometrically, the trapezoidal approximation formula results if we approximate this area by the sum of the trapezoidal areas shown in Figure 5.35.

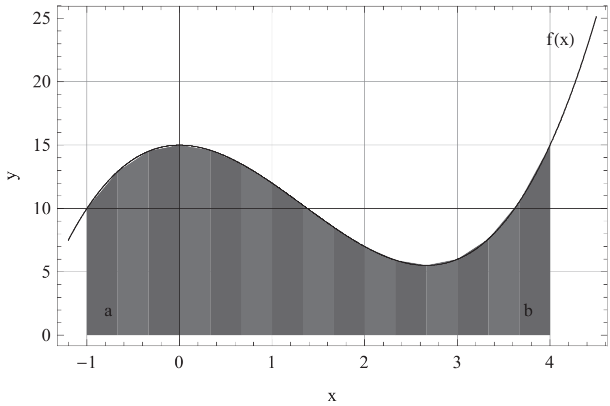


Figure 5.35. Trapezoidal approximation of a definite integral.

Example 5.42. Numerical Integration

Calculate the numerical value for ln(2) by using a definite integration. Compare the trapezoidal with the midpoint approximation.

Solution 5.42. Applying the trapezoidal rule we find the following approximation for the integral $\int_1^2 \frac{1}{x}\, dx$.

The following function implements the trapezoidal rule for a numerical integration in *Mathematica.* The first three arguments are the lower and upper boundary of the interval $[a, b]$ and the number of sub partitions n of this interval. The last argument represents the function used in the integrand.

```
trapezoidalApproximation[a_, b_, n_, f_] :=
  (b - a)/(2 n) (f[a] + f[b] + Fold[Plus, 0, Table[2. f[a + (b - a)/n i], {i, 1, n - 1}]])
```

The application to the function $\frac{1}{x} = \left(\frac{1}{\#}\right)$ & is given in the next line.

```
trapezoidalApproximation[1, 2, 6, (1/#) &]
```

```
0.694877
```

A table with different sub partitions of the interval is generated in the next line where i counts the number of sub partitions

```
TableForm[Table[{i, trapezoidalApproximation[1, 2, i, (1/#) &]},
  {i, 1, 10}], TableHeadings → {None, {"i", "A(i)"}}]
```

i	A(i)
1	$\frac{3}{4}$
2	0.708333
3	0.7
4	0.697024
5	0.695635
6	0.694877
7	0.694419
8	0.694122
9	0.693918
10	0.693771

It is obvious from the table that the value for ln(2) approaches some value near 0.693 .... which can be verified by using the *Mathematica* function NIntegrate[] for numerical integration.

NIntegrate$\left[\frac{1}{x}, \{x, 1, 2\}\right]$

0.693147

The approximation with the trapezoidal formula can be improved by increasing the number of partitions such as

```
trapezoidalApproximation[1, 2, 250, (1/#) &]
```

```
0.693148
```

With 250 partitions we reach a 5 digit accuracy of the result.

The midpoint approximation is defined by the following function

```
midpointApproximation[a_, b_, n_, f_] := Block[{Δx}, Δx = N[(b - a)/n];
  Δx (Fold[Plus, 0, Table[f[a + Δx (i + 1) - Δx/2], {i, 0, n - 1}]])]
```

Applying the midpoint formula to the function $1/x = (1/\#)$ & delivers for 10 partitions the value

```
midpointApproximation[1, 2, 10, (1/#) &]
```

```
0.692835
```

The following table gives the results for different partitions of the interval [1, 2]

```
TableForm[Table[{i, midpointApproximation[1, 2, i, (1/#) &]}, {i, 1, 10}],
 TableHeadings → {None, {"i", "A(i)"}}]
```

i	A(i)
1	0.666667
2	0.685714
3	0.689755
4	0.69122
5	0.691908
6	0.692284
7	0.692512
8	0.692661
9	0.692762
10	0.692835

Comparing both approximations we observe that the trapezoidal method approaches the "true" value from above and the midpoint method from below.

```
midpointApproximation[1, 2, 250, (1/#) &]
```

```
0.693147
```

An accuracy of 6 digits is react with 250 subdivisions of the total interval.▲

The value of ln(2) rounded to nine decimal places is

```
N[Log[2], 9]
```

```
0.693147181
```

so that the midpoint approximation in the example above produced a more accurate result than the trapezoidal approximation. To see why this should be so, we need to look at the midpoint approximation from another viewpoint. For differentiable functions, the midpoint approximation is sometimes called the tangent line approximation because over each sub interval the area of the rectangle used in the midpoint approximation is equal to the area of the trapezoidal whose upper boundary is the tangent line to $y = f(x)$ at the midpoint of the interval (Figure 5.36).

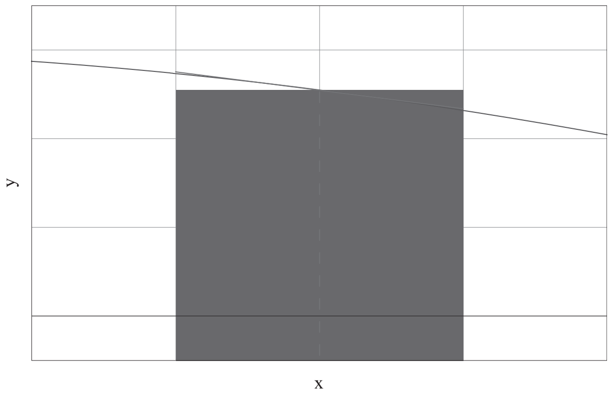


Figure 5.36. Approximation of the area below a function $f(x_k)$ by a rectangle and a tangent line.

The equality of these areas follows from the fact that the related triangles of the tangent are congruent.

In the following we will denote the midpoint and trapezoid approximations of $\int_a^b f(x)\,dx$ with n sub intervals by M_n and T_n, respectively, and we will denote the errors in these approximations by

$$\left|E_M\right| = \left|\int_a^b f(x)\,dx - M_n\right| \qquad \text{and} \qquad \left|E_T\right| = \left|\int_a^b f(x)\,dx - T_n\right|.$$

The following Theorem summarizes our observations found in the example above:

Theorem 5.24. *Errors for Midpoint and Trapezoidal Approximations*

Let f be continuous on $[a, b]$, and let $|E_M|$ and $|E_T|$ be the absolute error that result from the midpoint and trapezoidal approximation of $\int_a^b f(x)\,dx$ using n sub intervals.

a) If the graph of f is either concave up or concave down on (a, b), then $|E_M| < |E_T|$, that is the error from the midpoint approximation is less than that from the trapezoidal approximation.
b) If the graph of f is concave down on (a, b) then

$$T_n < \int_a^b f(x)\,dx < M_n.$$

c) If the graph of f is concave up on (a, b), then

$$M_n < \int_a^b f(x)\,dx < T_n.\blacksquare$$

Example 5.43. Comparison of concave up and down functions

Compare the numeric integrals for ln(2) and sin(1) by using the two different numerical methods.

Solution 5.43. Applying the two functions defined above shows the behavior given in the theorem for a concave down

```
TableForm[Table[{i, midpointApproximation[1, 2, i, (1/#) &],
   trapezoidalApproximation[1, 2, i, (1/#) &],
   Abs[midpointApproximation[1, 2, i, (1/#) &] -
     trapezoidalApproximation[1, 2, i, (1/#) &]]}, {i, 1, 10}],
 TableHeadings → {None, {"i", "A_M(i)", "A_T(i)", "|A_M(i)-A_T(i)|"}}]
```

i	$A_M(i)$	$A_T(i)$	$\lvert A_M(i)-A_T(i)\rvert$
1	0.666667	$\frac{3}{4}$	0.0833333
2	0.685714	0.708333	0.022619
3	0.689755	0.7	0.0102453
4	0.69122	0.697024	0.00580392
5	0.691908	0.695635	0.00372703
6	0.692284	0.694877	0.00259302
7	0.692512	0.694419	0.00190723
8	0.692661	0.694122	0.0014613
9	0.692762	0.693918	0.00115519
10	0.692835	0.693771	0.000936043

and a concave up function

```
TableForm[Table[{i, midpointApproximation[0, 1, i, (Cos[#]) &],
   trapezoidalApproximation[0, 1, i, (Cos[#]) &],
   Abs[midpointApproximation[0, 1, i, (Cos[#]) &] -
     trapezoidalApproximation[0, 1, i, (Cos[#]) &]]}, {i, 1, 10}],
 TableHeadings → {None, {"i", "A_M(i)", "A_T(i)", "|A_M(i)-A_T(i)|"}}]
```

i	$A_M(i)$	$A_T(i)$	$\lvert A_M(i)-A_T(i)\rvert$
1	0.877583	$\frac{1}{2}$ (1 + Cos[1])	0.107431
2	0.850301	0.823867	0.0264338
3	0.845379	0.833665	0.0117142
4	0.843666	0.837084	0.00658257
5	0.842875	0.838664	0.00421086
6	0.842446	0.839522	0.00292347
7	0.842187	0.840039	0.00214752
8	0.842019	0.840375	0.00164403
9	0.841904	0.840605	0.0012989
10	0.841822	0.84077	0.00105206

▲

Warning: Do not conclude from the above example that the midpoint approximation is always better than the trapezoidal approximation; for some values of *n*, the trapezoidal approximation can be more accurate over an interval on which the function changes concavity.

Over an interval on which the integrand does not change concavity Theorem 5.14 guarantees that a definite integral is better approximated by the midpoint approximation than by the trapezoidal approximation and that the value of the definite integral lies between these two approximations. The numerical evidence reveals that $E_T \approx -2\,E_M$ is satisfied for the errors. This suggests that

$$3\int_a^b f(x)\,dx = 2\int_a^b f(x)\,dx + \int_a^b f(x)\,dx = \\ 2\,(M_n + E_M) + (T_n + E_T) = (2\,M_n + T_n) + (2\,E_M + E_T) \approx 2\,M_n + T_n. \tag{5.140}$$

That is

$$\int_a^b f(x)\,dx \approx \frac{1}{3}\,(2\,M_n + T_n) \tag{5.141}$$

Using the midpoint and trapezoidal approximation formulas in Table 5.2, we can derive a formula for this kind of approximation. For convenience, we partition the interval $[a, b]$ into $2\,n$ sub intervals, each of length $(b-a)/(2\,n)$. As before, label the endpoints of these sub intervals by $a = x_0,\ x_1,\ x_2, \ldots,\ x_{2\,n} = b$. Then $x_0,\ x_2,\ x_4, \ldots,\ x_{2\,n}$ define a partition of $[a, b]$ into n equal sub intervals, and the midpoints of these sub intervals are $x_1,\ x_3,\ x_5,\ \ldots,\ x_{2\,n-1}$, respectively. Using $y_i = f(x_i)$, we have

$$M_n = \frac{(b-a)}{n}[y_1 + y_3 + \ldots + y_{2\,n-1}] = \frac{b-a}{2\,n}[2\,y_1 + 2\,y_3 + \ldots + 2\,y_{2\,n-1}] \tag{5.142}$$

$$T_n = \frac{(b-a)}{2\,n}[y_0 + 2\,y_2 + 2\,y_4 + \ldots + y_{2\,n-2} + y_{2\,n}]. \tag{5.143}$$

Now define $S_{2\,n}$ by

$$S_{2\,n} = \\ \frac{1}{3}\,(2\,M_n + T_n) = \frac{1}{3}\left(\frac{b-a}{2\,n}\right)[y_0 + 4\,y_1 + 2\,y_2 + 4\,y_3 + \ldots + 2\,y_{2\,n-2} + 4\,y_{2\,n-1} + y_{2\,n}]. \tag{5.144}$$

The approximation

$$\int_a^b f(x)\,dx \approx S_{2\,n} \tag{5.145}$$

as given in (5.144) is known as **Simpson's rule**. We denote the absolute error in this approximation by

$$\left|E_S\right| = \left|\int_a^b f(x)\,dx - S_{2\,n}\right|. \tag{5.146}$$

Having the functions for the midpoint and the trapezoidal approximation available it is now easy to introduce a function for Simpson's rule

```
simpsonApproximation[a_, b_, n_, f_] :=
 1/3 (2 midpointApproximation[a, b, n, f] +
    trapezoidalApproximation[a, b, n, f])
```

The application on the two examples discussed above give for ln(2)

```
simpsonApproximation[1, 2, 10, (1/#) &]
```

```
0.693147
```

and for sin(1) we find

```
simpsonApproximation[0, 1, 10, (Cos[#]) &]
```

```
0.841471
```

The geometric background of Simpson's rule is the generalization of both the midpoint and trapezoidal approximation for a definite integral by replacing the linear segments by a quadratic function $y = A x^2 + B x + C$, thus capturing some sense of the concavity of the function.

5.5.6 Tests and Exercises

The following two subsections serve to test your understanding of the last section. Work first on the test examples and then try to solve the exercises.

5.5.6.1 Test Problems

T1. How do you define and calculate the area of the region between the graphs of two continuous functions? Give an example.

T2. How do you define and calculate the volumes of solids by the method of slicing? Give an example.

T3. Describe the method of cylindrical shells. Give an example.

T4. How do you locate the center of mass of a thin flat plate of material? Give an example.

T5. How do you find the length of the graph of a smooth function over a closed interval? Give an example. What about functions that do not have continuous first derivatives?

T6. What is a center of mass?

5.5.6.2 Exercises

E1. Find the areas of the regions enclosed by the lines and curves in the following examples:

- **a.** $f(x) = 2x - x^2$ and $f(x) = -3$,
- **b.** $f(x) = x^2 - 2$ and $f(x) = 2$,
- **c.** $f(x) = x^4$ and $f(x) = 6x$,
- **d.** $f(x) = x^2$ and $f(x) = -x^2 + 4x$,
- **e.** $f(x) = 7 - x^2$ and $f(x) = x^2 + 4$,
- **f.** $f(x) = x\sqrt{a^2 - x^2}$, $a > 0$, and $f(x) = 0$.

E2. Find the areas of the regions enclosed by the lines and curves in the following examples:

- **a.** $x = y^2$ and $x = y + 2$,
- **b.** $x = 2y^2$, $x = 0$, and $y = 3$,
- **c.** $y^2 - 4x = 4$ and $4x - y = 16$,
- **d.** $x = y^2 - 1$ and $x = |y|\sqrt{1 - y^2}$,
- **e.** $x - y^{2/3} = 0$ and $x + y^4 = 2$,
- **f.** $x = y^3 - y^2$ and $x = 2y$.

Find the area of the propeller-shaped region enclosed by the curves $x - y^{1/3} = 0$ and $x - y^{1/5} = 0$.

E4. Find the area of the “triangular” region in the first quadrant bounded on the left by the y-axis and on the right by the curves $y = \sin(x)$ and $y = \cos(x)$.

E5. Show that if f is odd on $[-a, a]$ then

$$\int_{-a}^{a} f(x)\, dx = 0. \tag{1}$$

Test the result with $f = \sin(x)$ and $a = \pi/2$.

E6. Use the shell method to find the volumes of the solids generated by revolving the regions bounded by the curves and lines in the following examples about the y-axis.

a. $y = x$, $y = -x/3$, $x = 2$,
b. $y = 2 - x^2$, $y = x^2$, $x = 0$,
c. $y = 2x - 1$, $y = \sqrt{x}$, $x = 0$,
d. $y = x^2$, $y = 2 - x$, $x = 0$, for $x \geq 0$.

E7. Use the shell method to find the volumes of the solids generated by revolving the regions bounded by the curves and lines in the following examples about the x-axis.

a. $x = \sqrt{y}$, $x = -y$, $y = 2$,
b. $x = 2y - y^2$, $x = 0$,
c. $y = x$, $y = 2x$, $y = 2$,
d. $y = \sqrt{x}$, $y = 0$, $y = 2 - x$

E8. Find the lengths of the curves in the following examples:

a. $x = 1 - t$, $y = 2 + 3t$, $-2/3 \leq t \leq 1$,
b. $x = t^3$, $y = 3t^2/2$, $0 \leq t \leq \sqrt{3}$,
c. $x = \cos(t)$, $y = t + \sin(t)$, $0 \leq t \leq \pi$,
d. $x = \frac{1}{3}(2t + 3)^{3/2}$, $y = t + t^2/2$, $0 \leq t \leq 3$,
e. $x = 8\cos(t) + 8t\sin(t)$, $y = 8\sin(t) - 8t\cos(t)$, $0 \leq t \leq \pi/2$.

E9. Find the areas of the surfaces generated by revolving the curves in the following examples about the indicated axes. If you have a CAS, you may want to graph these curves to see what they look like.

a. $y = \frac{x^3}{9}$, $0 \leq x \leq 2$; x – axis,
b. $x = 2\sqrt{4 - y}$, $0 \leq y \leq \frac{15}{4}$; y – axis,
c. $y = \sqrt{x + 1}$, $1 \leq x \leq 5$; x – axis,
d. $x = \frac{y^3}{3}$, $0 \leq y \leq 1$; y – axis.

E10 The line segment joining the origin to the point (h, r) is revolved about the x-axis to generate a cone of height h and base radius r. Find the cone’s surface area with the parametric equations $x = ht$, $y = rt$, $0 \leq t \leq 1$. Check your result with the geometry formula: Area $= \pi r$ (slant height).

A

Notations

This section collects some notations used in the book to make the *Mathematica* expressions compatible with the mathematical notation.

```
<< "BarCharts`"; << "Histograms`"; << "PieCharts`"
```

The following notation introduces a general sequence

$$\{a_n\}_{n=1}^{+\infty}. \tag{B.1}$$

To have the same notation in *Mathematica* available the following steps are needed

```
<< "Notation`"
```

Here the notation is defined as an equivalent with a table

```
Notation[{f_}^{a_}_{n_=m_} ⟺ Table[f_, {n_, m_, a_}]]
```

The corresponding alias is set her as Seq

```
AddInputAlias[{□}^{□}_{□=□}, "Seq"]
```

Options

This section defines options used in the notebooks to change the formating and style of plots.

This line switches off the messages generated by *Mathematica*

```
Off[General::"spell1"]; Off[General::"spell"]; Off[Solve::"ifun"];
```

The following line changes the global text style settings

```
$TextStyle = {FontFamily → "Arial", FontSize → 12};
```

The next line sets options for the Plot function

```
SetOptions[Plot, GridLines → Automatic,
  Frame → True, PlotStyle -> RGBColor[0.501961, 0, 0.25098]];
```

Here are options for ListPlot set

```
SetOptions[ListPlot, GridLines → Automatic,
  Frame → True, PlotStyle -> RGBColor[0.501961, 0, 0.25098]];
```

The options for ParametricPlot are changed

```
SetOptions[ParametricPlot, GridLines → Automatic,
  Frame → True, PlotStyle -> RGBColor[0.501961, 0, 0.25098]];
```

The options for FilledPlot are changed

```
SetOptions[FilledPlot, GridLines → Automatic,
  Frame → True, PlotStyle -> RGBColor[0.501961, 0, 0.25098]];
```

References

[1] Abell, M. L. & Braselton, J. P. Mathematica by example. 4th ed. (Elsevier, Amsterdam, 2009).

[2] Ayres, F. & Mendelson, E. Schaum's outlines calculus. 5th ed. (McGraw-Hill, New York, 2009).

[3] Banner, A. D. The calculus lifesaver. All the tools you need to excel at calculus (Princeton University Press, Princeton, 2007).

[4] Baumann, G. Classical mechanics and nonlinear dynamics. 2nd ed. (Springer, New York, NY, 2005).

[5] Baumann, G. Electrodynamics, quantum mechanics, general relativity, and fractals. 2nd ed. (Springer, New York, NY, 2005).

[6] Berresford, G. Brief applied calculus. 4th ed. (Houghton Mifflin Co., Boston MA, 2006).

[7] Bittinger, M. L. & Ellenbogen, D. Calculus and its applications. 9th ed. (Pearson/Addison Wesley, Boston, 2008).

[8] Bleau, B. L. Forgotten calculus. A refresher coursewith applications to economics and business and the optional use of the graphing calculator. 3rd ed. (Barron' s Educational Series, Hauppauge N.Y., 2002).

[9] Goldstein, L. J., Lay, D. C. & Schneider, D. I. Calculus & its applications (Pearson Education, Upper Saddle River NJ, 2006).

[10] Hass, J. & Weir, M. D. Thomas' calculus. Early transcendentals (Pearson/Addison Wesley, Boston, 2008).

[11] Huettenmueller, R. Precalculus demystified (McGraw-Hill, New York, 2005).

[12] Hughes-Hallett, D. Calculus. 4th ed. (J. Wiley, Hoboken N.J., 2005).

[13] Hughes-Hallett, D., Gleason, A. M. & Marks, E. J. Applied calculus. 3rd ed. (Wiley, Hoboken, NJ, 2006).

[14] Kelley, W. M. The complete idiot' s guide to calculus. 2nd ed. (Alpha, Indianapolis IN, 2006).

[15] Kelley, W. M. The humongous book of calculus problems. Translated for people who don' t speak math!! (Alpha Books, New Yoprk, NY, 2006).

[16] Larson, R. Brief calculus. An applied approach. 8th ed. (Houghton Mifflin, Boston MA, 2007).

[17] Larson, R. Calculus. An applied approach. 8th ed. (Houghton Mifflin Co., Boston MA, 2007).

[18] Larson, R. Elementary linear algebra. 6th ed. (Houghton Mifflin, Boston MA, 2008).

[19] Larson, R., Hostetler, R. P. & Edwards, B. H. Calculus with analytic geometry. 8th ed. (Houghton Mifflin, Boston, 2007).

[20] Lial, M. L., Greenwell, R. N. & Ritchey, N. P. Calculus with applications. 9th ed. (Pearson/Addison Wesley, Boston MA, 2008).

[21] Rumsey, D. Pre - calculus for dummies. 1st ed. (Wiley Pub. Inc., Indianapolis IN, 2008).

[22] Ruskeepää, H. Mathematica navigator. Mathematics, statistics, and graphics. 3rd ed. (Elsevier/Academic Press, Amsterdam, Boston, 2009).

[23] Ryan, M. Calculus for dummies (Wiley, Hoboken, 2003).

[24] Simmons, G. F. Calculus with analytic geometry. 2nd ed. (McGraw-Hill, New York, NY 1996).

[25] Stewart, J. Single variable Calculus. Concepts and contexts. 4th ed. (Brooks/Cole Cengag Learning, Belmont CA, 2009).

[26] Varberg, D. E., Purcell, E. J. & Rigdon, S. E. Calculus (Pearson Prentice Hall, Upper Saddl River N.J., 2007).

[27] Washington, A. J. Basic technical mathematics with calculus. 9th ed. (Pearson/Prentice Hall Upper Saddle River N.J., 2009).

[28] Wellin, P. R., Gaylord, R. J. & Kamin, S. N. An introduction to programming with Mathematica. 3rd ed. (Cambridge Univ. Press, Cambridge, 2005).

Index

D

E