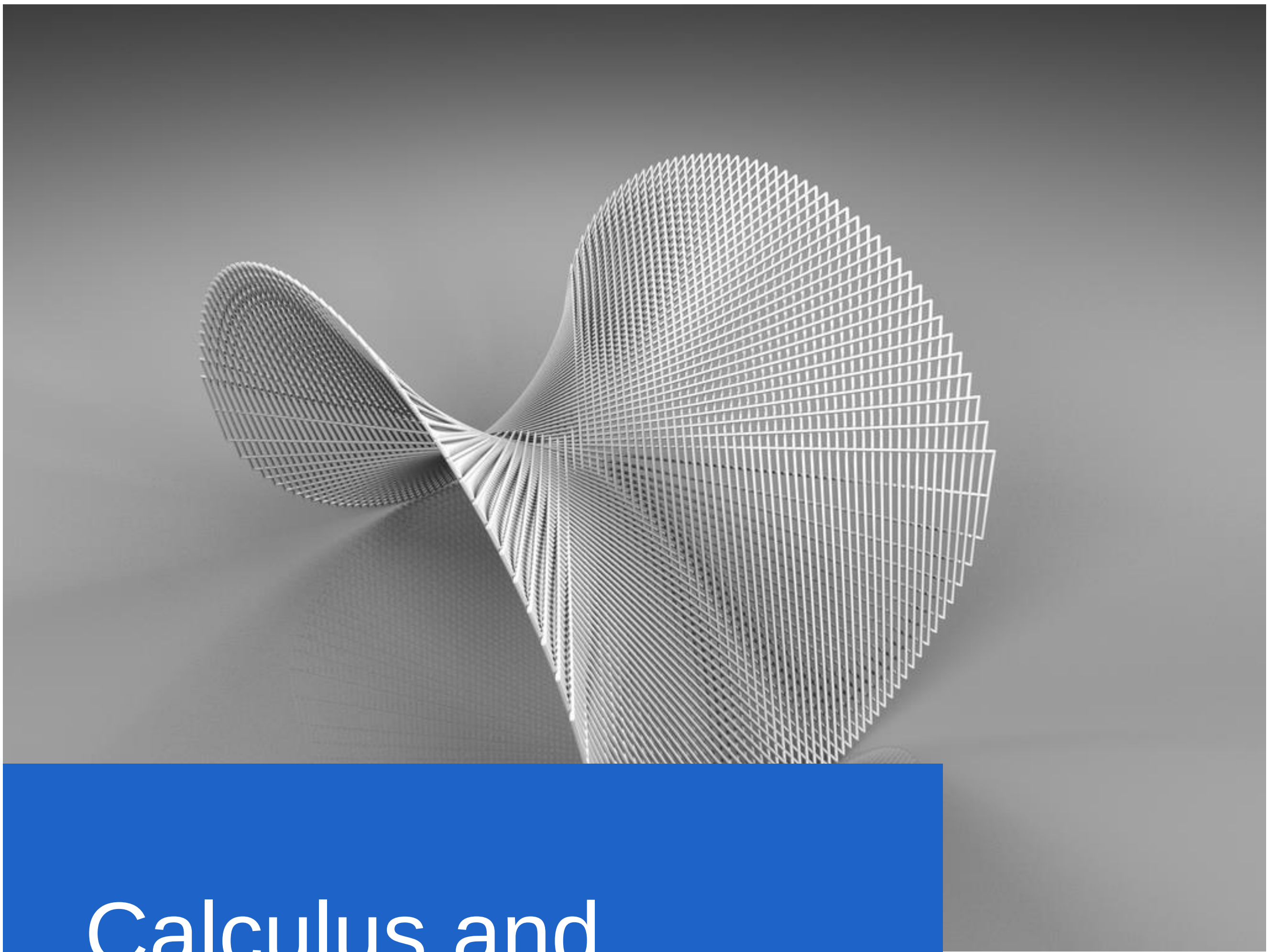
FACULTY OF
BIOSCIENCE ENGINEERING



Calculus and Analysis

Prof. dr. ir. Jan Baetens
M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

Academic year 2022–2023 – version 1

GHENT
UNIVERSITY

Calculus and Analysis

Prof. dr. ir. Jan M. Baetens

M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

The cover photo represents a hyperbolic paraboloid whose standard equation is given by

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}\,.$$

Preface

The purpose of this course is to present mathematics as the science of deductive reasoning and not as the art of manipulation. Unfortunately, many students feel mathematics is incomprehensible and is riddled with complex and abstract jargon. Our goal is to impose a lasting understanding of and appreciation for calculus on the student. Our course is intended to give the student an understanding of what calculus is truly about. It does not take more intelligence than that of a parrot to be able to go through a list of theorems and equations; but only when one understands their origins can one correctly and confidently apply them in the real world.

The over-emphasis on the calculator and foremostly the computer is definitely a point of confusion for the student. The computer is only a time-saving machine whose usefulness depends on the knowledge of the user. We do admit the computer is a remarkable machine, and we will make use of it whenever appropriate, yet it is this fascination that gives students a false sense of what they are doing. The confidence gained from all the correct answers leads to an inseparable dependence where the student is absolutely helpless without it.

Throughout the textbook we constantly refer to science and engineering. The purpose of this is to show how the scientific method applies to all disciplines and to understand that mathematics is an expression of one's observations and hypothesis. For that reason, several examples and exercises were chosen because of their relevance in reality, such that the reader can get a good feel of why and how this course is so important for future engineers. Note that because of its engineering viewpoint, we always indicate the dimensions of the used base quantities, being mass [M], time [T], temperature $[\Theta]$ and length [L]. Throughout this course the icon in the margin indicates that there's a supporting You Tube video available. The QR-code below takes you directly to the appropriate You Tube video online. At the end of every chapter one can find an extensive list of exercises linked to the topics discussed in the corresponding chapter.

Even though much time and efforts have been spent in compiling this text, it cannot be free of errors, and the authors would be grateful if these would be reported to them so that the quality of this text can be improved even further.

It goes without saying that many people have contributed to this course in addition to its authors, namely, Demir Ali Köse, Janos Coquyt, Tinne De Boeck, Diego De Gusem, Lander De Visscher, Wannes Dewulf, Jeroen Galle, Jelle Hustinx, Hanna Jaspaert, Linde Lambrecht, Robin Simoens, Ward Van Belle, Caitlin Vanden Bussche, Victor Vanthilt and Hilder Vernieuwe.

Finally, we are grateful that Ben Orlin, author of the book *Math with Bad Drawings* and the blog `https://mathwithbaddrawings.com/`, granted us permission to include his cartoons at the end of some of

the chapters.

Ghent, September 10, 2020

The authors

Contents

II Single variable calculus 237

1

Introduction

Imagine the following situation.

Sam and Alex are travelling in the car, but its speedometer is broken. Still, Alex wants to know how fast they are going, so he asks Sam. The latter says that they covered 1.2 kilometres in the last minute, so he argues that they are driving 72 km/h. Alex is not, however, satisfied with this answer because he does not want to know the average speed [LT^{-1}] for the last minute, or even the last second, rather he wants to know the speed right now. As they are approaching a road sign, Sam says that they will measure it up there. He observes that they were AT the sign for zero seconds, and the distance was zero meters, so their speed is:

$$\frac{0\text{ m}}{0\text{ s}} = \frac{0}{0}\,,$$

and he wisely says that he does not know. He argues that he needs to know some distance over some time, so keeping the time should zero cannot be done.

Actually, Alex wants to know their instantaneous speed, and this might seem pretty amazing, but it is not easy to work out the speed of a car at any point in time. Even the speedometer of a car just shows us an average of how fast we were going for the last (very short) amount of time.

Now, consider we drop a ball from the roof top terrace of the main building at Campus Ledeganck. For the sake of simplicity, we use the following simplified formula to find the distance d [L], measured in metres, fallen:

$$d = 5\,t^2\,,$$

where t [T] is time, measures in seconds. Clearly, after one second, the distance fallen is five metres, but how fast is that? We know

$$\text{speed} = \frac{\text{distance}}{\text{time}}\,,$$

so at one second we get a speed of 5 m/s, but as in the previous situation this constitutes an average speed. If we like to know the instantaneous speed, we run in exactly the same problem as before, as we get for the speed at $t = 1\text{s}$:

$$\text{speed} = \frac{0 \text{ m}}{0 \text{ s}}.$$

Let us try to circumvent this problem by inventing a time Δt so short it will not matter. Let us work out the difference in distance between t and $t+\Delta t$. At 1 second the ball has fallen

$$5\,t^2 = 5\cdot(1)^2 = 5 \text{ m}.$$

At $t+\Delta t$ seconds the ball has fallen

$$\begin{aligned} 5\,t^2 &= 5\,(1+\Delta t)^2 \text{ m}, \\ &= 5\,(1+2\Delta t+(\Delta t)^2) \text{ m}, \\ &= 5+10\Delta t+5(\Delta t)^2 \text{ m}. \end{aligned}$$

Consequently, the difference in distance between t and $t+\Delta t$ is

$$10\Delta t+5(\Delta t)^2\text{m}\,,$$

while we get the corresponding speed by dividing this change in distance by the time elapse Δt:

$$\begin{aligned} \text{speed} &= \frac{10\Delta t+5(\Delta t)^2\text{m}}{\Delta t \text{ s}}, \\ &= 10+5\Delta t\text{m/s}. \end{aligned}$$

Now if we want Δt to be so small it will not matter, we shrink it to zero and get 10 m/s.

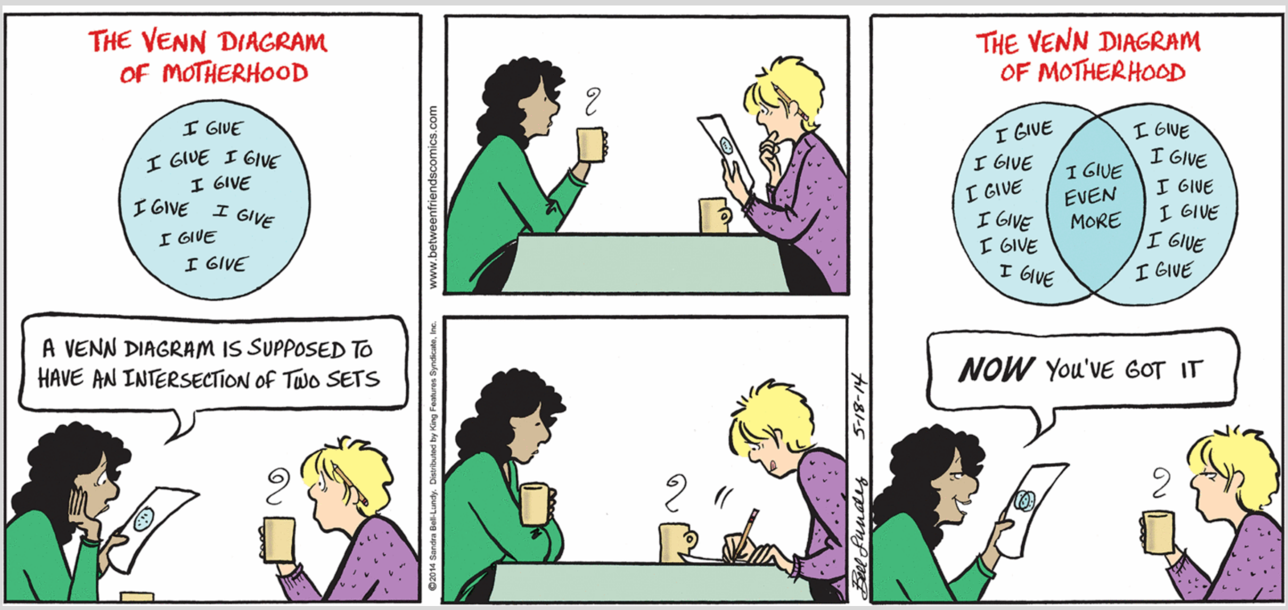
Without really paying attention to it, we just used calculus to cut time and distance into such small pieces that a pure answer came out. The fundamental idea of calculus is to study change by studying instantaneous change, by which we mean changes over tiny intervals of time. It turns out that such changes tend to a be lot simpler to analyse than changes over finite intervals of time.

The goal of this course is to get a comprehensive understanding of what calculus exactly is, and even more importantly, what we can do with it.

In Part I we present the preliminaries that one should master before even trying to move on to the study of calculus itself. The latter is the subject of Parts II and III. More precisely, Part II introduces differential and integral calculus of functions of one variable, while multivariable functions are covered in Part III.

PART I

PRECALCULUS



The only way to learn mathematics is to do mathematics.

— Paul Halmos —

2

Sets and numbers

To make sure that everyone has an understanding of the basic concepts that are at the basis of subsequent chapters, we begin this part with a brief summary of set theory and some of the associated vocabulary and notations we use throughout the text.

2.1 Sets

2.1.1 Logic operators

Throughout this course, and especially when defining new mathematical objects, we will often make use of notation that contains one or more logic operators. An overview of them is given in Table 2.1. Note the difference between $X \Rightarrow Y$ and $X \Leftrightarrow Y$. $X \Rightarrow Y$ states that if X is true, Y is also true, but if Y is true, X is not necessarily true. $X \Leftrightarrow Y$, on the other hand, states that X and Y are both true or both not true and thus equivalent. For instance, we may write

$$\text{Garfield is a cat} \ \wedge \ \text{all cats are mammals} \ \Rightarrow \ \text{Garfield is a mammal}\,,$$

though

$$\text{Garfield is a mammal} \ \not\Rightarrow \ \text{Garfield is a cat}\,!$$

Table 2.1: Overview of important logic operators. X and Y denote logic statements, which are either true or false.

Notation	Reads as
$X \Rightarrow Y$	X implies Y; if X then Y
$X \Leftrightarrow Y$	X if and only if Y
$X \wedge Y$	X and Y
$X \vee Y$	X or Y
$\forall x$	for all (elements) x
$\exists x$	there exists an (element) x
$\exists!\, x$	there exists just one (element) x
$.\vert.$	. so that .
$:.$	it holds that .

2.1.2 Definition and representation of sets

We start with a definition of a set.

Definitie 2.1 (Set)
A **set** (*verzameling*) is a well-defined collection of objects which are called the elements of the set.

In this definition, 'well-defined' means that it is possible to determine if something belongs to the collection or not, without prejudice. For example, the collection of letters that make up the word 'smolko' is well defined and is a set, but the collection of the worst math teachers in the world is not well defined, and so is not a set. In general, there are three ways to describe sets.

1. The **verbal method**: use a sentence to define a set.
2. The **roster method**: begin with a left brace '{', list each element of the set only once and then end with a right brace '}'.
3. The **set-builder method**: a combination of the verbal and roster methods.

For example, let S be the set described verbally as the set of letters that make up the word 'smolko'. A roster description of S would be $S = \{s, m, o, l, k\}$. Note that sets do not allow for repeated elements while they are also orderless, so $\{k, l, m, o, s\}$ is also a roster description of S. A set-builder description of S is:

$$S = \{x \,|\, x \text{ is a letter in the word 'smolko'}\}.$$

In this notation we call x a **dummy variable** and 'x is a letter in the word 'smolko'' the **predicate**. The way to read this is: 'The set of elements x such that x is a letter in the word 'smolko.' Clearly m is in S and q is not in S, i.e. $m \in S$ and $q \notin S$. Moreover, the empty set is written as $A = \{\}$ or $A = \emptyset$, and a set containing a single element is reffered to as a **singleton** (*singleton*).

Graphically, sets are typically represented by means of so-called **Venn diagrams** (*Venn-diagram*), enclosed areas in the plane. For instance, Figure 2.1 shows the Venn diagram of the set $S = \{s, m, o, l, k\}$.

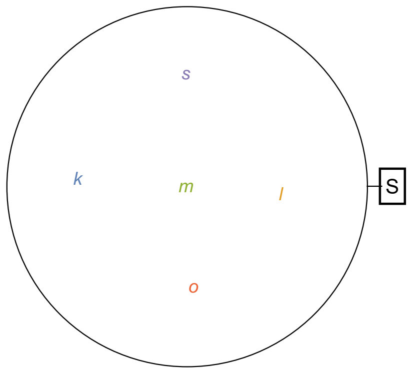


Figure 2.1: Venn diagram of the set $S = \{s, m, o, l, k\}$.

The fact that a set sensu stricto does not bear an ordering constitutes an inconvenience in applications where one may often easily envisage an order relation to order the elements in a particular set. In this way, one arrives at a so-called ordered set, which is formally defined below.

Definitie 2.2 (Ordered set)
An **ordered set** (*geordende verzameling*) is a set S, together with a relation $<$ such that

1. For any $x, y \in S$, exactly one of $x < y$, $x = y$, or $y < x$ holds.
2. If $x < y$ and $y < z$, then $x < z$.

We write $x \leq y$ if $x < y$ or $x = y$. Besides, we may define $>$ and $\geq$ in a similar way.

For example, the set of countries can be ordered by landmass, so India > Lichtenstein. Likewise, the dictionary is the ordered set of words where the order is the so-called lexicographic ordering. In the remainder of this course, we will, however, mostly be interested in ordered sets of numbers.

2.1.3 Set operations

Having defined sets, we can now devise some operations that can be performed with them. Let us consider two sets A and B. First, we define that A and B are equal if and only if they have the same elements, that is

$$(A = B) \quad \Leftrightarrow \quad \big(\forall x \mid (x \in A) \Leftrightarrow (x \in B)\big) \,.$$

The equality of two sets can be expressed alternatively upon introducing the concept of **subsets** (*deelverzameling*). We say that A is a subset of the set B, if all elements of A are in B; that is if and only if $\forall x : x \in A \Rightarrow x \in B$. We write this as $A \subseteq B$. It immediately follows that two sets A and B are equal if and only if $A \subseteq B$ and $B \subseteq A$. A **proper subset** (*strikte deelverzameling*) A of a set B, denoted $A \subset B$, is a subset that is strictly contained in B and so necessarily excludes at least one element of B.

Equality

Robert Recorde, a Welsh mathematician, introduced the equal sign in 1557. He motivated his choice, by stating: *And to avoid the tedious repetition of these words: is equal to: I will set as I do often in work use, a pair of parallels, or Gemowe lines of one length, thus: =, because no 2 things, can be more equal.*

Mathematically, we may write

$$(A \subset B) \quad \Leftrightarrow \quad \big((A \subseteq B) \wedge (A \neq B)\big).$$

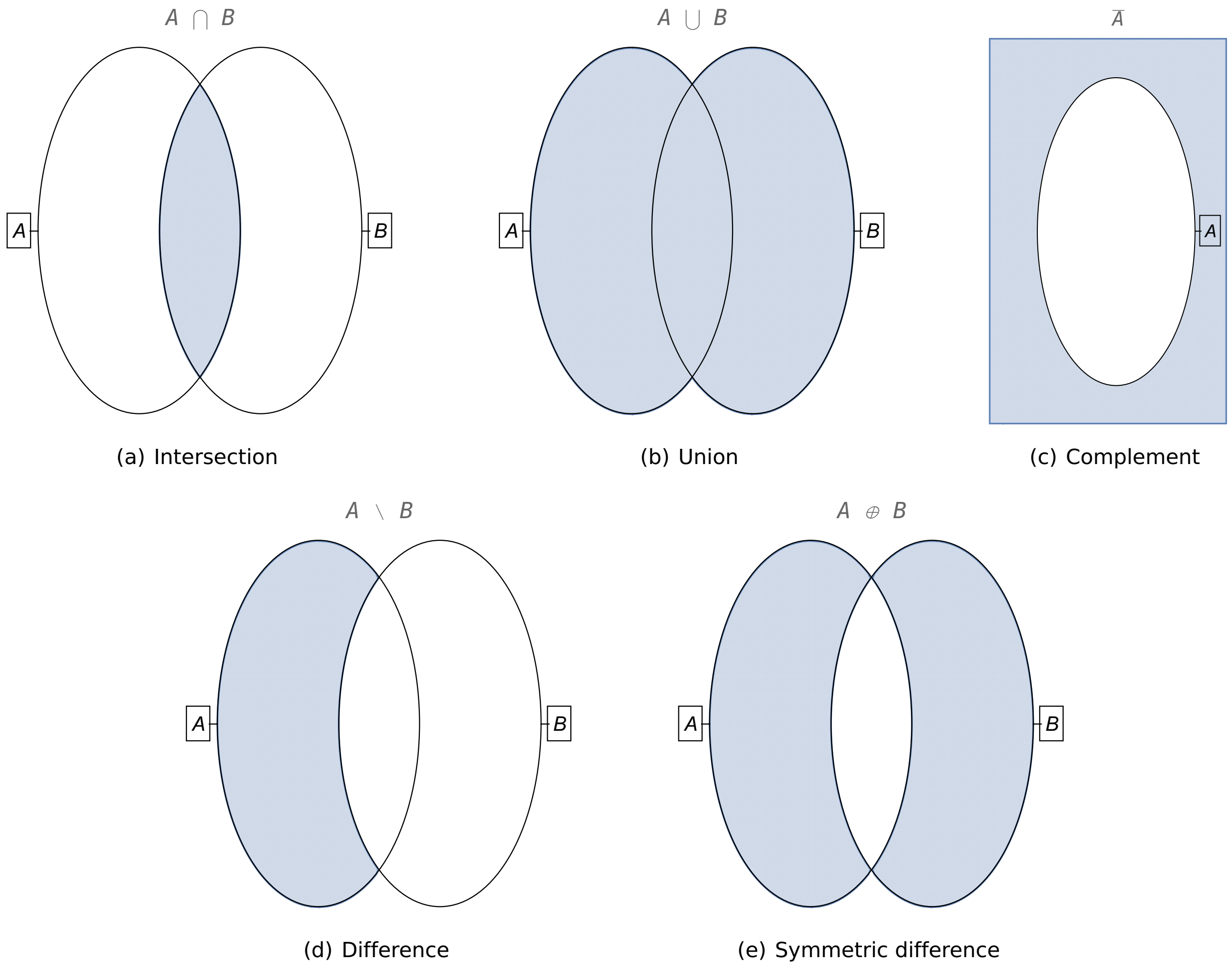
For instance, the regular polygons make up a proper subset of the set of polygons.

Finally, we can define a **universal set** (*universum*), often denoted by U, which contains all objects, including itself. Actually, in the set-builder description of the exemplary set $S = \{s, m, o, l, k\}$, it is implicitly understood that the predicate should be interpreted with respect to the letters available in the European alphabet. This alphabet may hence be understood as the universal set in which the predicate must be interpreted. In the case of ambiguity this can made more explicit in the set-builder description as

$$S = \{x \in \text{European alphabet} \,|\, x \text{ is a letter in the word 'smolko'}\}.$$

Let us now envisage the following operations on the sets A and B, subsets of the universal set U, which are illustrated in Figure 2.2.

- **Intersection** (*doorsnede*):
  $$A \cap B = \{x \,|\, (x \in A) \wedge (x \in B)\}.$$
  This yields the common elements of A and B. Two sets are called **disjoint** (*disjunct*) if $A \cap B = \varnothing$. This operation generalises directly to n sets.
- **Union** (*unie*):
  $$A \cup B = \{x \,|\, (x \in A) \vee (x \in B)\}.$$
  This yields the set of elements that belong to either of the two sets. This operation generalises directly to n sets.
- **Complement** (*complement*):
  $$\overline{A} = \{x \in U \,|\, x \notin A\}.$$
  This yields the set of elements in the universal set U that do not belong to a set A.
- **Difference** (*verschil*):
  $$A \setminus B = \{x \,|\, (x \in A) \wedge (x \notin B)\}.$$
  This yields the set of elements that belong to set A but not to set B. There is also a relationship between the set difference and set complement, which can be understood from the following:
  $$\begin{aligned} A \setminus B &= \{x \,|\, (x \in A) \wedge (x \notin B)\} \\ &= \{x \,|\, (x \in A \wedge x \in U) \wedge (x \notin B)\} \\ &= \{x \,|\, (x \in A) \wedge (x \in U \wedge x \notin B)\} \\ &= \left\{x \,|\, (x \in A) \wedge \left(x \in \overline{B}\right)\right\} \\ &= A \cap \overline{B}. \end{aligned}$$



(a) Intersection (b) Union (c) Complement
(d) Difference (e) Symmetric difference

Figure 2.2: Set operations involving two sets A and B.

- **Symmetric difference** (*symmetrisch verschil*):

$$A \oplus B = \{x \mid (x \in A \cup B) \wedge (x \notin A \cap B)\}.$$

  This yields the set of elements that belong to either one or the other set but not both.

- **Cartesian product** (*Cartesisch product*):

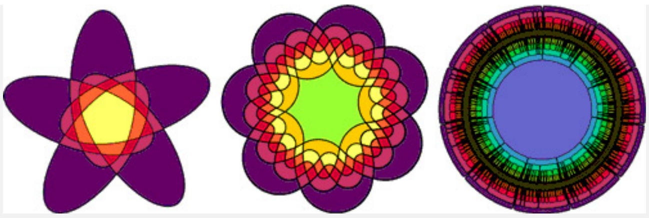
$$A \times B = \{(a, b) \mid (a \in A) \wedge (b \in B)\}.$$

  This yields the set of ordered pairs of (a, b) of all elements a and b, that belong to set A and B, respectively. When taking the Cartesian product of A with itself, i.e. $A \times A$, this is also denoted as A^2.

Venn-diagrams

Venn diagrams were named after John Venn (1843–1923), who studied and standardised these diagrams. A major issue that he tried to tackle, was finding symmetrical diagrams of partially multiple overlapping sets. Venn only got as far as 4 sets and it took until 1975 for mathematicians to extend this to 5 and more. The figure below illustrates this for 5, 7 and 11 sets, respectively.

Venn-diagrams



2.1.4 Set properties

Below we list the most important properties of sets, most of which can be understood intuitively or using a Venn diagram representation.

- **Commutativity** (*commutativiteit*) of intersection and union:
$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A.$$
- **Associativity** (*associativiteit*) of intersection and union:
$$A \cap (B \cap C) = (A \cap B) \cap C \quad \text{and} \quad A \cup (B \cup C) = (A \cup B) \cup C.$$
- **Distributivity** (*distributiviteit*) with respect to intersection and union:
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)\,.$$
- **Identity laws**:
$$A \cup \varnothing = A \quad \text{and} \quad A \cap U = A.$$
- **Complement laws**:
$$A \cup \overline{A} = U \quad \text{and} \quad A \cap \overline{A} = \varnothing.$$

For instance, if we are looking for the words that are common to Dutch, English and German, it does not matter that we first determine the words that are common to Dutch and English and then look which of those also are used in German, or start by first determining the words that are common to German and English and finally verify which of those are also used in Dutch.

For completeness, we also mention the **De Morgan's laws** (*regels van De Morgan*):

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{and} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

When considering ordered sets, we are often interested in the elements lying at their boundaries.

Definitie 2.3 (Upper bound, lower bound, supremum, infimum)
Let $A \subset S$, where S is an ordered set.

1. If there exists a $b \in S$ such that $x \leq b$ for all $x \in A$, then we say A is **bounded above** (*naar boven begrensd*) and b is an **upper bound** (*bovengrens*) of A.
2. If there exists a $a \in S$ such that $x \geq a$ for all $x \in A$, then we say A is **bounded below** (*naar beneden begrensd*) and a is a **lower bound** (*ondergrens*) of A.

3. If there exists an upper bound β of A such that whenever b is any upper bound for A we have $\beta \leq b$, then β is called the **least upper bound** (*kleinste bovengrens*) or the **supremum** of A. We write
$$\sup A = \beta.$$
This is illustrated in Figure 2.3.

4. Similarly, if there exists a lower bound α of A such that whenever a is any lower bound for A we have $\alpha \geq a$, then α is called the **greatest lower bound** (*grootste ondergrens*) or the **infimum** of A. We write
$$\inf A = \alpha.$$

When a set A is both bounded above and bounded below, we say simply that A is **bounded** (*begrensd*).

It should be stressed that the supremum (or infimum) is automatically unique (if it exists). Indeed, if β and β' are both suprema of A, then $\beta \leq \beta'$ and $\beta' \leq \beta$, because both β and β' are the least upper bounds, so it must hold that $\beta = \beta'$. For instance, let $S := \{a, b, c, d, e\}$ be ordered as $a < b < c < d < e$, and let $A := \{a, c\}$. Then c, d, and e are upper bounds of A, and c is the – unique – least upper bound or supremum of A. Likewise, the infimum of $\{2, 3, 4\}$ is 2.

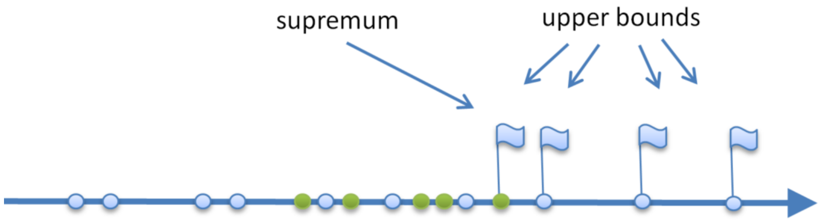


Figure 2.3: Upper bounds and the supremum of an ordered set $A \subset S$, where the ordering is endowed by the axis and the elements of A are shown as green dots.

Finally, it should be noted that a supremum or infimum for A (even if they exist) need not be in A. If $\sup A \in A$, then we also denote it by $\max A$ and call it the **maximum** (*maximum*) of A, and likewise if $\inf A \in A$, then we also denote it by $\min A$ and call it the **minimum** (*minimum*) of A. For

Example 2.1

Let us consider the set
$$A = \left\{ \frac{1}{n} \middle| n = 1, 2, 3, \ldots \right\}.$$
Then it holds that $\sup A = 1$ and since this supremum belongs to A, we say that $\max A = 1$. On the other hand, $\inf A = 0$ does not belong to A, so A has no minimum.

We conclude this section with an important property that a set may have as it is at the basis of some of the proofs in the upcoming chapters.

Definitie 2.4 (Least-upper-bound property)
An ordered set S has the **least-upper-bound property** (*supremumeigenschap*) if every nonempty subset $A \subset S$ that is bounded above has a least upper bound, that is $\sup A$ exists in S.

The least-upper-bound property is sometimes called the **completeness property**.

2.2 The set of real numbers

2.2.1 Definition

Throughout your mathematical upbringing, you have encountered several famous sets of numbers:

- The set of **natural numbers** (*natuurlijke getallen*): $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.
- The set of **integers** (*gehele getallen*): $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.
- The set of **rational numbers** (*rationale getallen*):

$$\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a, b \in \mathbb{Z} \wedge b \neq 0 \right\}.$$

Essentially, rational numbers are the ratios of integers, provided the denominator is not zero. For instance,

$$\frac{3}{4} = 0.75, \qquad \text{and} \qquad \frac{1}{3} = 0.333333\ldots$$

are just two exemplary rational numbers. Looking at those, it is clear that another way to describe the rational numbers is:

$$\mathbb{Q} = \{x \,|\, x \text{ possesses a repeating or terminating decimal representation}\}.$$

Indeed, it can be proofed that any decimal number with a repeating or terminating decimal representation can be written as a ratio of integers, so as a rational number.

There are of course numbers with a decimal that neither repeats nor terminates, e.g.

$$\pi = 3.141592654\ldots, \qquad \text{and} \qquad 0.123456789101112123\ldots$$

Such numbers are called **irrational numbers** (*irrationale getallen*) and they form the set of the irrational numbers, denoted $\mathbb{I}$. Now, we can define a new set, namely the set of so-called **real numbers** (*reële getallen*) as follows:

$$\mathbb{R} = \mathbb{I} \cup \mathbb{Q}.$$

Figure 2.4 shows how the sets of natural, rational and real numbers are nested. It clearly holds that the set of natural numbers is a proper subset of the one of the integers, which on its turn is again a proper subset of the set of rational numbers, and so on. Besides, this Venn diagram emphasizes that the sets of rational and irrational numbers are disjoint.

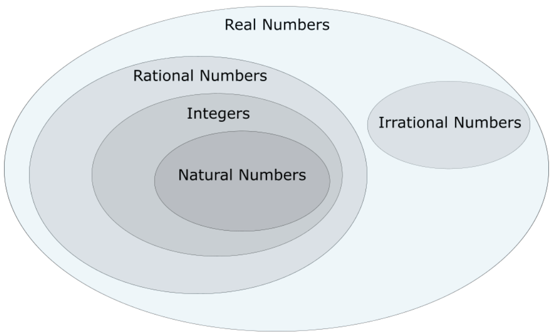
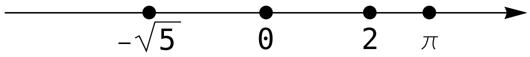


Figure 2.4: Venn diagram of the sets of natural, integer, rational, irrational and real numbers.

The set $\mathbb{R}$ may be visualized as a line because its elements can be ordered using an order relation. More precisely, the real numbers can be identified with the points on an infinitely long line once its origin, unit of length and orientation have been chosen:



With every real number x corresponds one point on this line, and vice versa, every point on this line represents one real number. This line is called the **real number line** (*reële getallenas*).

In addition to the set of real numbers, we often make reference to one of the following subsets for the sake of brevity:

$$
\begin{aligned}
\mathbb{R}_0 &= \mathbb{R}\backslash\{0\}, \\
\mathbb{R}^+ &= \{x \mid x \in \mathbb{R} \wedge x \geq 0\}, \\
\mathbb{R}^- &= \{x \mid x \in \mathbb{R} \wedge x \leq 0\}, \\
\mathbb{R}_0^+ &= \{x \mid x \in \mathbb{R} \wedge x > 0\}, \\
\mathbb{R}_0^- &= \{x \mid x \in \mathbb{R} \wedge x < 0\}.
\end{aligned}
$$

Moreover, it is possible to extend $\mathbb{R}$ with two more elements, namely positive infinity ($+\infty$) and negative infinity ($-\infty$), which are defined as:

$$\forall x \in \mathbb{R}: \quad -\infty < x < +\infty\,,$$

and which do not belong to $\mathbb{R}$. Doing so, one arrives at the set of **extended real numbers** (*uitgebreide reële getallen*):

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}\,.$$

For completeness, it should be mentioned that there is a further extension of the set of real numbers possible to the set of **complex numbers** (*complexe getallen*), defined by

$$\mathbb{C} = \left\{a + ib \mid (a, b \in \mathbb{R}) \wedge \left(i^2 = -1\right)\right\}\,.$$

This extension allows us, for instance, to compute the square root of a negative number. The complex numbers are discussed in Section 2.3. Throughout this course we will, however, mostly restrict our discussion to the set of real numbers.

History of complex numbers

A long history preceded the development of the set of complex numbers. Three milestones are listed below.

1. Negative numbers
   - China (202 BC – AD 220): *Nine Chapters on the Mathematical Art (Jiu zhang suan-shu)*
   - Red counting rods were used to denote gains (positive coefficients) and black rods for losses (negative).
2. The number 0
   - Ancient civilisations such as the Greeks and Romans had no real concept of zero as a number, although the Babylonians left blanks.
   - ca. 650 AD : Indian mathematics introduced 0, which was eventually spread to the Arab and Asian nations.
   - In Europe 0 was introduced (along with the rest of the Hindu-Arabic number system) mainly via the *Liber Abaci* by Fibonacci. Initially, there was a strong opposition against new numeric system by people clinging on

History of complex numbers

the old Roman system.

3. The number i
   - Italy, 16th century: Studied during the discovery of algebraic solutions for the roots of cubic and quartic polynomials by Italian mathematicians

2.2.2 Real number arithmetic

2.2.2.1 Addition and multiplication

In the set of real numbers, we can define to main operations namely, addition ($+$) and multiplication ($\cdot$). If a, b and c are three real numbers, we have the following five axioms:

1. **Algebraic closure** (*algebraïsch gesloten*):
$$a+b \in \mathbb{R} \quad \text{and} \quad a \cdot b \in \mathbb{R}\,.$$
2. **Commutativity** (*commutativiteit*):
$$a+b = b+a \quad \text{and} \quad a \cdot b = b \cdot a\,.$$
3. **Associativity** (*associativiteit*):
$$a+(b+c) = (a+b)+c \quad \text{and} \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c\,.$$
4. **Identity property**:
$$a+0 = 0+a = a \quad \text{and} \quad a \cdot 1 = 1 \cdot a = a\,.$$
5. **Inverse property**:
$$a+(-a) = (-a)+a = 0 \quad \text{and} \quad a \cdot a^{-1} = a^{-1} \cdot a = 1$$

Essentially, the identity property indicates that 0 is the **neutral element** (*neutraal element*) of the addition operation and 1 is neutral element of the multiplication operation, while the inverse property shows that there is always an **opposite element** (*tegengesteld element*) in the case of addition and an **inverse element** (*invers element*) in the case of multiplication. Finally, according to the sixth axiom of real numbers, multiplication **distributes** over addition:

$$a \cdot (b+c) = a \cdot b + a \cdot c \quad \text{and} \quad (a+b) \cdot c = a \cdot c + b \cdot c\,.$$

The operations of subtraction and division are not listed above because they fail to possess many of the aforementioned properties. More precisely, subtraction and division are not commutative, nor associative, as for instance $4-1 \neq 1-4$ and likewise $(4-1)-2 \neq 4-(1-2)$.

Since any set that has the operations of addition and multiplication defined on it and that satisfies the preceding axioms is called a **field** (*veld*), the sets of rational and real numbers constitute fields. On the other hand, the set of integers is not a field, as it does not contain multiplicative inverses. For example, there is no $x \in \mathbb{Z}$ such that $2x = 1$.

Clearly, we may extend the notion of a field to an **ordered field** (*geordend veld*) if the underlying set is ordered. Hence, the sets of rational and real numbers are ordered fields.

Throughout this course we will sometimes be confronted with problems involving the summation of many numbers, e.g.

$$a_0 + a_1 + a_2 + \cdots + a_n\,,$$

where n is some natural number. Clearly, writing such sums can become quite cumbersome, so a shorthand notation thereof has been established using the **capital sigma notation**; that is

$$a_0 + a_1 + a_2 + \cdots + a_n = \sum_{i=0}^{n} a_i\,,$$

where i is the **summation index** (*index*) and a_i is a **generic term** (*algemene term*) in the summation. This expression should be read as the sum of the numbers a_i for i going from 0 to n. Here, the index i starts at 0, but this is not a requisite.

We can formulate the following useful properties of summations:

- Sum or difference of summations:
$$\sum_{i=0}^{n} a_i \pm \sum_{i=0}^{n} b_i = \sum_{i=0}^{n} (a_i \pm b_i)\ .$$
- Splitting a summation:
$$\sum_{i=0}^{n} a_i = \sum_{i=0}^{m} a_i + \sum_{i=m+1}^{n} a_i\,, \text{ where } m < n\,.$$
- Scalar multiplication:
$$c \cdot \sum_{i=0}^{n} a_i = \sum_{i=0}^{n} c \cdot a_i\,, \text{where } c \in \mathbb{R}\,.$$
- Constant summation:
$$\sum_{i=0}^{n} c = c \cdot (n+1)\,, \text{where } c \in \mathbb{R}\,.$$

Moreover, we recall (without proof) two useful identities, that will be used in later chapters and involve the sum of the first n (squared) natural numbers:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \qquad \text{and} \qquad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.1}$$

Similarly to the capital sigma notation for summation, we can use the **capital pi notation** for the product of n numbers:

$$\prod_{i=0}^{n} a_i = a_0 \cdot a_1 \cdot a_2 \cdots a_n\,.$$

Using the capital pi notation, it becomes easy to define the so-called **factorial** (*faculteit*) of a natural number n, denoted $n!$:

$$n! = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

From here on, we will mostly drop the dot-notation to denote multiplication and use a blank space instead. Only in case of ambiguity (e.g. when multiplying two numbers), we will hold on the dot-notation for clarity.

2.2.2.2 Exponentiation

In addition to the four elementary operations in $\mathbb{R}$, namely addition, subtraction, multiplication and division, we can also define **exponentiation** (*machtsverheffing*). It involves two numbers, the **base** (*grondtal*) $b \in \mathbb{R}_0$ and the **exponent** (*exponent*) n and is written as b^n. When n is a strictly positive integer, exponentiation corresponds to repeated multiplication of the base: that is, b^n is the product of multiplying n bases:

$$b^n = \underbrace{b\,b\,b\cdots b}_{n \text{ factors}}\,.$$

This expression should be read as b raised to the power of n. For negative powers, we have

$$b^{-n} = \frac{1}{b^n}\,.$$

By convention, it holds that any non-zero number raised to the 0 power is 1, i.e. $b^0 = 1$ if $b \neq 0$. The expression b^2 is often called the square of b or b squared, while b^3 is frequently called the cube of b or b cubed.

The exponent does not necessarily have to be an integer, it can as well be a rational number, such as $1/n$, where $n \in \mathbb{N}_0$. More precisely, we can have

$$x = b^{\frac{1}{n}}\,,$$

which should be interpreted as the number x for which the n-th power equals b. This implies that $b^{1/n}$ is a solution to the equation

$$x^n = b\,.$$

Alternatively, $b^{1/n}$ is often written using the radical symbol as $\sqrt[n]{b}$. It is called the principal n-th root of b. If n is even and b is positive, then $x^n = b$ has two real solutions because even powers of real numbers are always positive. These solutions are the positive and negative n-th roots, i.e.

$$\sqrt[n]{a} \qquad \text{and} \qquad -\sqrt[n]{a}\,.$$

If b is negative, the equation has no solution in the set real numbers for even n. On the other hand, if n is odd, then $x^n = b$ has exactly one real solution $b^{1/n}$ that is positive if b is positive and negative if b is negative. Finally, taking a positive real number b to a rational exponent m/n, where m is an integer and n is a positive integer, and considering principal roots only, yields

$$b^{\frac{m}{n}} = \left(b^m\right)^{\frac{1}{n}} = \sqrt[n]{b^m} = \left(\sqrt[n]{b}\right)^m\,.$$

The following basic identities hold for the operation of exponentiation for every $a, b \in \mathbb{R}$ and $p, q \in \mathbb{Q}$:

- $a^p\,a^q = a^{p+q}\,,$
- $(a^p)^q = a^{p\,q}\,,$
- $(a\,b)^p = a^p\,b^p\,,$

and recalling that $1/a^p = a^{-p}$, we also have:

- $\dfrac{a^p}{a^q} = a^{p-q}$, if $a \neq 0$,
- $\left(\dfrac{a}{b}\right)^p = \dfrac{a^p}{b^p}$, if $b \neq 0$.

However, be aware that exponentiation is not commutative, nor associative, unlike addition and multiplication. Indeed, it is clear that $2^3 = 8 \neq 3^2 = 9$ and likewise

$$\left(2^3\right)^4 = 8^4 = 4096 \neq 2^{(3^4)} = 2^{81} = 2417851639229258349412352\,.$$

Example 2.2

Yearly, there are about three consecutive generations of box moths in Belgium. Suppose the number of box moths (*Cydalima perspectalis*) in generation i can be described as follows:

$$B_i = r\,B_{i-1}\,,$$

where r [T^{-1}] represents the growth rate of the box moth population. Assuming that we know the initial number of box moths (generation 0), we can compute the number of box moths in generation 1 as

$$B_1 = r\,B_0\,.$$

Then, we can compute the number of box moths in generation 2:

$$B_2 = r\,B_1 = r\,r\,B_0 = r^2\,B_0\,,$$

and so on.

The total number of box moths that has seen daylight up to and including generation n, can then be written as

$$\begin{aligned} T_n &= \sum_{i=0}^{n} B_i\,, \\ &= B_0 + B_1 + B_2 + \cdots + B_n\,, \\ &= B_0 + r\,B_0 + r^2\,B_0 + \cdots + r^n\,B_0\,. \end{aligned}$$

For instance, if $B_0 = 5$, $r = 1.1$ and $n = 5$, we find $T_5 \approx 39$ individuals.

For what concerns the square and cube of the sum and difference of two real numbers, we can infer the following well-known identities:

$$\begin{aligned} (a+b)^2 &= a^2 + 2ab + b^2 \\ (a-b)^2 &= a^2 - 2ab + b^2 \\ (a+b)(a-b) &= a^2 - b^2 \\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3 \\ (a-b)^3 &= a^3 - 3a^2b + 3ab^2 - b^3 \\ (a+b)(a^2 - ab + b^2) &= a^3 + b^3 \\ (a-b)(a^2 + ab + b^2) &= a^3 - b^3 \end{aligned}$$

2.2.2.3 Arithmetic in $\overline{\mathbb{R}}$

Obviously, the rules of arithmetic that apply to $\mathbb{R}$ apply to $\overline{\mathbb{R}}$ as well, but we need to introduce a few more rules involving a real number $a \in \mathbb{R}$ and/or $+\infty$ and/or $-\infty$. With regard to addition, we have

$$\begin{aligned} a+(+\infty) &= (+\infty)+a = +\infty\,, \\ a+(-\infty) &= (-\infty)+a = -\infty\,, \\ (+\infty)+(+\infty) &= +\infty\,, \\ (-\infty)+(-\infty) &= -\infty\,, \end{aligned}$$

while for subtraction we have

$$\begin{aligned} a-(+\infty) &= -\infty\,, \\ a-(-\infty) &= +\infty\,, \\ (+\infty)-a &= +\infty\,, \\ (-\infty)-a &= -\infty\,, \\ (+\infty)-(-\infty) &= +\infty\,, \\ (-\infty)-(+\infty) &= -\infty\,. \end{aligned}$$

For multiplication, we accordingly have (considering $a \in \mathbb{R}_0$)

$$a\cdot(+\infty) = (+\infty)\cdot a = \begin{cases} +\infty\,, & \text{if} \quad a>0\,, \\ -\infty\,, & \text{if} \quad a<0\,, \end{cases}$$

and

$$a\cdot(-\infty) = (-\infty)\cdot a = \begin{cases} -\infty\,, & \text{if} \quad a>0\,, \\ +\infty\,, & \text{if} \quad a<0\,. \end{cases}$$

Likewise, for products involving $+\infty$ and/or $-\infty$ only:

$$\begin{aligned} (+\infty)\cdot(+\infty) &= +\infty\,, \\ (-\infty)\cdot(-\infty) &= +\infty\,, \\ (+\infty)\cdot(-\infty) &= (-\infty).(+\infty) = -\infty\,. \end{aligned}$$

And for the division, we have

$$\frac{a}{+\infty} = \frac{a}{-\infty} = 0\,,$$

irrespective of the sign of $a \in \mathbb{R}_0$, and

$$\frac{+\infty}{a} = \begin{cases} +\infty\,, & \text{if} \quad a>0\,, \\ -\infty\,, & \text{if} \quad a<0\,, \end{cases}$$

$$\frac{-\infty}{a} = \begin{cases} -\infty\,, & \text{if} \quad a>0\,, \\ +\infty\,, & \text{if} \quad a<0\,. \end{cases}$$

And finally, for what concerns principal n-th roots, for $n \in \mathbb{N}_0$:

$$\begin{aligned} \sqrt[n]{+\infty} &= +\infty\,, \\ \sqrt[2n+1]{-\infty} &= -\infty\,. \end{aligned}$$

2.2.3 Completeness of $\mathbb{R}$

We have already seen that the set of real numbers is an ordered field. Even more important is that this ordered field is in fact the only complete ordered field. Basically, this means that if an alien were to construct a mathematical system on the basis of Axioms (1)-(6) from Section 2.2.2 endowed with an ordering and satisfying the least-upper-bound-property (Definition 2.4), the alien's system would differ from the real number system we devised only in that the alien might use different symbols for the real numbers and $+$, $\cdot$ and $<$.

This all gives rise to the following intuitive theorem.

Theorem 2.1 (Supremum)
If a nonempty set S of real numbers is bounded above, then $\sup S$ is the unique real number β such that

1. *$x \leq \beta$ for all x in S;*
2. *if $\epsilon > 0$ (no matter how small), there is an x_0 in S such that $x_0 > \beta - \epsilon$.*

Essentially, this theorem can be interpreted geometrically as follows: β is the supremum of S if no point of S is to the right of β, but there is at least one point of S to the right of any number less than β.

Proof To prove the theorem, we first show that $\beta = \sup S$ has properties 1) and 2). Since β is an upper bound of S, it must satisfy 1). Since any real number α less than β can be written as $\beta - \epsilon$ with $\epsilon = \beta - \alpha > 0$, 2) is just another way of saying that no number less than β is an upper bound of S. Hence, $\beta = \sup S$ satisfies 1) and 2).

Now we show that there cannot be more than one real number with properties 1) and 2). Suppose that $\beta_1 < \beta_2$ and β_2 has property 2); thus, if $\epsilon > 0$, there is an x_0 in S such that $x_0 > \beta_2 - \epsilon$. Then, by taking $\epsilon = \beta_2 - \beta_1$, we see that there is an x_0 in S such that

$$x_0 > \beta_2 - (\beta_2 - \beta_1) = \beta_1,$$

so β_1 cannot have property 1). Therefore, there cannot be more than one real number that satisfies both 1) and 2). □

Obviously, we can state a similar theorem for what concerns the infimum.

Theorem 2.2 (Infimum)
If a nonempty set S of real numbers is bounded below, then $\inf S$ is the unique real number α such that

1. *$x \geq \alpha$ for all x in S;*
2. *if $\epsilon > 0$ (no matter how small), there is an x_0 in S such that $x_0 < \alpha + \epsilon$.*

Proof As above. □

The real numbers also have the property that it is possible to exceed any positive number, no matter how large, by adding an arbitrary positive number, no matter how small, to itself sufficiently many times.

Theorem 2.3 (Archimedean property)
If ρ and ϵ are positive, then $n\epsilon > \rho$ for some $n \in \mathbb{N}_0$.

Proof The proof of this theorem is by contradiction. If the statement is false, ρ is an upper bound of the set

$$S = \{x\},$$

where $x = n\epsilon$ and $n \in \mathbb{N}_0$. Therefore, since the set of real numbers is complete, S has a supremum β and it holds that

$$n\epsilon \le \beta, \tag{2.2}$$

for all $n \in \mathbb{N}_0$. Since $n+1$ is in $\mathbb{N}_0$ whenever n is, Equation (2.2) implies that

$$(n+1)\epsilon \le \beta$$

and therefore

$$n\epsilon \le \beta - \epsilon$$

for all $n \in \mathbb{N}_0$. Hence, $\beta - \epsilon$ is an upper bound of S. Since $\beta - \epsilon < \beta$, this contradicts the definition of β.□

2.2.4 Intervals in $\mathbb{R}$

Segments of the real number line are called **intervals** (*interval*) of numbers. Table 2.2 gives a summary of the interval notation for real numbers. If the endpoint is included in the interval, we use closing square brackets, '[' or ']', when defining the interval and use a filled dot to indicate membership in the interval on the real number line. Otherwise, we use opening square brackets, ']' or '[', and a circle to indicate that the endpoint is not part of the set. If the interval does not have finite endpoints, we use the symbols $-\infty$ to indicate that the interval extends infinitely to the left and $+\infty$ to indicate that the interval extends infinitely to the right. Since infinity is a concept, and not a number, we always use opening square brackets when using these symbols in interval notation.

It should not be forgotten that any interval in $\mathbb{R}$ corresponds with a certain set of real numbers, so that we may apply the set operations introduced in Section 2.1.3 directly to intervals. For example, if $A = [-5, 3[$ and $B =]1, +\infty[$, then we easily find $A \cap B =]1, 3[$ and $A \cup B = [-5, +\infty[$. Likewise, we find $A \setminus B = [-5, 1]$.

Having introduced the set of real numbers and intervals in $\mathbb{R}$, we are now ready to give a name to some special points in this set, which will turn out useful in the subsequent chapters.

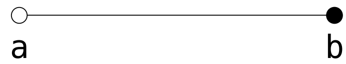
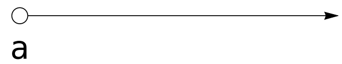
Definitie 2.5 (Boundary and interior points, open, closed and bounded sets)
Let S be a set of points in $\mathbb{R}$. A point P in $\mathbb{R}$ is a **boundary point** (*randpunt*) of S if all open disks centred at P contain both points in S and points not in S.

A point P in S is an **interior point** (*inwendig punt*) of S if there is an open interval centred at P that contains only points in S.

A point P in $\mathbb{R}$ is an **accumulation or limit point** (*ophopingspunt*) of S if the intersection between every open interval containing P and S contain infinitely many points.

A set S is **open** (*open*) if every point in S is an interior point.

Table 2.2: Interval notation for two real numbers a and b for which it holds that $a < b$.

Set of real numbers	Interval notation	Region on the real number line
$\{x \mid a < x < b\}$	$]\,a, b\,[$	
$\{x \mid a \leq x < b\}$	$[a, b\,[$	
$\{x \mid a < x \leq b\}$	$]\,a, b]$	
$\{x \mid a \leq x \leq b\}$	$[a, b]$	
$\{x \mid x < b\}$	$]-\infty, b\,[$	
$\{x \mid x \leq b\}$	$]-\infty, b]$	
$\{x \mid x > a\}$	$]\,a, +\infty\,[$	
$\{x \mid x \geq a\}$	$[a, +\infty\,[$	
$\mathbb{R}$	$]-\infty, +\infty\,[$	

A set S is **closed** (*gesloten*) if it contains all of its boundary points.

A set S is **bounded** (*begrensd*) if there is an $M > 0$ such that the open interval, centred at the origin with width M, contains S. A set that is not bounded is unbounded.

2.3 Complex numbers

We leave a detailed discussion of complex numbers to the course 'Algebra', and restrict here to a basic introduction to complex which suffices for the scope of this course.

2.3.1 Definition

Consider the polynomial $p(x) = x^2 + 1$. The zeros of p are the solutions to $x^2 + 1 = 0$, or $x^2 = -1$. This equation has no real solutions, but we can formally extract the square roots of both sides to get $x = \pm\sqrt{-1}$. The quantity $\sqrt{-1}$ is usually relabeled i, the so-called **imaginary unit** (*imaginaire eenheid*). The number i, while not a real number, plays along well with real numbers, and acts very much like any other radical expression. For instance, $3(2i) = 6i$, $7i - 3i = 4i$, $(2-7i)+(3+4i) = 5-3i$, and so forth. The key property that distinguishes i from the real numbers is the fact that

$$i^2 = -1\,.$$

Hence, if c is a real number with $c \geq 0$, then we can write

$$\sqrt{-c} = i\sqrt{c}\,.$$

Having defined the imaginary unit, we are now in the position to define the complex numbers.

Definitie 2.6 (Complex number)
A **complex number** (*complex getal*) is a number of the form

$$a + bi\,,$$

where a and b are real numbers and i is the **imaginary unit** (*imaginaire eenheid*).

Do not forget that a or b could be zero, which means numbers like $3i$ and 6 are also complex numbers. In other words, do not forget that the complex numbers include the real numbers, so 0 and $\pi - \sqrt{21}$ are both considered complex numbers (See Figure 2.5).

2.3.2 Complex number arithmetic

The arithmetic of complex numbers is as you would expect, as long as you remember that $i^2 = -1$.

Example 2.3
Perform the indicated operations. Write your answer in the form $a + bi$.

1. $(1-2i)-(3+4i)$
2. $(1-2i)(3+4i)$
3. $\dfrac{1-2i}{3-4i}$
4. $\big(x-(1+2i)\big)\big(x-(1-2i)\big)$

Solution

1. We combine like terms to get $(1-2i)-(3+4i) = 1-2i-3-4i = -2-6i$.

2. Using the distributive property, we get

$$(1-2i)(3+4i) = 3+4i-6i-8i^2\,.$$

Since $i^2 = -1$, we get $3+4i-6i-8i^2 = 11-2i$.

3. First we deal with the denominator $3-4i$ as we would any other denominator containing a radical, and multiply both numerator and denominator by $3+4i$. Doing so produces

$$\frac{1-2i}{3-4i}\,\frac{3+4i}{3+4i} = \frac{(1-2i)(3+4i)}{(3-4i)(3+4i)} = \frac{11-2i}{25} = \frac{11}{25} - \frac{2}{25}\,i.$$

4. We can rely on the fact that $(a-b)(a+b) = a^2-b^2$ and see that

$$\begin{aligned}\big(x-(1+2i)\big)\big(x-(1-2i)\big) &= \big((x-1)-2i\big)\big((x-1)+2i\big)\\ &= \big((x-1)^2-(2i)^2\big)\\ &= x^2-2x+5.\end{aligned}$$

A couple of remarks about the last example are in order. First, the **conjugate** (*(complex) toegevoegde*) of a complex number $a+bi$ is the number $a-bi$. The notation commonly used for conjugation is a bar; that is

$$\overline{a+bi} = a-bi\,.$$

For example, $\overline{3+2i} = 3-2i$ and $\overline{6} = 6$. The properties of the conjugate are summarized below, for z and w complex numbers.

- $\overline{\overline{z}} = z$
- $\overline{z}+\overline{w} = \overline{z+w}$
- $\overline{z}\,\overline{w} = \overline{zw}$
- $\left(\overline{z}\right)^n = \overline{z^n}$, for any natural number n
- z is a real number if and only if $\overline{z} = z$.

To form the **opposite** (*tegengestelde*) of a complex number, take the opposite of each part:

$$-(a+bi) = -a+(-b)i = -a-bi.$$

For example, the opposite of $6-2i$ is $-6+2i$.

Although we will only rarely have to resort to complex numbers throughout this course, their importance in engineering cannot be underestimated since complex numbers have essential applications in a variety of scientific and related areas such as signal processing, control theory, electromagnetism, fluid dynamics, quantum mechanics, cartography, and vibration analysis. For that reason, you will often encounter them in more advanced mathematics courses, such as differential equations. Besides, using complex numbers one can construct arty and intriguing graphics, like the so-called Julia set depicted in Figure 2.5.

Quaternions

The quaternions are a number system that extends the complex numbers, dating back to the 1840s only. They are generally represented as:

$$a + b\,\mathbf{i} + c\,\mathbf{j} + d\,\mathbf{k},$$

where a, b, c, and d are real numbers, and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the fundamental quaternion units for which it holds that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$. Amongst other things, the spin of an electron can be described using quaternions.

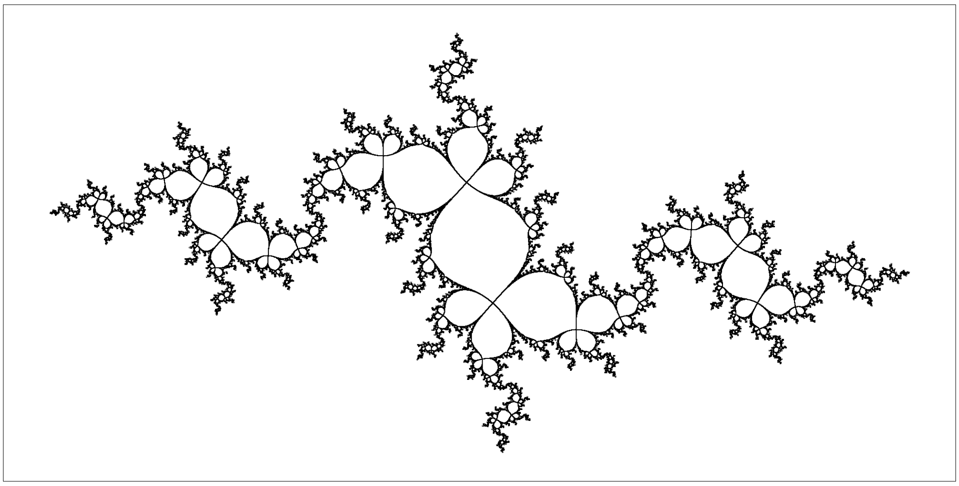


Figure 2.5: Exemplary Julia set.

2.4 Exercises

Sets and Logic operators

Assignment 2.1 — Determine the negation of the expressions below.

(a) $\exists\, x \in \mathbb{Z},\ \forall\, y \in \mathbb{Z}:\ x < y$

(b) $\forall\, y \in \mathbb{Z},\ \exists\, x \in \mathbb{Z}:\ x < y$

(c) $\forall\, x \in \mathbb{Z},\ \exists\, y \in \mathbb{Z}:\ x + y = 0$

(d) $\exists\, y \in \mathbb{Z},\ \forall\, x \in \mathbb{Z}:\ x + y = 0$

Assignment 2.2 — Define the following sets with words and write them out completely.

(a) $A = \{a \mid a \in \mathbb{N} \ \wedge\ 2 < a < 6\}$

(b) $B = \{x \mid x \in \mathbb{Q}^+ \ \wedge\ 2x^2 + x - 6 = 0\}$

(c) $C = \{x \mid x \in \mathbb{Z}^+ \ \wedge\ x^2 - 5 = 0\}$

Assignment 2.3 — Write in a concise manner that A is a set of;

(a) all even numbers bigger than 100.

(b) all pairs of integers whose first and second elements are even and odd, respectively.

(c) all integers, different from zero, that are multiples of 3.

(d) all positive rational numbers whose square root is greater than 3.

(e) all numbers that when divided by 6 result in a remainder of 2.

Assignment 2.4 — Fill in the correct symbols. Choose from $\subset, \not\subset, =, \neq, \in, \notin, \ni, \not\ni$. Multiple answers might be possible.

(a) $\{1,3,5,7,9,11,\dots\} \dots \{x \in \mathbb{N} \mid x \text{ is an even number}\}$

(b) $\{x \mid x \text{ is a rose}\} \dots \{x \mid x \text{ is a flower}\}$

(c) $\{1,3,5,7,9\} \dots 2$

(d) $\{1\} \dots \{1,3,5,7,9\}$

(e) $\{1,3\} \dots \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(f) $\{1,3\} \dots \{1,3,5,7,9\}$

(g) $\{1,3,5,7,9\} \dots \varnothing$

(h) $\{1\} \dots \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(i) $\{1,3,\{5,7,9\}\} \dots 5$

Assignment 2.5 — Assume $A = \{1, \{1\}, \{2\}\}$. Which from the following statements is true?

(a) $1 \in A$

(b) $\{1\} \in A$

(c) $\{1\} \subseteq A$

(d) $\{\{1\}\} \subseteq A$

(e) $2 \in A$

(f) $\{\{2\}\} \subseteq A$

(g) $\{\{2\}\} \subset A$

(h) $\{2\} \subseteq A$

Assignment 2.6 — Given $U = \{1, 2, 3, \ldots, 9, 10\}$, assume $A = \{1, 2, 3, 4, 5\}$, $B = \{1, 2, 3, 4\}$, $C = \{3, 5, 7\}$ and $D = \{2, 4, 6, 8\}$. Describe each of the following sets:

(a) $(A \cup B) \cap C$

(b) $A \cup (B \cap C)$

(c) $\overline{C} \cup \overline{D}$

(d) $\overline{C \cap D}$

(e) $(A \cup B) \setminus C$

(f) $A \cup (B \setminus C)$

(g) $(B \setminus C) \setminus D$

(h) $B \setminus (C \setminus D)$

(i) $(A \cup B) \setminus (C \cap D)$

Assignment 2.7 — Simplify the following expressions:

(a) $A \cap (B \setminus A)$

(b) $(A \setminus B) \cup (A \cap B)$

(c) $\overline{A} \cup \overline{B} \cup \left(A \cap B \cap \overline{C}\right)$

(d) $(A \cap B) \cup \left(A \cap B \cap \overline{C} \cap D\right) \cup \left(\overline{A} \cap B\right)$

(e) $(A \cap B) \cup \left(B \cap \left((C \cap D) \cup \left(C \cap \overline{D}\right)\right)\right)$

Assignment 2.8 — Prove the associative properties with respect to union and intersection for three sets.

Assignment 2.9 — Prove the following properties.

(a) $A \cup B = B \iff A \subset B \iff A \cap B = A \iff A \setminus B = \varnothing$

(b) $A \setminus B = A \iff A \cap B = \varnothing \iff B \setminus A = B$

(c) $(A \setminus B) \cap A = A \setminus B$

(d) $(A \setminus B) \cap B = \varnothing$

(e) $(A \cup B) \setminus B = A \setminus B$

(f) $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$

Assignment 2.10 — Use set notation to define the shaded areas in Figure 2.6.

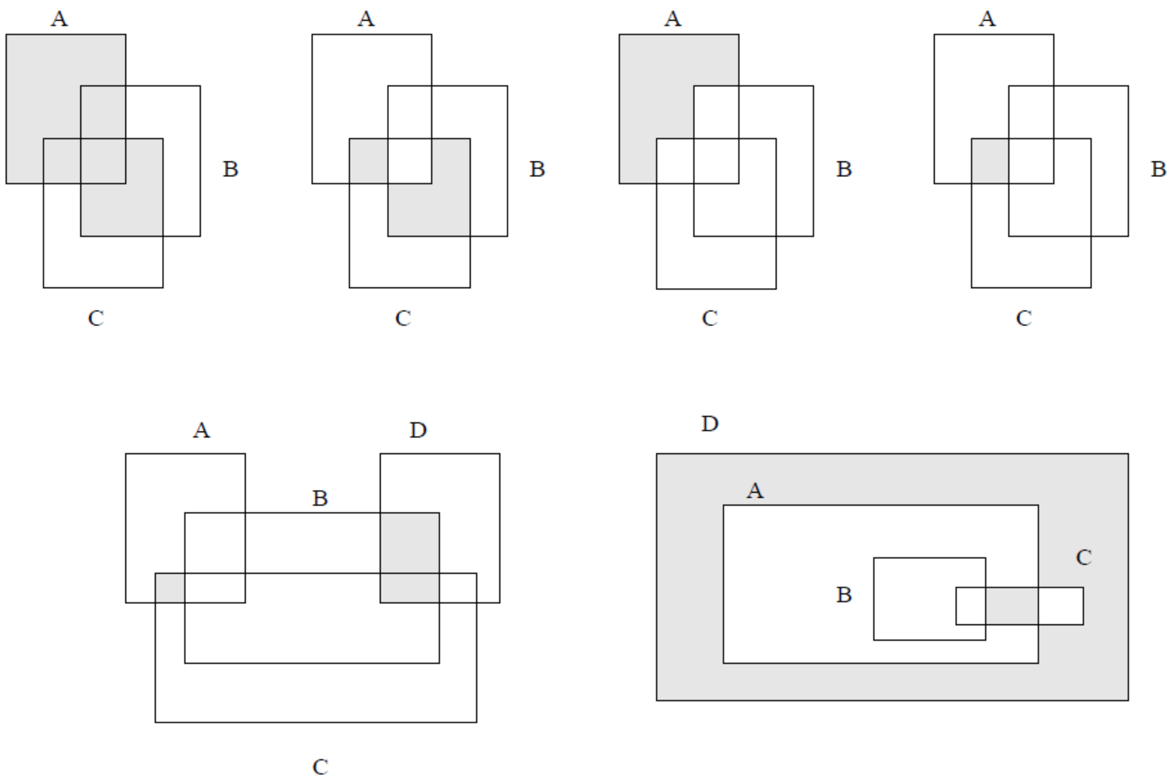


Figure 2.6: Shaded areas used in exercise 2.10.

Assignment 2.11 — Assume $W \subset \mathbb{R}$ and $b \in \mathbb{R}$. Give a correct expression with logical operators for the expressions below.

(a) b is an upper limit of set W.

(b) b is not an upper limit of set W.

(c) W is an upwardly bounded set.

(d) W is not an upwardly bounded set.

Assignment 2.12 —
Determine the supremum and infimum of the following subsets of $\mathbb{R}$.

(a) $\left\{\dfrac{2^n}{2^n+1}, n \in \mathbb{N}\right\}$ (b) $\left\{\dfrac{2n-1}{n+2}, n \in \mathbb{N}_0\right\}$ (c) $\left\{\dfrac{n}{n+1}, n \in \mathbb{N}\right\}$

Intervals in $\mathbb{R}$

Assignment 2.13 — Determine the infimum, supremum, minimum, maximum, boundary points, and interior points of A.

(a) $A = \{1\} \cup [2,5[\cup]5,7[$

(b) $A = \{-3\} \cup]0,4] \cup [7,+\infty[$

(c) $A =]-\infty,-2[\cup]-2,2[\cup [3,4[\cup]5,9]$

The set of real numbers

Assignment 2.14 — Which of the numbers below are rational or irrational?

(a) 5.369

(b) $\dfrac{12}{7}$

(c) $\sqrt{13}$

(d) $1.232345456767\ldots$

(e) $3.0236363636\ldots$

(f) $\sqrt{121}$

Assignment 2.15 — Rewrite the following expressions using a sum or multiplication sign.

(a) $x + x^2 + x^3 + x^4 + \ldots + x^{99}$

(b) $\sqrt{3} + \sqrt{5} + \sqrt{7} + \sqrt{9} + \ldots + \sqrt{51}$

(c) $\dfrac{1}{a+1} \cdot \dfrac{4}{a+2} \cdot \dfrac{9}{a+3} \cdot \dfrac{16}{a+4} \cdots \dfrac{169}{a+13}$

Assignment 2.16 — Calculate the following sums.

(a) $\displaystyle\sum_{j=0}^{3} 2^j$

(b) $\displaystyle\sum_{j=0}^{3} j^2$

(c) $\displaystyle\sum_{j=0}^{4} \frac{24}{j!}$

(d) $\displaystyle\sum_{j=1}^{4} \frac{(-1)^j}{j}$

(e) $\displaystyle\sum_{j=1}^{5} (2j-1)^2$

(f) $\displaystyle \mathrm{i}\sum_{j=1}^{100} -7j^2$

(g) $\displaystyle\sum_{j=1}^{90} \left(-2j^2 + 3j - 5\right)$

(h) $\displaystyle\sum_{\substack{0<k<10 \\ k \text{ is oneven}}} k^2$

Assignment 2.17 — Simplify each sum or product to an expression without sigma or pi notation.

(a) $\displaystyle\sum_{j=1}^{n} (3j-2)$

(b) $\displaystyle\sum_{j=1}^{n} \left((3j-5)\, \frac{1}{n^2} \right)$

(c) $\displaystyle\sum_{j=1}^{n} (3j-4)^2$

(d) $\displaystyle\sum_{j=1}^{n} \left((j-2)^2\, \frac{1}{n^3} \right)$

(e) $\displaystyle\prod_{j=1}^{n} j^3$

(f) $\displaystyle\prod_{k=2}^{n} \left(1 - \frac{1}{k^2} \right)$

Assignment 2.18 — Calculate or simplify the following algebraic forms.

(a) $\left(64\, a^{6m}b^{12n}c^{18p}\right)^{\frac{1}{6}}$

(b) $\dfrac{1-x}{1-\sqrt{x}}$

(c) $\left(1-\sqrt{2}-\sqrt{3}\right)^2$

(d) $b\sqrt{\dfrac{4a}{b^4}}-\sqrt{\dfrac{9a}{b^2}}+\dfrac{1}{b}\sqrt{\dfrac{a}{4}}+2b\sqrt{\dfrac{25a}{b^4}}$

(e) $\left(\sqrt{\dfrac{x+1}{x-1}}\right)\left(\sqrt[3]{\dfrac{x-1}{x+1}}\right)$

(f) $\sqrt[3]{a^3+\dfrac{3}{2}a^2b+\dfrac{3}{4}ab^2+\dfrac{1}{8}b^3}$

(g) $\dfrac{a^2+b^2}{(b-a)^2}\,\dfrac{(a-b)^3}{a+b}\,\dfrac{(-a-b)^2}{a^2-b^2}$

(h) $\left(x^{\frac{1}{3}}-x^{-\frac{1}{3}}\right)^3+3\left(x^{\frac{1}{3}}-x^{-\frac{1}{3}}\right)$

(i) $\dfrac{\left(x^2\right)^3x^{-4}\sqrt[3]{x^5}}{\sqrt[3]{x^2}\sqrt[3]{\sqrt[4]{\left(x^2\right)^3}}}$

(j) $\left(\dfrac{16^{-2}a^{\frac{1}{2}}b^{-3}}{81^{-1}a^{-\frac{1}{2}}b^3}\right)\sqrt{ab^{\frac{9}{4}}\left(ab^{\frac{3}{2}}\right)^{\frac{1}{2}}}$

Complex numbers

Assignment 2.19 — Determine and simplify

- $z+w$
- zw
- z^2
- z^{-1}
- $\dfrac{z}{w}$
- $\dfrac{w}{z}$
- $\bar{z}$
- $z\bar{z}$
- $(\bar{z})^2$

rewrite each pair of complex numbers in standard form: $a+bi$.

(a) $z=2+3i,\quad w=4i$

(b) $z=1+i,\quad w=-i$

(c) $z=3-5i,\quad w=2+7i$

(d) $z=\sqrt{2}-\sqrt{2}i,\quad w=\sqrt{2}+\sqrt{2}i$

(e) $z=1-\sqrt{3}i,\quad w=-1-\sqrt{3}i$

(f) $z=-\dfrac{\sqrt{2}}{2}+\dfrac{\sqrt{2}}{2}i,\quad w=-\dfrac{\sqrt{2}}{2}-\dfrac{\sqrt{2}}{2}i$

Assignment 2.20 — Write the following numbers in standard form: $a+bi$.

(a) $(4+8i)+(15-12i)$

(b) $(2+4i)-(6-7i)$

(c) $(2+3i)+(-5+i)$

(d) $(2+i)^2$

(e) $\overline{(5+6i)}(5+6i)$

(f) $\dfrac{1}{5+2i}$

(g) $\dfrac{1+i}{2+3i}$

(h) $\dfrac{1+2i}{3-4i}+\dfrac{2-i}{5i}$

Mathematics knows no races or geographic boundaries; for mathematics, the cultural world is one country.

— David Hilbert —

3 Functions

3.1 The Cartesian coordinate plane

In order to visualize the pure excitement that is calculus, we need to unite algebra and geometry. Simply put, we must find a way to draw algebraic things. Let us start with possibly the greatest mathematical achievement of all time: the **Cartesian coordinate plane** (*Cartesisch coördinatenstelsel*). So named in honour of René Descartes.

Imagine two real number lines crossing at a right angle at 0. The horizontal number line is usually called the ***x*-axis** (*x-as*), while the vertical number line is usually called the ***y*-axis** (*y-as*). For example, consider the point P in Figure 3.1. To use the numbers on the axes to label this point, we project the point P to the x- (respectively y-) axis. We then describe the point P using the **ordered pair** (*geordend koppel*) $(2,-4)$. The first number in the ordered pair is called the **abscissa** (*abscis*) or ***x*-coordinate** and the second is called the **ordinate** (*ordinaat*) or ***y*-coordinate**. When we speak of the Cartesian coordinate plane, we mean the set of all possible ordered pairs (x, y) as x and y take values from the real numbers. The ordered pair $(2,-4)$ comprise the **Cartesian coordinates** (*Cartesische coördinaten*) of the point P. In practice, the distinction between a point and its coordinates is blurred. We can think of $(2,-4)$ as instructions on how to reach P from the **origin** (*oorsprong*) $(0,0)$.

The axes divide the plane into four regions called **quadrants** (*kwadrant*). They are labelled with Roman numerals and proceed counter-clockwise around the plane . If a point other than the origin happens to lie on the axes, we typically refer to that point as lying on the positive or negative x-axis (if $y = 0$) or on the positive or negative y-axis (if $x = 0$). Such points do not belong to any of the four quadrants.

Using Cartesian coordinates, we can introduce the three main types of **symmetry** (*symmetrie*), namely symmetry about the x-axis, symmetry about the y-axis, and finally, symmetry about the origin.

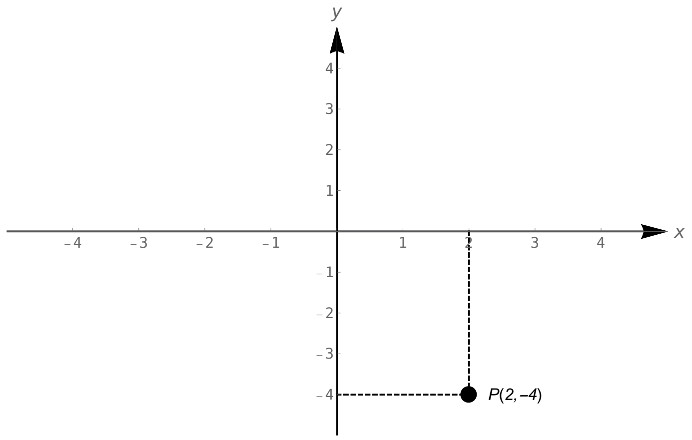


Figure 3.1: The point P located in the Cartesian coordinate plane.

3.2 Functions

3.2.1 Relations

Definitie 3.1 (Relation)
A **relation** (*relatie*) is a set of points in the plane. Hence, a relation R in $\mathbb{R}$ is a subset of the Cartesian product $\mathbb{R}^2$.

Since relations are sets, we can describe them using the techniques presented in Chapter 2. That is, we can describe a relation verbally, using the roster method, or using set-builder notation. Most frequently, the latter kind of description is preferred and a relation is defined using a specific predicate that depends on x and y, $P(x,y)$, i.e.

$$R = \left\{(x,y) \in \mathbb{R}^2 \mid P(x,y)\right\}.$$

The predicate $P(x,y)$ is the rule that allows us to select the ordered pairs (x,y) that make up the relation. Here, we call x the **argument** (*argument*) of the relation R and y its corresponding **image** (*beeld*). Since the elements in a relation are points in the plane, we often try to describe the relation graphically as well. Doing so produces the **graph** (*grafiek*) of the relation R.

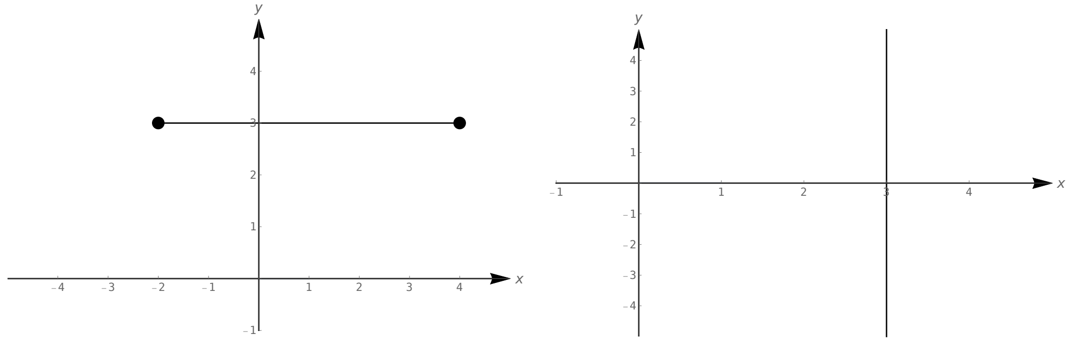
Example 3.1
Graph the following relations.

1. $B = \{(x,3) \mid -2 \leq x \leq 4\}$
2. $C = \{(3,y) \mid y \text{ is a real number}\}$

Solution

1. In words, $\{(x,3) \mid -2 \leq x \leq 4\}$ reads 'the set of points $(x,3)$ such that $-2 \leq x \leq 4$. Plotting several representative points should convince you that B describes the horizontal line segment from the point $(-2,3)$ up to and including the point $(4,3)$ (Figure 3.2(a)).
2. The relation C is described as the set of points $(3,y)$ such that y is a real number. All of these points have an x-coordinate of 3, but the y-coordinate is free to be whatever it wants to be,

without restriction. Hence, all the points of C lie on the vertical line $x = 3$ (Figure 3.2(b)).



(a) Set $B = \{(x,3) \,|\, -2 \le x \le 4\}$ (b) Set $C = \{(3,y) \,|\, y \text{ is a real number}\}$

Figure 3.2: Graphs of different relations.

The relation C in the previous example lead us to our final way to describe relations: **algebraically** (*algebraïsch*). We can more succinctly describe the points in C as those points which satisfy the equation $x = 3$. Let us now study the graphs of equations in a more general setting. For that purpose, we rely on the so-called fundamental graphing principle.

Definitie 3.2 (Fundamental graphing principle)
The graph of an equation is the set of points which satisfy the equation.

It is at this point that we gain some insight into the word 'relation'. If the equation to be graphed contains both x and y, then the equation itself is what is relating the two variables. For instance, in the next example, we consider the graph of the equation $x^2 + y^3 = 1$ by graphing the relation $R = \{(x,y) \,|\, x^2 + y^3 = 1\}$.

Example 3.2

Graph the equation $x^2 + y^3 = 1$.

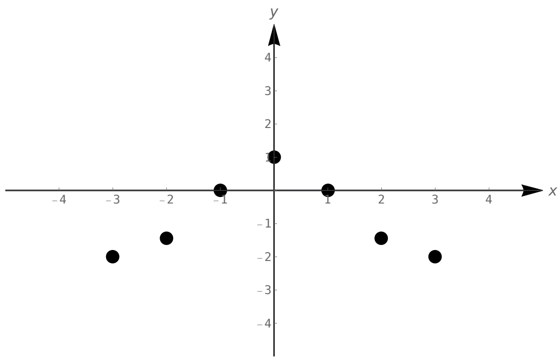
Solution

To efficiently generate points on the graph of this equation, we first solve this equation for y:

$$\begin{array}{rrcl} & x^2 + y^3 & = & 1 \\ \Leftrightarrow & y^3 & = & 1 - x^2 \\ \Leftrightarrow & \sqrt[3]{y^3} & = & \sqrt[3]{1 - x^2} \\ \Leftrightarrow & y & = & \sqrt[3]{1 - x^2}. \end{array}$$

We now substitute a value in for x, determine the corresponding value y, and plot the resulting point (x,y) in the Cartesian coordinate plane. We first generate a table of points which are on the graph of the equation. These points are then plotted in the plane as shown below.

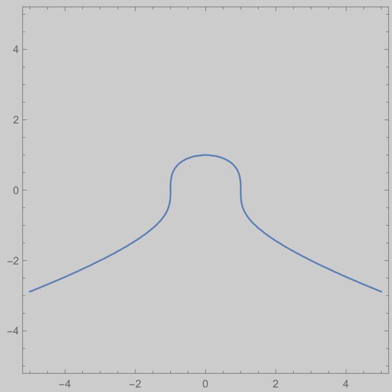
x	y
-3	-2
-2	$-\sqrt[3]{3}$
-1	0
0	1
1	0
2	$-\sqrt[3]{3}$
3	-2



Alternatively, we can construct a graph of this equation in Mathematica using the built-in function **`ContourPlot`**:

In[1]:=
```
ContourPlot[x^2+y^3==1,{x,-5,5},{y,-5,5}]
```

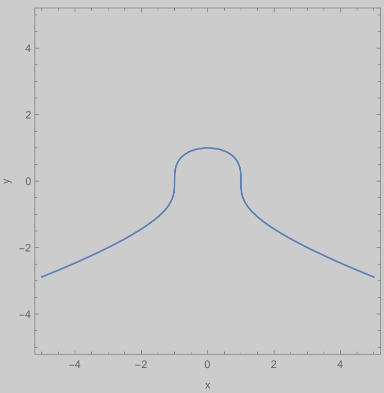
Out[1]=



Of course, we should add frame labels to this plot, which can be done as follows.

In[2]:=
```
ContourPlot[x^2+y^3==1,{x,-5,5},{y,-5,5}, FrameLabel →{"x","y"}]
```

Out[2]=



The places where the graph of an equation crosses or touches the axes are called the **intercepts** (*intercept*).

Another fact which you may have noticed about the graph in Example 3.2 is that it seems to be symmetric about the y-axis. To actually prove this analytically, we assume (x, y) is a generic point on the graph of the equation. That is, we assume $x^2 + y^3 = 1$ is true. To show that the graph as a whole is symmetric about the y-axis, we need to show that $(-x, y)$ satisfies the equation $x^2 + y^3 = 1$, too.

Substituting $(-x, y)$ into the equation gives

$$\begin{aligned} (-x)^2 + (y)^3 &\stackrel{?}{=} 1 \\ \Leftrightarrow \quad x^2 + y^3 &\stackrel{\checkmark}{=} 1. \end{aligned}$$

Since we are assuming that the original equation $x^2 + y^3 = 1$ is true, we have shown that $(-x, y)$ satisfies the equation and hence is on the graph. In this way, we can check whether the graph of a given equation possesses any of the symmetries introduced in Section 3.1.

- About the y-axis: substitute $(-x, y)$ into the equation and verify whether or not the result is equivalent to the original equation.
- About the x-axis: substitute $(x, -y)$ into the equation and verify whether or not the result is equivalent to the original equation.
- About the origin: substitute $(-x, -y)$ into the equation and verify whether or not the result is equivalent to the original equation.

3.2.2 Functions in $\mathbb{R}$

3.2.2.1 Definition

Definitie 3.3 (Function)
A relation in which each x-coordinate in $\mathbb{R}$ is matched with at most only one y-coordinate is said to describe y as a **function** (*functie*) of x. A function f that maps x to y is denoted as

$$f : x \mapsto f(x),$$

with $y = f(x)$.

The notation $y = f(x)$ (read: y equals f of x) means that the pair (x, y) belongs to the set of pairs defining the function f. x is called the **argument** or **input** (*input*) of the function f and $f(x)$ is called the **value** taken by the function when evaluated at a point x, or its **output** (*output*) or **image**. In the framework of applications, x is often called the **independent variable** (*onafhankelijke veranderlijke*), while y is called the **dependent variable** (*afhankelijke veranderlijke*). Loosely speaking, a function may be envisaged as a black box that returns for each input a corresponding output (Figure 3.3).

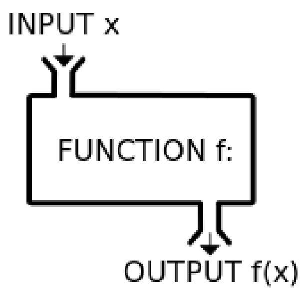


Figure 3.3: Describing a function as a black box.

To (graphically) determine whether or not a relation is function, we can use the following theorem, which is an immediate consequence of Definition 3.3.

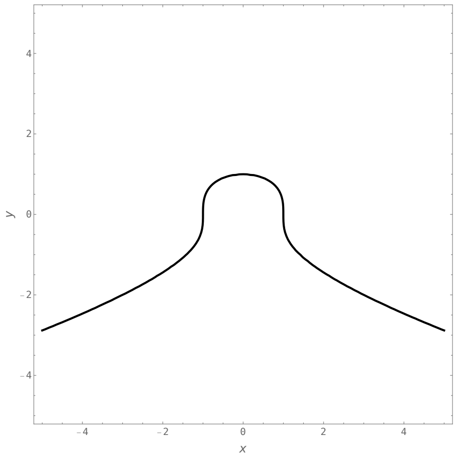
> **Theorem 3.1 (Vertical line test)**
> *A set of points in the plane represents y as a function of x if and only if no two points lie on the same vertical line.*

Proof This is a direct consequence of Definition 3.3 □

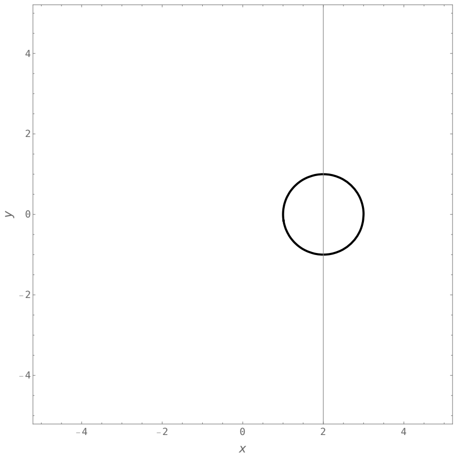
In other words, a relation R constitutes a function if and only if every line parallel to the y-axis intersects the graph of R in at most one point.

Example 3.3

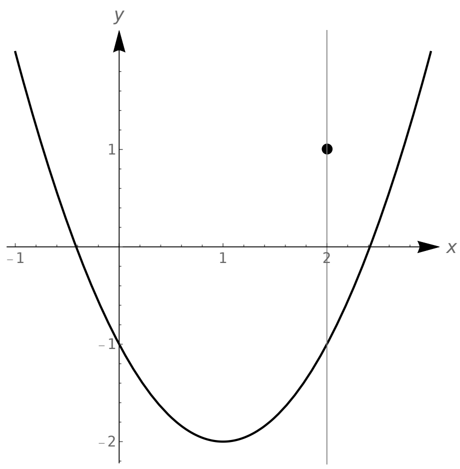
Determine graphically which of the following relations actually represents a function.



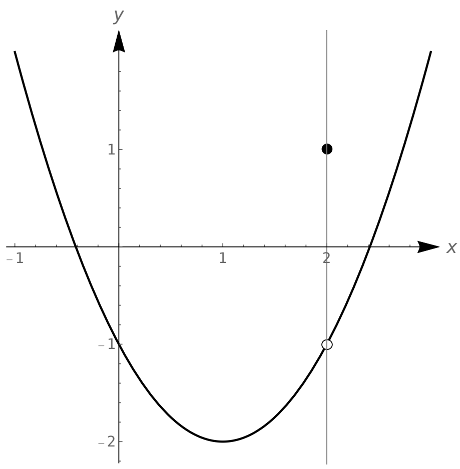
(a) Graph of relation A



(b) Graph of relation B



(c) Graph of relation C



(d) Graph of relation D

Solution

(a) In the graph of A, every vertical line crosses the graph at most once, so A does represent y as a function of x.

(b) Looking at the graph of B, we can easily imagine a vertical line crossing the graph more than once. Hence, B does not represent y as a function of x.

(c) In C, there is a point on the curve with x-coordinate 2 just below $(2, 1)$, which means that both $(2,1)$ and this point on the curve lie on the vertical line $x = 2$. Hence, the graph of C fails the vertical line test, so y is not a function of x here.

(d) In D notice that the point with x-coordinate 2 on the curve has been omitted, leaving an open circle there. Hence, the vertical line $x = 2$ crosses the graph of D only at the point $(2,1)$. Indeed, any vertical line will cross the graph at most once, so we have that the graph of D passes the vertical line test. Thus it describes y as a function of x.

Finally, it is important to note that a function for which the dependent variable can be written explicitly in terms of the independent variable, i.e. as $y = f(x)$, is called an **explicit function** (*expliciete functie*). On the other hand, if the function is defined by

$$F(x,y) = 0\,,$$

which does not contain y explicitly at one side of the equation, it is called an **implicit function** (*impliciete functie*). For instance,

$$x^2 + y^2 - 4 = 0$$

constitutes an implicit function, which defines two explicit functions, namely

$$y = \sqrt{4-x^2}, \qquad \text{and} \qquad y = -\sqrt{4-x^2}\,.$$

3.2.2.2 Function graphs

It is often useful to draw the graph of a function for getting a global view of its properties. Formally, we define the graph of a function as follows.

Definitie 3.4 (Fundamental graphing principle for functions)
The graph G of a function f is the set of points which satisfy the equation $y = f(x)$. More formally,

$$G = \left\{(x, f(x)) \mid x \in \mathbb{R} \wedge y = f(x)\right\}\,.$$

The x-coordinates of the x-intercepts of the graph of $y = f(x)$ can be found by solving $f(x) = 0$. For this reason, they are called the **zeros** (*nulpunt*) of f.

3.2.2.3 Domain, codomain and range

When defining a function f as

$$f : x \mapsto f(x)\,,$$

we not only need to specify how f maps an argument x to its image y, but also the sets to which x and $y = f(x)$ belong. In other words, we need to specify what are the possible in- and outputs of our black box (Figure 3.3). We call the set of possible inputs the **domain** (*domein*) of the function, denoted as $\operatorname{dom} f$. The set of possible outputs is called the **codomain** (*codomein*). Suppose X is the domain of a function f and Y is its codomain, then we can define this function using arrow notation to make explicit the domain and codomain:

$$f : X \to Y\,,$$

which should be read as f maps domain X to codomain Y. The set to which f actually maps the domain, is called the **range** (*bereik*) or **image** (*beeld*) of f, denoted by $\operatorname{im} f$. The range is thus defined as

$$\operatorname{im} f = \left\{ y = f(x) \mid x \in X \wedge y \in Y \right\} .$$

These concepts are illustrated in Figure 3.4. The difference between range and codomain may seem subtle, but can be very important, as will be shown in Example 3.4.

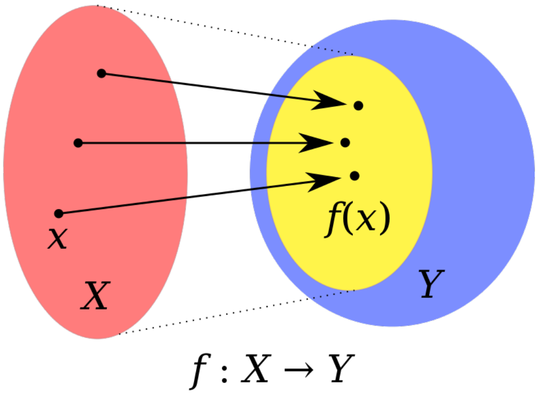


Figure 3.4: The domain X, codomain Y and range $\left\{ f(x) \mid x \in X \right\}$ of a function f.

Throughout this course we will focus on so-called **real functions** (*reële functie*), which are real-valued functions of a real variable. Such functions can be written as

$$f : \mathbb{R} \to \mathbb{R} \quad : \quad x \mapsto f(x) .$$

In most cases, we can relatively easily determine the domain, codomain and range of a function f by investigating the formula defining it. Doing so, we sometimes run into functions whose domain consists of two ore more consecutive open intervals, such as for

$$f(x) = \frac{1}{x} ,$$

for which $\operatorname{dom} f = \mathbb{R} \setminus \{0\} =]-\infty, 0[\cup]0, +\infty[$. The points at which a function in such a case is not defined are called the function's **singularities** (*singulariteiten*). So, $f(x) = \frac{1}{x}$ has a singularity at $x = 0$.

Example 3.4

Determine the domain, codomain and range of the following functions

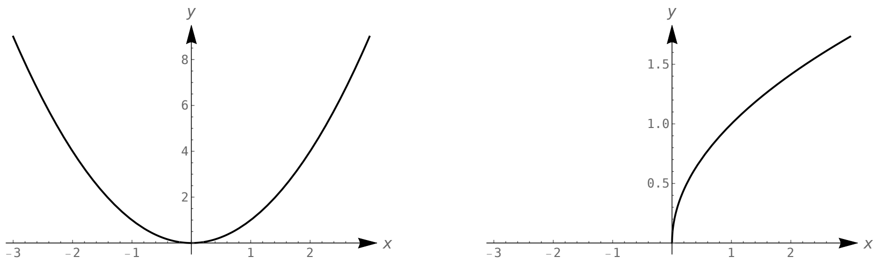
1. $f : \mathbb{R} \to \mathbb{R} \quad : \quad x \mapsto x^2 ,$
2. $g : \mathbb{R} \to \mathbb{R}^+ \quad : \quad x \mapsto x^2 ,$
3. $h : \mathbb{R}^+ \to \mathbb{R}^+ \quad : \quad x \mapsto x^{\frac{1}{2}} .$

Solution

1. From the function definition of f, we infer that both the domain and codomain are $\mathbb{R}$. Yet, since f does not map to any negative number, the range of f is $\mathbb{R}^+$. Its graph is shown in Figure 3.5(a).

2. From the function definition of g, we infer that the domain is $\mathbb{R}$, but its codomain is $\mathbb{R}^+$. Again, since g does not map to any negative number, the range of g is $\mathbb{R}^+$ and its graph is the same as the of f shown in Figure 3.5(b).

3. From the function definition of h, we infer that both the domain and codomain are $\mathbb{R}^+$. Since the mapping is done by the square root, this has important implications. Firstly, the square root of a negative number does not exist in $\mathbb{R}$, which sets a restriction on the domain. Secondly, the square root of $x \in \mathbb{R}^+$ maps x to $\sqrt{x}$ and $-\sqrt{x}$. Hence, if the codomain would be defined as $\mathbb{R}$, h would not be a function (see Definition 3.3)! From this we can also infer that the range of h is $\mathbb{R}^+$. Its graph is shown in Figure 3.5(b).



(a) Graph of f and g (b) Graph of h

Figure 3.5: Graphs of the functions f, g and h in Example 3.4.

3.2.3 Function arithmetic

It seems natural that functions should have their own arithmetic which is consistent with the arithmetic of real numbers. The following definitions allow us to add, subtract, multiply and divide functions using the arithmetic we already know for real numbers. Suppose f and g are functions and $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$, then we can define the following operations on $\operatorname{dom} f \cap \operatorname{dom} g$:

- The **sum** (*som*) of f and g:
$$(f+g)(x) = f(x) + g(x).$$
- The **difference** (*verschil*) of f and g:
$$(f-g)(x) = f(x) - g(x).$$
- The **product** (*product*) of f and g:
$$(fg)(x) = f(x)g(x).$$
- The **quotient** (*quotiënt*) of f and g:
$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)},$$
provided $g(x) \neq 0$.

Note that while the formula $(f+g)(x) = f(x) + g(x)$ looks suspiciously like some kind of distributive property, it is nothing of the sort; the addition on the left hand side of the equation is function addition, and we are using this equation to define the output of the new function $f+g$ as the sum of the real number outputs from f and g.

Example 3.5

Let $f(x) = 6x^2 - 2x$ and $g(x) = 3 - \frac{1}{x}$.

1. Find the domain of $g - f$. Then find and simplify a formula for $(g-f)(x)$.
2. Find the domain of g/f. Then find and simplify a formula for $\left(\frac{g}{f}\right)(x)$.

Solution

1. To find the domain of $g-f$ we need to find the domain of g and of f separately, then find the intersection of these two sets. Owing to the denominator in the expression $g(x) = 3 - \frac{1}{x}$, we get that the domain of g is $\mathbb{R}_0$. Since $f(x) = 6x^2 - 2x$ is valid for all real numbers, we have no further restrictions. Thus the domain of $g-f$ matches the domain of g, namely, $\mathbb{R}_0$.

   Moving along, we need to simplify a formula for $(g-f)(x)$. In this case, we get common denominators and attempt to reduce the resulting fraction. Doing so, we get

$$\begin{aligned}(g-f)(x) &= g(x) - f(x) \\ &= \left(3 - \frac{1}{x}\right) - \left(6x^2 - 2x\right) \\ &= \frac{-6x^3 + 2x^2 + 3x - 1}{x}.\end{aligned}$$

2. First, we find the domain of g and f separately, and find the intersection of these two sets. In addition, since $\left(\frac{g}{f}\right)(x) = \frac{g(x)}{f(x)}$, we are introducing a new denominator, namely $f(x)$, so we need to guard against this being 0 as well. The domain of g is $\mathbb{R}_0$ and the domain of f is $\mathbb{R}$. Setting $f(x) = 0$ gives $6x^2 - 2x = 0$ or $x = 0$ or $x = \frac{1}{3}$. So, the domain of g/f is $\mathbb{R} \setminus \{0, \frac{1}{3}\}$.

   Next, we find and simplify a formula for $\left(\frac{g}{f}\right)(x)$.

$$\begin{aligned}\left(\frac{g}{f}\right)(x) &= \frac{g(x)}{f(x)} \\ &= \frac{3 - \frac{1}{x}}{6x^2 - 2x} = \frac{3x-1}{(6x^2-2x)x} \\ &= \frac{3x-1}{2x^2(3x-1)} \qquad \left(\text{Assuming } x \neq \frac{1}{3} \text{ and } x \neq 0\right) \\ &= \frac{1}{2x^2}\end{aligned}$$

In addition to operations on functions, we can also compose functions. Function composition is defined below.

Definitie 3.5 (Composite function)
Suppose f and g are two functions. The **composite** (*samenstelling*) of g with f, denoted $g \circ f$, is defined by

$$(g \circ f)(x) = g(f(x)),$$

provided $x \in \operatorname{dom} f$ and $f(x) \in \operatorname{dom} g$.

The quantity $g \circ f$ is also read g composed with f or, more simply g of f. At its most basic level, Definition 3.5 tells us to obtain the formula for $(g \circ f)(x)$, we replace every occurrence of x in the formula for $g(x)$ with the formula we have for $f(x)$. If we take a step back and look at this from a procedural, inputs and outputs perspective, Defintion 3.5 tells us the output from $g \circ f$ is found by taking the output from f, $f(x)$, and then making that the input to g. The result, $g(f(x))$, is the output from $g \circ f$. This is illustrated in Figure 3.6 for a setting where $f : x \mapsto x^2$ and $g : x \mapsto x + 1$. Clearly, the notion of function composition can easily be generalised to an arbitrary number of functions. For instance, suppose f, g and h are three functions, then we may consider

$$\big((h \circ g) \circ f\big)(x) = h\Big(g\big(f(x)\big)\Big).$$

In the expression $g(f(x))$, the function f is often called the inside function while g is often called the outside function. There are two ways to go about evaluating composite functions - inside out and outside in - depending on which function we replace with its formula first. Both ways are demonstrated in the following example.

Example 3.6

Let $f(x) = x^2 - 4x$, $\quad g(x) = 2 - \sqrt{x+3}$ and $h(x) = \dfrac{2x}{x+1}$.

Find and simplify the indicated composite functions. State the domain of each.

1. $(g \circ f)(x)$
2. $\big(h \circ (g \circ f)\big)(x)$

Solution

1. By definition, $(g \circ f)(x) = g(f(x))$. We now illustrate two ways to approach this problem.

   - Inside out: We insert the expression $f(x)$ into g first to get

   $$(g \circ f)(x) = g(f(x)) = g\left(x^2 - 4x\right) = 2 - \sqrt{\left(x^2 - 4x\right) + 3} = 2 - \sqrt{x^2 - 4x + 3}.$$

   - Outside in: We use the formula for g first to get

   $$(g \circ f)(x) = g(f(x)) = 2 - \sqrt{f(x) + 3} = 2 - \sqrt{\left(x^2 - 4x\right) + 3} = 2 - \sqrt{x^2 - 4x + 3}.$$

   We get the same answer as before.

   To find the domain of $g \circ f$, we need to find the elements in the domain of f whose outputs $f(x)$ are in the domain of g. We accomplish this by determining the domain before we simplify the formula for the composite function. To that end, we examine $(g \circ f)(x) = 2 - \sqrt{\left(x^2 - 4x\right) + 3}$. To ensure that the argument of the square root is positive, it should hold that $x^2 - 4x + 3 \geq 0$. We find the zeros of $x^2 - 4x + 3$ to be $x = 1$ and $x = 3$. Consequently, the domain of $g \circ f$, is $]-\infty, 1] \cup [3, +\infty[$. Figure 3.7 shows the graph of $f(x)$, $g(x)$ and $(g \circ f)$.

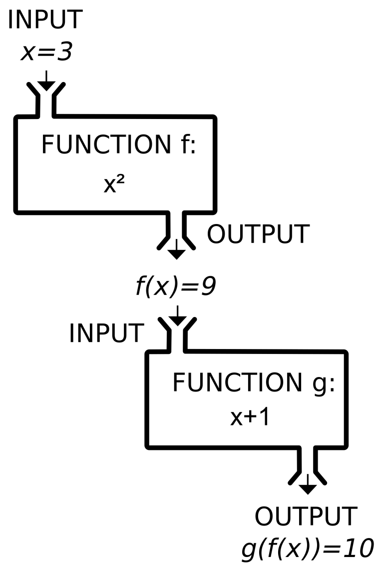


Figure 3.6: Function composition $(g \circ f)(x)$, where $f : x \mapsto x^2$ and $g : x \mapsto x+1$.

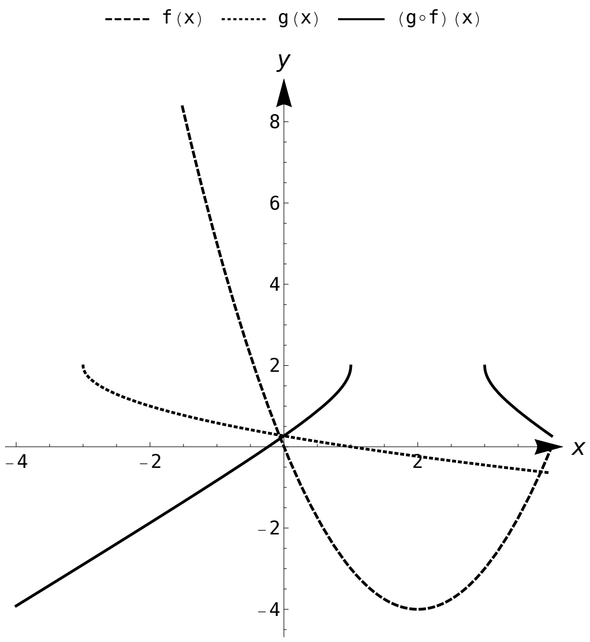


Figure 3.7: Graph of $f(x)$, $g(x)$ and $(g \circ f)$ in Example 3.6.

2. The expression $\big(h \circ (g \circ f)\big)(x)$ indicates that we first find the composite, $g \circ f$ and compose the function h with the result. We know already that $(g \circ f)(x) = 2 - \sqrt{x^2 - 4x + 3}$. We now proceed as usual.

   - Inside out: We insert the expression $(g \circ f)(x)$ into h first to get

$$\begin{aligned}\big(h \circ (g \circ f)\big)(x) &= h\big((g \circ f)(x)\big) = h\left(2 - \sqrt{x^2 - 4x + 3}\right) \\ &= \frac{2\left(2 - \sqrt{x^2 - 4x + 3}\right)}{\left(2 - \sqrt{x^2 - 4x + 3}\right) + 1} = \frac{4 - 2\sqrt{x^2 - 4x + 3}}{3 - \sqrt{x^2 - 4x + 3}}.\end{aligned}$$

- Outside in: We use the formula for $h(x)$ first to get

$$\begin{aligned}\big(h\circ(g\circ f)\big)(x) &= h\big((g\circ f)(x)\big)=\frac{2\big((g\circ f)(x)\big)}{\big((g\circ f)(x)\big)+1}\\ &= \frac{2\left(2-\sqrt{x^2-4x+3}\right)}{\left(2-\sqrt{x^2-4x+3}\right)+1}=\frac{4-2\sqrt{x^2-4x+3}}{3-\sqrt{x^2-4x+3}}.\end{aligned}$$

To find the domain of $(h\circ(g\circ f))$, we look at the step before we began to simplify,

$$(h\circ(g\circ f))(x)=\frac{2\left(2-\sqrt{x^2-4x+3}\right)}{\left(2-\sqrt{x^2-4x+3}\right)+1}.$$

For the square root, we need $x^2-4x+3\geq 0$, which requires $]-\infty,1]\cup[3,+\infty[$. Next, we set the denominator to zero and solve: $\left(2-\sqrt{x^2-4x+3}\right)+1=0$. We get $\sqrt{x^2-4x+3}=3$, and, after squaring both sides, we have $x^2-4x+3=9$. To solve $x^2-4x-6=0$, we use the quadratic formula and get $x=2\pm\sqrt{10}$. Hence we must exclude these numbers from the domain of $h\circ(g\circ f)$. Consequently our final domain for $h\circ(f\circ g)$ is

$$\left]-\infty,2-\sqrt{10}\right[\cup\left]2-\sqrt{10},1\right]\cup\left[3,2+\sqrt{10}\right[\cup\left]2+\sqrt{10},\infty\right[.$$

.

From this example, we learn that function composition is not commutative, so $(g\circ f)(x)\neq(f\circ g)(x)$, though it is associative, i.e. it holds that

$$\big((h\circ g)\circ f\big)(x)=\big(h\circ(g\circ f)\big)(x).$$

Also note the importance of finding the domain of the composite function before simplifying.

3.2.4 Function properties

When graphing functions, we will typically investigate whether or not there is some kind of symmetry to its graph, whether or not it is periodic, and so on.

3.2.4.1 Injections, surjections and bijections

Injections, surjections and bijections are classes of functions distinguished by the manner in which arguments and images are mapped to each other.

Definitie 3.6 (Injective, surjective and bijective)
A function $f: X\to Y$ is called

- **injective** (*injectief*) (one-to-one) if each element of the codomain is mapped to by at most one element of the domain. Mathematically, we may write:

$$\forall x_1,x_2\in\operatorname{dom}f: f(x_1)=f(x_2)\Rightarrow x_1=x_2.$$

This implies that two different elements belonging to the function's domain cannot be mapped to the same element in its codomain, so every straight line parallel to the x-axis intersects the graph of f in at most one point. An injective function is an **injection** (*injectie*);

- **surjective** (*surjectief*) (onto) if each element of the codomain is mapped to by at least one element of the domain. That is, the range and the codomain of the function are equal. This implies that every straight line parallel to the x-axis intersects the graph of f in at least one point. A surjective function is a **surjection** (*surjectie*); and
- **bijective** (*bijectief*) (one-to-one correspondence) if each element of the codomain is mapped to by exactly one element of the domain. So, the function is both injective and surjective. This implies that every straight line parallel to the x-axis intersects the graph of f in at exactly one point. A bijective function is a **bijection** (*bijectie*).

The four possible combinations of injective and surjective features are illustrated in Table 3.1. An injective function does not need to be surjective because not all elements of the codomain may be associated with arguments, and likewise a surjective function does not need to be injective as some images may be associated with more than one argument.

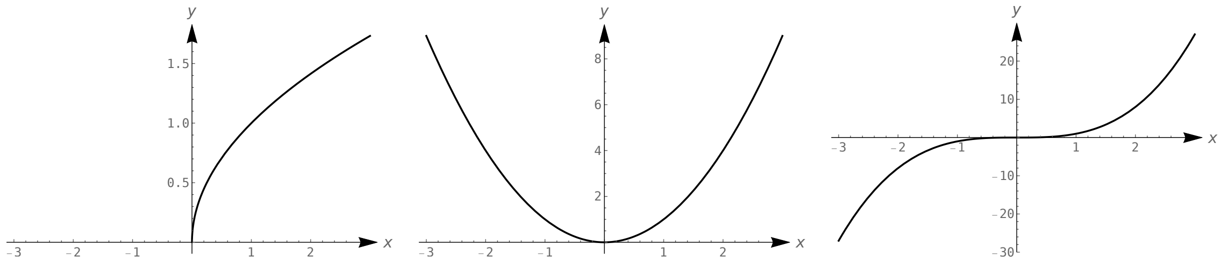
Example 3.7

Determine whether the following real functions are injections, surjections and/or bijections.

1. $f : \mathbb{R}^+ \to \mathbb{R} : x \mapsto \sqrt{x}$
2. $g : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$
3. $h : \mathbb{R} \to \mathbb{R} : x \mapsto x^3$

Solution — Figure 3.8 depicts the graphs of the considered functions, which were generated in Mathematica.

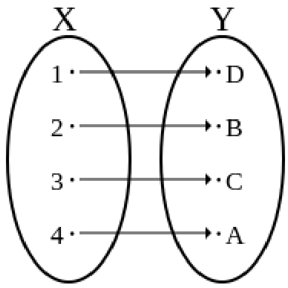
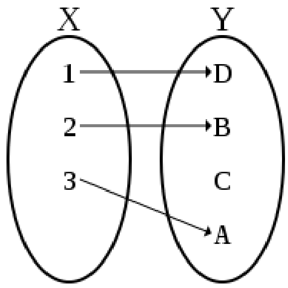
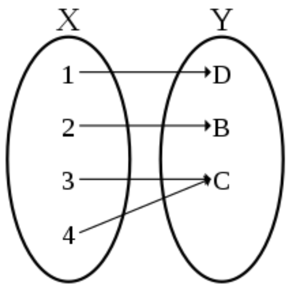
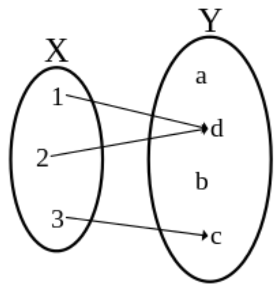
1. For any two elements x_1 and x_2 of this function's domain, it holds that if $\sqrt{x_1} = \sqrt{x_2}$, then $x_1 = x_2$. This means that every straight line parallel to the x-axis intersects with the function's graph in at most one point (Figure 3.8(a)), so the function f is an injection. Since the range of this function is restricted to $\mathbb{R}^+$, whereas its codomain is $\mathbb{R}$, this function is non-surjective, and hence it cannot be a bijection.
2. Since we have for any two elements x_1 and x_2 of this function's domain that if $x_1^2 = x_2^2$ then $x_1 = \pm x_2$, this function is non-injective. Besides, it is neither a surjection because its range is $\mathbb{R}^+$, whereas its codomain is $\mathbb{R}$. Consequently, it is not a bijection (Figure 3.8(b)).
3. For any two elements x_1 and x_2 of this function's domain, it holds that if $x_1^3 = x_2^3$, then $x_1 = x_2$. So the function h is an injection. Moreover, it is also a surjection because its range is $\mathbb{R}$; that is every straight line parallel to the x-axis intersects with the function's graph in at least one point. Since h is both an injection and surjection, it is a bijection (Figure 3.8(c)).



(a) Graph of f (b) Graph of g (c) Graph of h

Figure 3.8: Graphs of the functions f, g and h in Example 3.7.

Table 3.1: The four possible combinations of injective and surjective features.

	Surjective	Non-surjective
Injective		
Non-injective		

3.2.4.2 Symmetry

Of the three symmetries discussed in Section 3.2.1, only two are of significance to functions: symmetry about the y-axis and symmetry about the origin.

Definitie 3.7 (Even/odd functions)
A function f is

- an **even** (*even*) function if and only if $f(-x) = f(x)$ for all $x \in \operatorname{dom} f$. The graph of f is symmetric about the y-axis.
- an **odd** (*oneven*) function if and only if $-f(-x) = f(x)$, or, equivalently, $f(-x) = -f(x)$ for all $x \in \operatorname{dom} f$. The graph of f is symmetric about the origin.

Example 3.8

Determine analytically if the following functions are even, odd, or neither even nor odd. Verify your result with Mathematica.

1. $f(x) = \dfrac{5}{2-x^2}$
2. $g(x) = \dfrac{5x}{2-x^2}$

Solution

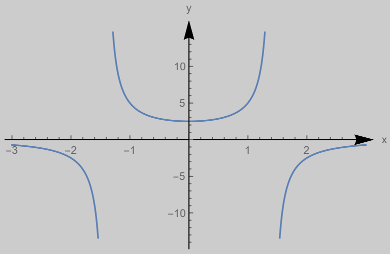
1. The first step is to replace x with $-x$ and simplify.

$$f(-x) = \frac{5}{2-(-x)^2} = \frac{5}{2-x^2} = f(x)$$

Hence, f is even. This conclusion can be verified in Mathematica using the function **`Plot`**.

In[3]:= `Plot[5/(2-x^2),{x,-3,3},AxesLabel→{"x","y"},AxesStyle→Arrowheads[{0,0.05`}]]`

Out[3]=

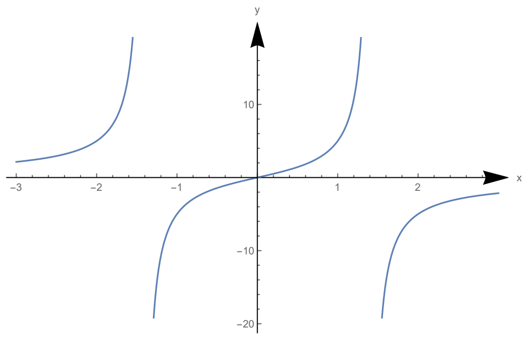


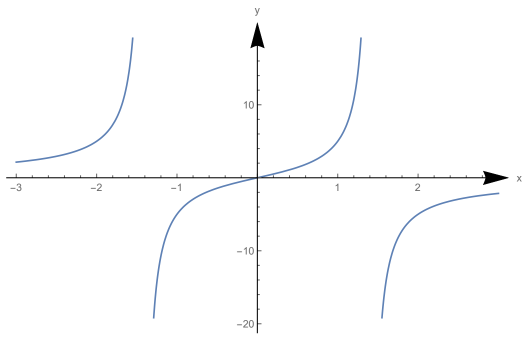
Here, the option **`AxesLabel`** was used to label the axes, while using the option `Arrowheads` we add arrowheads to these axes.

2. Again, we replace x with $-x$ and simplify.

$$g(-x) = \frac{5(-x)}{2-(-x)^2} = \frac{-5x}{2-x^2} = -g(x)$$

Clearly, g is odd. This is confirmed by the graph of this function generated in Mathematica:





3.2.4.3 Periodicity

Definitie 3.8 (Periodic function)
A function f is said to be **periodic** (*periodiek*) with **period** (*periode*) P ($P \in \mathbb{R}_0^+$), if

$$f(x+P) = f(x),$$

for all $x \in \operatorname{dom} f$. If there exists a least positive constant P with this property, it is called the **fundamental period**.

A function with period P will repeat on intervals of length P, and these intervals are referred to as **periods**.

3.2.4.4 Function behaviour

As you shall see in Chapters 4 and 5, each family of functions has its own unique attributes and we will study them all in great detail. The purpose of this section is to lay the foundation for that further study by investigating aspects of function behaviour which apply to all functions. To start, we will examine the concepts of **increasing** (*stijgend*), **decreasing** (*dalend*) and **constant** (*constant*). Before defining the concepts algebraically, it is instructive to first look at them graphically. For that purpose, consider the graph of a function f in Figure 3.9.

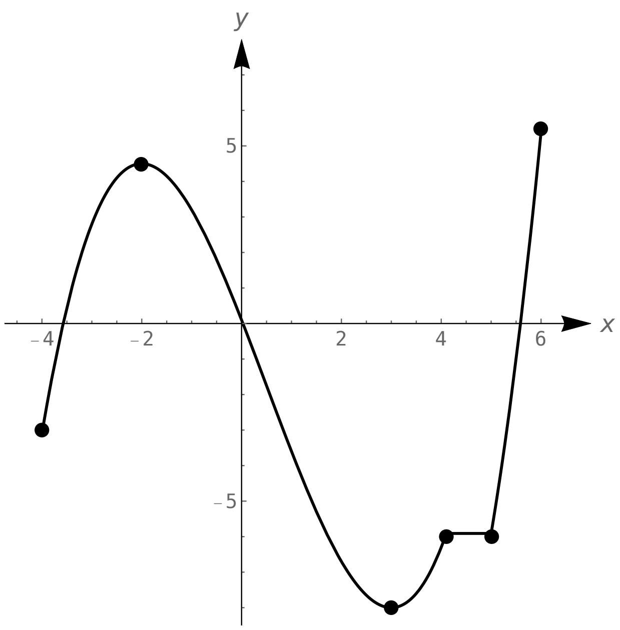


Figure 3.9: The graph of $y = f(x)$.

For the x values between -4 and -2 (inclusive), the y-coordinates on the graph are increasing, as we move from left to right. Hence the function f is increasing on the interval $[-4,-2]$. Analogously, we say that f is decreasing on the interval $[-2,3]$, increasing once more on the interval $[3,4]$, constant on $[4,5]$, and finally increasing once again on $[5,6]$.

Let us now introduce more formal algebraic definitions of what it means for a function to be increasing, decreasing or constant.

Definitie 3.9 (Function behaviour)
Suppose f is a function defined on an interval $I \subset \operatorname{dom} f$. We say f is:

- **increasing** on I if and only if $\forall x_1, x_2 \in I \mid x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$.
- **decreasing** on I if and only if $\forall x_1, x_2 \in I \mid x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2)$.
- **constant** on I if and only if $\forall x_1, x_2 \in I \mid x_1 < x_2 \Rightarrow f(x_1) = f(x_2)$.

If the order $\leq$ in the definition of an increasing function is replaced by $<$, we say that f is **strictly increasing** (*strikt stijgend*) on the interval I, and likewise for a **strictly decreasing** (*strikt dalend*) function. Clearly, if f is either strictly increasing or decreasing on an interval I, it must hold that f is an injective function .

We say that functions are **monotonically** (*monotoon*) increasing or decreasing on the interval I if they are entirely non-decreasing or entirely non-increasing, respectively. For instance, a function that increases monotonically does not exclusively have to increase, it simply must not decrease (Figure 3.10(a)). On the other hand, a function is **strictly monotone** (*strikt monotoon*) on the interval I if it is either strictly increasing or decreasing on that interval.

Acknowledging that functions are just a special type of set, it is of course meaningful to discuss the boundedness of functions, just as we did in Section 2.1 for sets.

Definitie 3.10 (Boundedness of functions)
Let f be a real function and $A \subset \operatorname{dom} f$. We call f

(i) **bounded above** on A if $\exists b \in \mathbb{R} \mid \forall x \in A \mid f(x) \leq b$.

(ii) **bounded below** on A if $\exists a \in \mathbb{R} \mid \forall x \in A \mid a \leq f(x)$.

(iii) **bounded** on A if $\exists r \in \mathbb{R} \mid \forall x \in A \mid |f(x)| \leq r$.

Intuitively, the graph of a bounded function stays within a horizontal band, while the graph of an unbounded function does not.

In conjunction with the supremum and infimum of ordered sets (Definition 2.3) this definition leads to the following properties for two real-valued functions $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ that are bounded on A.

(i) If $f(x) \leq g(x)$ for all $x \in A$, then $\sup_{x \in A} f(x) \leq \sup_{x \in A} g(x)$.

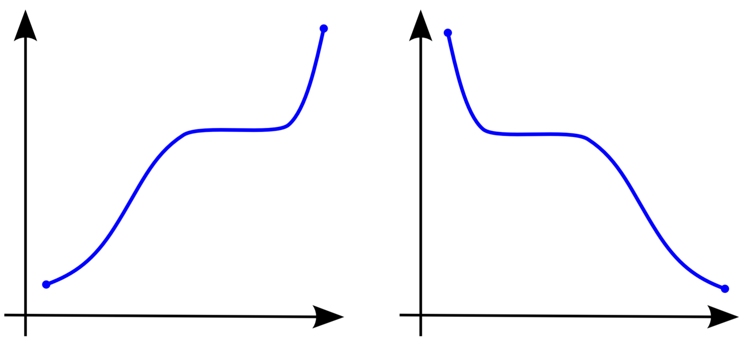
(ii) If $f(x) \leq g(y)$ for all $x \in A$ and all $y \in A$, then $\sup_{x \in A} f(x) \leq \inf_{y \in A} g(y)$.

Besides, it can be shown that the following hold:

(i) $\sup\{f(x) + g(x) \mid x \in A\} \leq \sup\{f(x) \mid x \in A\} + \sup\{g(x) \mid x \in A\}$;

(ii) $\inf\{f(x) + g(x) \mid x \in A\} \geq \inf\{f(x) \mid x \in A\} + \inf\{g(x) \mid x \in A\}$.

Now let us turn our attention to a few of the points on the graph in Figure 3.9. Clearly, the point $(-2, 4.5)$ does not have the largest y-value of all of the points on the graph of f but $(-2, 4.5)$ is on the top of the hill between $x = -4$ and $x = 3$. We say that the function f has a **local maximum** (*lokaal maximum*) at the point $(-2, 4.5)$, because the y-coordinate 4.5 is the largest y-value (hence, function value) on the curve near $x = -2$. Similarly, we say that the function f has a **local minimum** (*lokaal minimum*) at the point $(3, -8)$, since the y-coordinate -8 is the smallest function value near $x = 3$.



(a) (b)

Figure 3.10: Graph of a monotonically increasing (a) and decreasing (b) function.

If we look at the entire graph, we see that the largest y-value (the largest function value) is 5.5 at $x = 6$. In this case, we say the **maximum** (*maximum*) of f is 5.5, sometimes also called the absolute or global maximum. Similarly, the **minimum** (*minimum*) of f is -8. This is also sometimes referred to as the absolute or global maximum.

We formalize these concepts in the following definitions.

Definitie 3.11 (Local and global extrema)
Suppose f is a function with $f(a) = b$.

- We say f has a **local maximum** at $a \in \operatorname{dom} f$ if and only if

$$\exists \delta \in \mathbb{R}_0^+ : \forall x \in]a-\delta, a+\delta[\cap \operatorname{dom} f : f(x) \leq f(a).$$

  The value $f(a) = b$ is called a local maximum value of f in this case.

- We say f has a **local minimum** at $a \in \operatorname{dom} f$ if and only if

$$\exists \delta \in \mathbb{R}_0^+ : \forall x \in]a-\delta, a+\delta[\cap \operatorname{dom} f : f(x) \geq f(a).$$

  The value $f(a) = b$ is called a local minimum value of f in this case.

- The value b is called the **(absolute or global) maximum** of f if $\forall x \in \operatorname{dom} f : b \geq f(x)$. We may write

$$b = \max f.$$

- The value b is called the **(absolute or global) minimum** of f if $\forall x \in \operatorname{dom} f : b \leq f(x)$. We may write

$$b = \min f.$$

It is important to note that not every function will have all of these features. Indeed, it is possible to have a function with no local or absolute extrema at all!

3.2.5 Transformations

We may change or transform the graphs of functions by making certain modifications to their formulas. The transformations we will study fall into three broad categories: shifts, reflections and scalings. Suppose the graph in Figure 3.11 is the complete graph of a function f.

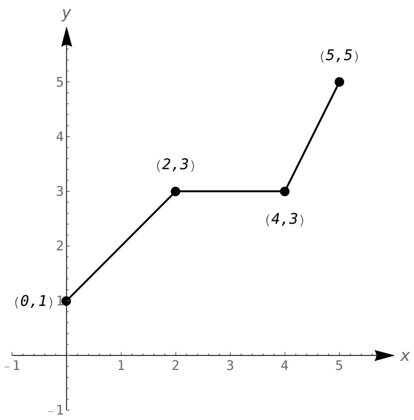


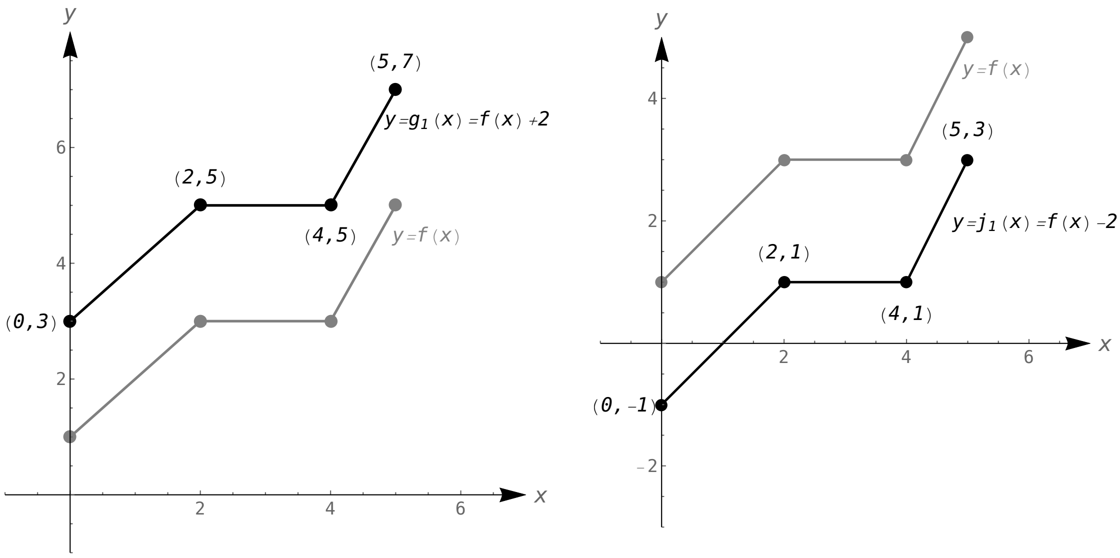
Figure 3.11: The graph of $y = f(x)$.

Suppose we wanted to graph the function defined by the formula $g_1(x) = f(x) + 2$. In order to graph g_1, we need to graph the points $(x, g_1(x))$. For example, using the points indicated on the graph of f, we can make the following table.

x	$f(x)$	$g_1(x) = f(x) + 2$
0	1	3
2	3	5
4	3	5
5	5	7

Hence, to obtain the graph of g_1, we just add 2 to the y-coordinate of each point on the graph of f (Figure 3.12(a)). Geometrically, we are 'shifting the graph up 2 units'. It is important to note that the domain of f and the domain of g are the same, but that the range of f is $[1, 5]$ while the range of g_1 is $[3, 7]$. In general, **shifting a function vertically** (*verticale verschuiving*) like this will leave the domain unchanged, but could very well affect the range.

You can easily imagine what would happen if we wanted to graph the function $j_1(x) = f(x) - 2$. Geometrically, we would then shift the graph down 2 units (Figure 3.12(b)).



(a) $y = g_1(x) = f(x) + 2$

(b) $y = j_1(x) = f(x) - 2$

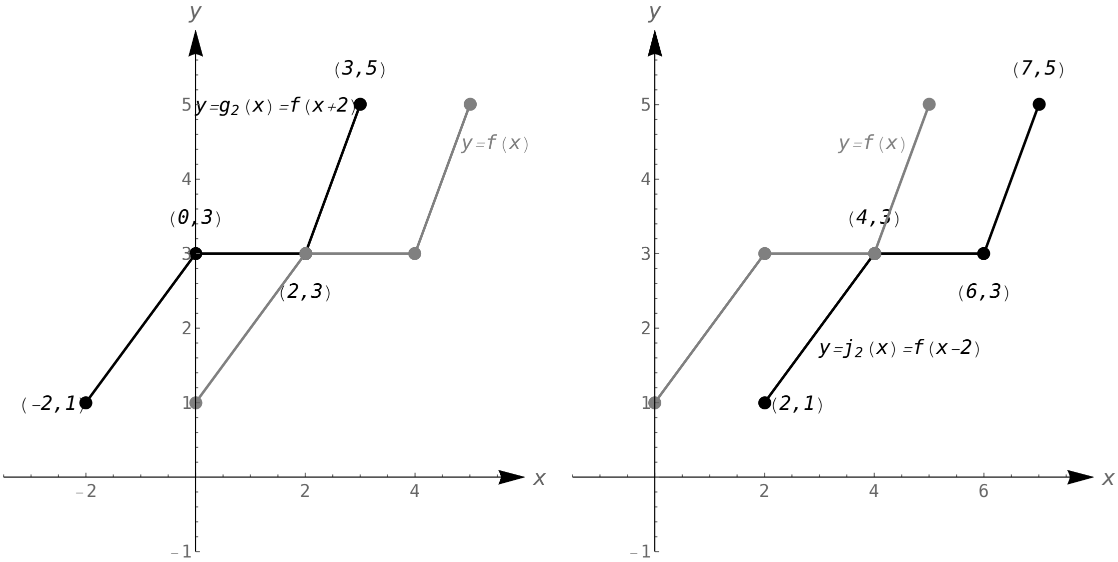
Figure 3.12: Vertical shifts of the graph of $y = f(x)$: two units up (a) and down (b).

Now, what happens if we add to or subtract from the input of the function? For instance, suppose we wanted to graph $g_2(x) = f(x + 2)$. We know, for instance, $f(0) = 1$. To determine the corresponding point on the graph of g_2, we need to figure out what value of x we must substitute into $g_2(x) = f(x + 2)$ so that the quantity $x + 2$, works out to be 0. Solving $x + 2 = 0$ gives $x = -2$, so $(-2, 1)$ is on the graph of g_2. Continuing in this fashion, we get the following table.

x	$g_2(x)$	$=$	$f(x+2)$
-2	$g_2(-2)$	$=$	$f(0) = 1$
0	$g_2(0)$	$=$	$f(2) = 3$
2	$g_2(2)$	$=$	$f(4) = 3$
3	$g_2(3)$	$=$	$f(5) = 5$

In summary, the points $(0,1)$, $(2,3)$, $(4,3)$ and $(5,5)$ on the graph of $y=f(x)$ give rise to the points $(-2,1)$, $(0,3)$, $(2,3)$ and $(3,5)$ on the graph of $y=g_2(x)$, respectively. In general, if (a,b) is on the graph of $y=f(x)$, then $(a-2,b)$ is on the graph of $y=g_2(x)$. The point $(a-2,b)$ is exactly 2 units to the left of the point (a,b) so the graph of $y=g_2(x)=f(x+2)$ is obtained by shifting the graph $y=f(x)$ to the left 2 units (Figure 3.13(a)).

Note that while the ranges of f and g_2 are the same, the domain of g_2 is $[-2,3]$ whereas the domain of f is $[0,5]$. In general, when we **shift the graph horizontally** (*horizontale verschuiving*), the range will remain the same, but the domain could change. Similarly, if we set out to graph $j_2(x)=f(x-2)$, we would effect a shift to the right 2 units (Figure 3.13(b)).



(a) $y=g_2(x)=f(x+2)$ (b) $y=j_2(x)=f(x-2)$

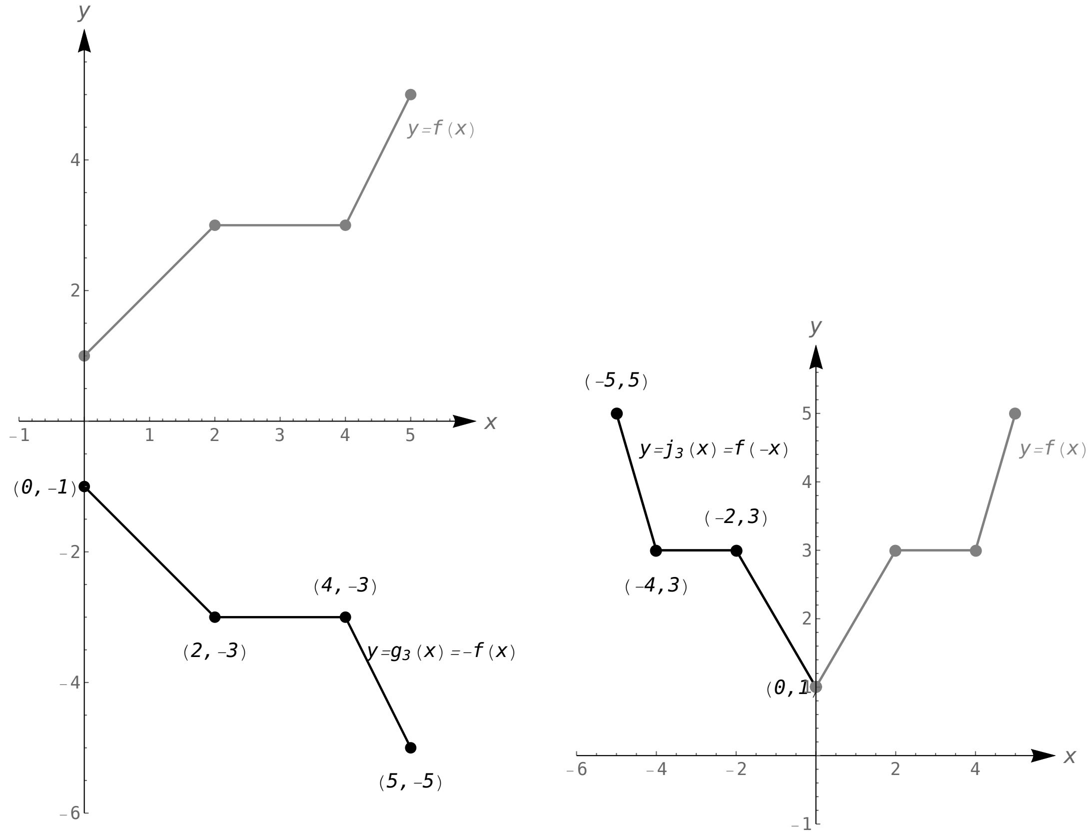
Figure 3.13: Horizontal shifts of the graph of $y=f(x)$: two units to the left (a) and right (b).

We now turn our attention to **reflections** (*spiegeling*). We know from Section 3.1 that the graph of $y=-f(x)$ is the graph of f reflected across the x-axis (Figure 3.14(a)). Similarly, the graph of $y=f(-x)$ is the graph of f reflected across the y-axis (Figure 3.14(b)).

Finally, we turn our attention to our last class of transformations known as **scalings** (*schaling*). A thorough discussion of scalings can get complicated because they are not as straightforward as the previous transformations. The transformations covered so far are known as **rigid transformations** (*directe isometrie*) because they do not change the shape of the graph, only its position and orientation in the plane. If, however, we wanted to make a new graph twice as tall as a given graph, we would be changing the shape of the graph. This type of transformation is called **non-rigid** (*indirecte isometrie*). Not only will it be important for us to differentiate between modifying inputs versus outputs, we must also pay close attention to the magnitude of the changes we make.

Suppose we wish to graph the function $g_4(x)=2f(x)$. From its graph, we can build a table of values for g_4 as before.

x	$f(x)$	$g_4(x)=2f(x)$
0	1	2
2	3	6
4	3	6
5	5	10



(a) $y = g_3(x) = -f(x)$ (b) $y = j_3(x) = f(-x)$

Figure 3.14: Reflections of the graph of $y = f(x)$: across x-axis (a) and y-axis (b).

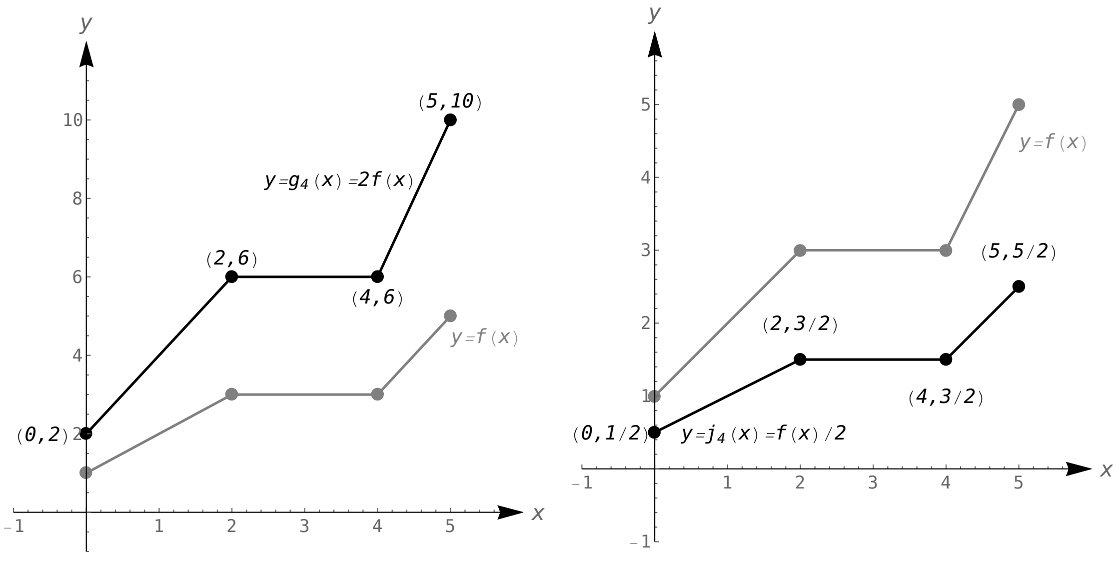
If (a, b) is on the graph of f, then $(a, 2b)$ is on the graph of g_4. In other words, to obtain the graph of g_4, we multiply all of the y-coordinates of the points on the graph of f by 2. This is known as a vertical scaling by a factor of 2 (Figure 3.15(a)). Likewise, if we wish to graph $y = \frac{1}{2}f(x)$, we multiply all of the y-coordinates of the points on the graph of f by $\frac{1}{2}$. This creates a vertical scaling by a factor of $\frac{1}{2}$ (Figure 3.15(b)).

In general, suppose f is a function and $a > 0$, then to obtain the graph of $y = af(x)$, we have to vertically scale the graph of f by a factor of a.

- If $a > 1$, we say the graph of f has undergone a **vertical stretching (expansion, dilation)** by a factor of a.
- If $0 < a < 1$, we say the graph of f has undergone a **vertical shrinking (compression, contraction)** by a factor of $\frac{1}{a}$.

In terms of inputs and outputs, multiplying the outputs from a function by positive number a causes the graph to be vertically scaled by a factor of a. It is natural to ask what would happen if we multiply the inputs of a function by a positive number. This leads us to our last transformation.

Suppose we want to graph $g_5(x) = f(2x)$. If we want to determine the point on g_5 which corresponds to the point $(2, 3)$ on the graph of f, we set $2x = 2$ so that $x = 1$. Substituting $x = 1$ into $g_5(x)$, we obtain $g_5(1) = f(2 \cdot 1) = f(2) = 3$, so that $(1, 3)$ is on the graph of g_5. Continuing in this fashion, we obtain the following table.



(a) $y = g_4(x) = 2f(x)$ (b) $y = j_4(x) = \frac{1}{2}f(x)$

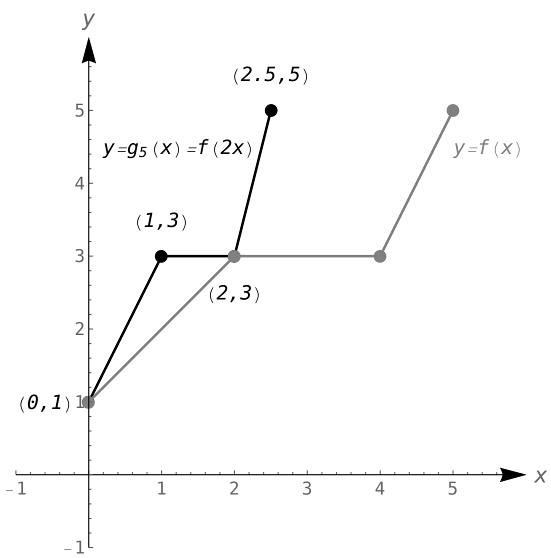
Figure 3.15: Vertical scalings of the graph of $y = f(x)$: stretching by a factor 2 (a); shrinking by a factor 1/2 (b).

x	$g_5(x)$	$=$	$f(2x)$
0	$g_5(0)$	$=$	$f(0) = 1$
1	$g_5(1)$	$=$	$f(2) = 3$
2	$g_5(2)$	$=$	$f(4) = 3$
$\frac{5}{2}$	$g_5(5/2)$	$=$	$f(5) = 5$

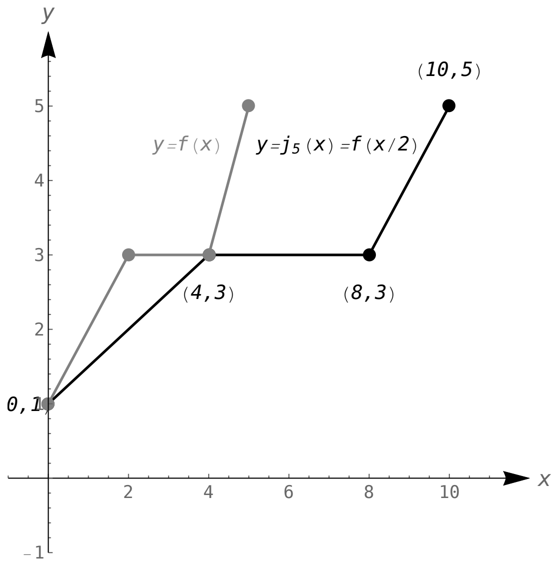
In general, if (a, b) is on the graph of f, then $\left(\frac{a}{2}, b\right)$ is on the graph of g. This results in a horizontal scaling by a factor of $\frac{1}{2}$ (Figure 3.16(a)). If, on the other hand, graphing $y = f\left(\frac{1}{2}x\right)$, results in a horizontal scaling by a factor of 2 (Figure 3.16(b)).

In general, suppose f is a function and $b > 0$, then to obtain the graph of $y = f(bx)$, we have to horizontally scale the graph of f by a factor of $\frac{1}{b}$.

- If $0 < b < 1$, we say the graph of f has undergone a **horizontal stretching (expansion, dilation)** by a factor of $\frac{1}{b}$.
- If $b > 1$, we say the graph of f has undergone a **horizontal shrinking (compression, contraction)** by a factor of b.



(a) $y = g_5(x) = f(2x)$



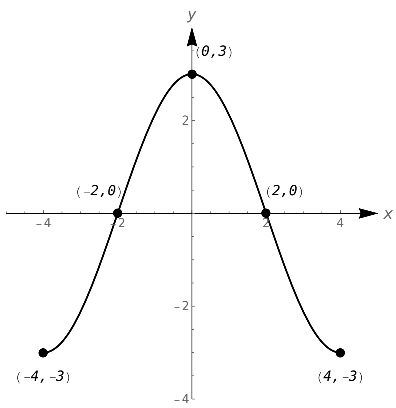
(b) $y = j_5(x) = f(\frac{x}{2})$

Figure 3.16: Horizontal scalings of the graph of $y = f(x)$: shrinking by a factor 2 (a); streching by a factor 1/2 (b).

Example 3.9

Below is the complete graph of $y = f(x)$. Use it to graph

$$g(x) = \frac{4-3f(1-2x)}{2}.$$



Solution

We track the five key points $(-4,-3)$, $(-2,0)$, $(0,3)$, $(2,0)$ and $(4,-3)$ indicated on the graph of f to their new locations. We first rewrite $g(x)$ as

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

Let us first focus on $f(-2x+1)$. To get from $f(x)$ to $f(-2x+1)$, we need a horizontal shift with one unit, i.e. $f(x+1)$, followed by a horizontal shrinking by a factor of 2, i.e. $f(2x+1)$, followed on its turn by a reflection across the y-axis. So, we set $-2x+1$ equal to the x-coordinates of the key points and solve. For example, solving $-2x+1=-4$, we get $x=\frac{5}{2}$. We summarize the results in the table below.

a	-4	-2	0	2	4
$x = \frac{a-1}{2}$	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$

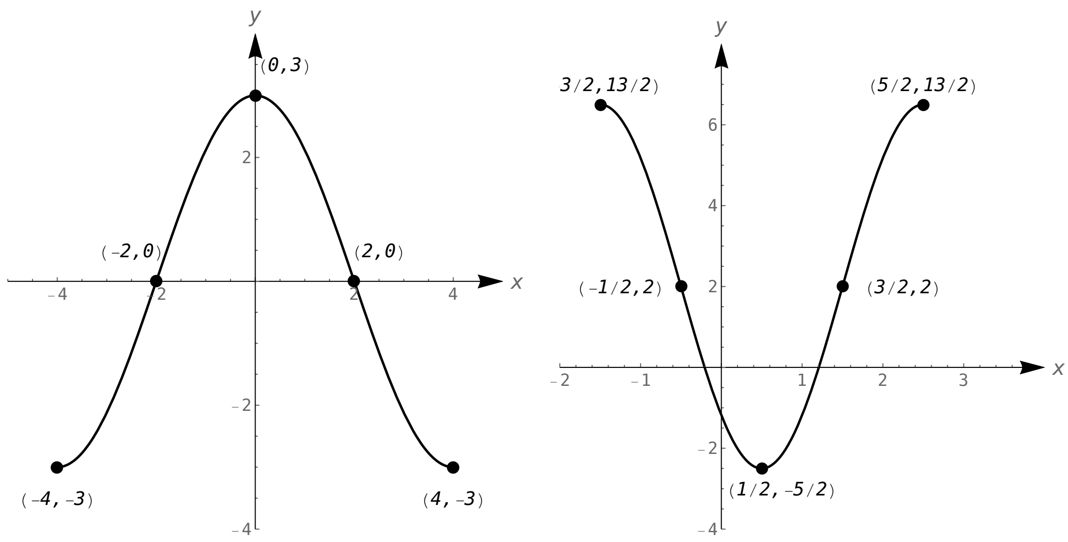
Next, we take each of the x values and substitute them into $g(x) = -\frac{3}{2}f(-2x+1)+2$ to get the corresponding y-values. Substituting $x = \frac{5}{2}$, and using the fact that $f(-4) = -3$, we get

$$g\left(\frac{5}{2}\right) = -\frac{3}{2}f\left(-2\left(\frac{5}{2}\right)+1\right)+2 = -\frac{3}{2}f(-4)+2 = -\frac{3}{2}(-3)+2 = \frac{9}{2}+2 = \frac{13}{2}.$$

We see that the output from f is first multiplied by $-\frac{3}{2}$. Thinking of this as a two step process, multiplying by $\frac{3}{2}$ and then by -1, we have a vertical stretching by a factor of $\frac{3}{2}$ followed by a reflection across the x-axis. Adding 2 results in a vertical shift up 2 units. Continuing in this manner, we get the following table.

x	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$
$g(x)$	$\frac{13}{2}$	2	$-\frac{5}{2}$	2	$\frac{13}{2}$

To graph g, we plot each of the points in the table above and connect them in the same order and fashion as the points to which they correspond (Figure 3.17(b)). The reader is strongly encouraged to graph the series of functions which shows the gradual transformation of the graph of f into the graph of g.



(a) $f(x)$ (b) $g(x) = \frac{4-3f(1-2x)}{2}$

Figure 3.17: The graph of the original function $f(x)$ (a), alongside the transformed function $g(x)$ (b).

3.2.6 Piecewise-defined functions

In many applications, one will encounter functions that are defined on a sequence of intervals. Such functions are referred to as **piecewise-defined functions**, or **piecewise functions** (*stuksgewijze functie*) for short. For instance,

$$f(x) = \begin{cases} (x+1)^2, & \text{if} \quad x < -1, \\ -x, & \text{if} \quad -1 \leq x < 1, \\ \sqrt{x-1}, & \text{if} \quad x \geq 1, \end{cases} \tag{3.1}$$

is a piecewise function. Its graph is given in Figure 3.18

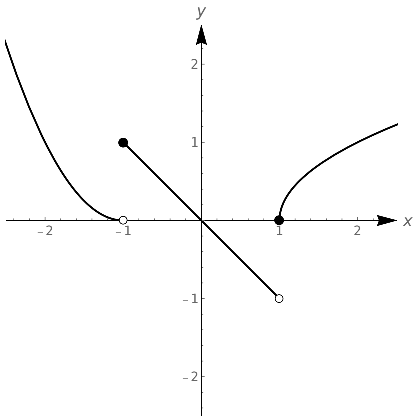


Figure 3.18: The graph of Equation (3.1).

3.2.7 Function families

Throughout the remainder of this course we will focus our attention on the so-called **elementary functions** (*elementaire functie*), which are functions that are compositions of a finite number of arithmetic operations, exponentials, logarithms, constants, and solutions of algebraic equations. Two important families can be distinguished among the elementary functions, namely the **algebraic** (*algebraïsche*) and **transcendental functions** (*transcendente functie*).

An algebraic function is a function that can be defined as the root of a polynomial equation. Quite often algebraic functions are algebraic expressions using a finite number of terms, involving only the algebraic operations addition, subtraction, multiplication, division, and raising to a fractional power. Examples of such functions are:

- power functions, e.g. $f(x) = 2x^3$,
- polynomial functions, e.g. $f(x) = 1 + x + x^3$,
- rational functions, e.g.
$$f(x) = \frac{1+x}{1+x+x^3},$$
- irrational functions, e.g. $f(x) = \sqrt{1+x+x^3}$,
- and any compositions thereof, e.g.
$$f(x) = \frac{1+x}{\sqrt{1+x+x^3}}.$$

A transcendental function is a function that does not satisfy a polynomial equation, in contrast to an algebraic function. In other words, a transcendental function transcends algebra in that it cannot be expressed in terms of a finite sequence of the algebraic operations of addition, multiplication, and root extraction. Examples of transcendental functions include

- exponential functions, e.g. $f(x) = 2^x$,
- logarithmic functions, e.g. $f(x) = \ln(x)$,
- trigonometric functions, e.g. $f(x) = \sin(x)$,
- hyperbolic functions, e.g. $f(x) = \sinh(x)$,
- and most compositions thereof, e.g.
$$f(x) = \frac{2^x}{\sin(x)}\,.$$

Algebraic and transcendental functions are studied in detail in Chapter 4 and 5, respectively.

3.3 Absolute value functions

3.3.1 Definition and properties

Throughout this course we adopt the following definition of the absolute value.

Definitie 3.12 (Absolute value)
The **absolute value** (*absolute waarde*) of a real number x, denoted $|x|$, is given by

$$|x| = \begin{cases} -x, & \text{if} \quad x < 0, \\ x, & \text{if} \quad x \geq 0. \end{cases}$$

In this definition, we define $|x|$ using a piecewise-defined function. Other ways to define the absolute value are that $|x|$ is the distance from the real number x to 0 on the number line, or by the equation $|x| = \sqrt{x^2}$. We first remind ourselves of the properties of the absolute value.

Let a, b and x be real numbers and let n be an integer. Then the following arithmetic properties hold.

- **Product rule:**
$$|ab| = |a||b|,$$
- **Power rule:**
$$\left|a^n\right| = |a|^n,$$
whenever a^n is defined,
- **Quotient rule:**
$$\left|\frac{a}{b}\right| = \frac{|a|}{|b|},$$
provided $b \neq 0$.

Besides, we have that $|x| = 0$ if and only if $x = 0$, and

- for $c > 0$, $|x| = c$ if and only if $x = c$ or $-x = c$,
- for $c < 0$, $|x| = c$ has no solution.

Finally, we have the following important theorem.

Theorem 3.2 (The triangle inequality)
If a and b are any two real numbers, then

$$|a+b| \leq |a| + |b|. \tag{3.2}$$

The triangle inequality appears in various forms in many contexts. It is the most important inequality in mathematics. We will use it often.

3.3.2 Absolute value functions

Next, we turn our attention to graphing **absolute value functions** (*absolute waarde functie*). Our strategy in the next example is to make liberal use of Definition 3.12.

Example 3.10

Graph each of the following functions.

1. $f(x) = |x|$
2. $h(x) = \dfrac{|x|}{x}$

Find the zeros of each function and the x- and y-intercepts of each graph, if any exist. From the graph, determine the domain and range of each function, list the intervals on which the function is increasing, decreasing or constant, and find the relative and absolute extrema, if they exist.

Solution

1. To find the zeros of f, we set $f(x) = 0$. We get $|x| = 0$, which gives us $x = 0$. So we get $(0, 0)$ as our x-intercept. To find the y-intercept, we set $x = 0$, and find $y = f(0) = 0$, so that $(0, 0)$ is our y-intercept as well. Using Definition 3.12, we get

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

   Hence, for $x < 0$, we are graphing the line $y = -x$; for $x \geq 0$, we have the line $y = x$. In this way, we get the graph shown in Figure 3.19(a).

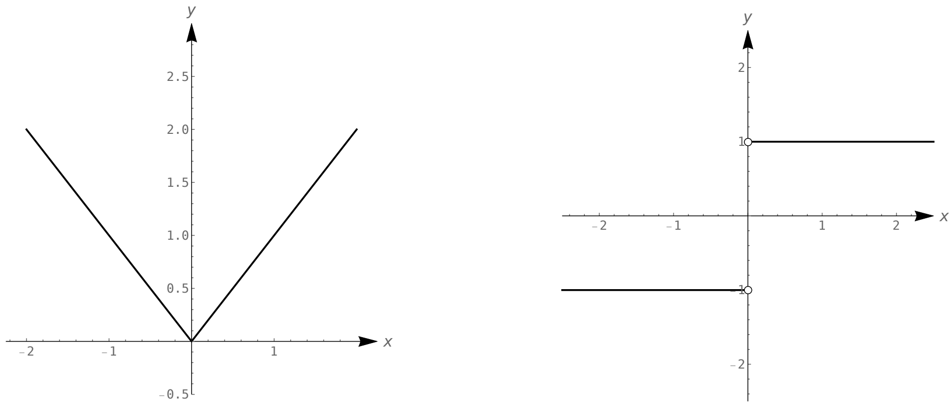
   By projecting the graph to the x-axis, we see that the domain is $\mathbb{R}$. Projecting to the y-axis gives us the range $[0, +\infty[$. The function is increasing on $[0, +\infty[$ and decreasing on $]-\infty, 0]$. The relative minimum value of f is the same as the absolute minimum, namely 0 which occurs at $(0, 0)$. There is no relative maximum value of f. There is also no absolute maximum value of f, since the y-values on the graph extend infinitely upwards.

2. We first note that, due to the fraction in the formula of $h(x)$ it should hold that $x \neq 0$. Thus the domain is $\mathbb{R}_0$. To find the zeros of h, we set $h(x) = \frac{|x|}{x} = 0$. This last equation implies $|x| = 0$, which implies $x = 0$. However, $x = 0$ is not in the domain of h, which means we have,

in fact, no x-intercepts. We have no y-intercepts either, since $h(0)$ is undefined. Re-writing the absolute value in the function gives

$$h(x) = \begin{cases} \dfrac{-x}{x}, & \text{if} \quad x < 0, \\ \dfrac{x}{x}, & \text{if} \quad x > 0, \end{cases} = \begin{cases} -1, & \text{if} \quad x < 0, \\ 1, & \text{if} \quad x > 0. \end{cases}$$

To graph this function, we graph two horizontal lines: $y = -1$ for $x < 0$ and $y = 1$ for $x > 0$. We have open circles at $(0, -1)$ and $(0, 1)$ because the domain of h excludes 0. The range consists of just two y-values: $\{-1, 1\}$. The function h is constant on $]-\infty, 0[$ and $]0, +\infty[$. The local minimum value of h is the absolute minimum value of h, namely -1; the local maximum and absolute maximum values for h also coincide: they both are 1. Every point on the graph of h is simultaneously a relative maximum and a relative minimum.



(a) (b)

Figure 3.19: Graph of $f(x) = |x|$ (a) and $h(x) = \frac{|x|}{x}$ (b).

For what concerns inequalities involving the absolute value, we have the following properties, which follow easily from Definition 3.12.

- If $c > 0$, then $|x| < c$ is equivalent to $-c < x < c$.
- If $c > 0$, then $|x| \leq c$ is equivalent to $-c \leq x \leq c$.
- If $c \leq 0$, then $|x| < c$ has no solution, while if $c < 0$, then $|x| \leq c$ has no solution.
- If $c \geq 0$, then $|x| > c$ is equivalent to $x < -c$ or $x > c$.
- If $c \geq 0$, then $|x| \geq c$ is equivalent to $x \leq -c$ or $x \geq c$.
- If $c < 0$, then $|x| > c$ and $|x| \geq c$ are true for all real numbers.

We can understand each of these statements graphically, so do not learn them by heart. For instance, if $c > 0$, the graph of $y = c$ is a horizontal line which lies above the x-axis through $(0, c)$. Essentially, to solve $|x| < c$, we are looking for the x values where the graph of $y = |x|$ is below the graph of $y = c$. Both graphs are shown in Figure 3.20. We know that the graphs intersect when $|x| = c$, which happens when $x = c$ or $x = -c$. We see that the graph of $y = |x|$ is below $y = c$ for x between $-c$ and c, and hence we get $|x| < c$ is equivalent to $-c < x < c$. The other properties can be shown similarly.

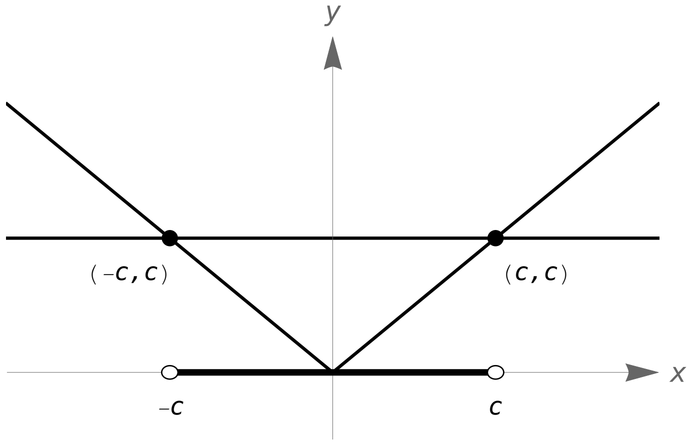


Figure 3.20: The graphs of $y = c$ ($c > 0$) and $y = |x|$.

3.4 Inverse functions

3.4.1 Definition and properties

We can define an **inverse function** (*inverse functie*) as follows.

Definitie 3.13 (Inverse function)
Consider a function $f : X \to Y$, defined by

$$f = \{(x, y) \mid x \in X \wedge y = f(x)\} .$$

Then, the relation

$$f^{-1} = \{(y, x) \mid x \in X \wedge y = f(x)\}$$

is the inverse relation f^{-1} of the function f (Figure 3.21). If and only if this inverse relation is a function on the range Y, this inverse relation f^{-1} is the **inverse of** f (*inverse van* f) and the function f is called **invertible** (*inverteerbaar*).

At this point it is important to recall that a relation constitutes a function if and only if each x-coordinate is matched with at most one y-coordinate (Definition 3.3).

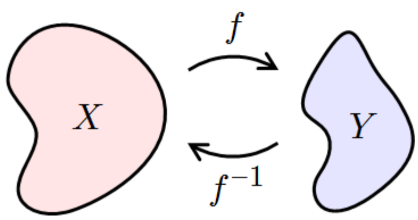


Figure 3.21: If f maps X to Y, then f^{-1} maps Y back to X.

Suppose now f and f^{-1} are inverse functions, then we obviously have the following properties.

- The range of f is the domain of f^{-1} and the domain of f is the range of f^{-1}, i.e.

$$\operatorname{im} f = \operatorname{dom} f^{-1} \quad \text{and} \quad \operatorname{dom} f = \operatorname{im} f^{-1}.$$

- $(f^{-1} \circ f)(x) = x$ for all x in dom f and $(f \circ f^{-1})(x) = x$ for all x in dom f^{-1}.
- $\left(f^{-1}\right)^{-1}(x) = f(x)$ for all x in dom f.
- $f(a) = b$ if and only if $f^{-1}(b) = a$.
- (a, b) is on the graph of f if and only if (b, a) is on the graph of f^{-1}.

The last property tells us that the graphs of inverse functions are reflections about the line $y = x$.

The following theorem guarantees that there exists exactly one inverse function f^{-1} for an invertible function f.

Theorem 3.3 (Uniqueness of an inverse function)
Let $f : X \to Y$ be a function. Then, if f has an inverse function, then that inverse function is unique.

Proof This theorem can be shown by starting from the assumption that we have two inverse functions of an invertible function $f : X \to Y$, namely $g_1 : Y \to X$ and $g_2 : Y \to X$. Then, from the definition of an inverse function, it must hold that

$$\begin{aligned} g_1 \circ f &= f \circ g_1 = Id\,, \\ g_2 \circ f &= f \circ g_2 = Id\,, \end{aligned}$$

where Id represents the **identity function** (*identieke functie*) $Id(x) = x$ that assigns every real number x to the same real number x.
Therefore,

$$f \circ g_1 = f \circ g_2$$

which we may compose by the left side to get

$$g_1 \circ (f \circ g_1) = g_1 \circ (f \circ g_2)\,,$$

or by relying on the associativity of function composition as

$$(g_1 \circ f) \circ g_1 = (g_1 \circ f) \circ g_2\,.$$

Since $g_1 \circ f = Id$, it we have that $Id \circ g_1 = Id \circ g_2$, which implies that $g_1 = g_2$. □

Let us now turn our attention to the function $f(x) = x^2$. Is f invertible? A likely candidate for the inverse is the function $g(x) = \sqrt{x}$. Checking the composition yields $(g \circ f)(x) = g(f(x)) = \sqrt{x^2} = |x|$, which is not equal to x for all x in the domain $]-\infty, \infty[$. For example, when $x = -2$, $f(-2) = (-2)^2 = 4$, but $g(4) = \sqrt{4} = 2$, which means g failed to return the input -2 from its output 4. Since a function matches a number with exactly one other number it is impossible to construct a function which takes 4 back to both $x = 2$ and $x = -2$ (Definition 3.3). Still, from a graphical standpoint, we know that if $y = f^{-1}(x)$ exists, its graph can be obtained by reflecting $y = x^2$ about the line $y = x$ (Figure 3.22).

We see that the graph of the supposed inverse function fails the vertical line test (Theorem 3.1), and as such, does not represent y as a function of x. The vertical line $x = 4$ on the graph on the right corresponds to the horizontal line $y = 4$ on the graph of $y = f(x)$. The fact that the horizontal line $y = 4$ intersects the graph of f twice means that two different inputs, namely $x = -2$ and $x = 2$, are matched with the same output, 4, which is the cause of all of the trouble. Consequently, a function f is invertible if and only if f is an injective function (one-to-one), and this makes that the corresponding inverse function f^{-1} is an injection as well.

Moreover, for strictly monotone real functions, the existence of an inverse function is guaranteed through the following theorem.

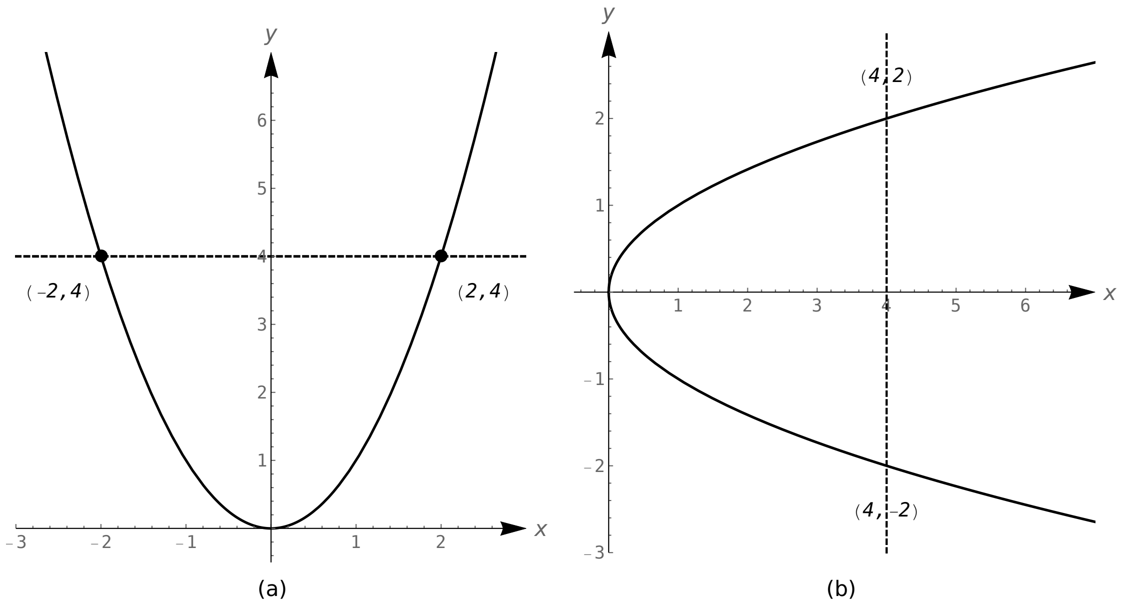


Figure 3.22: Graph of $y = f(x) = x^2$ (a) and the reflection of $y = x^2$ about the line $y = x$ (b).

Theorem 3.4 (Inverse of strictly monotone functions)
Let f be a real function defined on an open interval $I \subset \mathbb{R}$ and let f be strictly monotone on I. Then, f always has an inverse function f^{-1} and f^{-1} has the same behaviour as f.

The last statement in this theorem implies that if f is strictly increasing on I then so is f^{-1}, and likewise if f is strictly decreasing on I.

Proof The proof of this theorem follows easily once one acknowledges that a monotone function f must be a bijection because its monotonicity guarantees that every element in its image is reached for exactly one element in its domain. Suppose f is strictly increasing on $[a, b]$. Since f is a bijection, its inverse f^{-1} exists and it is also a bijection. Now, consider $u, v \in \operatorname{im} f$, let $u < v$ and put $a = f^{-1}(u)$ and $b = f^{-1}(v)$. Consequently, $f(a) = u$ and $f(b) = v$. If $b \leq a$, then $f(b) \leq f(a)$, or equivalently, $v \leq u$. This is a contradiction, so it must hold that $a < b$, which implies that $f^{-1}(u) < f^{-1}(v)$, from which we conclude that f^{-1} is strictly increasing.

A similar reasoning shows that that f^{-1} is strictly decreasing on I if f is strictly decreasing on this interval. □

To find the inverse of an invertible function, we may follow the following steps.

1. Write $y = f(x)$.
2. Interchange x and y.
3. Solve $x = f(y)$ for y to obtain $y = f^{-1}(x)$.

3.4.2 Domain restriction

Let us return to the function $f(x) = x^2$. We know that f is not one-to-one, and thus, is not invertible. However, if we restrict the domain of f, we can produce a new function g which is one-to-one. If we define $g(x) = x^2$ for $x \geq 0$, we can investigate the graph of this function (Figure 3.23). It is clear that g is an injective function, so we can try find its inverse. We proceed as follows

$$
\begin{array}{rrcl}
1. & y & = & g(x) \\
 & & = & x^2, \ x \geq 0 \\
2. & x & = & y^2, \ y \geq 0 \qquad \text{(Switch } x \text{ and } y\text{.)} \\
3. & y & = & \pm\sqrt{x}, y \geq 0 \\
\Rightarrow & y & = & \sqrt{x}
\end{array}
$$

We get $g^{-1}(x) = \sqrt{x}$. At first it looks like we will run into the same trouble as before, but when we check the composition, the domain restriction on g saves the day. We get

$$\left(g^{-1} \circ g\right)(x) = g^{-1}(g(x)) = g^{-1}\left(x^2\right) = \sqrt{x^2} = |x| = x,$$

since $x \geq 0$, and likewise

$$\left(g \circ g^{-1}\right)(x) = g\left(g^{-1}(x)\right) = g\left(\sqrt{x}\right) = \left(\sqrt{x}\right)^2 = x.$$

Graphing g and g^{-1} on the same set of axes shows that they are reflections about the line $y = x$.

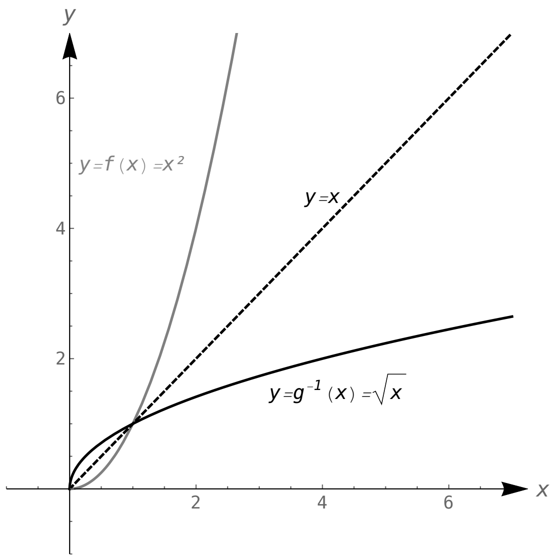


Figure 3.23: The graph of $g(x) = x^2$ for $x \geq 0$ and its inverse $g^{-1}(x) = \sqrt{x}$.

Example 3.11

Determine the inverse of the following function, if it exists, and check your answer graphically:

$$f(x) = x^2 - 2x + 4,$$

where $x \leq 1$.

Solution

We can reformulate the function definition as

$$f(x) = (x-1)^2 + 3\,,$$

from which we infer that it represents the parabola displayed in Figure 3.22, but shifted to the right by one unit and up three units. Moreover, since this function's domain is restricted to $x \leq 1$, we are selecting only the left half of the parabola. Consequently, f is an injective function and thus it is invertible.

We now derive the formula for $f^{-1}(x)$.

$$\begin{array}{lrcll}
1. & y & = & f(x) & \\
 & & = & (x-1)^2+3,\ x \le 1 & \\
2. & x-3 & = & (y-1)^2,\ y \le 1 & \text{(Switch } x \text{ and } y.) \\
\Leftrightarrow & \pm\sqrt{x-3} & = & y-1 & \\
\Rightarrow & y & = & 1-\sqrt{x-3} & \text{(Since } y \le 1.)
\end{array}$$

We have $f^{-1}(x) = 1-\sqrt{x-3}$. In order to check our answer graphically, we graph $y = f^{-1}(x)$ and $y = f(x)$ in Figure 3.24. From this figure it is clear that we obtained the correct inverse function since its graph is the reflection of the graph of f about the first bisector.

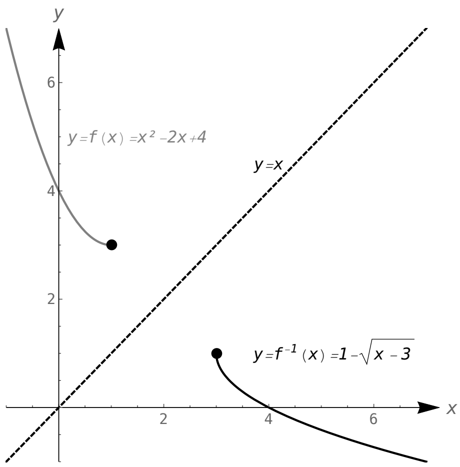
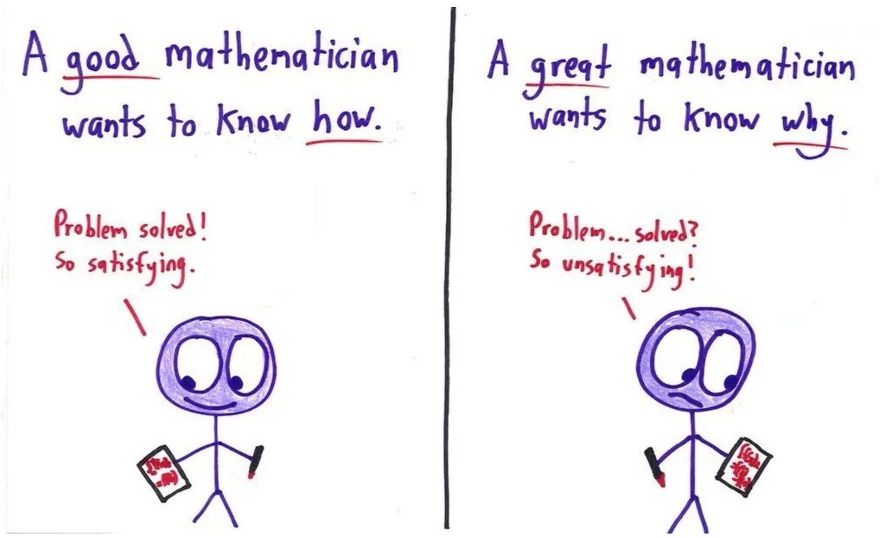


Figure 3.24: The graph of $f(x) = x^2-2x+4$ for $x \le 1$ and its inverse $f^{-1}(x) = 1-\sqrt{x-3}$.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

3.5 Exercises

Functions in $\mathbb{R}$

Assignment 3.1 — Which of the following graphs corresponds with a function?

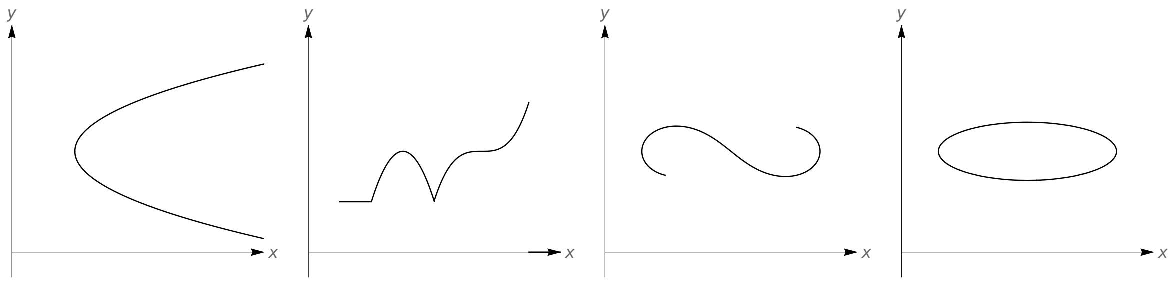


Figure 3.25: The graphs from exercise 3.1.

Function arithmetic

Assignment 3.2 — Consider the functions below and determine an expression for and the domain of a) $(f+g)(x)$, b) $(f-g)(x)$, c) $(fg)(x)$ and d) $\left(\frac{f}{g}\right)(x)$.

(a) $f(x) = x^3 - 1$ and $g(x) = \dfrac{x+1}{x-1}$

(b) $f(x) = \dfrac{x}{2}$ en $g(x) = \dfrac{2}{x}$

(c) $f(x) = x$ and $g(x) = \sqrt{x+1}$

Assignment 3.3 — Determine the following for the functions $f(x) = x+5$ and $g(x) = x^2 - 3$:

(a) $(f \circ g)(0)$

(b) $g(f(0))$

(c) $(f \circ f)(-5)$

(d) $g(g(2))$

Assignment 3.4 — Determine $g \circ f$ and $f \circ g$. Check the domain of each composite function as well.

(a) $f(x) = 5x$ and $g(x) = \dfrac{1}{x-2}$

(b) $f(x) = x^2$ and $g(x) = \sqrt{x}$

(c) $f(x) = x^3$ and $g(x) = \sqrt[3]{1-x}$

(d) $f(x) = x^2 - 4$ and $g(x) = |x-1|$

(e) $f(x) = |x|$ and $g(x) = \sqrt{4-x}$

Assignment 3.5 — Consider the function

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

Calculate $(g \circ g)(x)$ and determine the domain.

Function properties

Assignment 3.6 — Assume f to be even, g to be odd and both to be defined over $\mathbb{R}$. Which of the following functions is even, odd or neither?

(a) $f+g$

(b) fg

(c) f/g

(d) g/f

(e) f^2

(f) g^2

(g) $f \circ g$

(h) $g \circ f$

(i) $f \circ f$

(j) $g \circ g$

Assignment 3.7 — Consider the function $f(x)$ with domain $[0,2]$ and image $[0,1]$ (Figure 3.26).

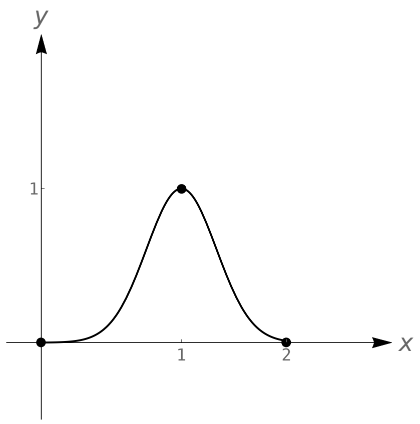


Figure 3.26: The graph from the function $y = f(x)$ from Exercise 3.7.

Sketch the graph of the transformations of $f(x)$ below and determine their domain and image.

(a) $y = f(x)+2$

(b) $y = f(x)-1$

(c) $y = f(x+2)$

(d) $y = f(x-1)$

(e) $y = -f(x)$

(f) $y = f(-x)$

(g) $y = f(4-x)$

(h) $y = 2f(x)$

(i) $y = -\dfrac{f(x)}{2}$

(j) $y = f(2x)$

(k) $y = f\left(\dfrac{x}{3}\right)$

(l) $y = 1-f(1-x)$

(m) $y = 1+f\left(-\dfrac{x}{2}\right)$

Piecewise-defined functions

Assignment 3.8 — Sketch the graphs of the piecewise functions below.

(a) $f(x) = \begin{cases} x^2, & \text{als} \quad x \le -2, \\ 3 - x, & \text{als} \quad -2 < x < 2, \\ 4, & \text{als} \quad x \ge 2. \end{cases}$

(b) $f(x) = \begin{cases} x + 5, & \text{als} \quad x \le -3, \\ \sqrt{9 - x^2}, & \text{als} \quad -3 < x \le 3, \\ -x + 5, & \text{als} \quad x > 3. \end{cases}$

(c) $g(x) = \begin{cases} x^2, & \text{als} \quad x \le -1, \\ \sqrt{1 - x^2}, & \text{als} \quad -1 < x \le 1, \\ x, & \text{als} \quad x > 1. \end{cases}$

(d) $f(x) = \begin{cases} \dfrac{1}{x}, & \text{als} \quad -6 < x < -1, \\ x, & \text{als} \quad -1 < x < 1, \\ \sqrt{x}, & \text{als} \quad 1 < x < 9. \end{cases}$

Assignment 3.9 — Determine the function $f(x)$ depicted in Figure 3.30.

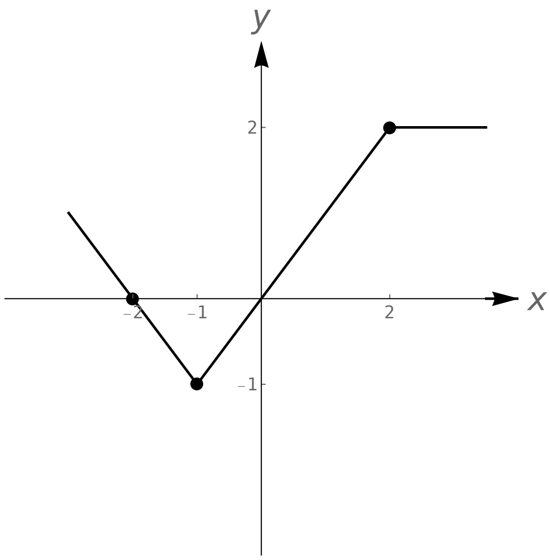


Figure 3.30: The graph of the piecewise function $f(x)$ from Exercise 3.9

Absolute value functions

Assignment 3.10 — Solve the equations below.

(a) $\frac{2}{3}|5-2x|-\frac{1}{2}=5$

(b) $|2-5x|=5|x+1|$

(c) $|x+3|=|2x+1|$

(d) $4|2x-1|-2=10$

(e) $7|x-2|+7=-2|x-2|+2$

(f) $|2x+3|+9\leq 7$

(g) $|3x-7|<2$

(h) $\left|\frac{x}{2}-1\right|\leq 1$

(i) $|x-|x||=10$

(j) $-2\left(|-9x|+|1-9x|\right)=-100$

(k) $|x|-|x-2|=2$

(l) $|x+1|>|x-3|$

(m) $|x-3|<2|x|$

(n) $|x|-|2-x|<2$

(o) $\frac{|3-5x|}{x-2}>6$

(p) $||x+2|-5|>3$

(q) $\left|\frac{1-2x}{1-x}\right|>2$

(r) $\left|\frac{x}{2+x}\right|<1$

(s) $||x|-|7-x||=21$

(t) $|1-x|+|2x-1|-|x+1||x-1|=x$

(u) $-3|x-1|+|3x-1|\leq x-1$

(v) $|5-x|+|3x-1|\geq x+2$

(w) $|3-|2-x||\leq 2x$

Assignment 3.11 — Sketch the graphs of the functions below.

(a) $f(x)=|x-2|$

(b) $f(x)=1+|x-2|$

(c) $f(x)=\frac{x+|x|}{2}$

(d) $f(x)=\frac{x-|x|}{2}$

(e) $f(x)=|x+2|-|x-3|+1$

(f) $f(x)=3|x+4|-4$

(g) $f(x)=||x|+3|$

(h) $f(x)=||x|-3|$

(i) $|x|+|y|=1$

Inverse functions

Assignment 3.12 — Determine the inverse f^{-1} of f if f is defined as below. Is f^{-1} a function? Verify this both algebraically and graphically.

(a) $f(x)=2x+8$

(b) $f(x)=1-\frac{4+3x}{5}$

(c) $f(x)=\frac{x+6}{x+5}$

(d) $f(x)=x^3+1$

(e) $f(x)=x^2+x$

(f) $f(x)=\sqrt{1-x^2}$

(g) $f(x)=3\sqrt{x-1}-4$

(h) $f(x)=x^2-6x+5$

Review exercises

Assignment 3.13 — For the following functions, determine the domain, codomain and image in each case. Then examine whether they are injective, surjective, and/or bijective. Also examine whether the functions are periodic, even/odd, increasing/decreasing, and/or bounded. If possible, determine any maxima and/or minima.

(a) $f(x) = \sqrt{4-x^2}$

(b) $f(x) = \sqrt{4+x^2}$

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

(d) $f(x) = \dfrac{1}{1+x^2}$

(e) $f(x) = x^5+1$

(f) $f(x) = x^4+1$

(g) $f(x) = \sqrt{1-x}$

(h) $f(x) = |x|$

(i) $f(x) = \dfrac{x}{x+1}$

(j) $f(x) = -\dfrac{4}{x^3}$

(k) $f(x) = 3-2\sqrt{x+2}$

(l) $f(x) = \sqrt{x^2+4x+4}$

Obvious is the most dangerous word in mathematics

— Eric Temple Bell —

4

Algebraic functions

4.1 Polynomial functions

4.1.1 Constant and linear functions

Many of the functions we already encountered in the preceding chapter were either constant or linear. Here, we first of all give a more formal definition of a linear function.

Definitie 4.1 (Linear function)
A **linear function** (*lineaire functie*) is a function of the form

$$f(x) = ax + b,$$

where a and b are real numbers with $a \neq 0$. The domain of a linear function is $\mathbb{R}$.

For the case $a = 0$, we get $f(x) = b$, which is referred to as a **constant function** (*constante functie*).

Recall that to graph a function f, we graph the equation $y = f(x)$. Hence, the graph of a linear function is a line with slope a and y-intercept $(0, b)$; the graph of a constant function is a horizontal line (a line with slope $a = 0$) and an y-intercept of $(0, b)$. For that reason, Definition 4.1 is therefore often referred to as the slope-intercept definition of a line in the plane. In general, given two points in the plane (x_1, y_1) and (x_2, y_2) the equation of straight line is given by

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1), \tag{4.1}$$

where

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

is the slope of the line.

4.1.2 Quadratic functions

Definitie 4.2 (Quadratic function)
A **quadratic function** (*kwadratische functie*) is a function of the form

$$f(x) = ax^2 + bx + c,$$

where a, b and c are real numbers with $a \neq 0$. The domain of a quadratic function is $\mathbb{R}$.

The most basic quadratic function is $f(x) = x^2$, which is shown in Figure 4.1. Its shape is called a **parabola** (*parabool*). The point $(0,0)$ is called the **vertex** (*top*) of the parabola.

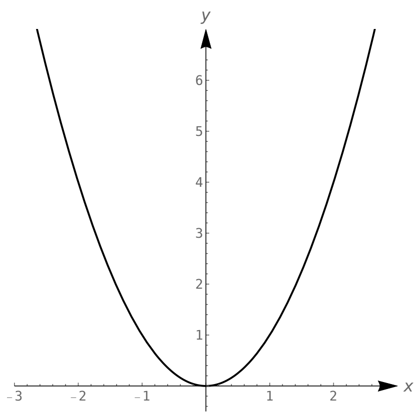


Figure 4.1: The graph of $y = x^2$.

Definition 4.2 uses the general form of a quadratic function, though a quadratic function f may as well be defined using its so-called standard form, being

$$f(x) = a(x-h)^2 + k\,,$$

where a, h and k are real numbers with $a \neq 0$. The vertex of the graph of $y = f(x)$ is in this notation given by (h,k). Any quadratic function can be rewritten in standard form by completing the square.

The graph of $y = a(x-h)^2 + k$ is a parabola opening upwards if $a > 0$, and opening downwards if $a < 0$. Moreover, the symmetry enjoyed by the graph of $y = x^2$ about the y-axis is translated to a symmetry about the vertical line $x = h = -\frac{b}{2a}$ which is the vertical line through the vertex. This line is called the **axis of symmetry** (*symmetrieas*) of the parabola.

Our next example is a classic application of quadratic functions.

Example 4.1

Donnie inherits a large parcel of land near Tielt from one of his relatives. The time is finally right for him to pursue his dream of farming some cannabis (*Cannabis sativa*). He wishes to build a rectangular pasture, and estimates that he has enough money for 200 linear metres of fencing material. If he makes the pasture adjacent to a stream (so no fencing is required on that side), what are the dimensions of the pasture which maximize the area? What is the maximum area? If an average cannabis plant needs 2.5 square metres of grazing area, how many cannabis can Donnie keep in his pasture? It is always helpful to sketch the problem situation (Figure 4.2).

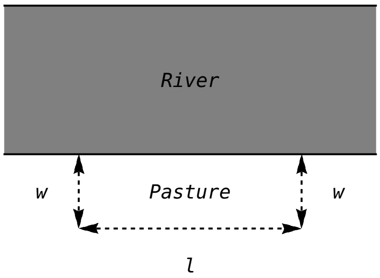


Figure 4.2: Donnie's pasture.

Solution

We are tasked to find the dimensions of the pasture which would give a maximum area. We let w [L] denote the width of the pasture and we let l [L] denote the length of the pasture. Since the units given to us in the statement of the problem are metres, we assume w and l are measured in metre. The area of the pasture, which we will call A, is related to w and l by the equation $A = wl$. Since w and l are both measured in metre, A [L^2] has units of square metre. We are given the total amount of fencing available is 200 metres, which means $w + l + w = 200$, or, $l + 2w = 200$. We now have two equations,

$$\begin{cases} A = wl\,, \\ l + 2w = 200\,. \end{cases}$$

In order to maximize A, we need to use the information given to write A as a function of just one variable, either w or l:

$$A = wl = w(200 - 2w) = 200w - 2w^2.$$

We now have A as a function of w:

$$A(w) = -2w^2 + 200w.$$

Before we go any further, we need to find the applied domain of A so that we know what values of w make sense in this problem situation. Since w represents the width of the pasture, $w > 0$. Likewise, l represents the length of the pasture, so $l = 200 - 2w > 0$. Solving this latter inequality, we find $w < 100$. Hence, the function we wish to maximize is $A(w) = -2w^2 + 200w$ for $0 < w < 100$. Since A is a quadratic function of w, we know that the graph of $y = A(w)$ is a parabola. Since the coefficient of w^2 is -2, we know that this parabola opens downwards. This means that there is a maximum value to be found, and we know it occurs at the vertex. Using the vertex formula, we find

$$\begin{cases} w = -\dfrac{200}{2(-2)} = 50\,, \\ A(50) = -2(50)^2 + 200(50) = 5000\,. \end{cases}$$

Since $w = 50$ lies in the applied domain, $0 < w < 100$, we have that the area of the pasture is maximized when the width is 50 metres. To find the length, we use $l = 200 - 2w$ and find $l = 200 - 2(50) = 100$, so the length of the pasture is 100 metres. The maximum area is $A(50) = 5000$, or 5000 m^2. If an average cannabis plant requires 2.5 square metres of pasture, Donnie can raise $\frac{5000}{2.5} = 2000$ such plants.

In practice, quadratic functions often pop up in inequalities, which we can solve graphically.

One of the classic applications of inequalities is the notion of **tolerances**. Recall that for real numbers

x and c, the quantity $|x-c|$ may be interpreted as the distance from x to c. Solving inequalities of the form $|x-c| \le d$ for $d \ge 0$ can then be interpreted as finding all numbers x which lie within d units of c. We can think of the number d as a tolerance and our solutions x as being within an accepted tolerance of c. We use this principle in the next example.

Example 4.2

The area A ([L^2]) of a square piece of particle board which measures x [L] centimetres on each side is $A(x) = x^2$. Suppose a manufacturer needs to produce a 24 centimetre by 24 centimetre square piece of particle board as part of a home office desk kit. How close does the side of the piece of particle board need to be cut to 24 centimetres to guarantee that the area of the piece is within a tolerance of 0.25 square centimetres of the target area of 576 square centimetres?

Solution

Mathematically, we express the desire for the area $A(x)$ to be within 0.25 square centimetres of 576 as $|A-576| \le 0.25$. Since $A(x) = x^2$, we get $|x^2-576| \le 0.25$, which is equivalent to $-0.25 \le x^2 - 576 \le 0.25$. Recalling the increasing property of the square root; that is if $0 \le a \le b$, then $\sqrt{a} \le \sqrt{b}$, we proceed

$$\begin{array}{rrcll}
 & -0.25 \le & x^2-576 & \le 0.25 & \\
\Leftrightarrow & 575.75 \le & x^2 & \le 576.25 & \text{(Add 576 across the inequalities.)} \\
\Leftrightarrow & \sqrt{575.75} \le & \sqrt{x^2} & \le \sqrt{576.25} & \text{(Take square roots.)} \\
\Leftrightarrow & \sqrt{575.75} \le & |x| & \le \sqrt{576.25} & (\sqrt{x^2} = |x|)
\end{array}$$

Consequently, we find the solution to $\sqrt{575.75} \le |x|$ to be $\left]-\infty, -\sqrt{575.75}\,\right] \cup \left[\sqrt{575.75}, +\infty\right[$ and the solution to $|x| \le \sqrt{576.25}$ to be $\left[-\sqrt{576.25}, \sqrt{576.25}\,\right]$. To solve $\sqrt{575.75} \le |x| \le \sqrt{576.25}$, we intersect these two sets to get $\left[-\sqrt{576.25}, -\sqrt{575.75}\,\right] \cup \left[\sqrt{575.75}, \sqrt{576.25}\,\right]$. Since x represents a length, we discard the negative answers and get $\left[\sqrt{575.75}, \sqrt{576.25}\,\right]$. This means that the side of the piece of particle board must be cut between $\sqrt{575.75} \approx 23.995$ and $\sqrt{576.25} \approx 24.005$ centimetres, a tolerance of (approximately) 0.005 centimetres of the target length of 24 centimetres.

Our last example in the section demonstrates how inequalities can be used to describe regions in the plane.

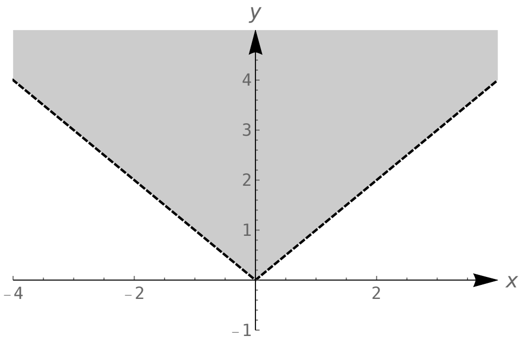
Example 4.3

Sketch the following relations.

1. $R = \{(x,y) : y > |x|\}$
2. $S = \{(x,y) : y \le 2 - x^2\}$

Solution

1. The relation R consists of all points (x,y) whose y-coordinate is greater than $|x|$. If we graph $y = |x|$, then we want all of the points in the plane above the points on the graph. Dotting the graph of $y = |x|$ to indicate that the points on the graph itself are not in the relation, we get the shaded region in Figure 4.3(a).
2. For a point to be in S, its y-coordinate must be less than or equal to the y-coordinate on the parabola $y = 2 - x^2$. This is the set of all points below or on the parabola $y = 2 - x^2$ (Figure 4.3(b)).



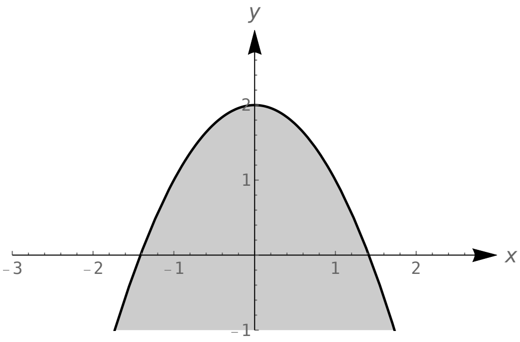


Figure 4.3: Graph of the relation R (a) and S (b).

Many quadratic equations $ax^2 + bx + c = 0$ cannot be solved by factoring them. This is generally true when the roots are not rational numbers. An other method of solving quadratic equations involves the use of the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{4.2}$$

where $a \neq 0$. In relation to quadratic equations, complex numbers come in when the value under the radical portion of the quadratic formula is negative. When this occurs, the equation has no roots in $\mathbb{R}$. The roots belong to $\mathbb{C}$, will be called **complex roots** (*complexe wortels*) (or imaginary roots), and are expressible as $a \pm bi$. When using the quadratic formula to solve a quadratic equation with real coefficients, there are three possibilities depending on the discriminant $D = b^2 - 4ac$:

1. Two different real roots if $D > 0$.
2. One real root if $D = 0$.
3. Two complex roots, complex conjugates, if $D < 0$.

Example 4.4

Solve the following quadratic equations.

1. $x^2 + 2x + 2 = 0$
2. $x^2 - 4x + 13 = 0$

Solution

1. Substituting in Equation (4.2), we get

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-2 \pm \sqrt{-4}}{2}.$$

Since the discriminant $D = b^2 - 4ac$ is negative, this equation has no solution in $\mathbb{R}$. But if you were to express the solution using imaginary numbers, the solutions would be

$$x = \frac{-2 \pm 2i}{2} = -1 \pm i.$$

2. Notice that after substitution in Equation (4.2), we are left with a negative value under the

square root radical.

$$x = \frac{4 \pm \sqrt{-36}}{2}$$

There are two complex roots

$$x = \frac{4 \pm 6i}{2} = 2 \pm 3i.$$

4.1.3 General polynomial functions

4.1.3.1 Definition

Three of the families of functions studied thus far – constant, linear and quadratic – belong to a much larger group of functions called polynomials. We begin our formal study of general polynomials with a definition and some examples.

Definitie 4.3 (Polynomial function)
A **polynomial function** (*veeltermfunctie*) is a function of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

where a_0, a_1, ..., a_n are real numbers, $a_n \neq 0$ and $n \in \mathbb{N}$. The domain of a polynomial function is $\mathbb{R}$.

Given $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ with $a_n \neq 0$, we say

- The natural number n is called the **degree** (*graad*) of the polynomial f.
- $a_n x^n$ is called the **leading term** (*hoogstegraadsterm*) of the polynomial f.
- The real number a_n is called the **leading coefficient** of the polynomial f.
- The real number a_0 is called the **constant term** (*constante term*) of the polynomial f.

Moreover, if $f(x) = a_0$, and $a_0 \neq 0$, we say f has degree 0, while if $f(x) = 0$, we say f has no degree.

In Part II, we will introduce the tools that are needed to graph polynomial functions and understand their behaviour. Anyhow, we will often have to determine the zeros of a polynomial equation. For that reason, we recall the most important facts about the factorization of polynomials.

4.1.3.2 Factorization of polynomials

Suppose we wish to find the zeros of $f(x) = x^3 + 4x^2 - 5x - 14$. Setting $f(x) = 0$ results in the polynomial equation $x^3 + 4x^2 - 5x - 14 = 0$. It is easy to see that $f(2) = 0$, but possible other zeros seem less obvious. Now, if $x = 2$ is a zero, there should be a factor of $(x-2)$ lurking around in the factorization of $f(x)$. In other words, we should expect that $x^3 + 4x^2 - 5x - 14 = (x-2)\,q(x)$, where $q(x)$ is some second degree polynomial. We can find this polynomial through **polynomial division** (*Euclidische deling*). Dividing $x^3 + 4x^2 - 5x - 14$ by $x - 2$ gives

$$
\begin{array}{r|l}
x^3 + 4x^2 - 5x - 14 & x-2 \\ \cline{2-2}
-\left(x^3 - 2x^2\right) & x^2+6x+7 \\ \cline{1-1}
6x^2 - 5x & \\
-\left(6x^2 - 12x\right) & \\ \cline{1-1}
7x - 14 & \\
-(7x - 14) & \\ \cline{1-1}
0 &
\end{array}
$$

This means $x^3+4x^2-5x-14 = (x-2)\left(x^2+6x+7\right)$, so to find the zeros of f, we now solve $(x-2)\left(x^2+6x+7\right)=0$. We get $x-2=0$ (which gives us our known zero, $x=2$) as well as $x^2+6x+7=0$. The latter leads to additional zeros, namely $x=-3\pm\sqrt{2}$.

First of all, we should remember what we may expect when dividing polynomials in general.

Definitie 4.4 (Polynomial division)
Suppose $d(x)$ and $p(x)$ are nonzero polynomials where the degree of p is greater than or equal to the degree of d. There exist two unique polynomials, $q(x)$ and $r(x)$, such that

$$p(x) = d(x)\,q(x) + r(x),$$

where either $r(x)=0$ or the degree of r is strictly less than the degree of d.

As you may recall, all of the polynomials in Definition 4.4 have special names. The polynomial p is called the **dividend** (*deeltal*); d is the **divisor** (*deler*); q is the **quotient** (*quotiënt*); r is the **remainder** (*rest*). If $r(x)=0$ then d is called a **factor** of p.

The additional finding that $x-2$ is a factor of $x^3+4x^2-5x-14$ as $x=2$ is a zero of $x^3+4x^2-5x-14=0$ can be generalized in the following theorem.

Theorem 4.1 (The factor theorem)
Suppose $p(x)$ is a nonzero polynomial. The real number c is a zero of $p(x)$ if and only if $x-c$ is a factor of $p(x)$.

For completeness we also mention the related remainder theorem.

Theorem 4.2 (The remainder theorem)
Suppose p is a polynomial of degree at least 1 and c is a real number. When $p(x)$ is divided by $x-c$ the remainder is $p(c)$.

Sometimes $x-c$ is not only a factor of a given polynomial $p(x)$, but as well of the quotient resulting from dividing $p(x)$ by $x-c$. In that case, we say that the **multiplicity** (*multipliciteit*) of the factor $x-c$ is 2, and hence the multiplicity of $x=c$ as a zero of $p(x)$ is 2 as well. Clearly, the multiplicity of a factor $x-c$ can be at most equal to the degree of the polynomial $p(x)$.

Clearly, a polynomial division can become quite tedious, but when dividing polynomials by quantities of the form $x-c$, we can rely on **Horner's approach** (*methode van Horner*). Essentially, this boils down to constructing a synthetic division tableau for the polynomial division problem. Let us rework our division problem using this tableau to see how it greatly streamlines the division process. To divide

$x^3+4x^2-5x-14$ by $x-2$, we write 2 in the place of the divisor and the coefficients of $x^3+4x^2-5x-14$ in for the dividend. Then bring down the first coefficient of the dividend.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & & & & \\ \hline & & & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & & & \\ \hline & 1 & & & \end{array}$$

Next, take the 2 from the divisor and multiply by the 1 that was 'brought down' to get 2. Write this underneath the 4, then add to get 6.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & & \\ \hline & 1 & & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & & \\ \hline & 1 & 6 & & \end{array}$$

Now take the 2 from the divisor times the 6 to get 12, and add it to the -5 to get 7.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & \\ \hline & 1 & 6 & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & \\ \hline & 1 & 6 & 7 & \end{array}$$

Finally, take the 2 in the divisor times the 7 to get 14, and add it to the -14 to get 0.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & 14 \\ \hline & 1 & 6 & 7 & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & 14 \\ \hline & 1 & 6 & 7 & \boxed{0} \end{array}$$

The first three numbers in the last row of our tableau are the coefficients of the quotient polynomial. Remember, we started with a third degree polynomial and divided by a first degree polynomial, so the quotient is a second degree polynomial. Hence the quotient is x^2+6x+7. The number in the box is the remainder. Synthetic division is our tool of choice for dividing polynomials by divisors of the form $x-c$. It is important to note that it works only for these kinds of divisors. Besides, when doing a synthetic division, do not forget to insert 0's in the division tableau to account for any missing powers of x.

In Mathematica, a polynomial can be factored over integers using the built-in function **Factor** as follows.

```
In[4]:= Factor[x^3 +4*x^2 -5*x -14]

Out[4]= (-2 +x) (7 +6x +x^2)
```

Of course, if we allow for irrational numbers in our factorisation, the last factor in $(-2+x)(7+6x+x^2)$ can be factored as $(x+3-\sqrt{2})(x+3+\sqrt{2})$. For that purpose, one can look for the roots of $7+6x+x^2=0$ in Mathematica using **Solve**.

The following theorem gives us an upper bound on the number of real zeros.

> **Theorem 4.3 (Zeros and multiplicity)**
> *Suppose f is a polynomial of degree $n \geq 1$. Then f has at most n real zeros, counting multiplicities.*

In many cases, however, the factorization of a polynomial f will lead to quadratic terms with complex zeros. In that case it is important to note that complex roots of a polynomial f occur as complex conjugate pairs, as indicated by the following theorem.

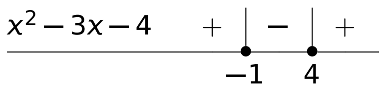
> **Theorem 4.4 (Conjugate pairs theorem)**
> *If f is a polynomial function with real number coefficients and z is a zero of f, then so is $\overline{z}$.*

Even though a polynomial f has complex roots, we can still write down a real factorization involving linear factors corresponding to the real zeros of f and irreducible quadratic factors that give the complex zeros of f, as formalized in the following theorem.

> **Theorem 4.5 (Real factorization theorem)**
> *Suppose f is a polynomial function with real coefficients. Then $f(x)$ can be factored into a product of linear factors corresponding to the real zeros of f and irreducible quadratic factors which give the complex zeros of f.*

The value of Theorem 4.5 is illustrated in the next example.

Once we determined the zeros of an equation, we will often have to determine on which intervals the corresponding function is positive and on which intervals it is negative. For that purpose, we will use a so-called sign chart. For instance, consider the quadratic function $f(x) = x^2 - 3x - 4$. The zeros of $x^2 - 3x - 4 = 0$ are $x = -1$ and $x = 4$, and by choosing a few test values in the resulting intervals $]-\infty, -1]$, $[-1, 4]$ and $[4, +\infty[$ we can determine the sign of the function value at those points. This leads to the following sign diagram:



where the dots indicate that the zeros belong to the function's domain, otherwise circles are used.

> **Splines**
>
> A spline is a function defined piecewise by polynomials and is used for addressing interpolating problems. The term spline comes from the flexible spline devices used by shipbuilders and draftsmen to draw smooth shapes.
>
> Een spline of degree n on the interval $[a, b]$ is a function S on that interval consisting of a concatenation of k polynomials S_i defined on the subinterval $[x_{i-1}, x_i]$ in $[a, b]$. Hence,
>
> $$S(x) = \begin{cases} S_0(x) & x \in [x_0, x_1] \\ S_1(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_{k-1}(x) & x \in [x_{k-1}, x_k], \end{cases}$$
>
> where $a = x_0 < x_1 < \ldots < x_{k-1} < x_k = b$ and $S_i(x)$ is a polynomial of a degree not higher than n.

4.2 Rational functions

4.2.1 Definition

If we add, subtract or multiply polynomial functions according to the function arithmetic rules, we will produce another polynomial function. If, on the other hand, we divide two polynomial functions, the result may not be a polynomial. Here we study we study rational functions - functions which are ratios of polynomials.

Definitie 4.5 (Rational function)
A **rational function** (*rationale functie*) is a function which is the ratio of polynomial functions. Said differently, h is a rational function if it is of the form

$$h(x) = \frac{p(x)}{q(x)},$$

where p and q are polynomial functions.

Note that according to this definition, all polynomial functions are also rational functions. By taking $q(x) = 1$. Rational functions are used in numerical analysis for interpolation and approximation of functions. Approximations in terms of rational functions are well suited for computer algebra systems and other numerical software because they can be evaluated straightforwardly. They are also used to approximate or model more complex equations in science and engineering including fields and forces in physics, spectroscopy in analytical chemistry, enzyme kinetics in biochemistry, medicine concentrations in vivo, wave functions for atoms and molecules, and so on.

Obviously, we have domain issues anytime the denominator of a fraction is zero. In the example below, we review this concept as well as some of the arithmetic of rational expressions.

Example 4.5

Find the domain of the following rational functions. Write them in the form $\frac{p(x)}{q(x)}$ for polynomial functions p and q and simplify.

1. $f(x) = \dfrac{2x-1}{x+1}$
2. $g(x) = 2 - \dfrac{3}{x+1}$
3. $h(x) = \dfrac{2x^2-1}{x^2-1} - \dfrac{3x-2}{x^2-1}$
4. $r(x) = \dfrac{2x^2-1}{x^2-1} \div \dfrac{3x-2}{x^2-1}$

Solution

1. To find the domain of f, we find the zeros of the denominator and exclude them from the domain. Setting $x+1=0$ results in $x=-1$. Hence, our domain is $\mathbb{R}\backslash\{-1\}$. The expression $f(x)$ is already in the form requested and when we check for common factors among the numerator and denominator we find none, so we are done.

2. Proceeding as before, we determine the domain of g by solving $x+1=0$. As before, we find the domain of g is $\mathbb{R}\backslash\{-1\}$. To write $g(x)$ in the form requested, we need to get a common denominator.

$$\begin{aligned} g(x) &= 2 - \frac{3}{x+1} &&= \frac{2(x+1)}{x+1} - \frac{3}{x+1} \\ &= \frac{(2x+2)-3}{x+1} &&= \frac{2x-1}{x+1} \end{aligned}$$

This formula is now completely simplified.

3. The denominators in the formula for $h(x)$ are both $x^2 - 1$ whose zeros are $x = \pm 1$. As a result, the domain of h is $\mathbb{R} \backslash \{-1, 1\}$. We now proceed to simplify $h(x)$:

$$\begin{aligned} h(x) &= \frac{2x^2-1}{x^2-1} - \frac{3x-2}{x^2-1} \\ &= \frac{2x^2-1-3x+2}{x^2-1} &&= \frac{2x^2-3x+1}{x^2-1} \\ &= \frac{(2x-1)(x-1)}{(x+1)(x-1)} &&= \frac{(2x-1)\cancel{(x-1)}}{(x+1)\cancel{(x-1)}} \\ &= \frac{2x-1}{x+1} \end{aligned}$$

Note that it is important to find the domain of h before simplifying the expression defining it. Otherwise, we would get that the domain of h is $\mathbb{R} \setminus \{-1\}$ instead of the correct $\mathbb{R} \setminus \{-1, 1\}$.

4. To find the domain of r, it may help to temporarily rewrite $r(x)$ as

$$r(x) = \frac{\dfrac{2x^2-1}{x^2-1}}{\dfrac{3x-2}{x^2-1}}.$$

We need to set all of the denominators equal to zero which means we need to solve not only $x^2 - 1 = 0$, but also $\frac{3x-2}{x^2-1} = 0$. We find $x = \pm 1$ for the former and $x = 2/3$ for the latter. Our domain is $\mathbb{R} \backslash \{-1, 2/3, 1\}$. We simplify $r(x)$:

$$\begin{aligned} r(x) &= \frac{2x^2-1}{x^2-1} \div \frac{3x-2}{x^2-1} &&= \frac{2x^2-1}{x^2-1} \cdot \frac{x^2-1}{3x-2} \\ &= \cancel{\frac{(2x^2-1)\cancel{(x^2-1)}}{\cancel{(x^2-1)}(3x-2)}} &&= \frac{2x^2-1}{3x-2} \end{aligned}$$

A few remarks about Example 4.5 are in order. Note that the expressions for $f(x)$, $g(x)$ and $h(x)$ work out to be the same. However, only two of these functions are actually equal. Recall that functions are ultimately sets of ordered pairs (Definition 3.3), so for two functions to be equal, they need, among other things, to have the same domain. Since $f(x) = g(x)$ and f and g have the same domain, they are equal functions. Even though the formula $h(x)$ is the same as $f(x)$, the domain of h is different than the domain of f, and thus they are different functions.

4.2.2 Graphs of rational functions

In Part II we will introduce the tools that are needed to graph rational functions and understand their behaviour. Still, here we already want to underline a few distinctive features of the graph of any rational function. For that purpose, consider the graph of the following rational function

$$f(x) = \frac{2x-1}{x+1}, \tag{4.3}$$

which is shown in Figure 4.4.

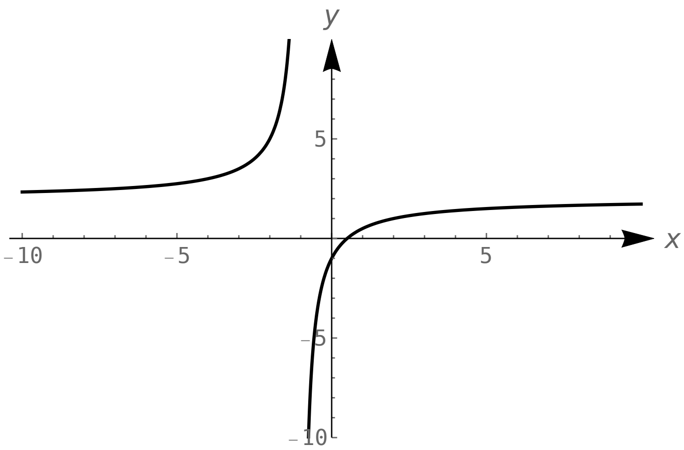


Figure 4.4: The graph of $f(x) = \frac{2x-1}{x+1}$.

First, note that the graph appears to break at $x = -1$. We know from Example 4.5 that $x = -1$ is not in the domain of f which means $f(-1)$ is undefined. We see that we can get near $x = -1$ from two directions.

As the x-values approach -1 from the left, the function values become larger and larger positive numbers. We express this symbolically by writing $x \underset{<}{\to} -1$, $f(x) \to +\infty$, or alternatively $x \to -1^-$, $f(x) \to +\infty$. Similarly, using analogous notation, we conclude that as $x \underset{>}{\to} -1$, $f(x) \to -\infty$, or equivalently $x \to -1^+$, $f(x) \to -\infty$. Here the > means approaching from above and < means approaching from below. For this type of unbounded behavior, we say the graph of $y = f(x)$ has a **vertical asymptote** (*verticale asymptoot*) of $x = -1$.

The other feature worthy of note about the graph of $y = f(x)$ is that it seems to level off on the left and right hand sides of the plot window. We see that as $x \to -\infty$, $f(x)$ approaches 2 coming from values larger than 2 (from above) and as $x \to +\infty$, $f(x)$ approaches 2 though coming from values smaller than 2 (from below). In this case, we say the graph of $y = f(x)$ has a **horizontal asymptote** (*horizontale asymptoot*) of $y = 2$. We formalize the concepts of vertical and horizontal asymptotes in Chapter 8. We then also introduce the notion of a slant asymptote.

Example 4.6

A mathematical model for the population P [–], in thousands, of a certain species of bacteria, t [+] days after it is introduced to an environment is given by

$$P(t) = \frac{100}{(5-t)^2},$$

for $0 \le t < 5$.

1. Find and interpret $P(0)$.
2. When will the population reach 100 000?
3. Determine the behavior of P as $t \underset{<}{\to} 5$. Interpret this result graphically and within the context of the problem.

Solution

1. Substituting $t = 0$ gives $P(0) = \frac{100}{(5-0)^2} = 4$, which means 4000 bacteria are initially introduced into the environment.
2. We remember that $P(t)$ is measured in thousands, so, 100 000 bacteria corresponds to $P(t) = 100$. Substituting for $P(t)$ gives
$$\frac{100}{(5-t)^2} = 100\,,$$
whose solution is $t = 4$ or $t = 6$. Of these two solutions, only $t = 4$ lies in our domain, so this is the solution we keep. Hence, it takes 4 days for the population of bacteria to reach 100 000.
3. To determine the behaviour of P as $t \underset{<}{\to} 5$, we can make a table

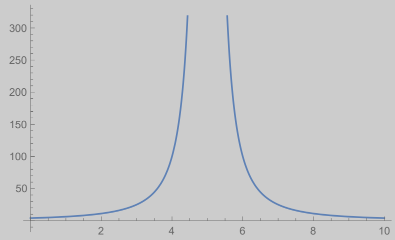
t	$P(t)$
4.9	10 000
4.99	1000 000
4.999	100 000 000
4.9999	10 000 000 000

In other words, as $t \underset{<}{\to} 5$, $P(t) \to +\infty$. The line $t = 5$ is a vertical asymptote of the graph of $y = P(t)$. This means that the population of bacteria is increasing without bound as we near 5 days, which cannot actually happen. For this reason, $t = 5$ is called the doomsday for this population. There is no way any environment can support infinitely many bacteria, so shortly before $t = 5$ the environment would collapse.

In Mathematica, we can verify the correctness of our reasoning above by plotting the studied function using the built-in function **Plot** as follows.

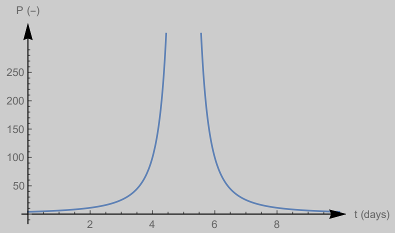
```
In[5]:= Plot[(100)/(5-t)^2,{t,0,10}]
```

Out[5]=



This plot becomes, however, more informative upon adding the axis labels and directions as follows.

In[6]:=
```
Plot[(100)/(5-t)^2,{t,0,10},     AxesLabel →{"t (days)","P (-)"},
AxesStyle→Arrowheads[{0,0.05}]]
```

Out[6]=


We conclude this section with one example illustrating the practical use of rational functions.

Example 4.7

The actual data relating the volume V [L^3] of a gas and its pressure P [M $L^{-1}T^{-2}$] used by Boyle and his assistant in 1662 to verify the gas law that bears his name is given below.

V	48	46	44	42	40	38	36	34	32	30	28	26
P	29.13	30.56	31.94	33.5	35.31	37	39.31	41.63	44.19	47.06	50.31	54.31

V	24	23	22	21	20	19	18	17	16	15	14
P	58.81	61.31	64.06	67.06	70.69	74.13	77.88	82.75	87.88	93.06	100.44

V	13	12
P	107.81	117.56

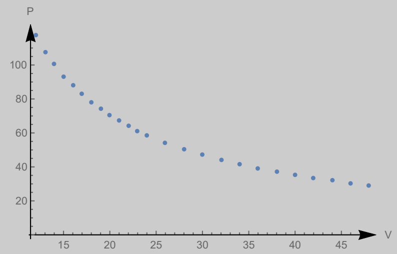
1. Use Mathematica to create a scatter plot for these data using V as the independent variable and P as the dependent variable. Does it appear from the graph that P is inversely proportional to V? Explain.
2. Assuming that P and V do vary inversely, use the data to approximate the constant of proportionality.

Solution

1. If P really does vary inversely with V, then $P = \frac{k}{V}$ for some constant k. To verify this, we create a scatterplot in Mathematica with the built-in function **`ListPlot`** as follows.

In[7]:=
```
data={{48,29.125`},{46,30.5625`},{44,31.9375`},{42,33.5`},{40,35.3125`},
{38,37},{36,39.3125`},{34,41.625`},{32,44.1875`},{30,47.0625`},
{28,50.3125`},{26,54.3125`},{24,58.8125`},{23,61.3125`},{22,64.0625`},
{21,67.0625`},{20,70.6875`},{19,74.125`},{18,77.875`},{17,82.75`},
{16,87.875`},{15,93.0625`},{14,100.4375`},{13,107.8125`},{12,117.5625`}};
ListPlot[data,AxesLabel→{"V","P"},AxesStyle→Arrowheads[{0,0.05`}]]
```

Out[7]=

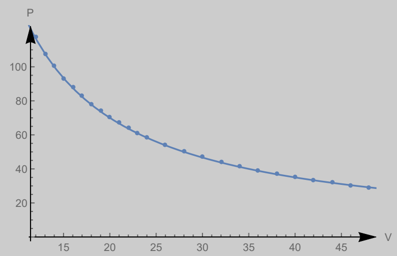


From the resulting plot, the points do seem to lie along a curve like $P = \frac{k}{V}$.

2. To determine the constant of proportionality, we note that from $P = \frac{k}{V}$, we get $k = PV$. Multiplying each of the volume numbers times each of the pressure numbers, we produce a number which is always approximately 1400. We suspect that $P = \frac{1400}{V}$. Graphing this function along with the data gives us good reason to believe our hypotheses that P and V are, in fact, inversely related. Again, we can do this in Mathematica using the built-in function **Show** to combine multiple plots in one window.

In[8]:=
```
Show[ListPlot[data,AxesLabel→{"V","P"},AxesStyle→Arrowheads[{0,0.05`}]],
Plot[1400/V,{V,0,50}]]
```

Out[8]=



4.3 Irrational functions

This section serves as a watershed for functions which are combinations of polynomial, and more generally, rational functions, with the operations of radicals, such as

$$f(x) = \sqrt{1-x^2}\,,$$

$$g(x) = \sqrt[4]{\frac{16x}{x^2-9}}\,,$$

and

$$h(x) = \sqrt[3]{x^3+3x^2-6x-8}\,.$$

It is business of Part II to introduce the tools to better understand the behaviour of such functions in all the detail. Here we restrict to the basics to help shore up the requisite skills needed for a good understanding of the subsequent parts of this course. In literature, functions containing radicals are sometimes referred to as **irrational functions** (*irrationale functie*).

It is worth remarking that, in the light of Section 3.4, we could define $f(x) = \sqrt[n]{x}$ functionally as the inverse of $g(x) = x^n$ with the stipulation that when n is even, the domain of g is restricted to $\mathbb{R}^+$. From what we know about $g(x) = x^n$ from Section 4.1, we can produce the graphs of $f(x) = \sqrt[n]{x}$ by reflecting the graphs of $g(x) = x^n$ across the line $y = x$. Figures 4.5(a)-4.5(c) show the graphs of $y = \sqrt{x}$, $y = \sqrt[4]{x}$ and $y = \sqrt[6]{x}$ obtained in this way from the graphs of $y = x^2$, $y = x^4$ and $y = x^6$, respectively. We can see the vertical steepening near $x = 0$ and the horizontal flattening as $x \to +\infty$. Likewise, Figures 4.5(d)-4.5(f) show the graphs of $y = \sqrt[3]{x}$, $y = \sqrt[5]{x}$ and $y = \sqrt[7]{x}$ obtained by reflecting the graphs of $y = x^3$, $y = x^5$ and $y = x^7$ about the axis $y = x$, respectively. These odd-indexed radical functions also follow a predictable trend - steepening near $x = 0$ and flattening as $x \to \pm\infty$.

Example 4.8

For the following functions, state their domains and determine their zero. Check your answer graphically using Mathematica.

1. $f(x) = 3x\sqrt[3]{2-x}$
2. $g(x) = \sqrt{2-\sqrt[4]{x+3}}$

Solution

1. As far as the domain is concerned, $f(x)$ has no denominators and no even roots, which means its domain is $\mathbb{R}$. Its zeros can be found as follows.

$$\begin{aligned} & f(x) &=& \ 0 \\ \Leftrightarrow\ & 3x\sqrt[3]{2-x} &=& \ 0 \\ \Leftrightarrow\ & 3x = 0 & \text{or} & \ \sqrt[3]{2-x} = 0 \\ \Leftrightarrow\ & x = 0 & \text{or} & \ 2-x = 0 \\ \Leftrightarrow\ & x = 0 & \text{or} & \ x = 2 \end{aligned}$$

The zeros 0 and 2 divide the real number line into three intervals, where the function may have a different behaviour. The graph of this function is shown in Figure 4.6.

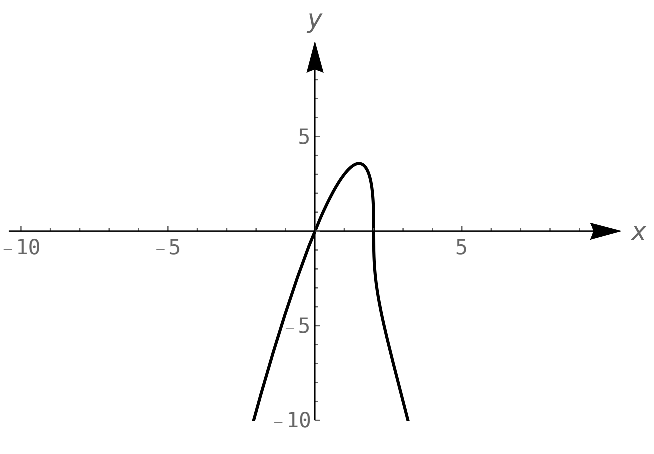


Figure 4.6: Graph of $f(x) = 3x\sqrt[3]{2-x}$.

2. In $g(x) = \sqrt{2-\sqrt[4]{x+3}}$, we have two radicals both of which are even indexed. To satisfy $\sqrt[4]{x+3}$, we require $x+3 \geq 0$ or $x \geq -3$. To satisfy $\sqrt{2-\sqrt[4]{x+3}}$, we need $2-\sqrt[4]{x+3} \geq 0$. Hence, we solve $2-\sqrt[4]{x+3} \geq 0$. If we let $r(x) = 2-\sqrt[4]{x+3}$, we know $x \geq -3$, so we concern

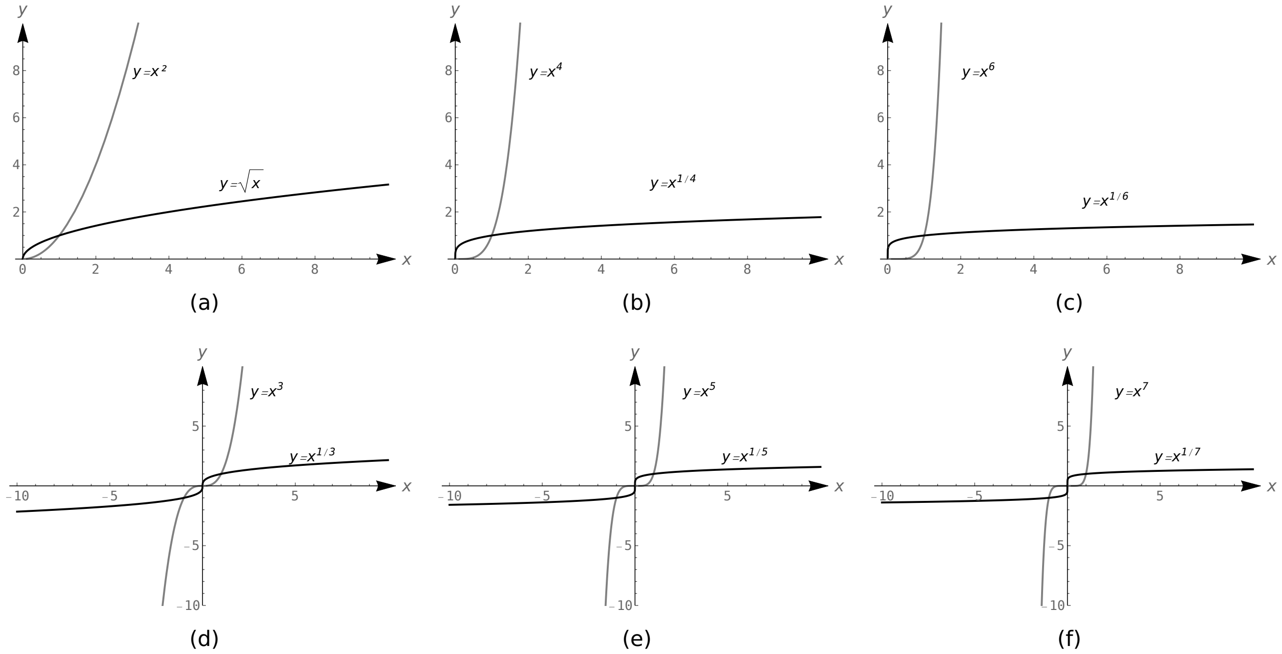


Figure 4.5: Graphs of $y = \sqrt{x}$ (a), $y = \sqrt[4]{x}$ (b), $y = \sqrt[6]{x}$ (c), $y = \sqrt[3]{x}$ (d), $y = \sqrt[5]{x}$ (e) and $y = \sqrt[7]{x}$ (f).

ourselves with only this portion of the number line. To find the zeros of r we set $r(x) = 0$ and solve $2 - \sqrt[4]{x+3} = 0$:

$$2 - \sqrt[4]{x+3} = 0 \quad \Leftrightarrow \quad \sqrt[4]{x+3} = 2\,,$$

which implies that $x + 3 = 2^4 = 16$, so $x = 13$. Since we raised both sides of an equation to an even power, we need to check to see if $x = 13$ is an extraneous solution. We find $x = 13$ does check since $2 - \sqrt[4]{x+3} = 2 - \sqrt[4]{13+3} = 2 - \sqrt[4]{16} = 2 - 2 = 0$. We find $2 - \sqrt[4]{x+3} \geq 0$ on $[-3, 13]$ so this is the domain of g. For what concerns g, we look for its zeros. Setting $g(x) = 0$ is equivalent to $\sqrt{2 - \sqrt[4]{x+3}} = 0$. After squaring both sides, we get $2 - \sqrt[4]{x+3} = 0$, whose solution we have found to be $x = 13$. Since we squared both sides, we double check and find $g(13)$ is, in fact, 0. The domain of g is $[-3, 13]$. This graph of this function is shown in Figure 4.7.

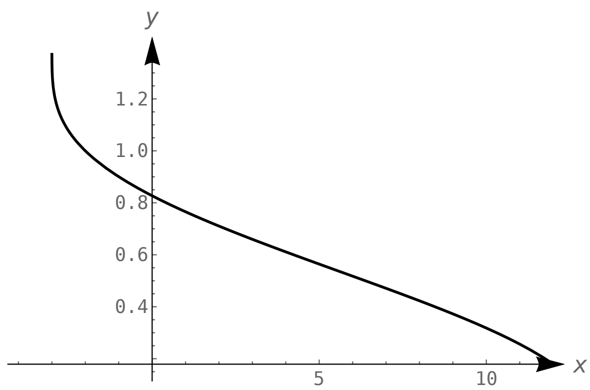


Figure 4.7: Graph of $g(x) = \sqrt{2 - \sqrt[4]{x+3}}$.

We conclude this section with an application in which an irrational function pops up naturally.

Example 4.9

Carl wishes to get high speed internet service installed in his remote bird observation post located 30 metres from the E34. The nearest junction box is located 50 metres downroad (Figure 4.8). Suppose it costs €15 per metre to run cable along the road and €20 per metre to run cable off of the road.

Express the total cost C of connecting the Junction Box to the Outpost as a function of x, the number of metres the cable is run along the E34 before heading off road directly towards the Outpost. Determine a reasonable applied domain for the problem.

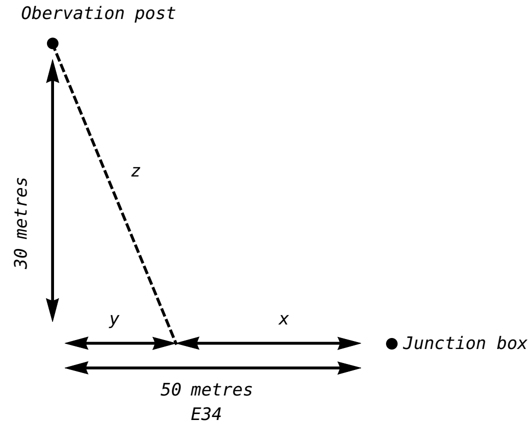


Figure 4.8: Map showing the observation post and junction box in Example 4.9.

Solution

The cost is broken into two parts: the cost to run cable along the E34 at €15 per metre, and the cost to run it off road at €20 per metre. Since x represents the metres of cable run along the E34, the cost for that portion is $15x$. From Figure 4.8, we see that the number of metres the cable is run off road is z, so the cost of that portion is $20z$.

Hence, the total cost is $C = 15x + 20z$. Our next goal is to determine z as a function of x. The diagram suggests we can use the Pythagorean theorem to get $y^2 + 30^2 = z^2$. But we also see $x + y = 50$ so that $y = 50 - x$. Hence, $z^2 = (50 - x)^2 + 900$. Solving for z, we obtain $z = \pm\sqrt{(50-x)^2 + 900}$. Since z represents a distance, we choose $z = \sqrt{(50-x)^2 + 900}$ so that our cost as a function of x only is given by

$$C(x) = 15x + 20\sqrt{(50-x)^2 + 900}.$$

From the context of the problem, we have $0 \leq x \leq 50$.

4.4 Conic sections

In this section, we will investigate the two-dimensional figures that are formed when a right circular cone is intersected by a plane.

History of conic sections

The Greek mathematician Menaechmus (c. 380-c. 320 BCE) is generally credited with discovering the shapes formed by the intersection of a plane and a right circular cone. Depending on how he tilted the plane when it intersected the cone, he formed different shapes at the intersection – beautiful shapes with near-perfect symmetry. It was also said that Aristotle may have had an intuitive understanding of these shapes, as he observed the orbit of the planet to be circular. He presumed that the planets moved in circular orbits around Earth, and for nearly 2000 years this was the commonly held belief.

It was not until the Renaissance movement that Johannes Kepler noticed that the orbits of the planet were not circular in nature. His law of planetary motion in the 1600s changed our view of the solar system forever. He claimed that the sun was at one end of the orbits, and the planets revolved around the sun in an oval-shaped path.

4.4.1 Overview

The name **conic section** (*kegelsnede*) is used to refer to any of the shapes that can be formed by intersecting a double-napped right circular cone with a plane. There are indeed several ways to intersect such a cone by a plane. Let us first consider a plane that does not contain the cone's vertex. Then, we can slice its top nappe, for instance, with a horizontal plane. This produces a circle (Figure 4.9(a)). Tilting this plane only slightly produces an ellipse (Figure 4.9(b)), while tilting the plane even further leads to a parabola (Figure 4.9(c)). If we continue increasing the tilting angle like this, the plane will at some point cut through both nappes, giving rise to a hyperbola (Figure 4.9(d)).

If the slicing plane contains the vertex of the cone, we get the so-called degenerate conics, namely a point (Figure 4.10(a)), a line (Figure 4.10(b)) or two intersecting lines (Figure 4.10(c)).

In the remainder of this section we will review the non-degenerate cases in detail. Then, in Chapter 5 we will show how any quadratic equation of the form

$$ax^2+bxy+cy^2+dx+ey+f=0$$

corresponds to a conic section.

4.4.2 Circles

In geometry, a circle is defined as follows.

Definitie 4.6 (Circle)
A **circle** (*cirkel*) with **centre** (*middelpunt*) (x_0,y_0) and **radius** (*straal*) $r>0$ is the set of all points (x,y) in the plane whose distance to (x_0,y_0) is r.

We express this definition algebraically using the distance formula as

$$r=\sqrt{(x-x_0)^2+(y-y_0)^2}\,.$$

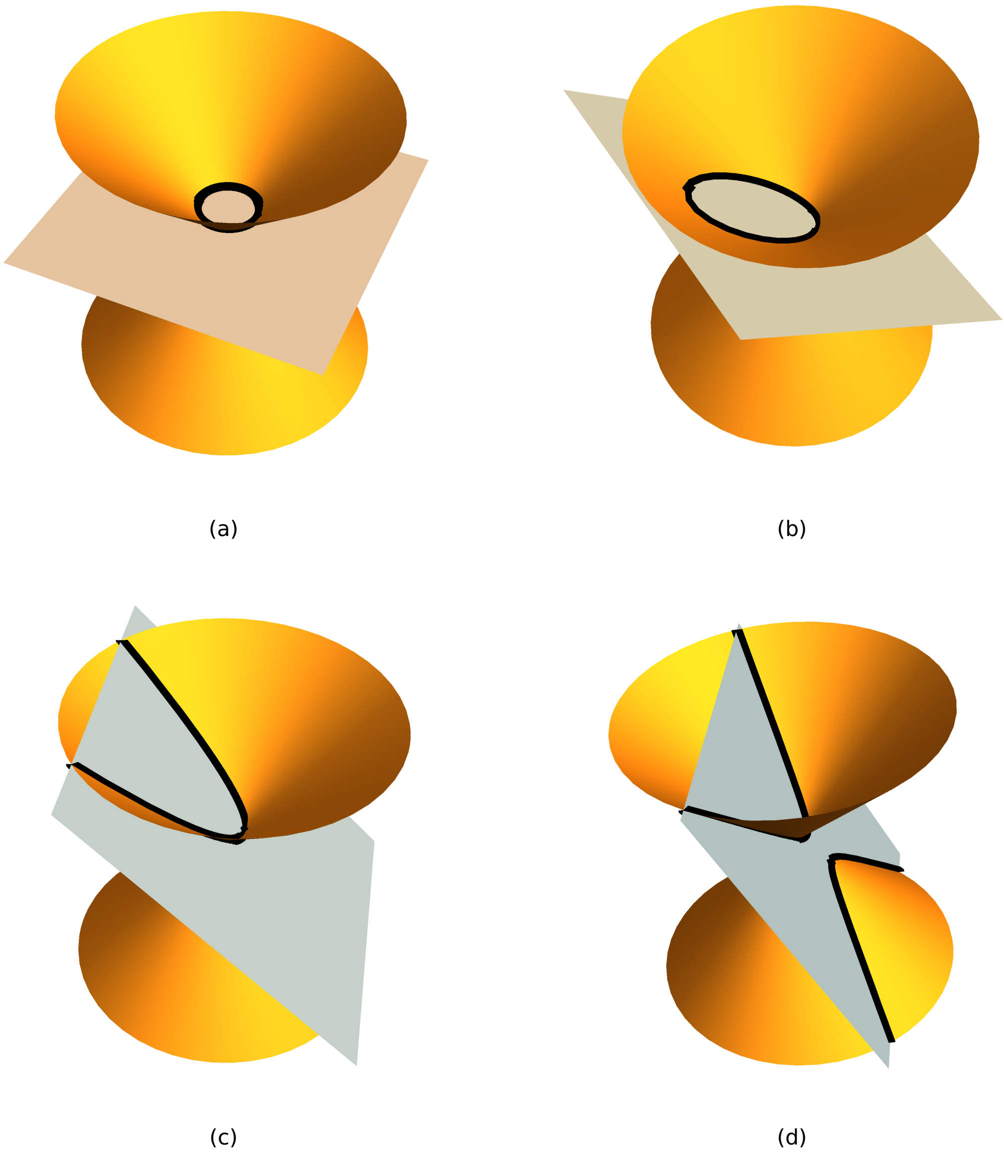


Figure 4.9: The intersection of a double-napped right circular cone with a plane with varying tilting angle: a circle (a), ellipse (b), parabola (c) and hyperbola (d).

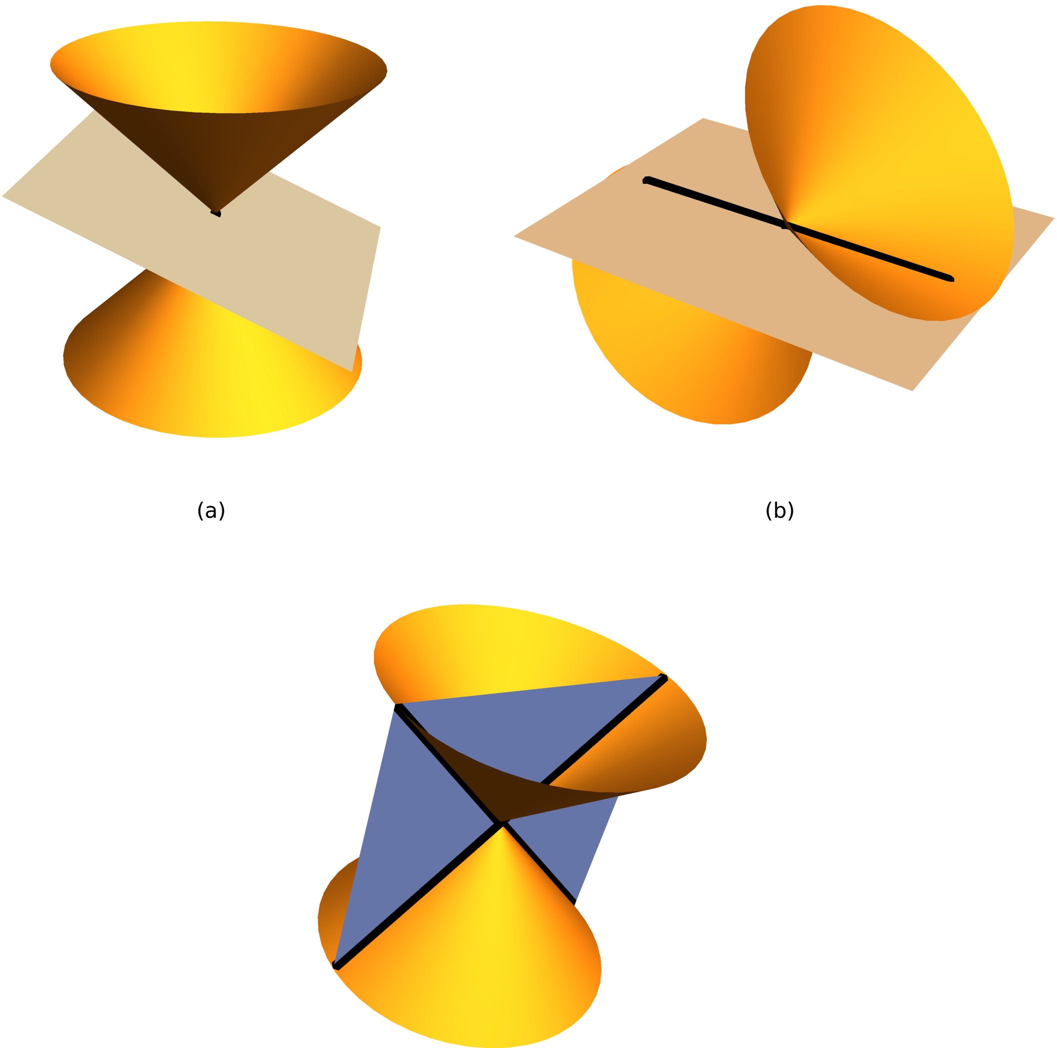
By squaring both sides of this equation, we get an equivalent equation (since $r > 0$) which gives us the standard equation of a circle with centre (x_0, y_0) and radius r:

$$(x - x_0)^2 + (y - y_0)^2 = r^2. \tag{4.4}$$

We close our introduction to circles with the most important circle in all of mathematics: the unit circle.

Definitie 4.7 (Unit circle)
The **unit circle** (*eenheidscirkel*) is the circle centred at $(0,0)$ with a radius of 1. The standard equation of the unit circle is

$$x^2 + y^2 = 1.$$



(a) (b) (c)

Figure 4.10: The intersection of a double-napped right circular cone with a plane containing the vertex of the cone and with varying tilting angle: a point (a), line (b) and two intersecting lines (c).

4.4.3 Ellipses

In the definition of a circle, (Definition 4.6), we fixed a point called the centre and considered all of the points which were at a fixed distance r from that one point. For the ellipse, we fix two distinct points and a distance d.

Definitie 4.8 (Ellipse)
Given two distinct points F_1 and F_2 in the plane and a fixed distance d, an **ellipse** (*ellips*) is the set of all points (x, y) in the plane such that the sum of each of the distances from F_1 and F_2 to (x, y) is d. The points F_1 and F_2 are called the **foci** (*brandpunten*) of the ellipse.

We may imagine taking a length of string and anchoring it to two points on a piece of paper. The curve traced out by taking a pencil and moving it so the string is always taut is an ellipse (Figure 4.11).

The centre of the ellipse is the midpoint of the line connecting the two foci. The **major axis** (*grote as*) of the ellipse is the line segment connecting two opposite ends of the ellipse which also contains the

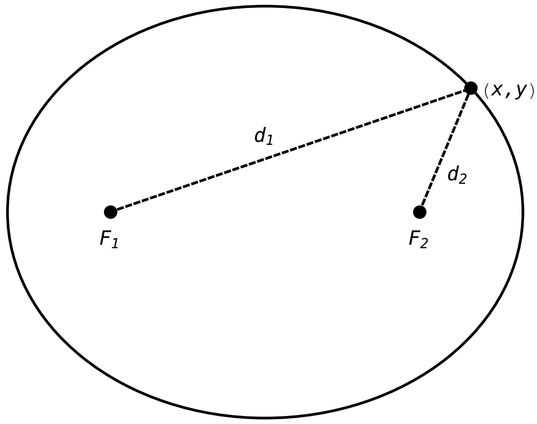


Figure 4.11: $d_1 + d_2 = d$ for all (x,y) on the ellipse.

centre and foci. The **minor axis** (*kleine as*) of the ellipse is the line segment connecting two opposite ends of the ellipse which contains the centre but is perpendicular to the major axis. The **vertices** (*top*) of an ellipse are the points of the ellipse which lie on the major axis. Notice that the centre is also the midpoint of the major axis (Figure 4.11) and that the major axis is the longer of the two axes through the centre, and likewise, the minor axis is the shorter of the two.

In order to derive the standard equation of an ellipse, we assume that the ellipse has its centre at $(0,0)$, its major axis along the x-axis, and has foci $(c,0)$ and $(-c,0)$ and vertices $(-a,0)$ and $(a,0)$ (Figure 4.12). We will label the y-intercepts of the ellipse as $(0,b)$ and $(0,-b)$. Moreover, we assume a, b, and c are all positive numbers.

Note that since $(a,0)$ is on the ellipse, it must satisfy the conditions of Definition 4.8. That is, the distance from $(-c,0)$ to $(a,0)$ plus the distance from $(c,0)$ to $(a,0)$ must equal the fixed distance d. Since all of these points lie on the x-axis, we get

$$\begin{aligned} (a+c)+(a-c) &= d \\ \Leftrightarrow \quad 2a &= d. \end{aligned}$$

We now use that fact $(0,b)$ is on the ellipse, along with the fact that $d = 2a$ to get

$$\begin{aligned} & \text{distance from } (-c,0) \text{ to } (0,b) + \text{distance from } (c,0) \text{ to } (0,b) &= 2a \\ \Leftrightarrow \quad & \sqrt{(0-(-c))^2+(b-0)^2}+\sqrt{(0-c)^2+(b-0)^2} &= 2a \\ \Leftrightarrow \quad & 2\sqrt{b^2+c^2} &= 2a \\ \Leftrightarrow \quad & \sqrt{b^2+c^2} &= a\,. \end{aligned}$$

From this, we get $a^2 = b^2 + c^2$, or $b^2 = a^2 - c^2$, which will prove useful later. Now consider an arbitrary point (x,y) on the ellipse. Applying Definition 4.8, we get

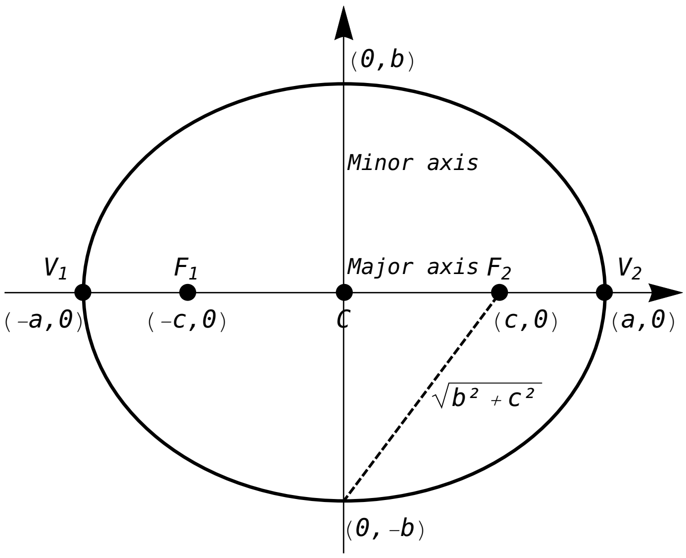


Figure 4.12: An ellipse with centre C $(0,0)$, foci F_1 $(-c,0)$, F_2 $(c,0)$ and vertices V_1 $(-a,0)$ and V_2 $(a,0)$.

$$\begin{array}{rrcll}
 & \sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2} & = & 2a & \\
\Leftrightarrow & \sqrt{(x+c)^2+y^2} & = & 2a-\sqrt{(x-c)^2+y^2} & \\
\Rightarrow & (x+c)^2+y^2 & = & 4a^2-4a\sqrt{(x-c)^2+y^2}+(x-c)^2+y^2 & \text{(Square both sides.)} \\
\Leftrightarrow & 4a\sqrt{(x-c)^2+y^2} & = & 4a^2+(x-c)^2-(x+c)^2 & \\
\Leftrightarrow & 4a\sqrt{(x-c)^2+y^2} & = & 4a^2-4cx & \\
\Leftrightarrow & a\sqrt{(x-c)^2+y^2} & = & a^2-cx & \\
\Rightarrow & a^2\left((x-c)^2+y^2\right) & = & a^4-2a^2cx+c^2x^2 & \text{(Square both sides.)} \\
\Leftrightarrow & a^2x^2-2a^2cx+a^2c^2+a^2y^2 & = & a^4-2a^2cx+c^2x^2 & \\
\Leftrightarrow & a^2x^2-c^2x^2+a^2y^2 & = & a^4-a^2c^2 & \\
\Leftrightarrow & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right). &
\end{array}$$

Recall now that $b^2 = a^2 - c^2$ so that

$$\begin{array}{rrcl}
 & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right) \\
\Leftrightarrow & b^2x^2+a^2y^2 & = & a^2b^2 \\
\Leftrightarrow & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} & = & 1\,.
\end{array}$$

This equation is for an ellipse centred at the origin. To get the formula for the ellipse centred at (x_0, y_0), we could use the transformations from Section 3.2.5 to obtain:

$$\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}=1\,. \tag{4.5}$$

Note that if $a > b$, then we have an ellipse whose major axis is horizontal, and hence, the foci lie to the left and right of the centre. If $b > a$, the roles of the major and minor axes are reversed, and the foci lie above and below the centre. Finally, it is worth mentioning that if $a = b$, we arrive at the standard equation of a circle. This indicates that it is fine to think of an ellipse as the deformation of a circle in which the circle is stretched farther in one direction than the other.

As with circles, an equation may be given which is an ellipse, but is not in the standard form of

Equation (4.5). In those cases, we will need to massage the given equation into the standard form by taking the following steps:

1. Group the same variables together on one side of the equation and position the constant on the other side.
2. Complete the square in both variables as needed.
3. Divide both sides by the constant term so that the constant on the other side of the equation becomes 1.

Example 4.10

Graph $x^2+4y^2-2x+24y+33=0$. Find the centre, the major and minor axes, the vertices, the endpoints of the minor axis, and the foci.

Solution

Since we have a sum of squares and the squared terms have unequal coefficients, it is a good bet we have an ellipse on our hands. We need to complete both squares, and then divide, if necessary, to get the right-hand side equal to 1. This ultimately leads to

$$\frac{(x-1)^2}{4}+\frac{(y+3)^2}{1}=1\,.$$

Now that this equation is in the standard form, we see that $x-x_0$ is $x-1$ so $x_0=1$, and $y-y_0$ is $y+3$ so $y_0=-3$. Hence, our ellipse is centred at $(1,-3)$. We see that $a^2=4$ so $a=2$, and $b^2=1$ so $b=1$. Consequently, the major axis will lie along the horizontal line $y=-3$, which means the minor axis lies along the vertical line $x=1$. The vertices are the points on the ellipse which lie along the major axis so in this case, they are the points $(-1,-3)$ and $(3,-3)$, and the endpoints of the minor axis are $(1,-2)$ and $(1,-4)$. To find the foci, we find $c=\sqrt{4-1}=\sqrt{3}$, which means the foci lie $\sqrt{3}$ units from the centre. Since the major axis is horizontal, the foci lie $\sqrt{3}$ units to the left and right of the centre, at $(1-\sqrt{3},-3)$ and $(1+\sqrt{3},-3)$ (Figure 4.13).

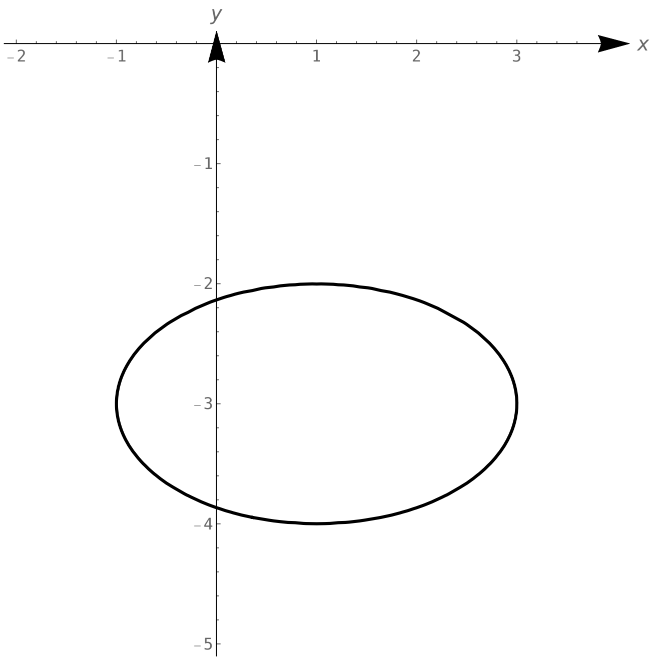


Figure 4.13: Graph of $x^2+4y^2-2x+24y+33=0$.

Johannes Kepler discovered that the orbits along which the planets travel around the Sun are ellipses with the Sun approximately at one focus (Figure 4.14). A key feature of such planetary orbits are their eccentricity, which is a measure of the roundness of an ellipse and can be quantified as below.

Definitie 4.9 (Eccentricity)
The **eccentricity** (*excentriciteit*) of an ellipse, denoted e, is the following ratio:

$$e = \frac{\text{distance from the centre to a focus}}{\text{distance from the centre to a vertex}}.$$

From this definition, we infer that for a circle $e = 0$, while we have for an ellipse that $e < 1$.

Finally, it is important to underline that ellipses have a reflective property. If we imagine the dashed lines in Figure 4.11 representing sound waves, then the waves emanating from one focus reflect off the top of the ellipse and head towards the other focus. Such geometry is exploited in the construction of so-called whispering galleries (Figure 4.15). If a person whispers at one focus, a person standing at the other focus will hear the first person as if they were standing right next to them. We explore this in our last example.

Example 4.11

Lisa and Jason want to exchange secrets from across a crowded whispering gallery. If the room is 40 metres high at the centre and 100 metres wide at the floor, how far from the outer wall should each of them stand so that they will be positioned at the foci of the ellipse?

Solution

It is most convenient to imagine this ellipse centred at $(0,0)$. Since the ellipse is 100 units wide and 40 units tall, we get $a = 50$ and $b = 40$. Hence, our ellipse has the equation

$$\frac{x^2}{50^2} + \frac{y^2}{40^2} = 1.$$

We are looking for the foci, and we get $c = \sqrt{50^2 - 40^2} = \sqrt{900} = 30$, so that the foci are 30 units from the centre. That means they are $50 - 30 = 20$ units from the vertices. Hence, Jason and Lisa should stand 20 metres from opposite ends of the gallery.

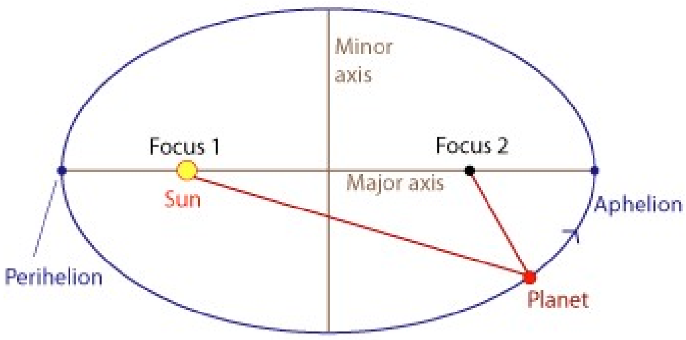


Figure 4.14: An elliptical orbit of a planet.

Figure 4.15: Whispering gallery at Grand Central station, New York, United states.

4.4.4 Parabolas

We have already learned in Section 4.1 that the graph of a quadratic function $f(x) = ax^2 + bx + c$ ($a \neq 0$) is called a parabola. We may also define parabolas in terms of distance.

Definitie 4.10 (Parabola)
Let F be a point in the plane and d be a line not containing F. A **parabola** (*parabool*) is the set of all points equidistant from F and d. The point F is called the **focus** (*brandpunt*) of the parabola and the line d is called the **directrix** (*richtlijn*) of the parabola.

Essentially, in Figure 4.16, each dashed line from the point F to a point on the curve has the same length as the dashed line from the point on the curve to the line d. The point V is the **vertex** (*top*). The vertex is the point on the parabola closest to the focus.

We want to use only the distance definition of parabola to derive the equation of a parabola and we should get an expression much like those studied in Section 4.1. Let p denote the directed distance from the vertex to the focus, which by definition is the same as the distance from the vertex to the directrix. For simplicity, assume that the vertex is $(0,0)$ and that the parabola opens upwards. Hence, the focus is $(0,p)$ and the directrix is the line $y = -p$. All this is presented schematically in Figure 4.17.

From the definition of parabola, we know that the distance from $(0,p)$ to (x,y) is the same as the

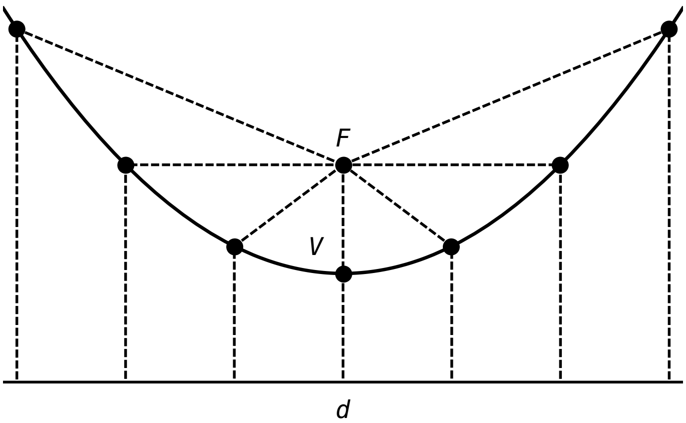


Figure 4.16: Geometric construction a a parabola.

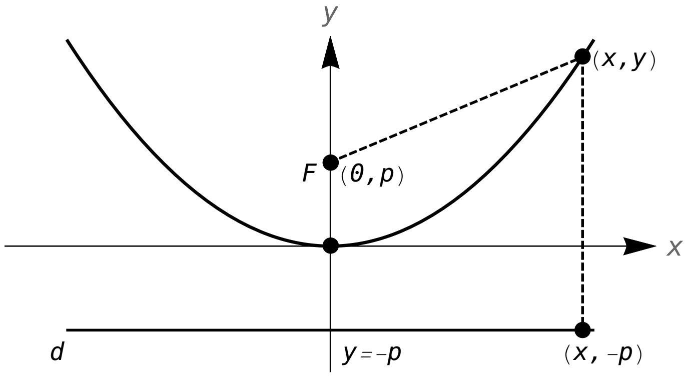


Figure 4.17: Parabola with vertex in $(0,0)$, directrix $d: y = -p$ and focus $F(0,p)$.

distance from $(x,-p)$ to (x,y). Using the distance formula, we get

$$\begin{array}{rrcll} & \sqrt{(x-0)^2+(y-p)^2} & = & \sqrt{(x-x)^2+(y-(-p))^2} & \\ \Leftrightarrow & \sqrt{x^2+(y-p)^2} & = & \sqrt{(y+p)^2} & \\ \Leftrightarrow & x^2+(y-p)^2 & = & (y+p)^2 & \text{(Square both sides.)} \\ \Leftrightarrow & x^2+y^2-2py+p^2 & = & y^2+2py+p^2 & \text{(Expand quantities.)} \\ \Leftrightarrow & x^2 & = & 4py. & \text{(Gather like terms.)} \end{array}$$

Solving for y yields $y = \frac{x^2}{4p}$, which is a quadratic function of the form found in Definition 4.2 with $a = \frac{1}{4p}$ and vertex $(0,0)$.

When $p < 0$, the parabola opens downwards. In our formulation, we say that p is a directed distance from the vertex to the focus: if $p > 0$, the focus is above the vertex; if $p < 0$, the focus is below the vertex. If we choose to place the vertex at an arbitrary point (x_0,y_0), we arrive at the standard equation of a vertical parabola using the transformations from Section 3.2.5:

$$(x-x_0)^2 = 4p(y-y_0)\,. \tag{4.6}$$

Notice that in this standard equation, only one of the variables, x, is squared. This is a quick way to distinguish an equation of a parabola from that of a circle or ellipse because in the equation of a circle or ellipse, both variables are squared.

If we interchange the roles of x and y, we can produce horizontal parabolas: parabolas which open to the left or to the right. The directrices of these would be vertical lines and the focus would either lie to the left or to the right of the vertex, as seen in Figure 4.18. The standard equation of a horizontal parabola with vertex (x_0,y_0) is

$$(y-y_0)^2 = 4p(x-x_0)\,. \tag{4.7}$$

If $p > 0$, the parabola opens to the right; if $p < 0$, it opens to the left.

As with circles and ellipses, not all parabolas will come to us in the forms in Equations (4.6) or (4.7). If we encounter an equation with two variables in which exactly one variable is squared, we can, however, put the equation into a standard form using the following steps.

1. Group the variable which is squared on one side of the equation and position the non-squared variable and the constant on the other side.
2. Complete the square if necessary and divide by the coefficient of the perfect square.
3. Factor out the coefficient of the non-squared variable.

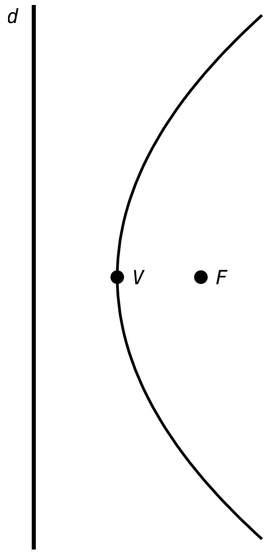


Figure 4.18: Horizontal parabola opening to the right.

Example 4.12

Consider the equation $y^2 + 4y + 8x = 4$. Put this equation into standard form and graph the parabola. Find the vertex, focus, and directrix.

Solution

We need a perfect square on the left-hand side of the equation and factor out the coefficient of the non-squared variable (x) on the other. We finally arrive at:

$$(y+2)^2 = -8(x-1)\,.$$

Now that the equation is in the form given in Equation (4.7), we see that $x - x_0$ is $x-1$ so $x_0 = 1$, and $y - y_0$ is $y+2$ so $y_0 = -2$. Hence, the vertex is $(1,-2)$. We also see that $4p = -8$ so that $p = -2$. Since $p < 0$, the focus will be the left of the vertex and the parabola will open to the left. The distance from the vertex to the focus is $|p| = 2$, which means the focus is 2 units to the left of 1, so if we start at $(1,-2)$ and move left 2 units, we arrive at the focus $(-1,-2)$. The directrix, then, is 2 units to the right of the vertex, so if we move right 2 units from $(1,-2)$, we would be on the vertical line $x = 3$. Moreover, the parabola is 8 units wide at the focus.

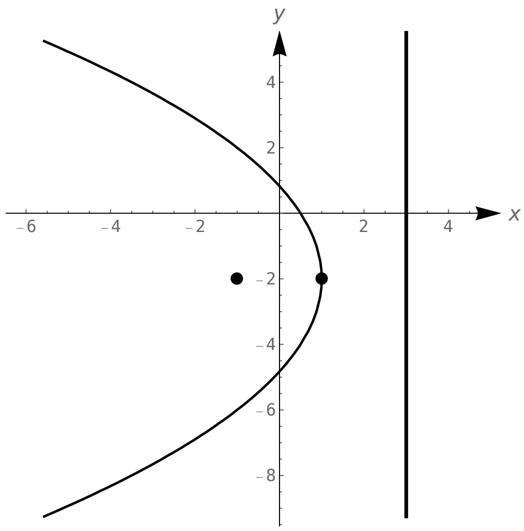
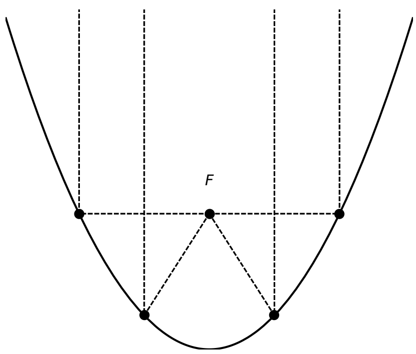


Figure 4.19: Graph of $y^2 + 4y + 8x = 4$.

Parabolas are used to model physical phenomena such as the trajectories of projectiles. Other applications of the parabola concern its reflective property. If we imagine the dashed lines in Figure 4.20(a) as emanating from the focus, we see that the waves are reflected off the parabola in a coherent fashion as in the case in a flash light. Here, the bulb is placed at the focus and the light rays are reflected off a parabolic mirror to give directional light. This reasoning also works the other way around if we imagine the dashed lines below as electromagnetic waves heading towards a parabolic dish. It turns out that the waves reflect off the parabola and concentrate at the focus which then becomes the optimal place for the receiver (Figure 4.20(b)).



(a)

(b)

Figure 4.20: Reflective property of parabola (a) and a parabolic antenna making use thereof (b).

Example 4.13

A satellite dish is to be constructed in the shape of a paraboloid of revolution. If the receiver placed at the focus is located 2 metres above the vertex of the dish, and the dish is to be 12 metres wide, how deep will the dish be?

Solution

One way to approach this problem is to determine the equation of the parabola suggested to us by this data. For simplicity, we will assume the vertex is $(0,0)$ and the parabola opens upwards. Our standard form for such a parabola is $x^2 = 4py$. Since the focus is 2 units above the vertex, we know $p = 2$, so we have $x^2 = 8y$. Since the parabola is 12 metres wide, we know the edge is 6 metres from a vertical line through the vertex. To find the depth, we are looking for the y value when $x = 6$. Substituting $x = 6$ into the equation of the parabola yields $6^2 = 8y$ or $y = 36/8 = 4.5$. Hence, the dish will be 4.5 metres deep.

Examples of parabolas occurring in nature are also manifold. For instance, rainbows and many natural bridges have a parabolic shape. Moreover, the trajectory traversed by jumping fish or aquatic mammals also approaches a parabola (Figure 4.21).

(a) (b) (c)

Figure 4.21: Parabolas in nature: rainbow (a), natural bridge (b) and trajectories of jumping dolphins.

4.4.5 Hyperbolas

In the definition of an ellipse, Definition 4.8, we fixed two points called foci and looked at points whose distances to the foci always added to a constant distance d. But what, if any, curve we would generate if we replaced added with subtracted. The answer is a hyperbola.

Definitie 4.11 (Hyperbola)
Given two distinct points F_1 and F_2 in the plane and a fixed distance d, a **hyperbola** (*hyperbool*) is the set of all points (x, y) in the plane such that the absolute value of the difference of each of the distances from F_1 and F_2 to (x, y) is d. The points F_1 and F_2 are called the **foci** (*brandpunten*) of the hyperbola.

Note that the hyperbola has two parts, called **branches** (*tak*). The **centre** (*middelpunt*) of the hyperbola is the midpoint of the line connecting the two foci. The **transverse axis** (*hoofdas*) of the hyperbola is the line segment connecting two opposite ends of the hyperbola which also contains the centre and foci. The **vertices** (*top*) of a hyperbola are the points of the hyperbola which lie on the transverse axis. In addition, there are lines called **asymptotes** (*asymptoot*) which the branches of the hyperbola approach for large x- and y-values (Figure 4.22). The **conjugate axis** (*nevenas*) of a hyperbola is the line through the centre which is perpendicular to the transverse axis. It contains two **imaginary vertices** (*imaginaire toppen*).

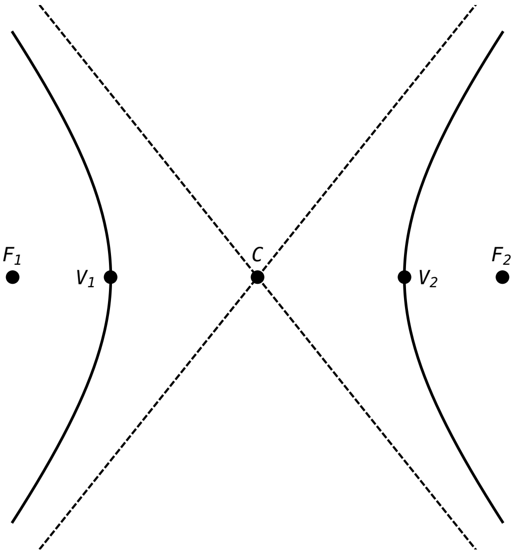


Figure 4.22: A hyperbola with centre C; foci F_1, F_2; and vertices V_1, V_2 and asymptotes (dashed)

Suppose now we wish to derive the equation of a hyperbola. For simplicity, we shall assume that the centre is $(0,0)$, the vertices are $(a,0)$ and $(-a,0)$ and the foci are $(c,0)$ and $(-c,0)$. We label the endpoints of the conjugate axis $(0,b)$ and $(0,-b)$. Although b does not enter into our derivation, we will have to justify this choice as you shall see later. As before, we assume a, b, and c are all positive numbers. Schematically we get the picture shown in Figure 4.23.

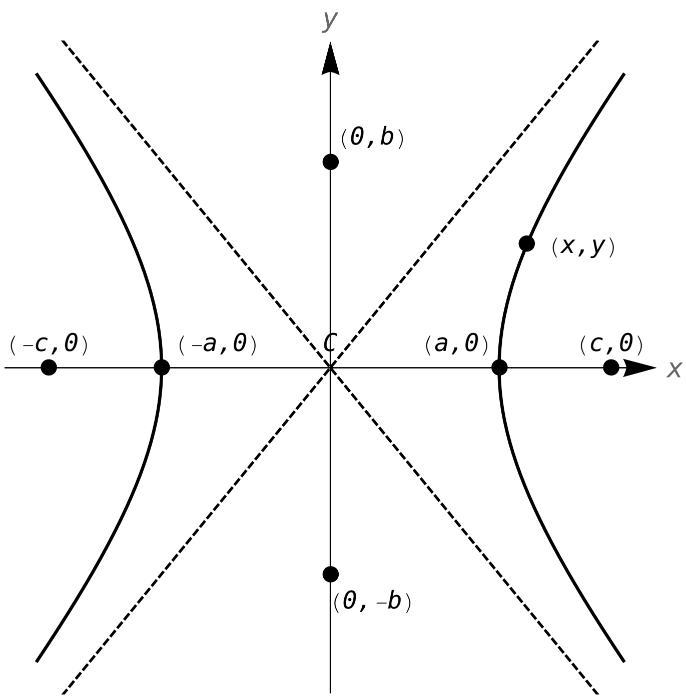


Figure 4.23: A hyperbola with centre in $(0,0)$; foci $F_1(-c,0)$, $F_2(c,0)$; and vertices $V_1(-a,0)$, $V_2(a,0)$, imaginary vertices $(0,-b)$ and $(0,b)$, and asymptotes (dashed).

Since $(a,0)$ is on the hyperbola, it must satisfy the conditions of Definition 4.11. That is, the distance from $(-c,0)$ to $(a,0)$ minus the distance from $(c,0)$ to $(a,0)$ must equal the fixed distance d. Since all these points lie on the x-axis, we get

$$\begin{aligned} & \text{distance from } (-c,0) \text{ to } (a,0) - \text{distance from } (c,0) \text{ to } (a,0) &=& \; d \\ \Leftrightarrow \quad & (a+c)-(c-a) &=& \; d \\ \Leftrightarrow \quad & 2a &=& \; d. \end{aligned}$$

Hence, d is the distance between the vertices V_1 and V_2.

Now consider a point (x, y) on the hyperbola. Applying Definition 4.11, we get

$$\begin{aligned} & \left|\text{distance from } (-c,0) \text{ to } (x,y) - \text{distance from } (c,0) \text{ to } (x,y)\right| &=& \; 2a \\ \Leftrightarrow \quad & \left|\sqrt{(x-(-c))^2+(y-0)^2}-\sqrt{(x-c)^2+(y-0)^2}\right| &=& \; 2a \\ \Leftrightarrow \quad & \left|\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}\right| &=& \; 2a\,. \end{aligned}$$

Following the same procedure as when deriving the standard formula of an ellipse (Equation (4.5)), we arrive at:

$$\left(a^2-c^2\right)x^2+a^2y^2 = a^2\left(a^2-c^2\right).$$

What remains is to determine the relationship between a, b and c. To that end, we note that since a and c are both positive numbers with $a < c$, we get $a^2 < c^2$ so that a^2-c^2 is a negative number. Hence, c^2-a^2 is a positive number. Let us rewrite the equation by solving for y^2/x^2 to get

$$\begin{aligned} & \left(a^2-c^2\right)x^2+a^2y^2 &=& \; a^2\left(a^2-c^2\right) \\ \Leftrightarrow \quad & -\left(c^2-a^2\right)x^2+a^2y^2 &=& \; -a^2\left(c^2-a^2\right) \\ \Leftrightarrow \quad & a^2y^2 &=& \; \left(c^2-a^2\right)x^2-a^2\left(c^2-a^2\right) \\ \Leftrightarrow \quad & \frac{y^2}{x^2} &=& \; \frac{\left(c^2-a^2\right)}{a^2}-\frac{\left(c^2-a^2\right)}{x^2}\,. \end{aligned}$$

As x and y attain very large values, the quantity $\frac{(c^2-a^2)}{x^2} \to 0$ so that $\frac{y^2}{x^2} \to \frac{(c^2-a^2)}{a^2}$. By setting $b^2 = c^2-a^2$ we get $\frac{y^2}{x^2} \to \frac{b^2}{a^2}$. This shows that $y \to \pm\frac{b}{a}x$ as $|x|$ grows large. Thus $y = \pm\frac{b}{a}x$ are the asymptotes to the graph. In our equation of the hyperbola we can substitute $a^2-c^2 = -b^2$ which yields

$$\begin{aligned} & \left(a^2-c^2\right)x^2+a^2y^2 &=& \; a^2\left(a^2-c^2\right) \\ \Leftrightarrow \quad & -b^2x^2+a^2y^2 &=& \; -a^2b^2 \\ \Leftrightarrow \quad & \frac{x^2}{a^2}-\frac{y^2}{b^2} &=& \; 1\,. \end{aligned}$$

The equation above is for a hyperbola whose centre is the origin and which opens to the left and right.

If the hyperbola were centred at a point (x_0, y_0), we would get the following:

$$\frac{(x-x_0)^2}{a^2}-\frac{(y-y_0)^2}{b^2} = 1\,. \tag{4.8}$$

If the roles of x and y were interchanged, then the hyperbola's branches would open upwards and downwards and we would get a vertical hyperbola. Its standard equation is given by

$$\frac{(y-y_0)^2}{b^2}-\frac{(x-x_0)^2}{a^2} = 1\,. \tag{4.9}$$

By convention, a always refers to the coefficient in the denominator of the term containing x^2, while b refers to the coefficient appearing in the denominator of the term containing y^2.

The distance from the centre to the foci, c, as seen in the derivation, can be found by the formula $c = \sqrt{a^2+b^2}$. Also note that we can quickly distinguish the equation of a hyperbola from that of a circle or ellipse because the hyperbola formula involves a difference of squares where the circle and ellipse formulas both involve the sum of squares.

Determining the location of a known event has many practical uses such as: locating the epicentre of an earthquake, an airplane crash site, the position of the person speaking in a large room, etc.. To determine the location of an earthquake's epicentre, seismologists use trilateration. A seismograph allows one to determine how far away the epicentre was; using three separate readings, the location of the epicentre can be approximated.

Example 4.14

Consider three microphones at positions A, B and C which all record a noise (a person's voice, an explosion, etc.) created at unknown location D. The microphone does not know when the sound was created, only when the sound was detected. How can the location be determined in such a situation?

Solution

If each location has a clock set to the same time, hyperbolas can be used to determine the location. Suppose the microphone at position A records the sound at exactly 12:00, location B records the time exactly 1 second later, and location C records the noise exactly 2 seconds after that. We are interested in the difference of times. Since the speed of sound is approximately 340 m/s, we can conclude quickly that the sound was created 340 meters closer to position A than position B. If A and B are a known distance apart (as shown in Figure 4.24(a)), then we can determine a hyperbola on which D must lie.

The difference of distances between A and B is 340 metres; this is also the distance between vertices of the hyperbola. So we know $2a = 340$. Positions A and B lie on the foci, so $2c = 1000$. From this we can find $b \approx 470$ and can sketch the hyperbola, with equation

$$\frac{x^2}{170^2} - \frac{y^2}{470^2} = 1\,,$$

whose graph is shown in Figure 4.24(b). We only care about the side closest to A because the sound was first heard at that location.

We can also find the hyperbola defined by positions B and C. In this case, $2b = 680$ as the sound travelled an extra 2 seconds to get to C. We still have $2c = 1000$, centring this hyperbola at $(-500, 500)$. We find $a \approx 367$. This hyperbola has equation

$$\frac{(y-500)^2}{340^2} - \frac{(x+500)^2}{367^2} = 1\,,$$

and is sketched in Figure 4.24(c). The intersection point of the two graphs is the location of the sound, at approximately $(188, -222.5)$.

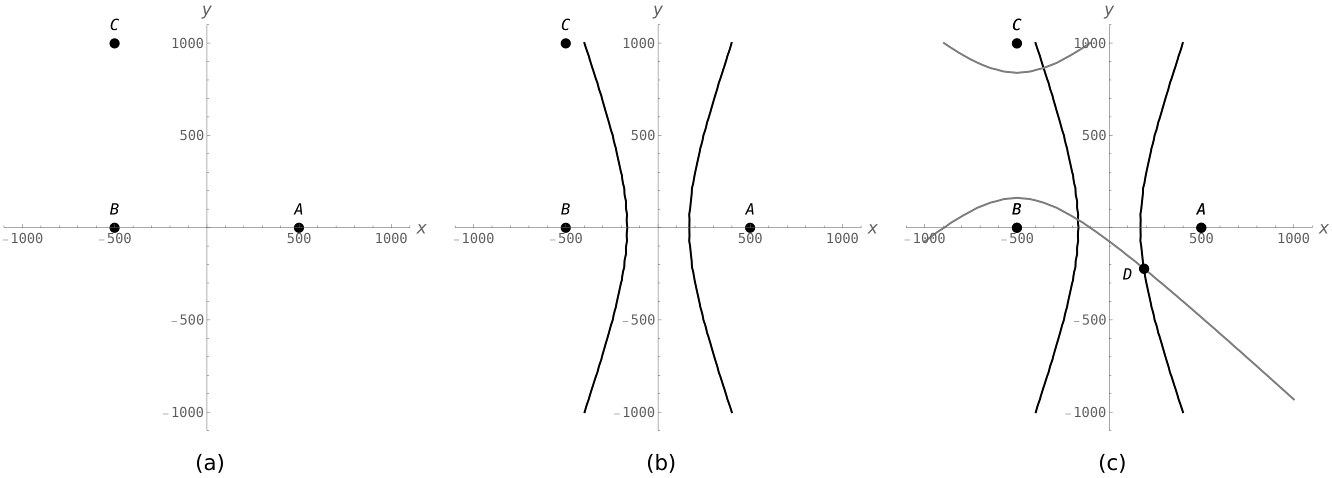
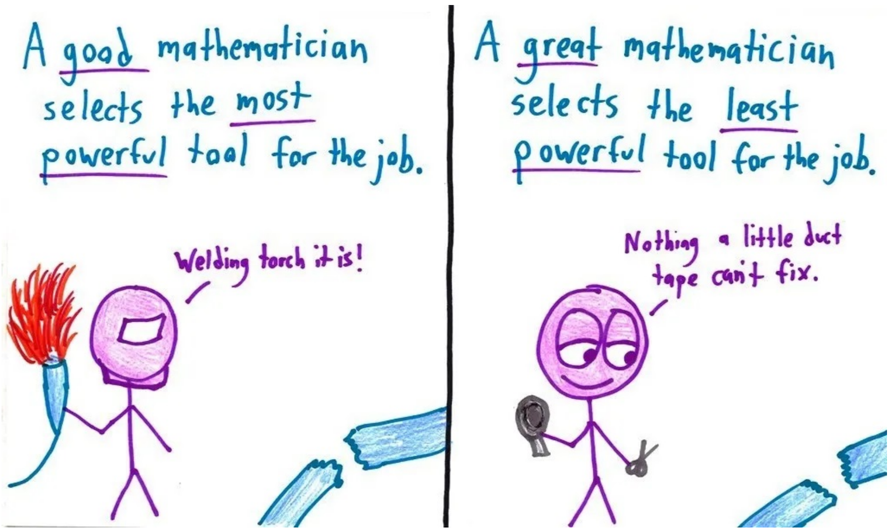


Figure 4.24: An example of trilateration using hyperbolas.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

4.5 Exercises

Polynomial functions

Assignment 4.1 — Draw and describe the following areas.

(a) $y \leq x-1$

(b) $|x|-4 < y < 2-x$

(c) $x^2 \leq y < x+2$

(d) $x^2+2x+y^2 < 8$

(e) $x^2+y^2 < 2x, \quad x^2+y^2 < 2y$

(f) $x^2+y^2-4x+2y > 4, \quad x+y > 1$

Assignment 4.2 — Factorize the polynomials below into real factors.

(a) $2x^6-128$

(b) $8x^3+12x^2+6x+1$

(c) $2x^3+3x^2+2x+3$

(d) $x^4-7x^3+18x^2-20x+8$

(e) $x^4+x^3-2x^2-4x-8$

(f) $2x^3-3x^2-3x+2$

(g) $8a^3-60a^2b+150ab^2-125b^3$

(h) $a^2-2a+1-b^2-4bc-4c^2$

(i) $x^3-4x^2y+7xy^2-4y^3$

(j) $2(x^2+6x+1)^2+5(x^2+6x+1)(x^2+1)+2(x^2+1)^2$

Assignment 4.3 — Write each polynomial as a product of real factors.

(a) $16x^4-8x^2+1$

(b) x^4-1

(c) $x^5-x^4-16x+16$

(d) $x^5+x^3+8x^2+8$

(e) $x^4+6x^3+9x^2$

(f) $x^6-3x^4+3x^2-1$

(g) $x^9-4x^7-x^6+4x^4$

Assignment 4.4 — Solve the equations below in $\mathbb{C}$ and determine both the real and complex decomposition.

(a) $x^3-3x^2+20=0$

(b) $2x^3-4x^2-10x+12=0$

(c) $x^6-16x^3+64=0$

(d) $8x^4-20x^3+18x^2-7x+1=0$

(e) $x^3-16x^2+48x+72=0$

(f) $4x^3-14x^2+8x+8=0$

(g) $x^5+6x^4+x^3-26x^2-32=0$

(h) $-2x^6-10x^5-16x^4-8x^3=0$

Assignment 4.5 — Determine a and b such that

(a) $2x^3+ax^2+bx-3$ has -1 and 3 as zeros.

(b) ax^3+19x^2+bx+8 is divisible by $x+2$ and $x+4$.

Assignment 4.6 — Solve the equation below in $\mathbb{R}$. Write the set of solutions as an interval.

(a) $-2x^3+19x^2-49x+20>0$

(b) $x^4-9x^2 \leq 4x-12$

(c) $3x^2+2x<x^4$

(d) $\dfrac{x^3+2x^2}{2}<x+2$

(e) $2x^4>5x^2+3$

Rational functions

Assignment 4.7 — Perform the divisions below.

(a) $\dfrac{1-5x^4+4x^5}{1-x}$

(b) $\dfrac{x^3-1}{x^2-2}$

(c) $\dfrac{x^2}{x^2+5x+3}$

(d) $\dfrac{x^3}{x^2+2x+3}$

(e) $\dfrac{2x^3-3x^2+4x-5}{x^2-6x+7}$

Assignment 4.8 — Solve the following rational equations in $\mathbb{R}$.

(a) $\dfrac{x^2-2x+1}{x^3+x^2-2x}=1$

(b) $\dfrac{x}{x^2-1}>0$

(c) $\dfrac{4x}{x^2+4}\geq 0$

(d) $\dfrac{3x^2-5x-2}{x^2-9}<0$

(e) $\dfrac{x^4-4x^3+x^2-2x-15}{x^3-4x^2}\geq x$

(f) $\dfrac{5x^3-12x^2+9x+10}{x^2-1}\geq 3x-1$

Assignment 4.9 — Determine the domain and intersections with the x-axis and y-axis of the rational functions below. Also determine any vertical and horizontal asymptotes.

(a) $f(x)=\dfrac{3x+2}{x^2+2x+2}$

(b) $f(x)=\dfrac{x^2-9}{x^3-x}$

(c) $f(x)=\dfrac{4}{x^3+x^2}$

(d) $f(x)=\dfrac{x^3+3x^2+6}{x^2+x-1}$

(e) $f(x)=\dfrac{x^2-x-12}{x^2+x-6}$

(f) $f(x)=\dfrac{x^2-x-6}{x+1}$

Irrational functions

Assignment 4.10 — Solve the following irrational equations in $\mathbb{R}$.

(a) $1+\dfrac{x+1}{\sqrt{x^2+2x}}=0$

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

(c) $\sqrt{x^2-3x+2}=|x|-2$

(d) $\sqrt[3]{x}+\sqrt[3]{x^2}+x>0$

(e) $\sqrt{3x+1}-\sqrt{x-4}=\sqrt{x+1}$

(f) $\sqrt{-x^2-x}\leq 2x+1$

Assignment 4.11 — In the theory of relativity, the mass m [M] of an object is not a constant quantity, but a variable quantity depending on the velocity v [LT^{-1}] of the object according to

$$m(v)=m_0\frac{c}{\sqrt{c^2-v^2}},$$

with $c=299792.458$ km/s (the speed of light) and m_0 [M] the mass of the object at rest. What speed must an object have in order for its mass to be twice its resting mass?

Assignment 4.12 — Determine the domain and intersections with the x-axis and y-axis of the irrational functions below. Also determine any vertical asymptotes.

(a) $f(x)=|5-\sqrt{8+2x}|$

(b) $f(x)=\sqrt[4]{\dfrac{16x}{x^2-9}}$

(c) $f(x)=\dfrac{5x}{\sqrt[3]{x^3+8}}$

(d) $f(x)=x^{\frac{2}{3}}(x-7)^{\frac{1}{3}}$

(e) $f(x)=\dfrac{1}{x\sqrt{x^2-4}}$

(f) $f(x)=\dfrac{\sqrt{x-3}}{\sqrt{2x+2}-\sqrt{x-1}-2}$

(g) $f(x)=\dfrac{\sqrt{2x^2-3x-2}}{\sqrt[3]{x^3+3x^2+7}-x-1}$

Assignment 4.13 — Determine which function belongs to which graph in Figure 4.25.

(a) $f(x)=\sqrt{\dfrac{1}{9-x^2}}$

(b) $f(x)=\sqrt{-2x+6}$

(c) $f(x)=\sqrt{x^3-3x^2+2x}$

(d) $f(x)=\sqrt{3x^2-x+2}$

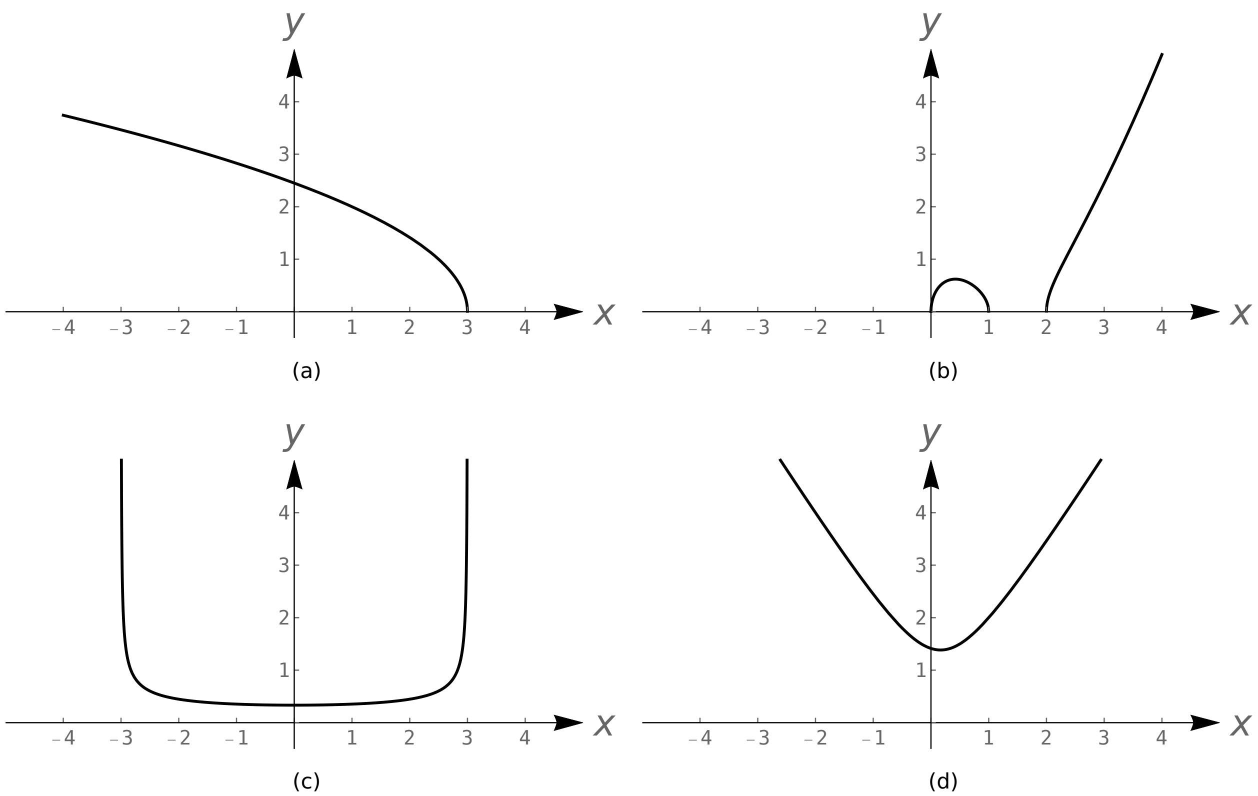


Figure 4.25: Graphs of the irrational functions in Exercise 16.

Assignment 4.14 — The function

$$f_t(x) = \frac{4}{\sqrt{2tx - x^2}},$$

with $t \in \mathbb{R}_0^+$ represents a family of irrational functions.

(a) Determine from f_t the domain, zero points, and asymptotes.

(b) Sketch the graphs for f_1, f_2 and f_3.

(c) Investigate whether two graphs for different values of t have an intersection point.

Conic sections

Assignment 4.15 — Determine the shape of the parabolas below. Find the top, the focal point, the axis of symmetry, the directrix, and the points of intersection with the x-axis and the y-axis. Draw the graph and determine the image.

(a) $y = x^2 - 4x + 3$

(b) $y = x^2 - 2x$

(c) $y^2 + 2y + 2x = 0$

(d) $x^2 + x + y = 0$

(e) $y = -3x^2 + 5x + 4$

Assignment 4.16 — Figure 4.26 shows the graph of $y = x^2$ and four shifts. For each graph, write the equation of the function that is shown.

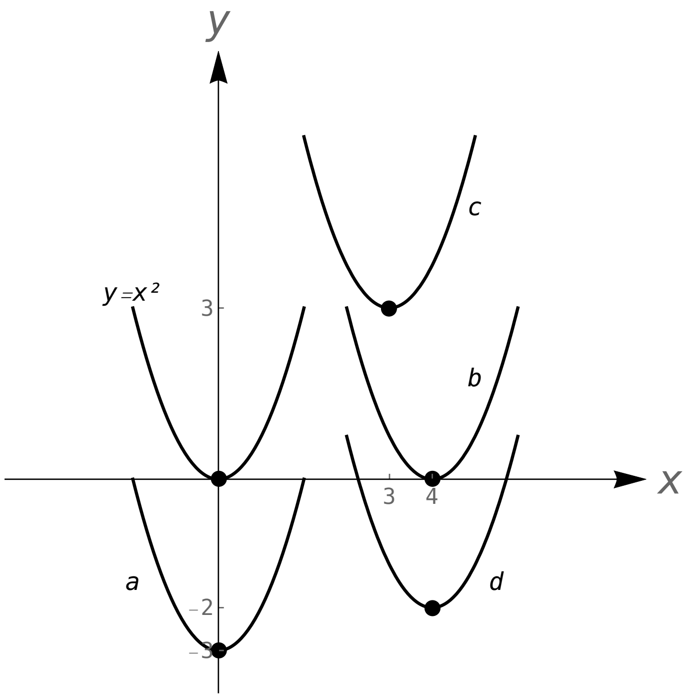


Figure 4.26: The graph of $y = x^2$ and four shifts.

Assignment 4.17 — Identify and sketch the graph of the conic sections with the following standard equations.

(a) $2x^2 + 3y^2 - 2 = 0$

(b) $2y^2 + 3x = 0$

(c) $x^2 - 2y^2 + 3 = 0$

(d) $3x^2 - 4y = 0$

(e) $-2x^2 - 4y^2 + 3 = 0$

(f) $2x^2 + 3y = 0$

(g) $y^2 - 3x = 0$

(h) $x^2 - y^2 - 3 = 0$

(i) $-2x^2 + 3y^2 + 3 = 0$

Assignment 4.18 — Determine the equation of the given conic sections.

(a) Ellipse with focal points in $(0,\pm 2)$ and length of half major axis equal to 3.

(b) Ellipse with focal points in $(0,1)$ en $(4,1)$ and with a distance from the center to the top that is twice as great as the distance from the center to the focal point.

(c) Ellipse with center in the origin, through $(3,1)$ and with vertical minor axis with length 4.

(d) Parabola with focal point in $(2,3)$ and top in $(2,4)$.

(e) Parabola through the origin, with focal point in $(0,-1)$ and upper tangent $y=0$.

(f) Parabola with focal point in $(-2,0)$ and directrix $x=2$.

(g) Hyperbola with center in origin and through $(1,5)$ and $(2,7)$.

(h) Hyperbola with focal points in $(0,\pm 2)$ and tops in $(0,\pm 1)$.

(i) Hyperbola with focal points in $(\pm 5,1)$ and asymptotes $x=\pm(y-1)$.

Assignment 4.19 — Identify and sketch the graph of the following conic sections.

(a) $x^2+y^2+2x-3y=0$

(b) $3x^2+3y^2+2x+7y=3$

(c) $3x^2-2y^2+3x+4y=0$

(d) $2x^2+3y^2-4x-12y+10=0$

(e) $2y^2+3x-4y+2=0$

(f) $2x^2-x-y+7=0$

(g) $3x^2+7y^2-14y+5=0$

(h) $2x^2-y^2-4x+3=0$

(i) $4x^2-y^2-4y=0$

(j) $9x^2+4y^2-18x+8y=23$

Assignment 4.20 — A hyperbolic mirror is used in some telescopes. Such a mirror has the property that an incoming light beam directed to one focal point will be reflected back to the other focal point. Use Figure 4.27 to construct the equation that models the hyperbolic mirror.

Assignment 4.21 — Long-range navigation (LORAN)is a radio navigation system developed during World War II. This system allows an aircraft to be controlled by maintaining a constant difference between the distance of the aircraft from two fixed points: a master station and a slave station. Determine an equation for the hyperbola in Figure 4.28 that describes this fixed difference.

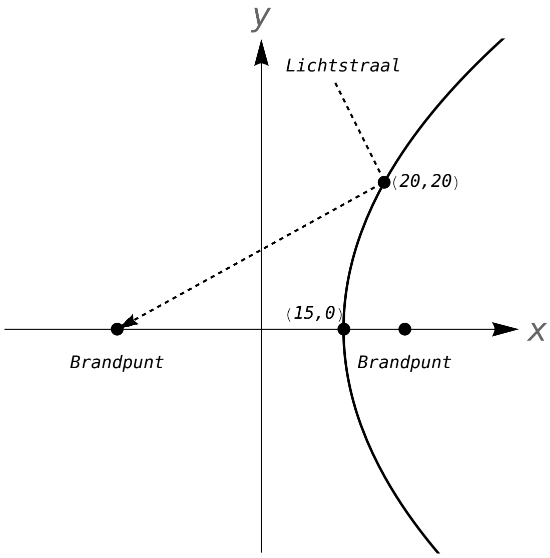


Figure 4.27: Reflection of an incoming light beam on a hyperbolic mirror.

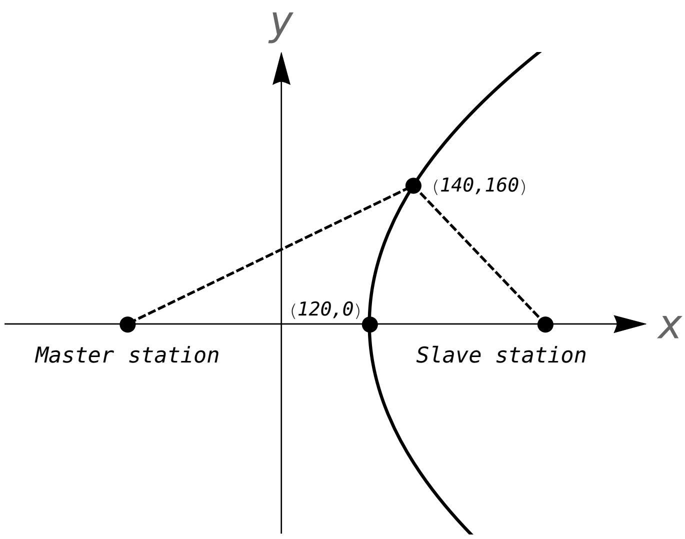


Figure 4.28: LORAN navigation

Review Exercises

Assignment 4.22 — Solve the inequalities below algebraically and graphically.

(a) $1 \le (2x-3)^2 \le 4$

(b) $3x^2 + 2x - 8 > 0$

(c) $\frac{3}{4}x^2 > 4(x-2)$

(d) $3(x-1)^2 > 4(x-1)$

(e) $5x + 4 \le 3x^2$

(f) $2 \le |x^2 - 9| < 9$

(g) $x^2 \le |4x-3|$

(h) $x|x+5| \ge -6$

(i) $x|x-3| < 2$

(j) $\dfrac{x^2}{x|x|+1} < \dfrac{1}{2}x$

Mathematics is a game played according to certain simple rules with meaningless marks on paper.

— David Hilbert —

5

Transcendental functions

5.1 Definition

A transcendental function is a function that does not satisfy a polynomial equation, in contrast to an algebraic function. In other words, a **transcendental function** (*transcendente functie*) transcends algebra in that it cannot be expressed in terms of a finite sequence of the algebraic operations of addition, multiplication, and root extraction. A function that is not transcendental is algebraic.

The most familiar transcendental functions are the logarithmic, exponential (with any non-trivial base), trigonometric, and hyperbolic functions, and the inverses of all of these. Less familiar are the special functions, such as the gamma, elliptic, and zeta functions. Besides, the generalized hypergeometric and Bessel functions are transcendental in general, but algebraic for some special parameter values.

5.2 Exponential and logarithmic functions

5.2.1 Definitions

5.2.1.1 Exponential functions

Up to this point, we have dealt with functions that involve terms of the form x^p where the base of the term, x, varies but the exponent of each term, p, remains constant. Here, we study functions of the form $f(x) = b^x$ where the base b is a constant and the exponent x is the variable. We start our exploration of these functions with $f(x) = 2^x$, whose graph is shown in Figure 5.1(a).

A few remarks about the graph of $f(x) = 2^x$ are in order. As $x \to -\infty$, the function $f(x) = 2^x$ takes on values that are increasingly closer to 0. In other words, as $x \to -\infty$, $f(x) \to 0^+$ and the x-axis is a

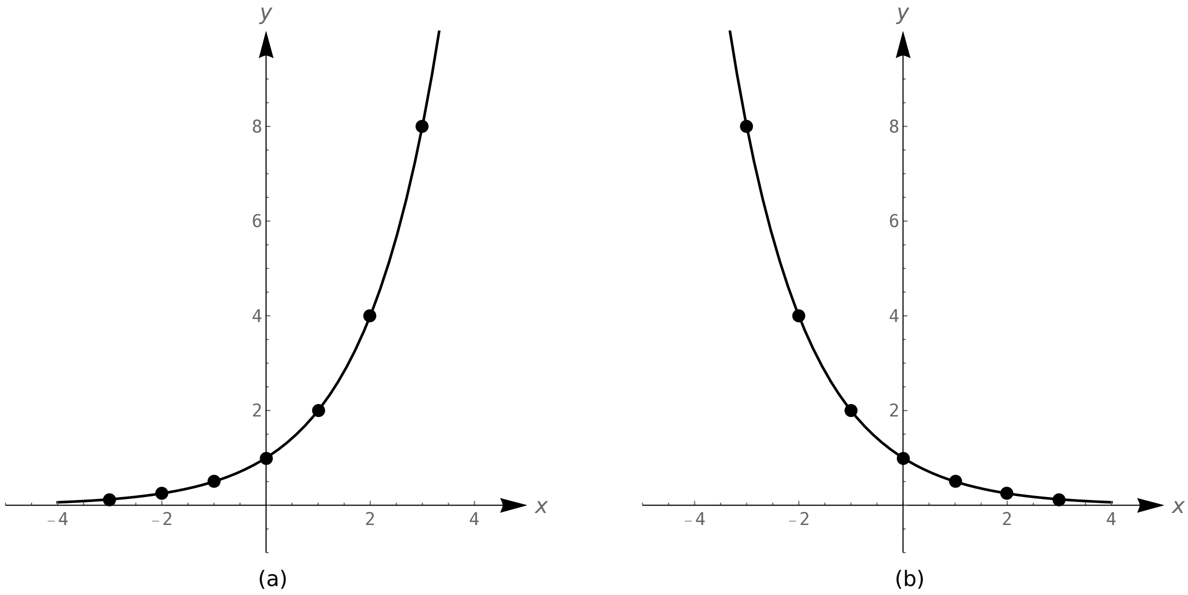


Figure 5.1: The graph of $y = f(x) = 2^x$ (a) and $y = g(x) = \left(\frac{1}{2}\right)^x = 2^{-x}$ (b).

horizontal asymptote. On the flip side, as $x \to +\infty$, we find $f(x) \to +\infty$. As a result, our graph suggests the range of f is $\mathbb{R}_0^+$. Besides, it is clear that f is injective and hence invertible, while $\operatorname{dom} f = \mathbb{R}$.

Here, we wish to study the family of functions $f(x) = b^x$, but which bases b make sense to study? We find that we run into difficulty if $b < 0$. For example, if $b = -2$, then the function $f(x) = (-2)^x$ has trouble, because, for instance, at $x = \frac{1}{2}, f(x) = \sqrt{-2}$ is not a real number. So we must restrict our attention to bases $b \geq 0$. What about $b = 0$? The function $f(x) = 0^x$ is undefined for $x \leq 0$ because we cannot divide by 0 and 0^0 is an indeterminate form. For $x > 0$, $0^x = 0$ so the function $f(x) = 0^x$ is the same as the function $f(x) = 0$ for $x > 0$. We know everything we can possibly know about this function, so we exclude it from our investigations. The only other base we exclude is $b = 1$, since the function $f(x) = 1^x = 1$ is, once again, a function we have already studied (see Chapter 4). Bearing this in mind, we are now ready to give a more formal definition of exponential functions.

Definitie 5.1 (Exponential function)
A function of the form

$$f(x) = b^x$$

where b is a strictly positive fixed real number ($b > 0$) and $b \neq 1$ is called a **base** b **exponential function** (*exponentiële functie met grondtal b*). Moreover, such a function is called exponentially increasing if $b > 1$ and exponentially decreasing if $0 < b < 1$.

Now, we could wonder what the graph of an exponential function with $0 < b < 1$ looks like. For instance, consider $g(x) = \left(\frac{1}{2}\right)^x$. Naively, we could certainly build a table of values and connect the points, but more wisely we could take a step back and note that $g(x) = \left(\frac{1}{2}\right)^x = \left(2^{-1}\right)^x = 2^{-x} = f(-x)$, where $f(x) = 2^x$. Thinking back to Section 3.2.5, the graph of $f(-x)$ is obtained from the graph of $f(x)$ by reflecting it across the y-axis (Figure 5.1(b)). We see that the domain and range of g match that of f, namely $\mathbb{R}$ and $\mathbb{R}_0^+$, respectively. Like f, g is also injective. Whereas f is always increasing, g is always decreasing. As a result, as $x \to -\infty$, $g(x) \to +\infty$, and on the flip side, as $x \to +\infty$, $g(x) \to 0^+$.

In literature, one very often comes across the wording **exponential growth** (*exponentiële groei*), but what exactly does it mean? Let us contrast exponential growth with linear growth in the following table.

x	$f(x)=2^x$	$h(x)=2x$
0	1	0
1	2	2
2	4	4
3	8	6
4	16	8
5	32	10

From this table we can infer that for these two functions, exponential growth dwarfs linear growth. More specifically, the former implies that original value from the range increases by the same percentage over equal increments found in the domain, whereas the latter refers to the original value from the range that increases by the same amount over equal increments found in the domain. For exponential growth, over equal increments, the constant multiplicative rate of change resulted in doubling the output whenever the input increased by one. For linear growth, the constant additive rate of change over equal increments resulted in adding 2 to the output whenever the input was increased by one.

Of all of the bases for exponential functions, two occur the most often in scientific circles. The first, base 10, is often called the **common base** (*tiendelige basis*). The second base is an irrational number, $e \approx 2.718$, called the **natural base**. The following examples give us an idea how these functions are used in the wild.

Example 5.1

The value of a tractor can be modelled by $V(x) = 25\left(\frac{4}{5}\right)^x$, where $x \geq 0$ is age of the vehicle in years and $V(x)$ is the value in thousands of euros.

1. Find and interpret $V(0)$.
2. Sketch the graph of $y = V(x)$ using transformations.
3. Find and interpret the horizontal asymptote of the graph of $y = V(x)$.

Solution

1. To find $V(0)$, we replace x with 0 to obtain $V(0) = 25\left(\frac{4}{5}\right)^0 = 25$. Since x represents the age of the tractor in years, $x = 0$ corresponds to the tractor being brand new. Since $V(x)$ is measured in thousands of euros, $V(0) = 25$ corresponds to a value of €25, 000. Putting it all together, we interpret $V(0) = 25$ to mean the purchase price of the tractor was €25 000.
2. To graph $y = 25\left(\frac{4}{5}\right)^x$, we start with the basic exponential function $f(x) = \left(\frac{4}{5}\right)^x$. Since the base $b = 4/5$ is between 0 and 1, the graph of $y = f(x)$ is decreasing. We plot the y-intercept $(0, 1)$ and two other points, $(-1, 5/4)$ and $(1, 4/5)$, and notice the horizontal asymptote $y = 0$ (Figure 5.2(a)). To obtain $V(x) = 25\left(\frac{4}{5}\right)^x$, we multiply the output from f by 25, which results in a vertical stretch by a factor of 25. We multiply all of the y-values in the graph by 25 and obtain the points $(-1, 125/4)$, $(0, 25)$ and $(1, 20)$. The horizontal asymptote remains 0. Finally, we restrict the domain to $\mathbb{R}^+$ to fit with the applied domain given to us (Figure 5.2(b)).

3. We see from the graph of V that its horizontal asymptote is $y=0$. This means as the tractor gets older, its value diminishes to 0.

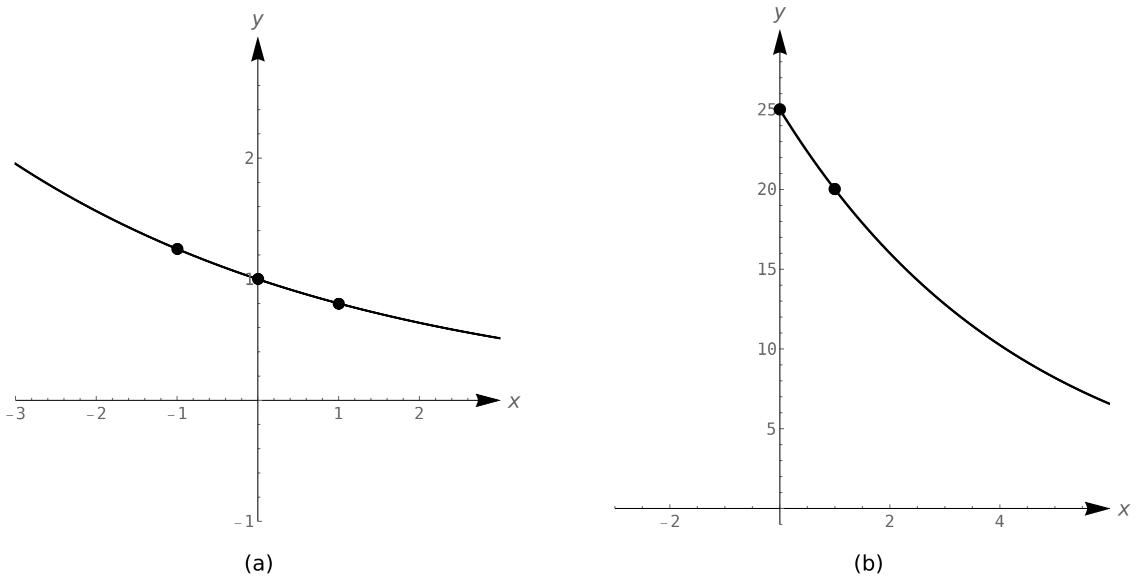


Figure 5.2: The graph of $y=f(x)=\left(\frac{4}{5}\right)^x$ (a) and $y=V(x)=25\left(\frac{4}{5}\right)^x$ (b) in Example 5.1.

The function in the previous example is often called a **decay curve**. In contrast, increasing exponential functions are used to model growth curves and we shall see several different examples of those later. We present another common decay curve in the following example. Although it may look more complicated than the previous example, it is actually just a basic exponential function which has been modified by a few transformations from Section 3.2.5.

Example 5.2

According to Newton's Law of cooling the temperature of coffee T [Θ] in degrees Celsius t [T] minutes after it is served can be modelled by

$$T(t)=21+50e^{-0.1t}.$$

1. Find and interpret $T(0)$.
2. Sketch the graph of $y=T(t)$ using transformations.
3. Find and interpret the horizontal asymptote of the graph.

Solution

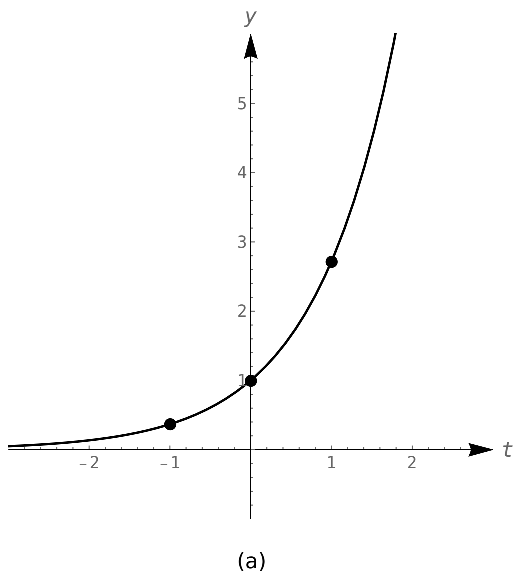
1. To find $T(0)$, we replace every occurrence of the independent variable t with 0 to obtain $T(0)=21+50e^{-0.1(0)}=71$. This means that the coffee was served at 71°C.
2. To graph $y=T(t)$ using transformations, we start with the basic function, $f(t)=e^t$. Since $e\approx 2.718>1$, the graph of f is an increasing exponential with y-intercept $(0,1)$ and horizontal asymptote $y=0$. The points $\left(-1,e^{-1}\right)\approx(-1,0.37)$ and $(1,e)\approx(1,2.72)$ are also

on the graph (Figure 5.3(a)). To use this information on $f(t) = e^t$, we rewrite $T(t)$ as

$$T(t) = 21 + 50f(-0.1t)\,.$$

Multiplication of the input to f, t, by -0.1 results in a horizontal expansion by a factor of 10 as well as a reflection about the y-axis. We divide each of the x-values of our points by -0.1 to obtain $\left(10, e^{-1}\right)$, $(0,1)$, and $(-10,e)$. Since none of these changes affected the y-values, the horizontal asymptote remains $y = 0$. Next, we see that the output from f is being multiplied by 50. This results in a vertical stretch by a factor of 50. We multiply the y-coordinates by 50 to obtain $\left(10, 50e^{-1}\right)$, $(0,50)$, and $(-10,50e)$. Obviously, the horizontal asymptote remains $y = 0$. Finally, we add 21 to all of the y-coordinates, which shifts the graph upwards to obtain $\left(10, 50e^{-1} + 21\right) \approx (10, 39.39)$, $(0,71)$, and $(-10, 50e + 21) \approx (-10, 156.91)$. Adding 21 to the horizontal asymptote shifts it upwards as well to $y = 21$. We connect these three points and, last but not least, we restrict the domain to match the applied domain $\mathbb{R}^+$ (Figure 5.3(b)).

3. From the graph, we see that the horizontal asymptote is $y = 21$. As $t \to +\infty$, the term $50e^{-0.1t}$ becomes very small. Hence, the graph of T is approaching the horizontal line $y = 21$ from above. This means that as time goes by, the temperature of the coffee is cooling to 21°C, presumably room temperature.



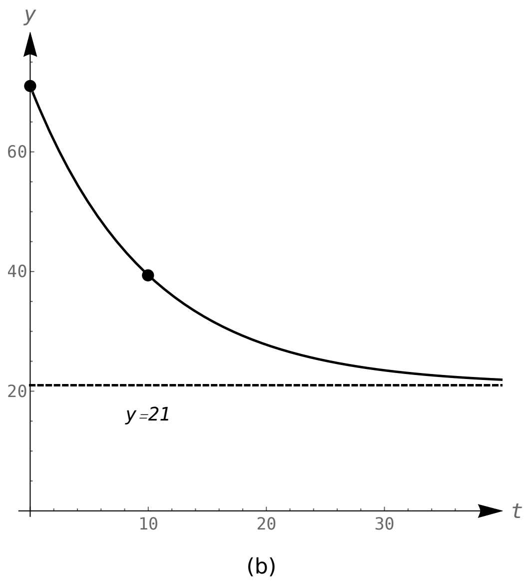


Figure 5.3: The graph of $y = f(t) = e^t$ (a) and $y = T(t) = 21 + 50e^{-0.1t}$ (b).

5.2.1.2 Logarithmic functions

As we have already remarked, the function $f(x) = b^x$ is injective, and hence invertible. We now turn our attention to these inverses, the logarithmic functions.

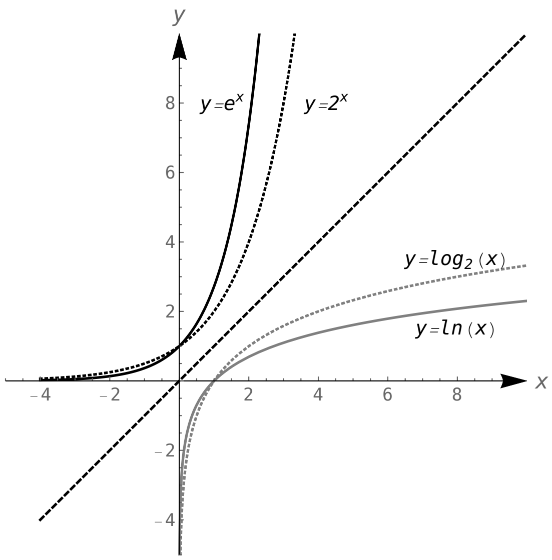
Definitie 5.2 (Logarithmic function)
The inverse of the exponential function $f(x) = b^x$ is called the **base b logarithm function** (*logaritmische functie met grondtal b*), and is denoted

$$f^{-1}(x) = \log_b(x)\,.$$

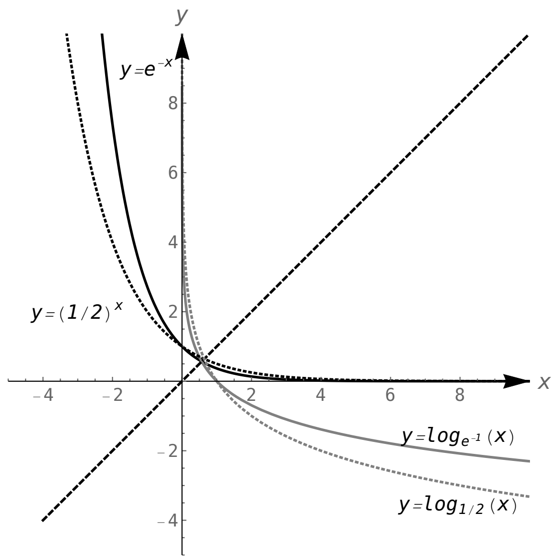
We read $\log_b(x)$ as log base b of x.

The **common logarithm** (*tiendelige logaritme, Briggse logaritme)* of a real number x is $\log_{10}(x)$ and is usually written $\log(x)$. The **natural logarithm** (*natuurlijke logaritme, Neperiaanse logaritme* of a real number x is $\log_e(x)$ and is usually written $\ln(x)$.

Since logarithmic functions are defined as the inverses of exponential functions, we can use the findings of Section 3.4 to tell us something about logarithmic functions. For example, we know that the domain of a logarithmic function is the range of an exponential function, namely $\mathbb{R}_0^+$, and that the range of a logarithmic function is the domain of an exponential function, namely $\mathbb{R}$. Since we know the basic shapes of $y = f(x) = b^x$ for the different cases of b, we can obtain the graph of $y = f^{-1}(x) = \log_b(x)$ by reflecting the graph of f across the line $y = x$ as shown below. The y-intercept $(0, 1)$ on the graph of f corresponds to an x-intercept of $(1, 0)$ on the graph of f^{-1}. The horizontal asymptotes $y = 0$ on the graphs of the exponential functions become vertical asymptotes $x = 0$ on the graphs of the logarithmic functions. All this is illustrated in Figure 5.4 for the functions $f_1(x) = e^x$, $f_2(x) = 2^x$, $f_3(x) = \left(\frac{1}{e}\right)^x$ and $f_4(x) = \left(\frac{1}{2}\right)^x$ and their corresponding inverses $f_1^{-1}(x) = \ln(x)$, $f_2^{-1}(x) = \log_2(x)$, $f_3^{-1}(x) = \log_{\frac{1}{e}}(x)$ and $f_4^{-1}(x) = \log_{\frac{1}{2}}(x)$, respectively.



(a)



(b)

Figure 5.4: The graph of $f_1(x) = e^x$ (a), $f_2(x) = 2^x$ (a), $f_3(x) = \left(\frac{1}{e}\right)^x$ (b) and $f_4(x) = \left(\frac{1}{2}\right)^x$ (b) and their corresponding inverses $f_1^{-1}(x) = \ln(x)$ (a), $f_2^{-1}(x) = \log_2(x)$ (a), $f_3^{-1}(x) = \log_{\frac{1}{e}}(x)$ (b) and $f_4^{-1}(x) = \log_{\frac{1}{2}}(x)$ (b), respectively.

Logarithms and the human psyche

Logarithms occur in several laws describing human perception. For instance, Hick's law proposes a logarithmic relation between the time individuals take to choose an alternative and the number of choices they have, while Fitts's law predicts that the time required to rapidly move to a target area is a logarithmic function of the distance to and the size of the target.

Interestingly, psychological studies found that individuals with little mathematics education tend to estimate quantities logarithmically, that is, they position a number on an unmarked line according to its logarithm, so that 10 is positioned as close to 100 as 100 is to 1000. Increasing education shifts this to a linear estimate that involves positioning 1000 10 times as far away.

Up until this point, restrictions on the domains of functions came from avoiding division by zero and keeping negative numbers from beneath even radicals. With the introduction of logarithmic functions, we now have another restriction. Since the domain of $f(x) = \log_b(x)$ is $\mathbb{R}_0^+$, the argument of the logarithmic function must be strictly positive.

5.2.2 Properties

As we shall see shortly, exponential functions inherit analogs of all of the properties of exponents you encountered in Chapter 2. First, we look at the consequence of exponential and logarithmic functions to be injective.

Let $f(x) = b^x$ and $g(x) = \log_b(x)$ where $b > 0$, $b \neq 1$. Then f and g are injective functions and

- $b^u = b^w$ if and only if $u = w$ for all real numbers u and w.
- $\log_b(u) = \log_b(w)$ if and only if $u = w$ for all real numbers $u > 0$, $w > 0$.

We now state the algebraic properties of exponential functions which will serve as a basis for the properties of logarithms. While these properties may look identical to the ones you learned in Chapter 2, they apply to real number exponents, not just rational exponents.

Theorem 5.1 (Algebraic properties of exponential functions)
Let $b > 0$, $b \neq 1$ and let u and w be real numbers, then

- ***Product rule***: $b^{u+w} = b^u\, b^w$
- ***Quotient rule***: $b^{u-w} = \dfrac{b^u}{b^w}$
- ***Power rule***: $(b^u)^w = b^{u\,w}$

To each of these properties of exponential functions corresponds an analogous property of logarithmic functions. We list these below in our next theorem.

Theorem 5.2 (Algebraic properties of logarithmic functions)
Let $b > 0$, $b \neq 1$ and let $u > 0$ and $w > 0$ be real numbers.

- ***Product rule***: $\log_b(u\,w) = \log_b(u) + \log_b(w)$
- ***Quotient rule***: $\log_b\left(\dfrac{u}{w}\right) = \log_b(u) - \log_b(w)$
- ***Power rule***: $\log_b\left(u^w\right) = w\log_b(u)$

From a purely functional approach, we can see the properties in Theorem 5.2 as an example of how inverse functions interchange the roles of inputs in outputs. For instance, the product rule for exponential functions given in Theorem 5.1, $f(u+w) = f(u)f(w)$, says that adding inputs results in multiplying outputs. Hence, whatever f^{-1} is, it must take the products of outputs from f and return them to the sum of their respective inputs. Since the outputs from f are the inputs to f^{-1} and vice-versa, we have that f^{-1} must take products of its inputs to the sum of their respective outputs. This is precisely what the product rule for logarithmic functions states in Theorem 5.2.

Example 5.3

Use the properties of logarithms to write the following as a single logarithm.

1. $\log_3(x-1)-\log_3(x+1)$
2. $\log(x)+2\log(y)-\log(z)$

Solution

1. The difference of logarithms requires the quotient rule:

$$\log_3(x-1)-\log_3(x+1)=\log_3\left(\frac{x-1}{x+1}\right).$$

2. We first apply the power rule, and then the product/quotient rule to get the following.

$$\begin{aligned}\log(x)+2\log(y)-\log(z) &= \log(x)+\log\left(y^2\right)-\log(z) && \text{(Power rule.)}\\ &= \log\left(xy^2\right)-\log(z) && \text{(Product rule.)}\\ &= \log\left(\frac{xy^2}{z}\right) && \text{(Quotient rule.)}\end{aligned}$$

We observe that using log properties to reassemble logarithms can increase the domain of the expression. For example, we leave it to the reader to verify the domain of $f(x)=\log_3(x-1)-\log_3(x+1)$ is $]1,+\infty[$ but the domain of $g(x)=\log_3\left(\frac{x-1}{x+1}\right)$ is $]-\infty,-1[\,\cup\,]1,+\infty[$. We will need to keep this in mind when we solve equations involving logarithms

In many cases it is convenient to change the base of the governing exponential or logarithmic functions. For that purpose, we may rely on the following theorem.

Theorem 5.3 (Change of base formulas)
Let $a,b>0$, and $a,b\neq 1$. Then, we have

- $a^x=b^{x\log_b(a)}$, *for all real numbers x;*
- $\log_a(x)=\dfrac{\log_b(x)}{\log_b(a)}$, *for all real numbers $x>0$.*

Proof The proof of Theorem 5.3 is a result of the properties studied earlier. For instance, if we start with $b^{x\log_b(a)}$ and use the power rule in the exponent to rewrite $x\log_b(a)$ as $\log_b(a^x)$ and then apply one of the inverse properties, we get

$$b^{x\log_b(a)}=b^{\log_b(a^x)}=a^x,$$

as required.

To verify the logarithmic form of the property, we also use the power rule and an inverse property. We note that

$$\log_a(x)\,\log_b(a)=\log_b\left(a^{\log_a(x)}\right)=\log_b(x),$$

and we get the result by dividing through by $\log_b(a)$. Note the inverse relationship between these two change of base formulas. To change the base of an exponential expression, we multiply the input by the factor $\log_b(a)$. To change the base of a logarithmic expression, we divide the output by the factor $\log_b(a)$. □

5.2.3 Exponential and logarithmic equations and inequalities

In this section we will briefly recall techniques for solving equations involving exponential or logarithmic functions. We first summarize below the two common ways to solve exponential equations.

1. Isolate the exponential function.
2. (a) If convenient, express both sides with a common base and equate the exponents.
   (b) Otherwise, take the natural log of both sides of the equation and use the power rule.

Likewise, the steps for solving an equation involving logarithmic functions are:

1. Isolate the logarithmic function.
2. (a) If convenient, express both sides as logs with the same base and equate the arguments of the log functions.
   (b) Otherwise, rewrite the log equation as an exponential equation.

Dual meaning of $\log(x)$

Throughout this text we adopted the notation $\ln(x)$ and $\log(x)$ to refer to the natural algorithm and common logarithm of x, respectively. In literature and on the Internet, however, one often finds that $\log(x)$ is used to refer to the natural logarithm of x, so beware when consulting other sources of information!

Example 5.4

Solve the following exponential and logarithmic equations.

1. $2^{3x} = 16^{1-x}$
2. $9 \cdot 3^x = 7^{2x}$
3. $75 = \dfrac{100}{1+3e^{-2t}}$
4. $\log_{117}(1-3x) = \log_{117}\left(x^2-3\right)$
5. $\log_7(1-2x) = 1-\log_7(3-x)$
6. $1+2\log_4(x+1) = 2\log_2(x)$

Solution

1. Since 16 is a power of 2, we can rewrite the equation as $2^{3x} = \left(2^4\right)^{1-x}$. Using properties of exponents, we get $2^{3x} = 2^{4(1-x)}$. Given the one-to-one property of exponential functions, we get $3x = 4(1-x)$, which gives $x = 4/7$.
2. We first note that we can rewrite the equation as $3^2 \cdot 3^x = 7^{2x}$ to obtain $3^{x+2} = 7^{2x}$. Since it is not convenient to express both sides as a power of 3 (or 7 for that matter) we use the natural log: $\ln\left(3^{x+2}\right) = \ln\left(7^{2x}\right)$. The power rule gives $(x+2)\ln(3) = 2x\ln(7)$. This equation is linear and can be solved for x:

$$\begin{aligned}
& (x+2)\ln(3) &=& \; 2x\ln(7) \\
\Leftrightarrow\; & x\ln(3)+2\ln(3) &=& \; 2x\ln(7) \\
\Leftrightarrow\; & 2\ln(3) &=& \; 2x\ln(7)-x\ln(3) \\
\Leftrightarrow\; & 2\ln(3) &=& \; x(2\ln(7)-\ln(3)) \\
\Leftrightarrow\; & x &=& \; \frac{2\ln(3)}{2\ln(7)-\ln(3)}
\end{aligned}$$

3. First, we isolate the exponential:

$$75 = \frac{100}{1+3e^{-2t}} \quad \Leftrightarrow \quad 75\left(1+3e^{-2t}\right) = 100$$
$$\Leftrightarrow \quad e^{-2t} = \tfrac{1}{9}\,.$$

Taking the natural log of both sides gives

$$\ln\left(e^{-2t}\right) = \ln\left(\frac{1}{9}\right) \quad \Leftrightarrow \quad t = \ln(3)\,.$$

4. Since we have the same base on both sides of this equation , we equate what is inside the logs to get $1-3x = x^2-3$. Solving $x^2+3x-4=0$ gives $x=-4$ and $x=1$. To check whether none of these is an extraneous solution, we substitute, for instance $x=1$, into our original equation to obtain $\log_{117}(-2) = \log_{117}(-2)$. While these expressions look identical, neither is a real number, which means $x=1$ is not in the domain of the original equation, and is not a solution. Similarly, we can verify that $x=-4$ is indeed a solution in the domain of the original equation.

5. We first collect the logarithms on the same side and then use the product rule to get

$$\log_7[(1-2x)(3-x)] = 1 \quad \Leftrightarrow \quad 7^1 = (1-2x)(3-x)\,,$$

which leads to $2x^2-7x-4=0$ whose solution is $x=-1/2$ or $x=4$. However, checking $x=4$ in the original equation produces $\log_7(-7) = 1-\log_7(-1)$, which is a clear domain violation.

6. We gather the logs to one side to get $1 = 2\log_2(x) - 2\log_4(x+1)$. Before we can combine the logarithms, however, we need a common base. Since 4 is a power of 2, we change the base

$$\log_4(x+1) = \frac{\log_2(x+1)}{\log_2(4)} = \frac{1}{2}\log_2(x+1)\,.$$

Hence, our original equation becomes

$$\begin{array}{rcll} & 1 & = & 2\log_2(x) - 2\left(\tfrac{1}{2}\log_2(x+1)\right) & \\ \Leftrightarrow & 1 & = & 2\log_2(x) - \log_2(x+1) & \\ \Leftrightarrow & 1 & = & \log_2\left(x^2\right) - \log_2(x+1) & \text{(Power rule.)} \\ \Leftrightarrow & 1 & = & \log_2\left(\dfrac{x^2}{x+1}\right) & \text{(Quotient rule.)} \end{array}$$

Rewriting this in exponential form, we get $\frac{x^2}{x+1} = 2$ or $x^2-2x-2=0$. Using the quadratic formula, we get $x = 1 \pm \sqrt{3}$. Yet, for what concerns the solution $x = 1-\sqrt{3}$, it holds that is negative so if substituted into the original equation, the term $2\log_2\left(1-\sqrt{3}\right)$ is undefined, and hence we should discard it as solution.

This example demonstrates the importance of checking for extraneous solutions when solving equations involving logarithms. These are easy to spot - any supposed solution which causes a negative number inside a logarithm needs to be discarded.

Just as we encountered for algebraic functions, we can also run into inequalities involving exponential or logarithmic functions.

Example 5.5

Solve the following inequalities.

1. $\dfrac{e^x}{e^x - 4} \leq 3$
2. $2^{x^2-3x} - 16 \geq 0$
3. $\dfrac{1}{\ln(x)+1} \leq 1$
4. $x \log(x+1) \geq x$

Solution

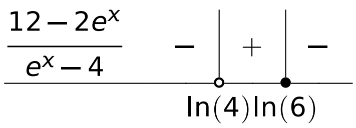
1. The first step we need to take to solve $\frac{e^x}{e^x-4} \leq 3$ is to get 0 on one side of the inequality. To that end, we subtract 3 from both sides and get .

$$\begin{aligned} \frac{e^x}{e^x-4} - 3 &\leq 0 \\ \Leftrightarrow \quad \frac{e^x}{e^x-4} - \frac{3\left(e^x-4\right)}{e^x-4} &\leq 0 \\ \Leftrightarrow \quad \frac{12-2e^x}{e^x-4} &\leq 0 \end{aligned}$$

We set $r(x) = \frac{12-2e^x}{e^x-4}$ and we note that r is undefined when its denominator $e^x - 4 = 0$, or when $e^x = 4$. Solving this gives $x = \ln(4)$, so the domain of r is $\mathbb{R} \setminus \{\ln(4)\}$. To find the zeros of r, we solve $r(x) = 0$ and obtain $12 - 2e^x = 0$. Solving for e^x, we find $e^x = 6$, or $x = \ln(6)$. When we build our sign diagram, finding test values may be a little tricky since we need to check values around $\ln(4)$ and $\ln(6)$. Recall that the function $\ln(x)$ is increasing which means $\ln(3) < \ln(4) < \ln(5) < \ln(6) < \ln(7)$. While the prospect of determining the sign of $r\big(\ln(3)\big)$ may be very unsettling, remember that $e^{\ln(3)} = 3$, so

$$r\big(\ln(3)\big) = \frac{12-2e^{\ln(3)}}{e^{\ln(3)}-4} = \frac{12-2(3)}{3-4} = -6$$

We determine the signs of $r\big(\ln(5)\big)$ and $r\big(\ln(7)\big)$ similarly and obtain the following sign diagram.



From this sign diagram, we find our answer to be $\left]-\infty, \ln(4)\right[\cup \left[\ln(6), +\infty\right[$.

2. Since we already have 0 on one side of the inequality, we set $r(x) = 2^{x^2-3x} - 16$. The domain of r is all real numbers, so in order to construct our sign diagram, we need to find the zeros of r. Setting $r(x) = 0$ gives

$$2^{x^2-3x} = 16 \quad \Leftrightarrow \quad 2^{x^2-3x} = 2^4 .$$

So, $x^2 - 3x = 4$ and the solutions of this equation are $x = 4$ and $x = -1$. From the sign diagram,

$$\begin{array}{c|c|c|c|c}
x^2-3x-4 & + & - & + \\
\hline
 & & -1 & 4 &
\end{array}$$

we see $r(x) \geq 0$ on $]-\infty,-1] \cup [4,+\infty[$ because 2^x is increasing everywhere and x^2-3x+4 is positive there.

3. We start solving this inequality by getting 0 on one side. Getting a common denominator yields

$$\frac{1}{\ln(x)+1} - \frac{\ln(x)+1}{\ln(x)+1} \leq 0 \quad \Leftrightarrow \quad \frac{\ln(x)}{\ln(x)+1} \geq 0.$$

We define $r(x) = \frac{\ln(x)}{\ln(x)+1}$ and set about finding the domain and the zeros of r. Due to the appearance of the term $\ln(x)$, we require $x > 0$. In order to keep the denominator away from zero, we solve $\ln(x)+1 = 0$ so $\ln(x) = -1$, so $x = e^{-1}$. Hence, the domain of r is $]0, e^{-1}[\cup]e^{-1}, +\infty[$. To find the zeros of r, we set $r(x) = 0$ so that $\ln(x) = 0$, and we find $x = e^0 = 1$. In order to determine test values for r, we need to find numbers between 0, e^{-1}, and 1 which have a base of e. Since $e \approx 2.718 > 1$, $0 < e^{-2} < e^{-1} < e^{-1/2} < 1 < e$. To determine the sign of $r\left(e^{-2}\right)$, we use the fact that $\ln\left(e^{-2}\right) = -2$, and find $r\left(e^{-2}\right) = \frac{-2}{-2+1} = 2$, which is positive. The rest of the test values are determined similarly.

$$\begin{array}{c|c|c|c}
\dfrac{\ln(x)}{\ln(x)+1} & + & - & + \\
\hline
 & e^{-1} \ (\circ) & 1 \ (\bullet) &
\end{array}$$

From our sign diagram, we find the solution to be $]0, e^{-1}[\cup [1, +\infty[$.

4. We begin by subtracting x from both sides to get $x\log(x+1) - x \geq 0$. We define $r(x) = x\log(x+1) - x$ and due to the presence of the logarithm, we require $x+1 > 0$, or $x > -1$. To find the zeros of r, we solve $r(x) = 0$ for x:

$$x\log(x+1) - x = 0 \quad \Leftrightarrow \quad x\big(\log(x+1) - 1\big) = 0\,.$$

This gives $x = 0$ or $\log(x+1) - 1 = 0$, where the latter means that $x = 9$. We select test values x so that $x+1$ is a power of 10, and we obtain $-1 < -0.9 < 0 < \sqrt{10} - 1 < 9 < 99$:

$$\begin{array}{c|c|c|c}
x\log(x+1)-x & + & - & + \\
\hline
 & 0 & 9 &
\end{array}$$

Our sign diagram gives the solution to be $]-1, 0] \cup [9, +\infty[$.

5.2.4 Applications

Exponential and logarithmic functions are used to model a wide variety of behaviours in the real world.

5.2.4.1 Growth models

The law of uninhibited - Malthusian - growth states as its premise that the instantaneous rate at which a population increases at any time is directly proportional to the population at that time. In other words, the more organisms there are at a given moment, the faster they reproduce. Formulating the law as stated results in a so-called differential equation, which is the logic of Mathematics III. Anyhow, solving this differential equation leads to the following model equation for the number of organisms N [-] at time t [T]:

$$N(t) = N_0 e^{kt}, \tag{5.1}$$

where $N(0) = N_0$ [–] is the initial number of organisms and $k > 0$ is the constant of proportionality, and represents the Malthusian growth rate [T^{-1}].

Example 5.6

In order to perform atherosclerosis research, epithelial cells are harvested from discarded umbilical tissue and grown in the laboratory. A technician observes that a culture of 12 000 cells grows to 5 000 000 cells in one week. Assuming that the cells follow the law of uninhibited growth, find a formula for the number of cells, N, in thousands, after t days.

Solution

Since N is to give the number of cells in thousands, we have $N_0 = 12$, so $N(t) = 12e^{kt}$. In order to complete the formula, we need to determine the growth rate k. We know that after one week, the number of cells has grown to five million. Since t measures days and the units of N are in thousands, this translates mathematically to $N(7) = 5000$. We get the equation $12e^{7k} = 5000$ which gives $k = \frac{1}{7}\ln\left(\frac{1250}{3}\right)$. Hence, we get

$$N(t) = 12e^{\frac{t}{7}\ln\left(\frac{1250}{3}\right)}.$$

Of course, in practice, we would approximate k to some desired accuracy, say $k \approx 0.8618$, which we can interpret as an 86.18% daily growth rate for the cells.

Obviously, the law of uninhibited growth will in most practical settings not hold because the availability of resources in the environment that are needed for organisms to grow and persist is limited. This effect can, however, be incorporated in a logistic - Verhulst - growth model, which incorporates that the rate of growth of a population varies jointly with the population itself as well as the room the population has to grow. More specifically, if a population behaves according to the assumptions of **logistic growth** (*logistische groei*), the number of organisms N [-] at time t [T] is given by

$$N(t) = \frac{L}{1 + Ce^{-kLt}}, \tag{5.2}$$

where $N(0) = N_0$ is the initial population, L [–] is the limiting population, C [–] is a measure of how much room there is to grow given by

$$C = \frac{L}{N_0} - 1.$$

and $k > 0$ is the constant of proportionality [T^{-1}].

The logistic function is used not only to model the growth of organisms, but is also to model the spread of disease and rumours.

Example 5.7

The number of people N, in hundreds, at a local community college who have heard the rumour 'Carl is afraid of Virginia Woolf' can be modelled using the logistic equation

$$N(t) = \frac{84}{1+2799e^{-t}},$$

where $t \geq 0$ is the number of days after April 1, 2009.

1. Find and interpret $N(0)$.
2. Find and interpret the end behaviour of $N(t)$.
3. How long until 4200 people have heard the rumour?

Solution

1. We find $N(0) = \frac{84}{1+2799e^0} = \frac{84}{2800} = \frac{3}{100}$. Since $N(t)$ measures the number of people who have heard the rumour in hundreds, $N(0)$ corresponds to 3 people. Since $t = 0$ corresponds to April 1, 2009, we may conclude that on that day, 3 people have heard the rumour.
2. We could simply note that $N(t)$ is written in the form of Equation (5.2), and identify $L = 84$. However, to see why the answer is 84, we proceed analytically. Since the domain of N is restricted to $t \geq 0$, the only end behaviour of significance is $t \to +\infty$. As we have seen before, as $t \to +\infty$, we have $2799e^{-t} \to 0^+$ and so $N(t) \approx 84$. Hence, as $t \to +\infty$, $N(t) \to 84$. This means that as time goes by, the number of people who will have heard the rumour approaches 8400.
3. To find how long it takes until 4200 people have heard the rumour, we set $N(t) = 42$. Solving $\frac{84}{1+2799e^{-t}} = 42$ gives $t = \ln(2799) \approx 7.937$. Consequently, it takes around 8 days until 4200 people have heard the rumour.

5.2.4.2 Newton's law of cooling

We first encountered Newton's Law of cooling in Example 5.2. In that example we had a cup of coffee cooling from 71°C to room temperature 21°C according to the formula $T(t) = 21 + 50e^{-0.1t}$, where t was measured in minutes. In this situation, we know the physical limit of the temperature of the coffee is room temperature and the differential equation which gives rise to our formula for $T(t)$ takes this into account. Newton's Law of Cooling states that the rate of cooling of the coffee at a given time t is directly proportional to how much of a temperature gap exists between the coffee at time t and room temperature, not the temperature of the coffee itself. Of course, if we take an item from the refrigerator and let it sit out in the kitchen, the object's temperature will rise to room temperature, and since the physics behind warming and cooling is the same, we combine both cases in the equation below.

The temperature T [Θ] of an object at time t [T] is given by the formula

$$T(t) = T_a + (T_0 - T_a)\,e^{-kt}, \tag{5.3}$$

where $T(0) = T_0$ [T^{-1}] is the initial temperature of the object, T_a [Θ] is the ambient temperature and $k > 0$ is the constant of proportionality.

Example 5.8

Recall from Example 5.2 that the temperature of coffee T (in degrees Celcius) t minutes after it is served can be modelled by $T(t) = 21 + 50e^{-0.1t}$. How long will the coffee be warmer than 50°C?

Solution

We need to find when $T(t) > 50$:

$$21 \ + \ 50e^{-0.1t} > 50 \quad \Leftrightarrow \quad 50e^{-0.1t} - 29 > 0\,.$$

Hence we set $r(t) = 50e^{-0.1t} - 29$. The domain of r is artificially restricted due to the context of the problem to $\mathbb{R}^+$. Solving r(t)=0 results in $e^{-0.1t} = \frac{29}{50}$ so that $t = -10\ln\left(\frac{29}{50}\right)$, or equivalently $t = 10\ln\left(\frac{50}{29}\right)$. Now, we construct the sign diagram:

$$\begin{array}{lcc|c} 50e^{-0.1t} - 29 & & + & - \\ \hline & & & \bullet \\ & & & 10\ln\left(\frac{50}{29}\right) \end{array}$$

This means it takes approximately 5.5 minutes for the coffee to cool to 50°C. Until then, the coffee is warmer than that.

5.2.4.3 Radioactive decay

Another real-world phenomenon that can be described by means of an exponential function is radioactive decay. There are elements which are unstable and emit energy spontaneously. In doing so, the amount of the element itself diminishes. The assumption behind the underlying model is that the rate of decay of an element at a particular time is directly proportional to the amount of the element present at that time. In other words, the more of the element there is, the faster the element decays. This is precisely the same kind of hypothesis which drives the law of uninhibited growth, and as such, the equation governing radioactive decay is similar to Equation (5.1) with the exception that the rate constant k is negative.

More precisely, the amount of a radioactive element A [M] at time t [T] is given by the formula

$$A(t) = A_0 e^{kt}, \tag{5.4}$$

where $A(0) = A_0$ is the initial amount of the element and $k < 0$ is the constant of proportionality, also referred to as the decay rate [T^{-1}]. In this context, one often uses the so-called **half-life** (*halfwaardetijd*), which is the time required for the amount of a radioactive element to reduce to half of its initial amount.

Example 5.9

Iodine-131 is a commonly used radioactive isotope used to help detect how well the thyroid is functioning. Suppose the decay of Iodine-131 follows the model given in Equation (5.4), and that the half-life of Iodine-131 is approximately 8 days. If 5 grams of Iodine-131 is present initially, find a function which gives the amount of Iodine-131, A, in grams, t days later.

Solution

Since we start with 5 grams initially, Equation (5.4) gives $A(t) = 5e^{kt}$. Since the half-life is 8 days, it takes 8 days for half of the Iodine-131 to decay, leaving half of it behind. Hence, $A(8) = 2.5$

which means $5e^{8k} = 2.5$. Solving, we get $k = \frac{1}{8}\ln(1/2) = -\ln(2)/8 \approx -0.08664$, which we can interpret as a loss of material at a rate of 8.664% daily. Hence, $A(t) = 5e^{-\frac{t\ln(2)}{8}} \approx 5e^{-0.08664t}$.

5.3 Trigonometric functions

5.3.1 Foundations of trigonometry

When two half-lines share a common initial point they form an **angle** (*hoek*) and the common initial point is called the **vertex** (*hoekpunt*) of the angle (Figure 5.5).

One commonly used system to measure angles is **degree measure** (*graden*). Quantities measured in degrees are denoted by the familiar '°' symbol. One complete revolution is 360°, and parts of a revolution are measured proportionately. Thus half of a revolution (a **straight angle**) (*gestrekte hoek*) measures $\frac{1}{2}(360°) = 180°$, a quarter of a revolution (a **right angle**) (*rechte hoek*) measures $\frac{1}{4}(360°) = 90°$ and so on. Recall that if an angle measures strictly between 0° and 90° it is called an **acute angle** (*scherpe hoek*) (Figure 5.5(a)) and if it measures strictly between 90° and 180° it is called an **obtuse angle** (*stompe hoek*) (Figure 5.5(b)). In practice, the distinction between the angle itself and its measure is blurred so that the sentence α is an angle measuring 42° is often abbreviated as $\alpha = 42°$.

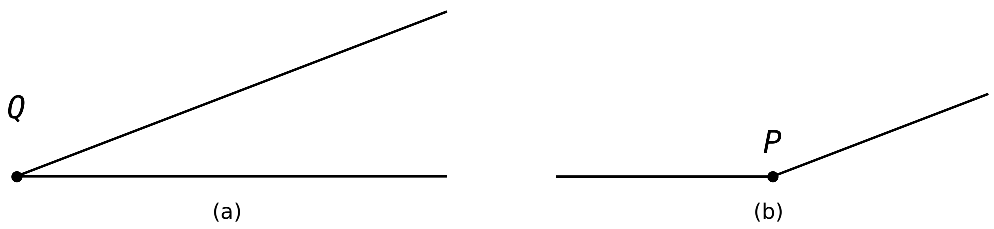


Figure 5.5: An acute angle with vertex Q (a) and obtuse angle with vertrex P (b).

Using our definition of the degree measure, we have that 1° represents the measure of an angle which constitutes $\frac{1}{360}$ of a revolution. Now, there are two ways to further subdivide degrees. The first, and most familiar, is **decimal degrees** (*decimale graden*). The second way to divide degrees is the **Degree - Minute - Second** (**DMS**) system. In this system, one degree is divided equally into sixty minutes, and in turn, each minute is divided equally into sixty seconds. Recall that two acute angles are called **complementary angles** (*complementaire hoeken*) if their measures add to 90°. Two angles, either a pair of right angles or one acute angle and one obtuse angle, are called **supplementary angles** (*supplementaire hoeken*) if their measures add to 180°. In Figure 5.6, the angles α and β are supplementary angles while the pair γ and δ are complementary angles.

When the direction of rotation matters, we will typically use an **oriented angle** (gerichte hoek). We imagine the angle being swept out starting from an **initial side** and ending at a **terminal side**. When the rotation is counter-clockwise, we say that the angle is **positive** (*positief*); when the rotation is clockwise, we say that the angle is **negative** (*negatief*).

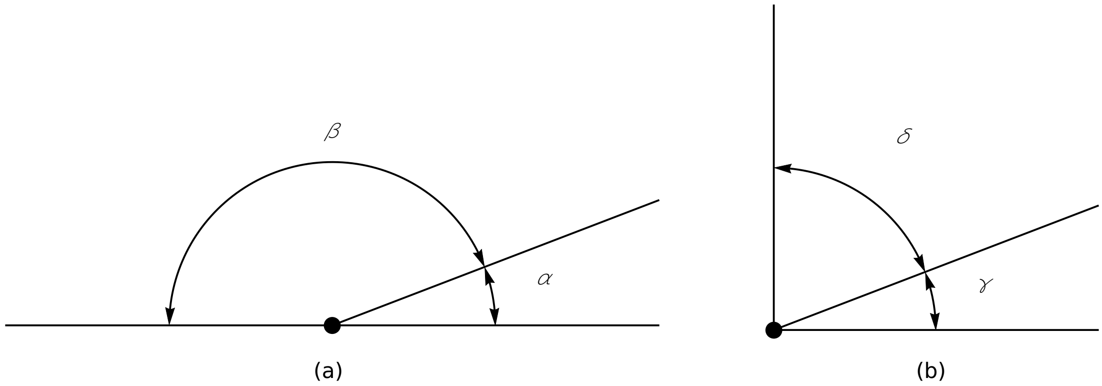


Figure 5.6: Supplementary (a) and complementary angles (b).

An angle is said to be in **standard position** if its vertex is the origin and its initial side coincides with the positive x-axis. Furthermore, two angles in standard position are called **coterminal** if they share the same terminal side. Such angles always differ by a multiple of 360° and since there are infinitely many integers, any given angle has infinitely many coterminal angles.

Remember that the real number π is defined to be the ratio of a circle's circumference C to its diameter d. It is a mathematical constant. In terms of the radius, we equivalently have $2\pi = \frac{C}{r}$. This tells us that for any circle, the ratio of its circumference to its radius is also always constant; in this case the constant is 2π. Suppose now we take a portion of the circle, so instead of comparing the entire circumference C to the radius, we compare some **arc** (*boog*) measuring s units in length to the radius (Figure 5.7). Let θ be the angle whose vertex is the centre of the circle and whose determining rays pass through the endpoints of the arc. Using proportionality arguments, it stands to reason that the ratio $\frac{s}{r}$ should also be a constant among all circles, and it is this ratio which defines the **radian measure** (*radialen*) of an angle.

We note that an angle with radian measure 1 means the corresponding arc length s equals the radius of the circle r, hence $s = r$. When the radian measure is 2, we have $s = 2r$; when the radian measure is 3, $s = 3r$, and so forth. Thus the radian measure of an angle θ tells us how many 'radius lengths' we need to sweep out along the circle to subtend the angle θ. Since one revolution sweeps out the entire circumference $2\pi r$, one revolution has radian measure $\frac{2\pi r}{r} = 2\pi$. Note that, by definition, the radian measure of an angle is a length divided by another length so that these measurements are actually dimensionless and are considered dimensionless numbers.

Since one revolution counter-clockwise measures 360° and the same angle measures 2π radians, we

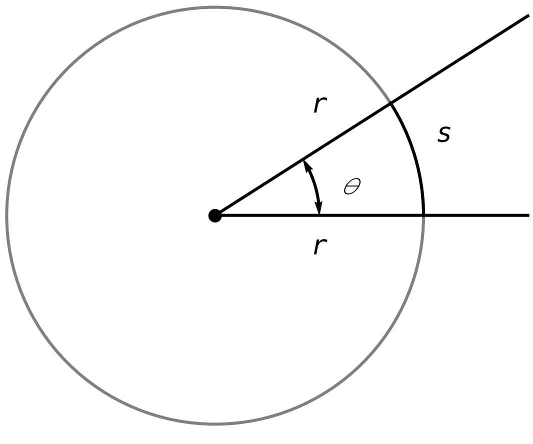


Figure 5.7: The radian measure of θ is $\frac{s}{r}$.

can use the proportion $\frac{2\pi \text{ radians}}{360°}$, or equivalently, $\frac{\pi \text{ radians}}{180°}$, as the conversion factor between degrees and radians. For example, to convert 60° to radians we find $\frac{\pi}{3}$ radians.

The merit of the radian measure lies in how easily angles in this measure can be identified with real numbers. Consider the unit circle, the angle θ in standard position, and the corresponding arc measuring s units in length. By definition, and the fact that the unit circle has radius 1, the radian measure of θ is $\frac{s}{r} = \frac{s}{1} = s$ so that we have $\theta = s$. In order to identify real numbers with oriented angles, we make good use of this fact by essentially wrapping the real number line around the unit circle and associating to each real number t an oriented arc on the unit circle with initial point $(1,0)$.

5.3.2 Circular motion

Suppose an object is moving as along a circular path of radius r from the point P to the point Q in an amount of time t [T] (Figure 5.8).

Here s represents a **displacement** (*verplaatsing*) so that $s > 0$ means the object is traveling in a counter-clockwise direction and $s < 0$ indicates movement in a clockwise direction. The average velocity of the object, denoted $\overline{v}$ $[T^{-1}]$, is defined as the average rate of change of the position of the object with respect to time. As a result, we have $\overline{v} = \frac{s}{t}$. The quantity $\overline{v}$ $[LT^{-1}]$ conveys two ideas: the direction in which the object is moving and how fast the position of the object is changing. The contribution of direction in the quantity $\overline{v}$ is either to make it positive (in the case of counter-clockwise motion) or negative (in the case of clockwise motion), so that the quantity $|\overline{v}|$ quantifies how fast the object is moving - it is the speed of the object. Measuring θ in radians we have $\theta = \frac{s}{r}$ thus $s = r\theta$ and

$$\overline{v} = \frac{s}{t} = \frac{r\theta}{t} = r\frac{\theta}{t}.$$

The quantity $\frac{\theta}{t}$ is called the **average angular velocity** (*gemiddelde hoeksnelheid*) of the object, denoted by $\overline{\omega}$. The quantity $\overline{\omega}$ is the average rate of change of the angle θ with respect to time. If $\overline{\omega}$ is constant throughout the duration of the motion, then it can be shown that the average velocities are the same as their instantaneous counterparts, v and ω, respectively. In this case, v is simply called the velocity of the object and is the instantaneous rate of change of the position of the object with respect to time and likewise for ω.

If the path of the object were uncurled from a circle to form a line segment, then the velocity of the object on that line segment would be the same as the velocity on the circle. For this reason, the quantity v is often called the linear velocity of the object in order to distinguish it from the angular velocity, ω. Consequently, for an object moving on a circular path of radius r with constant angular velocity ω, the (linear) velocity of the object is given by $v = r\omega$.

Example 5.10

Assuming that the surface of the Earth is a sphere, any point on the Earth can be thought of as an object travelling on a circle which completes one revolution in (approximately) 24 hours. The path traced out by the point during this 24 hour period is the Latitude of that point. Campus Coupure of Ghent University is at 51.05325° north latitude, and the radius of the earth at this Latitude is approximately 6365 km. Find the linear velocity of the campus as the world turns.

Solution

To use the formula $v = r\omega$, we first need to compute the angular velocity ω. The earth makes one revolution in 24 hours, and one revolution is 2π radians, so $\omega = \frac{2\pi \text{ radians}}{24 \text{ hours}} = \frac{\pi}{12 \text{ hours}}$. We are also assuming that we are viewing the rotation of the earth as counter-clockwise so $\omega > 0$. Hence, the

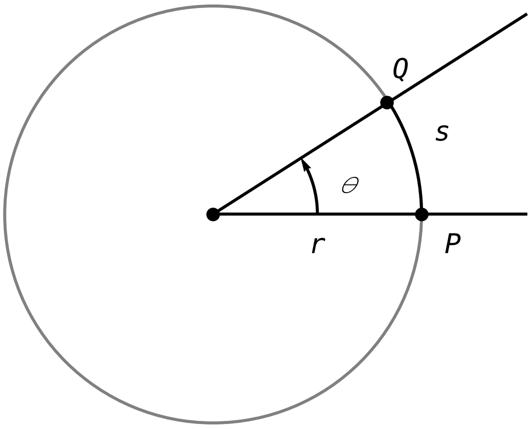


Figure 5.8: An object moving along a circular path of radius r from P to Q.

linear velocity is

$$v = 6365\,\text{km} \cdot \frac{\pi}{12\,\text{h}} \approx 1666\,\frac{\text{km}}{\text{h}}.$$

5.3.3 The six trigonometric functions

5.3.3.1 Cosine and sine on the unit circle

The sine and cosine of an acute angle are defined in the context of a right triangle. Consider for that purpose the generic right triangle with corresponding acute angle θ (Figure 5.9). The side with length a is called the **side of the triangle adjacent to** θ (*aanliggende rechthoekzijde*); the side with length b is called the **side of the triangle opposite** θ (*overstaande rechthoekzijde*); and the remaining side of length c (the side opposite the right angle) is called the **hypotenuse** (*hypothenusa*). We now imagine drawing this triangle in Quadrant I so that the angle θ is in standard position with the adjacent side to θ lying along the positive x-axis (Figure 5.10).

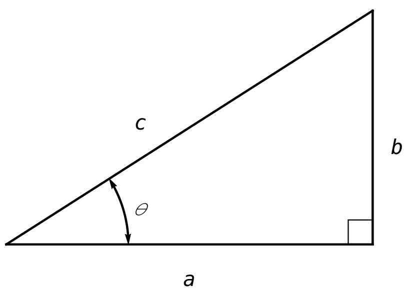


Figure 5.9: A right triangle with acute angle θ.

Then the **cosine** (*cosinus*) and **sine** (*sinus*) of θ are defined as

$$\cos(\theta) = \frac{a}{c} \tag{5.5}$$

and

$$\sin(\theta) = \frac{b}{c}, \tag{5.6}$$

so we have determined the cosine and sine of θ in terms of the lengths of the sides of the right triangle.

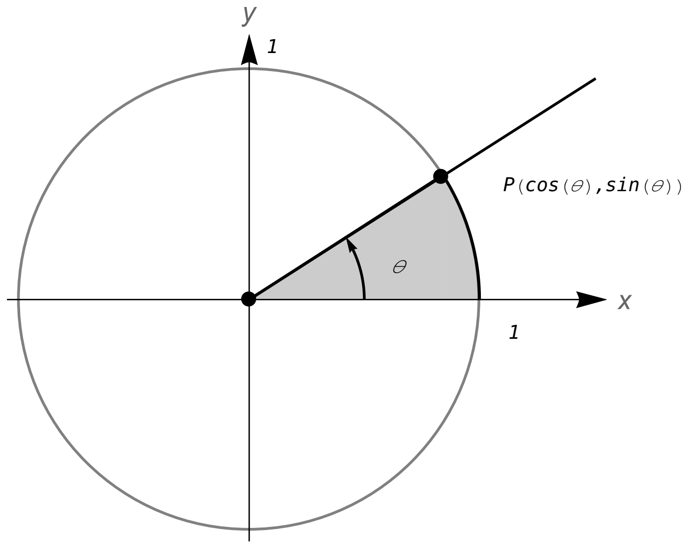


Figure 5.10: Position of a point P on the unit circle expressed in terms of $\cos(\theta)$ and $\sin(\theta)$.

We now imagine drawing this right triangle in Quadrant I so that the angle θ is in standard position with the adjacent side to θ lying along the positive x-axis and the point P lies on the unit circle (Figure 5.10). Given Equations (5.5) and (5.6), we see that the x-coordinate of the point P is the cosine of θ, while the y-coordinate of P is the sine of θ (Figure 5.10). Since, for each angle θ, there is only one associated value of $\cos(\theta)$ and only one associated value of $\sin(\theta)$, both the sine and cosine may be interpreted as functions. It can be verified easily that the gray area in Figure 5.10 equals $\theta/2$; that is halve the circular angle θ. Besides, given how we defined angles, it is easy to see that $\cos(\theta) = \cos(\theta + 2\,\pi)$ for all $\theta \in [0, 2\pi]$, from which it follows that the cosine function is periodic with period 2π. Similarly, we find that also the sine function is periodic with period 2π.

Having introduced the sine and cosine, we should recall one of the most important identities in trigonometry.

Theorem 5.4 (The Pythagorean identity)
For any angle θ, it holds that

$$\cos^2(\theta) + \sin^2(\theta) = 1\,. \tag{5.7}$$

We summarize the cosine and sine values for certain common angles in Table 5.1.

Table 5.1: Cosine and sine value of common angles

θ (degrees)	θ (radians)	$\cos(\theta)$	$\sin(\theta)$
0°	0	1	0
30°	$\frac{\pi}{6}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
45°	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
60°	$\frac{\pi}{3}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$
90°	$\frac{\pi}{2}$	0	1

To determine the cosines and sines of angles not given in Table 5.1 we may exploit the symmetry inherent in the unit circle. Suppose, for instance, we wish to know the cosine and sine of $\theta = 5\pi/6$. We plot θ in standard position below and, as usual, let $P(x, y)$ denote the point on the terminal side of θ which lies on the unit circle. Note that the terminal side of θ lies $\pi/6$ radians short of one half revolution (Figure 5.11(b)). Knowing from Table 5.1 that $\cos\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2}$ and $\sin\left(\frac{\pi}{6}\right) = \frac{1}{2}$. This means that the point on the terminal side of the angle $\frac{\pi}{6}$, when plotted in standard position, is $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$. From Figure 5.11(b), it is clear that the point $P(x, y)$ we seek can be obtained by reflecting that point about the y-axis. Hence, $\cos\left(\frac{5\pi}{6}\right) = -\frac{\sqrt{3}}{2}$ and $\sin\left(\frac{5\pi}{6}\right) = \frac{1}{2}$.

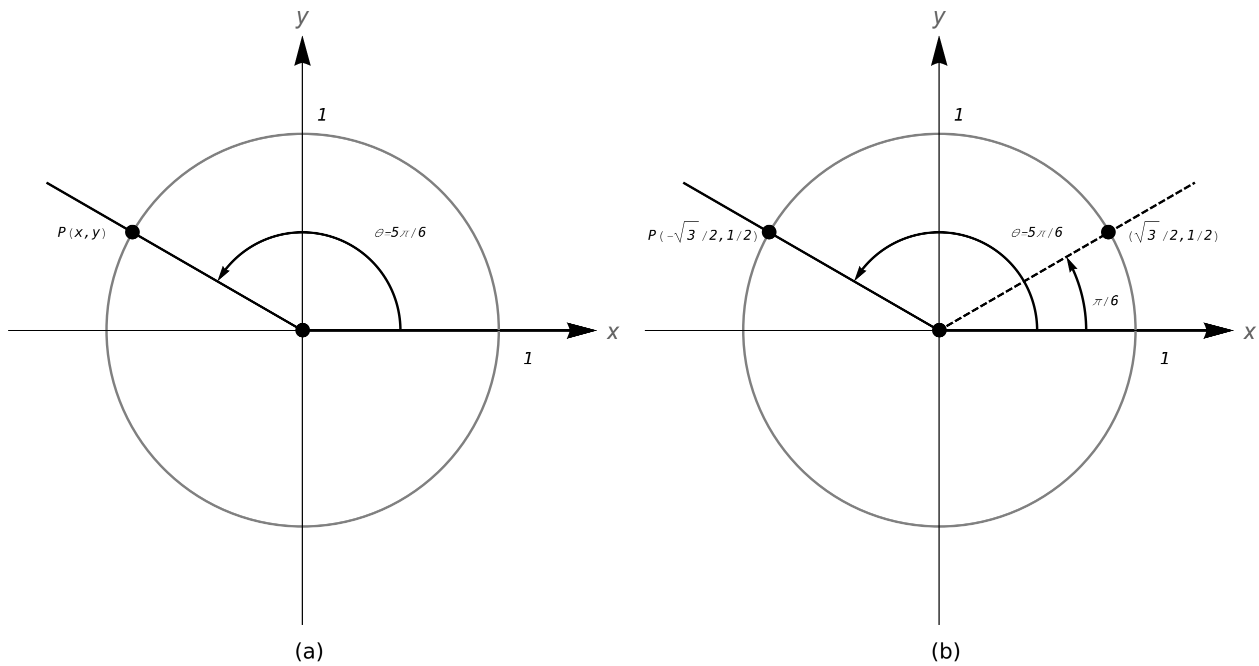


Figure 5.11: Determining the $\cos\left(\frac{5\pi}{6}\right)$ and $\sin\left(\frac{5\pi}{6}\right)$.

5.3.3.2 Beyond the unit circle

In defining the cosine and sine functions, we assigned to each angle a position on the unit circle. Here, we broaden our scope to include circles of radius r centred at the origin. Consider for the moment the acute angle θ drawn in Figure 5.12 in standard position. Let $Q(x, y)$ be the point on the terminal side of θ which lies on the circle $x^2 + y^2 = r^2$, and let $P(\widetilde{x}, \widetilde{y})$ be the point on the terminal side of θ which lies on the unit circle. Now consider dropping perpendiculars from P and Q to create two right triangles, ΔOPA and ΔOQB. These triangles are **similar** (*gelijkvormig*) , thus it follows that $\frac{x}{\widetilde{x}} = \frac{r}{1} = r$, so $x = r\widetilde{x}$ and, similarly, we find $y = r\widetilde{y}$. Since, by definition, $\widetilde{x} = \cos(\theta)$ and $\widetilde{y} = \sin(\theta)$, we get the coordinates of Q to be $x = r\cos(\theta)$ and $y = r\sin(\theta)$. By reflecting these points through the x-axis, y-axis and origin, we obtain the result for all other angles θ.

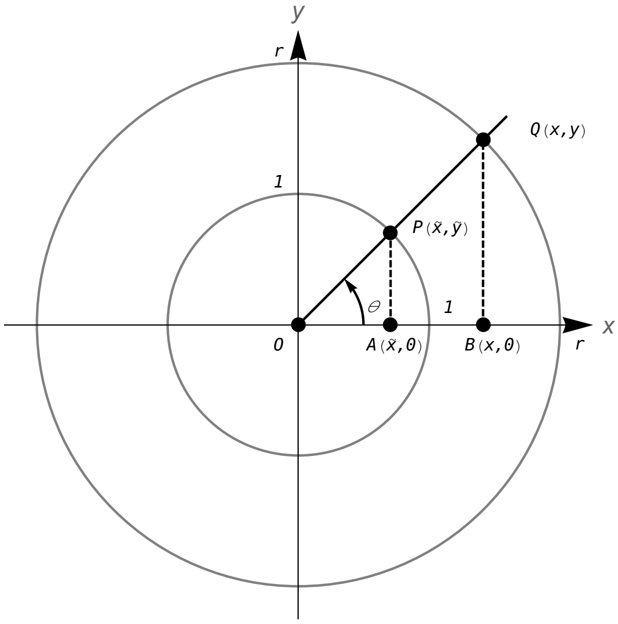


Figure 5.12: Relation between the coordinates of a point P on the unit circle and a point Q on the circle $x^2+y^2=r^2$, where both P and Q lie on the terminal side of θ.

The result is summarized in the following theorem.

Theorem 5.5 (Sine and cosine in the unit circle)
If $Q(x,y)$ is the point on the terminal side of an angle θ, plotted in standard position, which lies on the circle $x^2+y^2=r^2$ then $x=r\cos(\theta)$ and $y=r\sin(\theta)$. Moreover,

$$\cos(\theta)=\frac{x}{r}=\frac{x}{\sqrt{x^2+y^2}} \quad \textit{and} \quad \sin(\theta)=\frac{y}{r}=\frac{y}{\sqrt{x^2+y^2}}.$$

Proof See above. □

Theorem 5.5 also gives us what we need to describe the position of an object travelling in a circular path of radius r with constant angular velocity ω. Suppose that at time t, the object has swept out an angle measuring θ radians. If we assume that the object is at the point $(r,0)$ when $t=0$, the angle θ is in standard position. By definition, $\omega=\frac{\theta}{t}$ which we rewrite as $\theta=\omega t$. According to Theorem 5.5, the location of the object $Q(x,y)$ on the circle is found using the equations $x=r\cos(\theta)=r\cos(\omega t)$ and $y=r\sin(\theta)=r\sin(\omega t)$. Hence, at time t, the object is at the point $(r\cos(\omega t), r\sin(\omega t))$, where $\omega>0$ indicates a counter-clockwise direction and $\omega<0$ indicates a clockwise direction.

Example 5.11

Suppose we are in the situation of Example 5.10. Find the equations of motion of Campus Coupure as the earth rotates.

Solution

From Example 5.10, we take $r=6365$ km and and $\omega=\frac{\pi}{12\,\text{hours}}$. Hence, the equations of motion are $x=r\cos(\omega t)=6365\cos\left(\frac{\pi}{12}t\right)$ and $y=r\sin(\omega t)=6365\sin\left(\frac{\pi}{12}t\right)$, where x and y are measured in km and t is measured in hours.

5.3.3.3 The other trigonometric functions

Starting from the sine and cosine functions we introduced earlier, we may define four more trigonometric functions, being the **secant** (*secans*) of x:

$$\sec(x) = \frac{1}{\cos(x)}, \qquad \text{provided} \cos(x) \neq 0,$$

the **cosecant** (*cosecans*) of x:

$$\csc(x) = \frac{1}{\sin(x)}, \qquad \text{provided} \sin(x) \neq 0,$$

the **tangent** (*tangens*) of x:

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \qquad \text{provided} \cos(x) \neq 0,$$

and finally the **cotangent** (*cotangens*) of x:

$$\cot(x) = \frac{\cos(x)}{\sin(x)}, \qquad \text{provided} \sin(x) \neq 0.$$

Note that of the six trigonometric functions, only cosine and sine are defined for all angles. Table 5.2 lists the tangent and cotangent for certain common angles.

Table 5.2: Tangent and cotangent of common angles

x (degrees)	x (radians)	$\tan(x)$	$\cot(x)$
0°	0	0	undefined
30°	$\frac{\pi}{6}$	$\frac{\sqrt{3}}{3}$	$\sqrt{3}$
45°	$\frac{\pi}{4}$	1	1
60°	$\frac{\pi}{3}$	$\sqrt{3}$	$\frac{\sqrt{3}}{3}$
90°	$\frac{\pi}{2}$	undefined	0

By combining Equations (5.5) and (5.6) with the definition of the four other trigonometric functions, we can also express these in terms of the sides of a right triangle

$$\sec(\theta) = \frac{c}{a}, \qquad \csc(\theta) = \frac{c}{b}, \qquad \tan(\theta) = \frac{b}{a}, \qquad \cot(\theta) = \frac{a}{b},$$

where a, b and c are defined as in Figure 5.9.

Example 5.12

In order to determine the height of the maple tree (*Acer pseudoplatanus*) in the inner yard of Campus Coupure, two sightings from the ground, one 20 metres directly behind the other, are made. If the angles of inclination were 45° and 30°, respectively, how tall is the tree to the nearest foot?

Solution

Sketching the problem situation in Figure 5.13, we find ourselves with two unknowns: the height h of the tree and the distance x from the base of the tree to the first observation point.

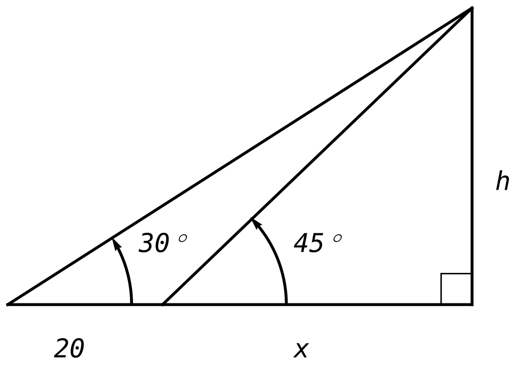


Figure 5.13: Determining the height of a maple tree.

Using the definition of the tangent function in terms of the sides of a right triangle, we get a pair of equations: $\tan(45°) = \frac{h}{x}$ and $\tan(30°) = \frac{h}{x+20}$. Since $\tan(45°) = 1$, the first equation gives $\frac{h}{x} = 1$, or $x = h$. Substituting this into the second equation gives $\frac{h}{h+20} = \tan(30°) = \frac{\sqrt{3}}{3}$. Clearing fractions, we get $3h = (h+20)\sqrt{3}$. The result is a linear equation for h, so we proceed to expand the right hand side and gather all the terms involving h to one side.

$$\begin{aligned} 3h &= (h+20)\sqrt{3} \\ \Leftrightarrow \quad (3-\sqrt{3})h &= 20\sqrt{3} \\ \Leftrightarrow \quad h &= \frac{20\sqrt{3}}{3-\sqrt{3}} \approx 27.32 \end{aligned}$$

Hence, the tree is approximately 27 metres tall.

5.3.4 Trigonometric identities

Here, we recall several collections of identities which have uses in this course and beyond.

5.3.4.1 Pythagorean identities

Given the four newly defined trigonometric functions, it makes sense to combine their definitions with the Pythagorean identity (Theorem 5.4) to derive new Pythagorean-like identities for the remaining four trigonometric functions. Assuming $\cos(x) \neq 0$, we may start with $\cos^2(x) + \sin^2(x) = 1$ and divide both sides by $\cos^2(x)$ to obtain $1 + \frac{\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}$. This reduces to

$$1 + \tan^2(x) = \sec^2(x)\,. \tag{5.8}$$

If $\sin(x) \neq 0$, we can divide both sides of the identity $\cos^2(x) + \sin^2(x) = 1$ by $\sin^2(x)$ to obtain

$$\cot^2(x) + 1 = \csc^2(x)\,. \tag{5.9}$$

These identities play an important role in not just trigonometry, but in calculus as well. We will use them later find the values of the trigonometric functions of an angle and solve equations and inequalities. In calculus, they are needed to simplify otherwise complicated expressions.

Example 5.13

Verify the following identities. Assume that all quantities are defined.

1. $\Big(\sec(\theta)-\tan(\theta)\big)\big(\sec(\theta)+\tan(\theta)\Big)=1$
2. $\dfrac{\sec(\theta)}{1-\tan(\theta)}=\dfrac{1}{\cos(\theta)-\sin(\theta)}$
3. $6\sec(\theta)\tan(\theta)=\dfrac{3}{1-\sin(\theta)}-\dfrac{3}{1+\sin(\theta)}$
4. $\dfrac{\sin(\theta)}{1-\cos(\theta)}=\dfrac{1+\cos(\theta)}{\sin(\theta)}$

Solution

In verifying identities, we typically start with the more complicated side of the equation and use known identities to transform it into the other side of the equation.

1. Expanding the left hand side of the equation gives:

$$\big(\sec(\theta)-\tan(\theta)\big)\big(\sec(\theta)+\tan(\theta)\big)=\sec^2(\theta)-\tan^2(\theta)\,,$$

which equals 1 according to Equation (5.8).

2. While both sides of our last identity contain fractions, the left side affords us more opportunities to use our identities. Substituting $\sec(\theta)=\frac{1}{\cos(\theta)}$ and $\tan(\theta)=\frac{\sin(\theta)}{\cos(\theta)}$, we get:

$$\begin{aligned}\frac{\sec(\theta)}{1-\tan(\theta)} &= \frac{\dfrac{1}{\cos(\theta)}}{1-\dfrac{\sin(\theta)}{\cos(\theta)}}=\frac{\left(\dfrac{1}{\cos(\theta)}\right)\cos(\theta)}{\left(1-\dfrac{\sin(\theta)}{\cos(\theta)}\right)\cos(\theta)}\\ &= \frac{1}{\cos(\theta)-\sin(\theta)}\,,\end{aligned}$$

which is exactly what we had set out to show.

3. The right hand side of the equation seems to hold more promise. We get common denominators and add:

$$\begin{aligned}\frac{3}{1-\sin(\theta)}-\frac{3}{1+\sin(\theta)} &= \frac{3(1+\sin(\theta))-3(1-\sin(\theta))}{(1+\sin(\theta))(1-\sin(\theta))}\\ &= \frac{(3+3\sin(\theta))-(3-3\sin(\theta))}{1-\sin^2(\theta)}\\ &= \frac{6\sin(\theta)}{1-\sin^2(\theta)}\,.\end{aligned}$$

Since we wish to transform this expression into $6\sec(\theta)\tan(\theta)$, we use a reciprocal and quotient identity and find

$$6\sec(\theta)\tan(\theta)=6\left(\frac{1}{\cos(\theta)}\right)\left(\frac{\sin(\theta)}{\cos(\theta)}\right).$$

In other words, we need to get cosines in our denominator. From Equation (5.7) we have $1-\sin^2(\theta)=\cos^2(\theta)$ so we get:

$$\begin{aligned}\frac{3}{1-\sin(\theta)}-\frac{3}{1+\sin(\theta)} &= \frac{6\sin(\theta)}{1-\sin^2(\theta)}=\frac{6\sin(\theta)}{\cos^2(\theta)}\\ &= 6\left(\frac{1}{\cos(\theta)}\right)\left(\frac{\sin(\theta)}{\cos(\theta)}\right)=6\sec(\theta)\tan(\theta)\,.\end{aligned}$$

4. It is debatable which side of the identity is more complicated. One thing which stands out is that the denominator on the left hand side is $1-\cos(\theta)$, while the numerator of the right hand side is $1+\cos(\theta)$. This suggests the strategy of starting with the left hand side and multiplying the numerator and denominator by the quantity $1+\cos(\theta)$:

$$\begin{aligned}\frac{\sin(\theta)}{1-\cos(\theta)} &= \frac{\sin(\theta)}{(1-\cos(\theta))}\cdot\frac{(1+\cos(\theta))}{(1+\cos(\theta))}\\ &= \frac{\sin(\theta)(1+\cos(\theta))}{1-\cos^2(\theta)}=\frac{\sin(\theta)(1+\cos(\theta))}{\sin^2(\theta)}\\ &= \frac{\cancel{\sin(\theta)}(1+\cos(\theta))}{\cancel{\sin(\theta)}\sin(\theta)}=\frac{1+\cos(\theta)}{\sin(\theta)}\,.\end{aligned}$$

5.3.4.2 Even/odd, cofunction and difference and sum identities

Our first set of identities is even/odd identities, which are summarized in the following theorem.

Theorem 5.6 (Even/Odd identities)
For all applicable angles θ, it holds that

- $\cos(-\theta)=\cos(\theta)$
- $\sin(-\theta)=-\sin(\theta)$
- $\tan(-\theta)=-\tan(\theta)$
- $\sec(-\theta)=\sec(\theta)$
- $\csc(-\theta)=-\csc(\theta)$
- $\cot(-\theta)=-\cot(\theta)$

Proof To proof these identities it suffices to show $\cos(-\theta)=\cos(\theta)$ and $\sin(-\theta)=-\sin(\theta)$. For that purpose, consider angles θ and $-\theta$ plotted in standard position, where $0\le\theta\le 2\pi$ (Figure 5.14). Let P and Q denote the points on the terminal sides of θ and $-\theta$, respectively, which lie on the unit circle. Hence, the coordinates of P are $(\cos(\theta),\sin(\theta))$ and the coordinates of Q are $(\cos(-\theta),\sin(-\theta))$. It follows that the points P and Q are symmetric about the x-axis, such that $\cos(-\theta)=\cos(\theta)$ and $\sin(-\theta)=-\sin(\theta)$. □

The even/odd identities can be used to derive the sum and difference identities for cosine.

Theorem 5.7 (Sum and difference identities for cosine)
For all angles α and β, it holds that

$$\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta)\,, \tag{5.10}$$

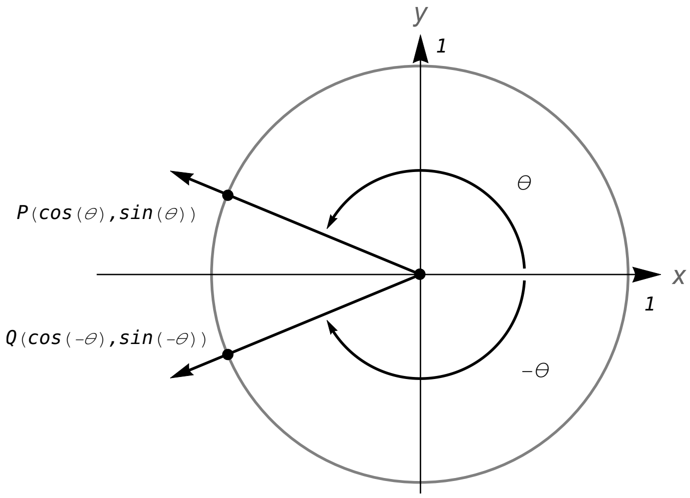


Figure 5.14: Even/Odd identity for cosine and sine.

$$\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta). \tag{5.11}$$

We can use the sum and difference identities for cosine to derive the so-called cofunction identities. The results are summarized in the following theorem.

Theorem 5.8 (Cofunction identities)
For all applicable angles θ, it holds that

- $\cos\left(\frac{\pi}{2} - \theta\right) = \sin(\theta)$
- $\sec\left(\frac{\pi}{2} - \theta\right) = \csc(\theta)$
- $\tan\left(\frac{\pi}{2} - \theta\right) = \cot(\theta)$
- $\sin\left(\frac{\pi}{2} - \theta\right) = \cos(\theta)$
- $\csc\left(\frac{\pi}{2} - \theta\right) = \sec(\theta)$
- $\cot\left(\frac{\pi}{2} - \theta\right) = \tan(\theta)$

Proof Consider for instance $\cos\left(\frac{\pi}{2} - \theta\right)$. Straightforward application of the difference identity for cosine yields:

$$\begin{aligned} \cos\left(\frac{\pi}{2} - \theta\right) &= \cos\left(\frac{\pi}{2}\right)\cos(\theta) + \sin\left(\frac{\pi}{2}\right)\sin(\theta) \\ &= (0)\big(\cos(\theta)\big) + (1)\big(\sin(\theta)\big) \\ &= \sin(\theta). \end{aligned}$$

Moreover, from this result we immediately get that

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

which says, in words, that the cosine of an angle is the sine of its complement. Now that these identities have been established for cosine and sine, the remaining trigonometric functions follow suit. □

With the cofunction identities in place, we are now in the position to derive the sum and difference formulas for sine. To achieve this, we convert to cosines using a cofunction identity, then expand using the difference formula for cosine

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

We can derive the difference formula for sine by rewriting $\sin(\alpha-\beta)$ as $\sin(\alpha+(-\beta))$ and using the sum formula and the even / odd identities.

Theorem 5.9 (Sum and difference identities for sine)
For all angles α and β, it holds that

$$\sin(\alpha+\beta) = \sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta), \tag{5.12}$$

$$\sin(\alpha-\beta) = \sin(\alpha)\cos(\beta)-\cos(\alpha)\sin(\beta). \tag{5.13}$$

Proof See above. □

Also for the tangent function, one may derive sum and difference identities.

Example 5.14

Derive a formula for $\tan(\alpha+\beta)$ in terms of $\tan(\alpha)$ and $\tan(\beta)$.

Solution

We can start expanding $\tan(\alpha+\beta)$ using a quotient identity and our sum formulas

$$\begin{aligned}
\tan(\alpha+\beta) &= \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} \\
&= \frac{\sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta)}.
\end{aligned}$$

Since $\tan(\alpha)=\frac{\sin(\alpha)}{\cos(\alpha)}$ and $\tan(\beta)=\frac{\sin(\beta)}{\cos(\beta)}$, it looks as though if we divide both numerator and denominator by $\cos(\alpha)\cos(\beta)$ we will have what we want

$$\begin{aligned}
\tan(\alpha+\beta) &= \frac{\sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta)} \cdot \frac{\frac{1}{\cos(\alpha)\cos(\beta)}}{\frac{1}{\cos(\alpha)\cos(\beta)}} \\
&= \frac{\frac{\sin(\alpha)\cancel{\cos(\beta)}}{\cos(\alpha)\cancel{\cos(\beta)}}+\frac{\cancel{\cos(\alpha)}\sin(\beta)}{\cancel{\cos(\alpha)}\cos(\beta)}}{\frac{\cancel{\cos(\alpha)}\cancel{\cos(\beta)}}{\cancel{\cos(\alpha)}\cancel{\cos(\beta)}}-\frac{\sin(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)}} \\
&= \frac{\tan(\alpha)+\tan(\beta)}{1-\tan(\alpha)\tan(\beta)}.
\end{aligned}$$

Naturally, this formula is limited to those cases where all of the tangents are defined.

Theorem 5.10 (Sum and difference identities for tangent)
For all applicable angles α and β, it holds that

$$\tan(\alpha+\beta) = \frac{\tan(\alpha)+\tan(\beta)}{1-\tan(\alpha)\tan(\beta)}, \tag{5.14}$$

$$\tan(\alpha-\beta) = \frac{\tan(\alpha)-\tan(\beta)}{1+\tan(\alpha)\tan(\beta)}. \tag{5.15}$$

Proof See above for sum identity. To find a formula for $\tan(\alpha-\beta)$, we can simply rewrite the difference as a sum: $\tan(\alpha+(-\beta))$. □

5.3.4.3 Double angle identities and power reduction formulas

Theorem 5.11 (Double angle identities)
For all applicable angles θ, it holds that

$$\cos(2\theta) = \cos^2(\theta)-\sin^2(\theta), \tag{5.16}$$

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta), \tag{5.17}$$

$$\tan(2\theta) = \frac{2\tan(\theta)}{1-\tan^2(\theta)}. \tag{5.18}$$

Proof Specialize the sum formulas for cosine (Theorem 5.7), sine (Theorem 5.9) and tangent (Theorem 5.10) to the case where $\alpha=\beta$. □

Note that the Pythagorean identity (Theorem 5.4) allows to express $\cos(2\theta)$ alternatively as

$$\cos(2\theta) = 2\cos^2(\theta)-1, \tag{5.19}$$

and

$$\cos(2\theta) = 1-2\sin^2(\theta). \tag{5.20}$$

In calculus, we will often be confronted with situations where it is useful to reduce the power of cosine and sine. Solving Equation (5.19) for $\cos^2(\theta)$ and the Equation (5.20) for $\sin^2(\theta)$ results in the aptly-named power reduction formulas below. These are also known as **Carnot's formulas** (*formules van Carnot*).

Theorem 5.12 (Power reduction formulas)
For all angles θ, it holds that

$$\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}, \tag{5.21}$$

$$\sin^2(\theta) = \frac{1-\cos(2\theta)}{2}. \tag{5.22}$$

Proof See above. □

Not only can these formulas be used to reduce the power of cosine, they can as well be used to derive expressions for the cosine, sine and tangent of half angle. To start, we apply the power reduction formula to $\cos^2\left(\frac{\theta}{2}\right)$

$$\cos^2\left(\frac{\theta}{2}\right) = \frac{1+\cos\left(2\left(\frac{\theta}{2}\right)\right)}{2} = \frac{1+\cos(\theta)}{2}\,.$$

5.3.4.4 Product to sum formulas and vice versa

The product to sum formulas, are easily verified by expanding each of the right hand sides in accordance with Theorems 5.7 and 5.9. They are of particular use in calculus, and we list them here for reference. These are also known as **Simpson's formulas** (*formules van Simpson*).

Theorem 5.13 (Product to sum formulas)
For all angles α and β, it holds that

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}\big(\cos(\alpha-\beta)+\cos(\alpha+\beta)\big)\,, \tag{5.23}$$

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}\big(\cos(\alpha-\beta)-\cos(\alpha+\beta)\big)\,, \tag{5.24}$$

$$\sin(\alpha)\cos(\beta) = \frac{1}{2}\big(\sin(\alpha-\beta)+\sin(\alpha+\beta)\big)\,. \tag{5.25}$$

Proof Prove as an exercise. □

Related to the product to sum formulas are the sum to product formulas. These are easily verified using the product to sum formulas, and as such, their proofs are left as exercises.

Theorem 5.14 (Sum to product formulas)
For all angles α and β, it holds that

$$\cos(\alpha)+\cos(\beta) = 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)\,, \tag{5.26}$$

$$\cos(\alpha)-\cos(\beta) = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)\,, \tag{5.27}$$

$$\sin(\alpha)\pm\sin(\beta) = 2\sin\left(\frac{\alpha\pm\beta}{2}\right)\cos\left(\frac{\alpha\mp\beta}{2}\right)\,. \tag{5.28}$$

Proof Prove as an exercise. □

5.3.5 Graphs of cosine and sine functions

5.3.5.1 Domain and range

In the same way we studied polynomial, rational, exponential, and logarithmic functions, we will study the trigonometric functions $f(x) = \cos(x)$ and $g(x) = \sin(x)$. The first order of business is to find the domains and ranges of these functions. Whether we think of identifying the real number x with the angle $\theta = x$ radians, or think of wrapping an oriented arc around the unit circle to find coordinates on the unit circle, it should be clear that both the cosine and sine functions are defined for all real numbers x. So dom $\cos(x) =$ dom $\sin(x) = \mathbb{R}$, and likewise im $\cos(x) =$ im $\sin(x) = [-1, 1]$.

5.3.5.2 Graphs

The even/odd identities in Theorem 5.6 tell us $\cos(-x) = \cos(x)$ for all real numbers x and $\sin(-x) = -\sin(x)$ for all real numbers x. This means $f(x) = \cos(x)$ is an even function, while $g(x) = \sin(x)$ is an odd function (see Chapter 3). Another important property of these functions is that $\cos(x+2\pi k) = \cos(x)$ and $\sin(x+2\pi k) = \sin(x)$, for all real numbers x and any integer k; that is the sine and cosine functions are periodic with period 2π. One last property of the functions $f(x) = \cos(x)$ and $g(x) = \sin(x)$ is worth pointing out: the graphs of both of these functions have no jumps, gaps, holes in the graph, asymptotes, corners or cusps.

To graph $y = \cos(x)$, we make a table using some of the common values of x in the interval $[0, 2\pi]$ (Table 5.3). This generates a portion of the cosine graph, which we call the **fundamental cycle** (*fundamentele cyclus*) of $y = \cos(x)$ (Figure 5.15(a)).

Table 5.3: Cosine and sine of some common angles.

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
$\cos(x)$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$	0	$\frac{\sqrt{2}}{2}$	1
$\sin(x)$	0	$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$	0

To get the graph of the cosine function for intervals stretching beyond the fundamental cycle, we way imagine copying and pasting this graph end to end infinitely in both directions (left and right) on the x-axis (Figure 5.15(c)).

We can plot the fundamental cycle of the graph of $y = \sin(x)$ similarly, with similar results (Figures 5.15(b) and 5.15(d)). It is of course no accident that the graphs of $y = \cos(x)$ and $y = \sin(x)$ are so similar. Using a cofunction identity (Theorem 5.8) along with the even property of cosine, we have

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

Recalling Section 3.2.5, we see from this formula that the graph of $y = \sin(x)$ is the result of shifting the graph of $y = \cos(x)$ to the right $\frac{\pi}{2}$ units.

Now that we know the basic shapes of the graphs of $y = \cos(x)$ and $y = \sin(x)$, we can rely on the tools from Section 3.2.5 to graph more complicated curves. To do so, we need to keep track of the movement of some key points on the original graphs, such as $x = 0$, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ and 2π.

By performing any of the transformations mentioned in Section 3.2.5 to the the basic graph of $f(x) = \cos(x)$ or $g(x) = \sin(x)$, we arrive at so-called **sinusoids** (*sinusoïde*) . Sinusoids can be characterized by four properties: period, amplitude, phase shift and vertical shift. More specifically, for the functions

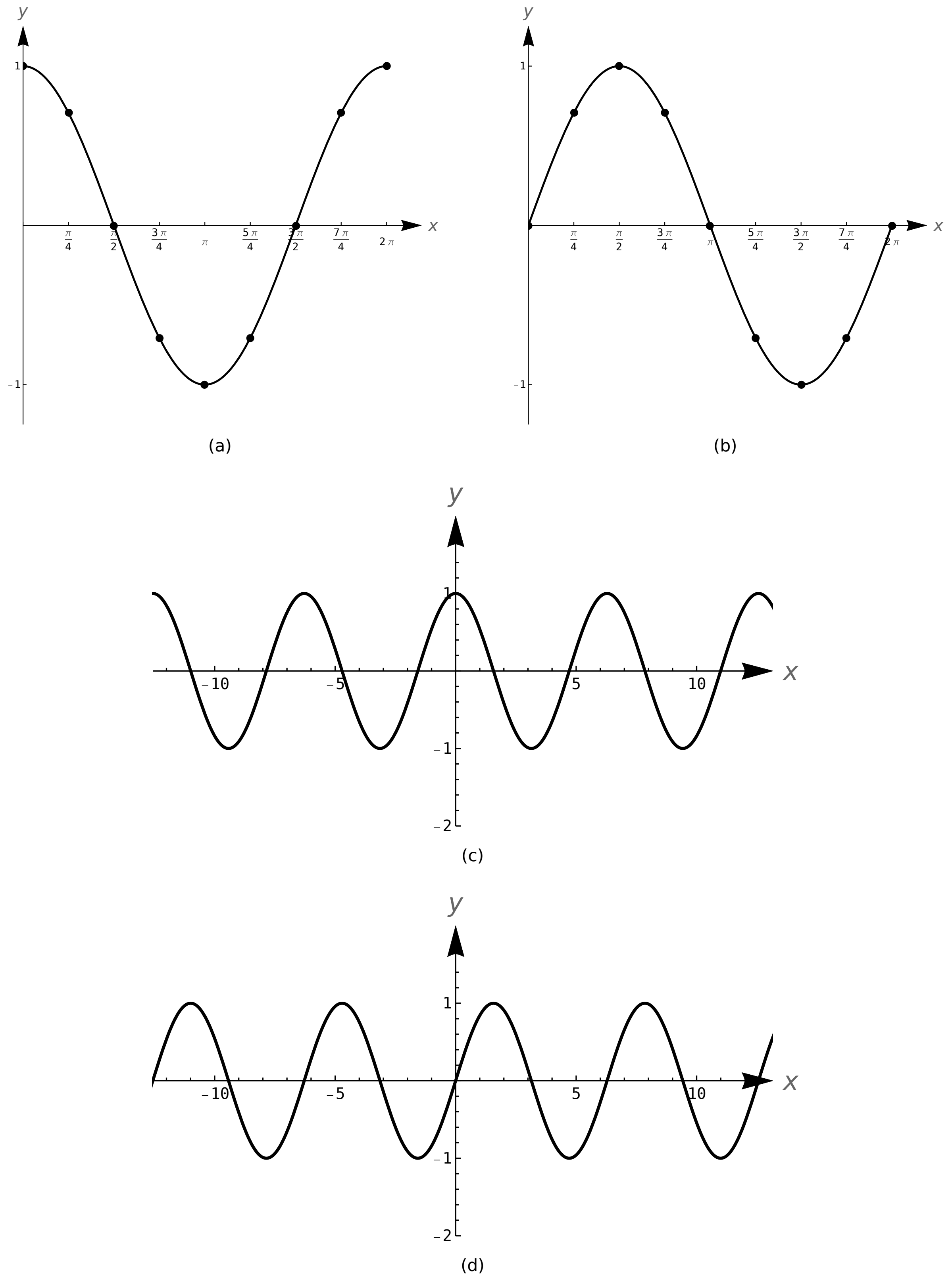


Figure 5.15: Graphs of the fundamental cycle (a,b) and four such cycles (c,d) of $y = \cos(x)$ (a,c) and $y = \sin(x)$ (b,d).

$$y = A\cos(\omega x + \phi) + B \quad \text{and} \quad y = A\sin(\omega x + \phi) + B,$$

or equivalently,

$$y = A\cos\left[\omega\left(x + \frac{\phi}{\omega}\right)\right] + B \quad \text{and} \quad y = A\sin\left[\omega\left(x + \frac{\phi}{\omega}\right)\right] + B,$$

we have:

- period $\dfrac{2\pi}{\omega}$
- amplitude $|A|$
- phase shift $-\dfrac{\phi}{\omega}$
- vertical shift B

for $\omega > 0$. Here, ϕ is called the **phase** (*fase*) of the sinusoid, while ω is nothing but the angular velocity. It is the number of cycles the sinusoid completes over a 2π interval.

5.3.6 Graphs of the other trigonometric functions

Before constructing the graphs of the other trigonometric functions, we first determine their domains and ranges.

5.3.6.1 Domain and range

Starting from the domain and range of the cosine and sine functions, we can now determine the domains and ranges of the other trigonometric functions.

For what concerns the function $f(x) = \sec(x) = \frac{1}{\cos(x)}$. We know that it is undefined whenever $\cos(x) = 0$. Since we know $\cos(x) = 0$ whenever $x = \frac{\pi}{2} + \pi k$ for integers k, the domain of this function, in set builder notation is

$$\text{dom } \sec(x) = \left\{x : x \neq \frac{\pi}{2} + \pi k, \forall k \in \mathbb{Z}\right\}.$$

Using interval notation to describe this set, we get

$$\text{dom } \sec(x) = \ldots \cup \left]-\frac{5\pi}{2}, -\frac{3\pi}{2}\right[\cup \left]-\frac{3\pi}{2}, -\frac{\pi}{2}\right[\cup \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\cup \left]\frac{\pi}{2}, \frac{3\pi}{2}\right[\cup \left]\frac{3\pi}{2}, \frac{5\pi}{2}\right[\cup \ldots$$

This is, however, a very cumbersome notation. In order to write this in a more compact way, we note that from the set-builder description of the domain, the kth point excluded from the domain, which we will call x_k, and which is given by $x_k = \frac{(2k+1)\pi}{2}$. Hence, the domain of $f(x) = \sec(x)$ consists of the intervals determined by successive points x_k:

$$]x_k, x_{k+1}[= \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[,$$

where $k = 0, \pm 1, \pm 2, \ldots$. The union of infinitely many such intervals can be written as

$$\operatorname{dom}\sec(x) = \bigcup_{k=-\infty}^{+\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[.$$

To determine the range of $f(x) = \sec(x)$, we appeal to the definition $\sec(x) = \frac{1}{\cos(x)}$ and recall that the range of $f(x) = \cos(x)$ is $[-1, 1]$. Since $f(x) = \sec(x)$ is undefined when $\cos(x) = 0$, we split our discussion into two cases: when $0 < \cos(x) \le 1$ and when $-1 \le \cos(x) < 0$. If the former case we can divide the inequality $\cos(x) \le 1$ by $\cos(x)$ to obtain $\sec(x) = \frac{1}{\cos(x)} \ge 1$. Moreover, we have that as $\cos(x) \to 0^+$, $\sec(x) \to +\infty$. If, on the other hand, if $-1 \le \cos(x) < 0$, then dividing by $\cos(x)$ causes a reversal of the inequality so that $\sec(x) = \frac{1}{\cos(x)} \le -1$. In this case, as $\cos(x) \to 0^-$, we get $\sec(x) \to -\infty$. Since $f(x) = \cos(x)$ admits all of the values in $[-1, 1]$, the function $f(x) = \sec(x)$ admits all of the values in $]-\infty, -1] \cup [1, +\infty[$. Using set-builder notation, the range of $f(x) = \sec(x)$ can be written as

$$\operatorname{im}\sec(x) = \{u : u \le -1 \vee u \ge 1\} = \{u : |u| \ge 1\}.$$

Similar arguments can be used to determine the domains and ranges of the remaining three circular functions: $\csc(x)$, $\tan(x)$ and $\cot(x)$. The reader is encouraged to do so. The results are summarized in Table 5.4.

Table 5.4: Domains and ranges of the trigonometric functions.

Function	Domain	Range
$f(x) = \sin(x)$	$\mathbb{R}$	$[-1, 1]$
$f(x) = \cos(x)$	$\mathbb{R}$	$[-1, 1]$
$f(x) = \sec(x)$	$\bigcup_{k=-\infty}^{\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[$	$\{u : \|u\| \ge 1\} =]-\infty, -1] \cup [1, +\infty[$
$f(x) = \csc(x)$	$\bigcup_{k=-\infty}^{\infty}]k\pi, (k+1)\pi[$	$\{u : \|u\| \ge 1\} =]-\infty, -1] \cup [1, +\infty[$
$f(x) = \tan(x)$	$\bigcup_{k=-\infty}^{\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[$	$\mathbb{R}$
$f(x) = \cot(x)$	$\bigcup_{k=-\infty}^{\infty}]k\pi, (k+1)\pi[$	$\mathbb{R}$

5.3.6.2 Graphs of secant and cosecant functions

Since $\sec(x) = \frac{1}{\cos(x)}$, we can use our table of values for the graph of $y = \cos(x)$ and take reciprocals. We already know from Table 5.4 that the domain of $f(x) = \sec(x)$ excludes all odd multiples of $\frac{\pi}{2}$. As $x \underset{<}{\to} \frac{\pi}{2}$, $\cos(x) \to 0^+$, so $\sec(x) \to +\infty$. Similarly, we find that as $x \underset{>}{\to} \frac{\pi}{2}$, $\sec(x) \to -\infty$; as $x \underset{<}{\to} \frac{3\pi}{2}$, $\sec(x) \to -\infty$; and as $x \underset{>}{\to} \frac{3\pi}{2}$, $\sec(x) \to +\infty$. This means we have a pair of vertical asymptotes to the graph of $y = \sec(x)$, $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$. Since $\cos(x)$ is periodic with period 2π, it follows that $\sec(x)$ is also. In Figure 5.16(a) we graph a fundamental cycle of $y = \sec(x)$, while Figure 5.16(c) shows four cycles. .

As one would expect, to graph $y = \csc(x)$ we begin with $y = \sin(x)$ and take reciprocals of the corresponding y-values. Here, we encounter issues at $x = 0$, $x = \pi$ and $x = 2\pi$. We graph the fundamental cycle of $y = \csc(x)$ in Figure 5.16(b) along with the dotted graph of $y = \sin(x)$ for reference. Since $y = \sin(x)$ and $y = \cos(x)$ are merely phase shifts of each other, so too are $y = \csc(x)$ and $y = \sec(x)$. This can be seen easily in Figure 5.16(d) showing four cycles.

5.3.6.3 Graphs of tangent and cotangent functions

Finally, we turn our attention to the graphs of the tangent and cotangent functions. When constructing a table of values for the tangent function, we see that $y = \tan(x)$ is undefined at $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$. As $x \underset{<}{\to} \frac{\pi}{2}$, $\sin(x) \to 1^-$ and $\cos(x) \to 0^+$, so that $\tan(x) = \frac{\sin(x)}{\cos(x)} \to +\infty$ producing a vertical asymptote at $x = \frac{\pi}{2}$. Using a similar analysis, we get that as $x \underset{>}{\to} \frac{\pi}{2}$, $\tan(x) \to -\infty$; as $x \underset{<}{\to} \frac{3\pi}{2}$, $\tan(x) \to +\infty$; and as $x \underset{>}{\to} \frac{3\pi}{2}$, $\tan(x) \to -\infty$. Plotting this information yields the graph in Figures 5.17(a) and 5.17(c).

From this graph, it appears as if the tangent function is periodic with period π. To prove that this is the case, we appeal to the sum formula for tangents. We have:

$$\tan(x+\pi) = \frac{\tan(x)+\tan(\pi)}{1-\tan(x)\tan(\pi)} = \frac{\tan(x)+0}{1-(\tan(x))(0)} = \tan(x),$$

which tells us the period of $\tan(x)$ is at most π. To show that it is exactly π, suppose p is a positive real number so that $\tan(x+p) = \tan(x)$ for all real numbers x. For $x = 0$, we have $\tan(p) = \tan(0+p) = \tan(0) = 0$, which means p is a multiple of π. The smallest positive multiple of π is π itself, so we have established the result. We take as our fundamental cycle for $y = \tan(x)$ the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

It should be no surprise that $y = \cot(x)$ behaves similarly to $y = \tan(x)$. It clearly appears from Figures 5.17(b) and 5.17(d) as if the period of $\cot(x)$ is π, and we leave it to the reader to prove this. We take as one fundamental cycle the interval $[0, \pi]$.

Example 5.15

Graph one cycle of the following functions. State the period of each.

1. $f(x) = 1 - 2\sec(2x)$
2. $g(x) = 2\cot\left(\frac{\pi}{2}x + \pi\right) + 1$

Solution

1. To graph $f(x) = 1 - 2\sec(2x)$, we first set the argument of secant, $2x$, equal to 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ and 2π and solve for x, to get 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$, π. Next, we substitute these x-values into $f(x)$. If $f(x)$ exists, we have a point on the graph in Figure 5.18(a); otherwise, we have found a vertical asymptote. In addition to these points and asymptotes, we have graphed the associated cosine curve (dashed curve), being $y = 1 - 2\cos(2x)$. Since one cycle is graphed over the interval $[0, \pi]$, the period is $\pi - 0 = \pi$.

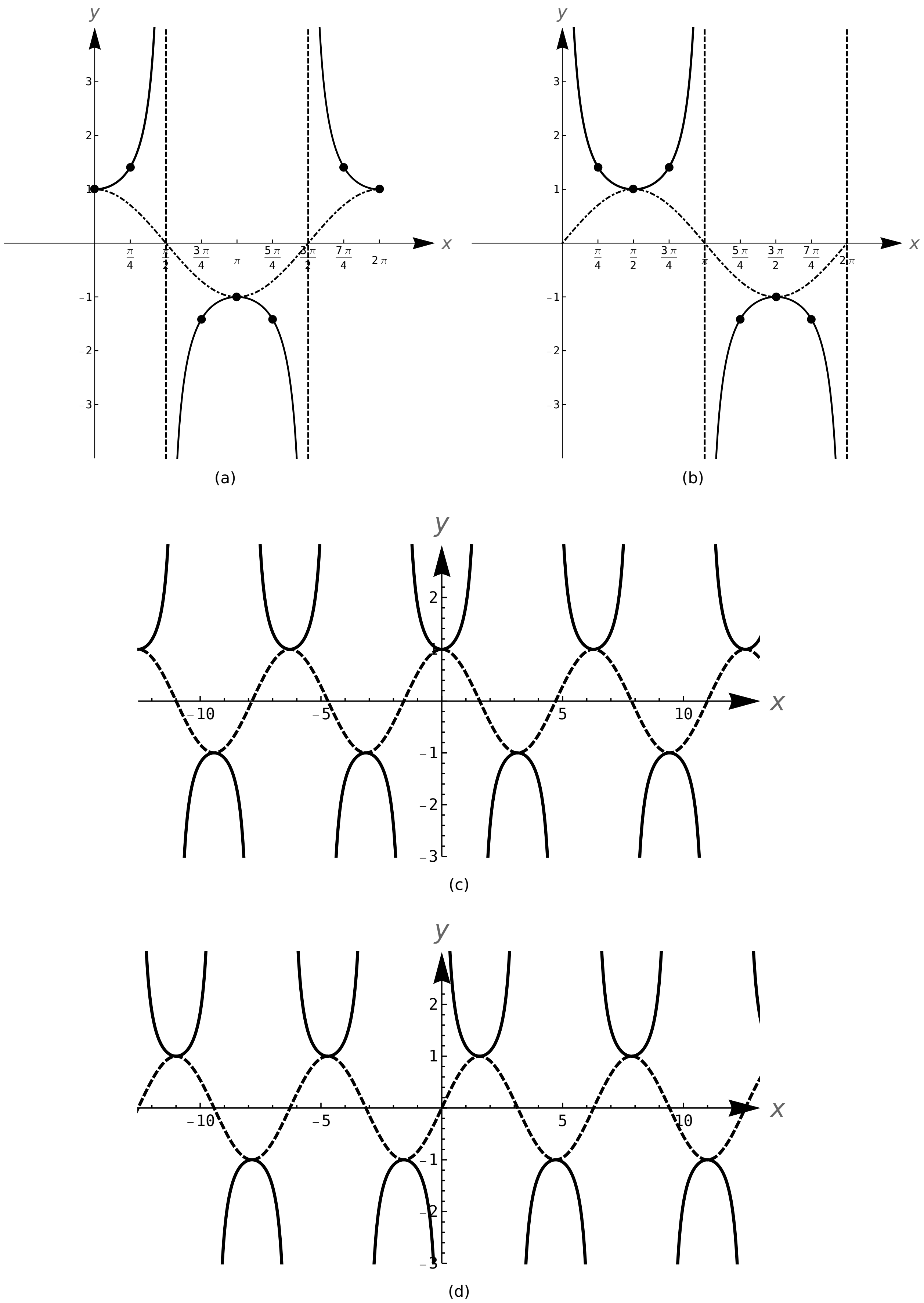


Figure 5.16: The fundamental cycle (a,b) and four cycles (c,d) of $y = \sec(x)$ (a,c) and $y = \csc(x)$ (b,d), superimposed on the graph of $y = \cos(x)$ and $y = \sin(x)$, respectively.

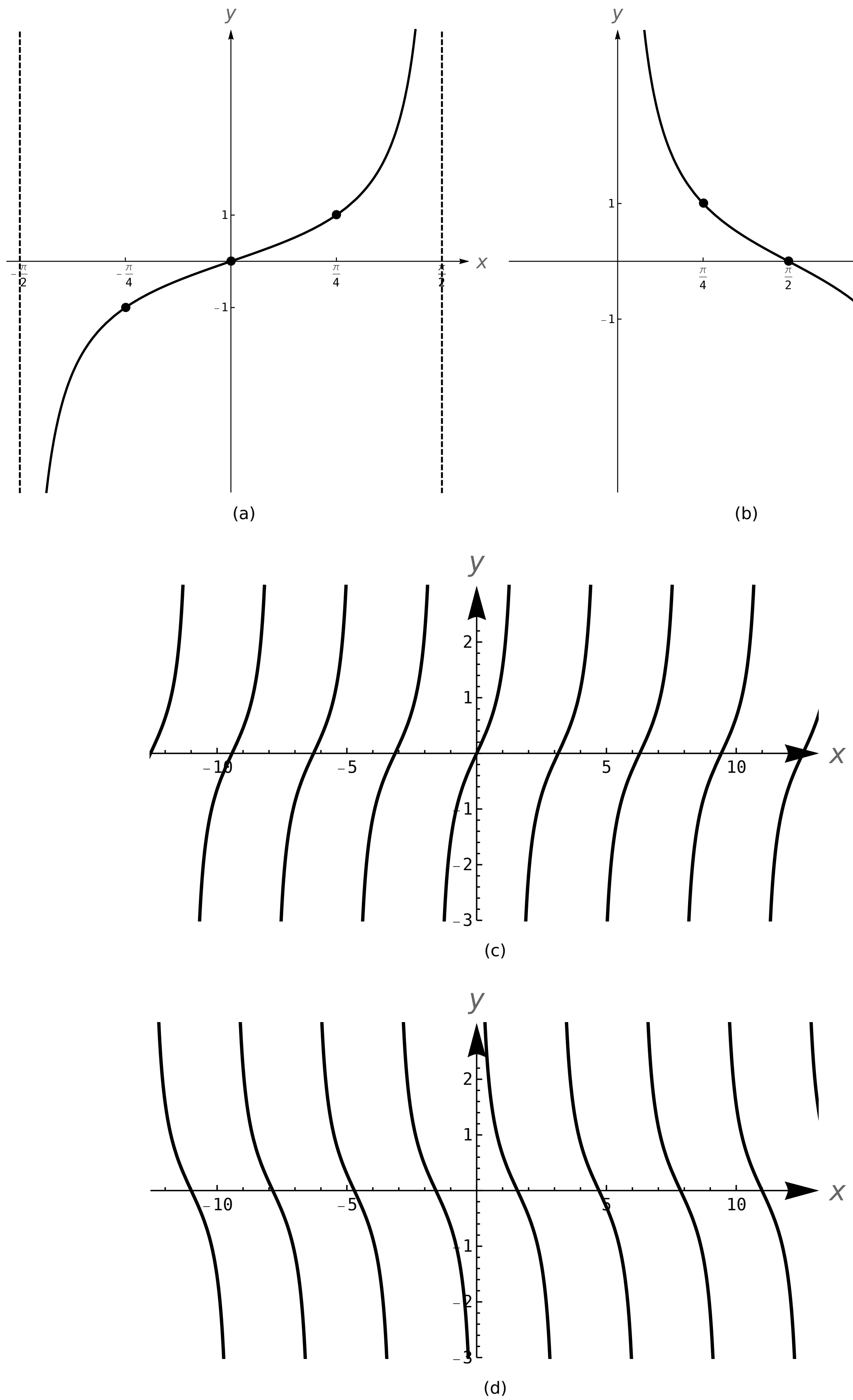


Figure 5.17: The fundamental cycle (a,b) and eight cycles (c,d) of $y = \tan(x)$ (a,c) and $y = \cot(x)$ (b,d).

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
$\sec(2x)$	1	undefined	−1	undefined	1
$f(x)$	−1	undefined	3	undefined	−1

2. We take 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ and π as quarter marks for constructing the fundamental cycle of the cotangent curve. To graph this function, we begin by setting $\frac{\pi}{2}x+\pi$ equal to each such mark and solve for x, to arrive at -2, $\frac{-3}{2}$, -1, $\frac{-1}{2}$ and 0. We now use these x-values to generate our graph (Figure 5.18(b)). We find the period to be $0-(-2)=2$.

x	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0
$\cot\left(\frac{\pi}{2}x+\pi\right)$	undefined	1	0	−1	undefined
$g(x)$	undefined	3	1	−1	undefined

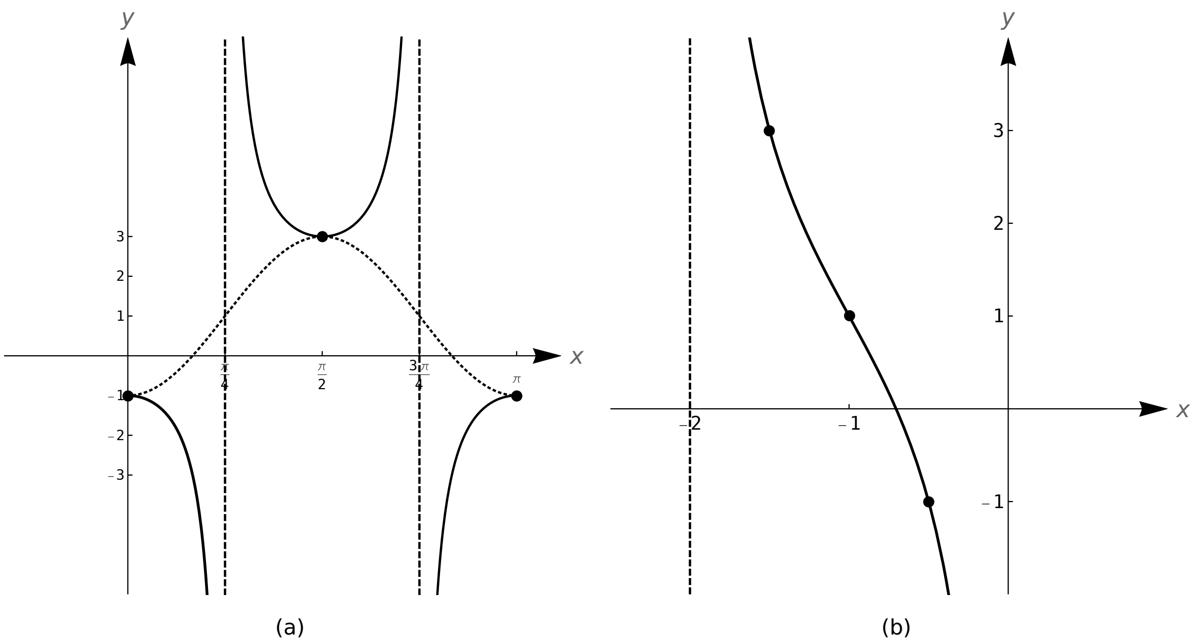


Figure 5.18: One cycle of $y=1-2\sec(2x)$ (a) and $y=2\cot(\frac{\pi}{2}x+\pi)+1$ (b) over $[0,2\pi]$.

5.3.6.4 Applications

Trigonometric functions are often used to solve problems that involve periodic behaviour, such as

- problems that involve circular movement on a repetitive nature;

- problems involving repetitive motion, such as spring motion (Figure 5.19(a)), oscillating waves and tides (Figure 5.19(b));
- problems involving environmental fluctuations (Figure 5.19(c)).

Example 5.16

According to the U.S. Naval Observatory website, the number of hours H of daylight that Fairbanks, Alaska received on the 21st day of the nth month of 2009 is given below. Here $t = 1$ represents January 21, 2009, $t = 2$ represents February 21, 2009, and so on.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Hours of Daylight	5.8	9.3	12.4	15.9	19.4	21.8	19.4	15.6	12.4	9.1	5.6	3.3

Find a sinusoid that models these data and use Mathematica to graph your answer along with the data.

Solution To get a feel for the data, we first plot it using the Mathematica built-in function **`ListPlot`**. The result is plotted in Figure 5.20.

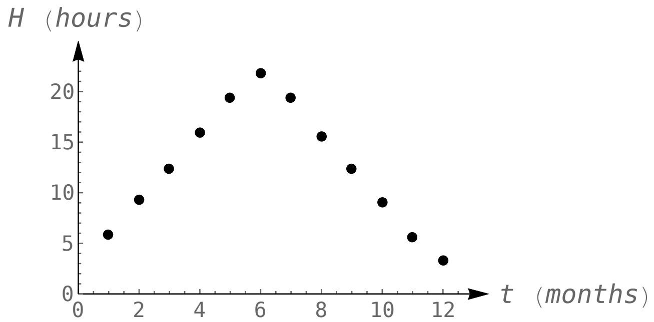
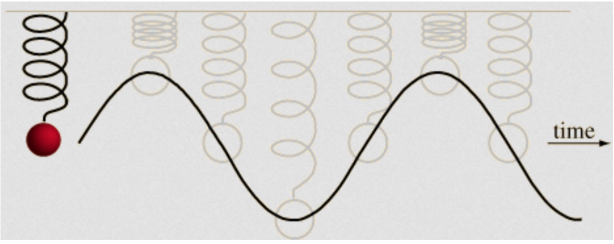


Figure 5.20: Hours of daylight on the 21st day of the month of 2009 in Fairbanks, Alaska

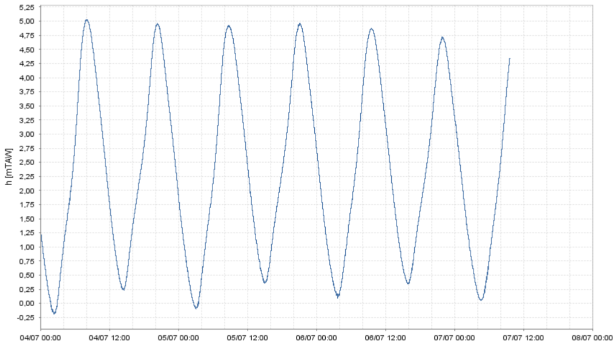
We do our best to find the constants A, ω, ϕ and B so that $H(t) = A\sin(\omega t + \phi) + B$ closely matches the data. We first go after the vertical shift B whose value determines the baseline. In a typical sinusoid, the value of B is the average of the maximum and minimum values. So here we take $B = \frac{3.3+21.8}{2} = 12.55$. Next is the amplitude A which is the displacement from the baseline to the maximum (and minimum) values. We find $A = 21.8 - 12.55 = 12.55 - 3.3 = 9.25$. At this point, we have $H(t) = 9.25\sin(\omega t + \phi) + 12.55$. Next, we go after the angular frequency ω. Since the data collected is over the span of a year (12 months), we take the period $T = 12$ months. This means $\omega = \frac{2\pi}{T} = \frac{2\pi}{12} = \frac{\pi}{6}$. The last quantity to find is the phase ϕ. It is easy to find the phase shift $-\frac{\phi}{\omega}$. Since we picked $A > 0$, the phase shift corresponds to the first value of t with $H(t) = 12.55$ (the baseline value). Here, we choose $t = 3$, since its corresponding H value of 12.4 is closer to 12.55 than the next value, 15.9, which corresponds to $t = 4$. Hence, $-\frac{\phi}{\omega} = 3$, so $\phi = -3\omega = -3\left(\frac{\pi}{6}\right) = -\frac{\pi}{2}$. We have

$$H(t) = 9.25\sin\left(\frac{\pi}{6}t - \frac{\pi}{2}\right) + 12.55,$$

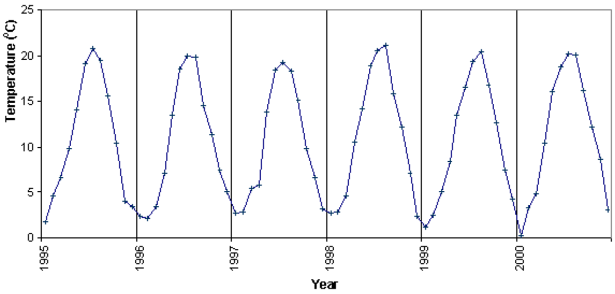
whose graph is shown together with the data in Figure 5.21.



(a) Motion of an object attached to a spring.



(b) Water height above sea level of the river Scheldt at the monitoring station Prosperpolder.



(c) Variation in annual cycle of river temperature at the Struma River in Bulgaria at the site of Boboshevo.

Figure 5.19: Applications of trigonometric functions illustrated.

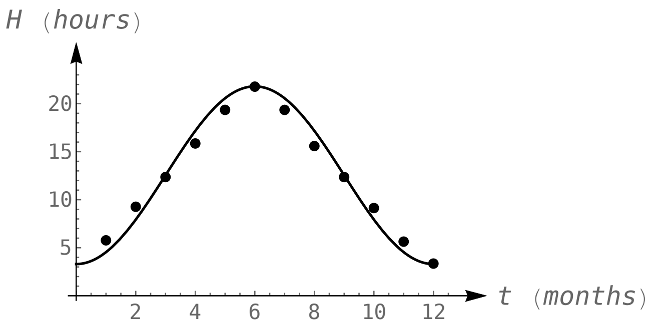


Figure 5.21: Hours of daylight on the 21st day of the month of 2009 in Fairbanks, Alaska

Forgotten trigonometric functions

In addition to the trigonometric functions we encountered in this chapter, there are a few functions that were common historically, but are nowadays sometimes forgotten, such as the chord, given by

$$\mathrm{crd}(\theta) = 2\sin\left(\frac{\theta}{2}\right),$$

the versine

$$\mathrm{versin}(\theta) = 1 - cos(\theta),$$

and the haversine

$$\mathrm{haversin}(\theta) = \frac{1}{2}\mathrm{versin}(\theta).$$

Still these functions are used in several fields. For instance, one period of a versine or haversine is commonly used in signal processing and control theory as the shape of a pulse.

5.4 Inverse trigonometric functions

In this section we concern ourselves with finding inverses of the trigonometric functions, which are also referred to as **arcus**, **antitrigonometric** or **cyclometric functions** (*cyclometrische functies*). Our immediate problem is that, owing to their periodic nature, none of the six trigoniometric functions is injective. To remedy this, we restrict the domains of the circular functions in the same way we restricted the domain of the quadratic function in Example 3.11 to obtain an injective function.

5.4.1 Inverse cosine and sine functions

We first consider $f(x) = \cos(x)$. Choosing the interval $[0, \pi]$ allows us to keep the range as $[-1, 1]$ as well as the properties of being smooth and continuous. Recall from Section 3.4 that the inverse of a function f is typically denoted f^{-1}. For this reason, some textbooks use the notation $f^{-1}(x) = \cos^{-1}(x)$ for the inverse of $f(x) = \cos(x)$. The obvious pitfall here is our convention of writing $(\cos(x))^2$ as $\cos^2(x)$, $(\cos(x))^3$ as $\cos^3(x)$ and so on. It is far too easy to confuse $\cos^{-1}(x)$ with $\frac{1}{\cos(x)} = \sec(x)$, so we will not use this notation. Instead, we use the notation $f^{-1}(x) = \arccos(x)$, read **arccosine** (*boogcosinus*) of x. As a consequence of the imposed domain restriction, we have that

$$\cos(\arccos(x)) = x\,,$$

provided $-1 \leq x \leq 1$, and

$$\arccos(\cos(x)) = x\,,$$

provided $0 \leq x \leq \pi$, which is important to keep in mind when solving equations involving (inverse) trigonometric functions.

To understand the arc in arccosine, recall that an inverse function, by definition, reverses the process of the original function. The function $f(x) = \cos(x)$ takes a real number input x, associates it with the angle $\theta = x$ radians, and returns the value $\cos(\theta)$. Digging deeper, we have that $\cos(\theta) = \cos(x)$ is the x-coordinate of the terminal point on the unit circle of an oriented arc of length $|x|$ whose initial point is $(1, 0)$. Hence, we may view the inputs to $f(x) = \cos(x)$ as oriented arcs and the outputs as x-coordinates on the unit circle. The function f^{-1}, then, would take x-coordinates on the unit circle and return oriented arcs, hence the arc in arccosine. In Figure 5.22 are the graphs of $f(x) = \cos(x)$ and $f^{-1}(x) = \arccos(x)$, where we obtain the latter from the former by reflecting it across the line $y = x$, in accordance with what we learned in Section 3.4.

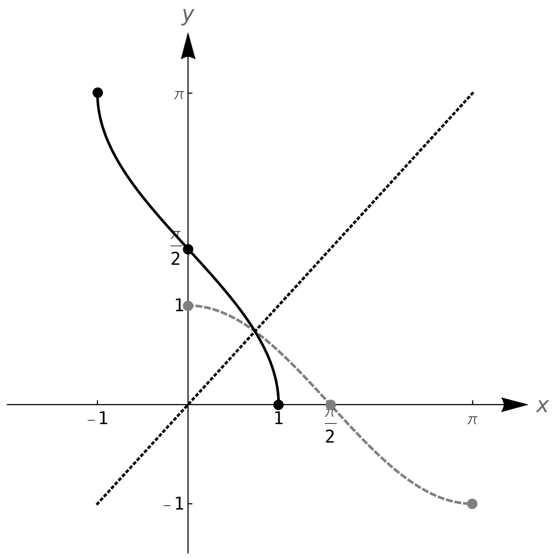


Figure 5.22: The graph of $y = \cos(x)$ for $x \in [0, \pi]$ (dashed) and $y = \arccos(x)$ for $x \in [-1, 1]$ (solid).

We restrict $g(x) = \sin(x)$ in a similar manner, although the interval of choice is $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (Figure 5.23). It should be no surprise that we call $g^{-1}(x) = \arcsin(x)$, which is read **arcsine** (*boogsinus*) of x. From Figure 5.23, we may infer that the arcsine is an odd function. Besides, due to the domain restriction, we have that

$$\sin(\arcsin(x)) = x\,,$$

provided $-1 \leq x \leq 1$ and

$$\arcsin(\sin(x)) = x\,,$$

provided $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$.

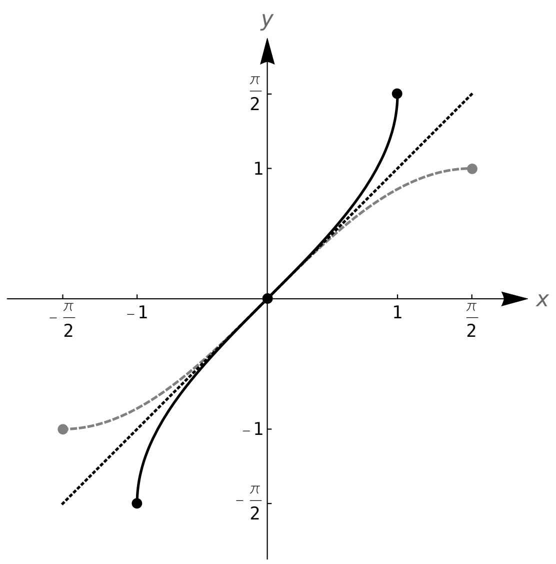


Figure 5.23: The graph of $y = \sin(x)$ for $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (dashed) and $y = \arcsin(x)$ for $x \in [-1, 1]$ (solid).

Example 5.17

Find the exact values of the following.

1. $\arccos\left(\frac{1}{2}\right)$
2. $\arccos\left(-\frac{\sqrt{2}}{2}\right)$
3. $\arcsin\left(-\frac{1}{2}\right)$
4. $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right)$
5. $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right)$
6. $\sin\left(\arccos\left(-\frac{3}{5}\right)\right)$

Solution

1. To find $\arccos\left(\frac{1}{2}\right)$, we need to find the real number x that lies between 0 and π for which it holds that $\cos(x) = \frac{1}{2}$. We know $x = \frac{\pi}{3}$ meets these criteria, so $\arccos\left(\frac{1}{2}\right) = \frac{\pi}{3}$.
2. The number $x = \arccos\left(-\frac{\sqrt{2}}{2}\right)$ lies in the interval $[0, \pi]$ with $\cos(x) = -\frac{\sqrt{2}}{2}$. Our answer is $\arccos\left(-\frac{\sqrt{2}}{2}\right) = \frac{3\pi}{4}$.
3. To find $\arcsin\left(-\frac{1}{2}\right)$, we seek the number x in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with $\sin(x) = -\frac{1}{2}$. The answer is $x = -\frac{\pi}{6}$ so that $\arcsin\left(-\frac{1}{2}\right) = -\frac{\pi}{6}$.
4. Since $0 \leq \frac{\pi}{6} \leq \pi$, we simply have that $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \frac{\pi}{6}$. However, in order to make sure we understand why this is the case, we can use the definition of arccosine. Working from the inside out, $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \arccos\left(\frac{\sqrt{3}}{2}\right)$. Now, $\arccos\left(\frac{\sqrt{3}}{2}\right)$ is the real number x with $0 \leq x \leq \pi$ and $\cos(x) = \frac{\sqrt{3}}{2}$. We find $x = \frac{\pi}{6}$, so that $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \frac{\pi}{6}$.
5. Since $\frac{11\pi}{6}$ does not fall between 0 and π, we are forced to work from the inside out starting with $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \arccos\left(\frac{\sqrt{3}}{2}\right)$. From the previous problem, we know $\arccos\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{6}$. Hence, $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \frac{\pi}{6}$.
6. We let $x = \arccos\left(-\frac{3}{5}\right)$ so that $\cos(x) = -\frac{3}{5}$ for some x where $0 \leq x \leq \pi$. Since $\cos(x) < 0$, we can narrow this down a bit and conclude that $\frac{\pi}{2} < x < \pi$, so that x corresponds to an angle in Quadrant II. In terms of x, then, we need to find $\sin\left(\arccos\left(-\frac{3}{5}\right)\right) = \sin(x)$. Using the Pythagorean identity, we get $\left(-\frac{3}{5}\right)^2 + \sin^2(x) = 1$ or $\sin(x) = \pm\frac{4}{5}$. Since x corresponds to a Quadrants II angle, we choose $\sin(x) = \frac{4}{5}$. Hence, $\sin\left(\arccos\left(-\frac{3}{5}\right)\right) = \frac{4}{5}$.

Most of the common errors encountered in dealing with the inverse trigonometric functions come from the need to restrict the domains of the original functions so that they are injective. One instance of this phenomenon is the fact that $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \frac{\pi}{6}$ as opposed to $\frac{11\pi}{6}$. This is the exact same phenomenon discussed in Section 3.4 when we saw $\sqrt{(-2)^2} = 2$ as opposed to -2.

5.4.2 Inverse tangent and cotangent functions

The next pair of functions we discuss are the inverses of tangent and cotangent, which are named **arctangent** (*boogtangens*) and **arccotangent** (*boogcotangens*), respectively. First, we restrict $f(x) = \tan(x)$ to its fundamental cycle on $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ to obtain $f^{-1}(x) = \arctan(x)$. Among other things, note that the vertical asymptotes $x = -\frac{\pi}{2}$ and $x = \frac{\pi}{2}$ of the graph of $f(x) = \tan(x)$ become the horizontal asymptotes $y = -\frac{\pi}{2}$ and $y = \frac{\pi}{2}$ of the graph of $f^{-1}(x) = \arctan(x)$ (Figure 5.24). Observe that the arctangent is an odd function and due to the domain restrition that

$$\tan\big(\arctan(x)\big) = x\,,$$

for all real numbers x, whereas

$$\arctan(\tan(x)) = x\,,$$

$-\frac{\pi}{2} < x < \frac{\pi}{2}$ only.

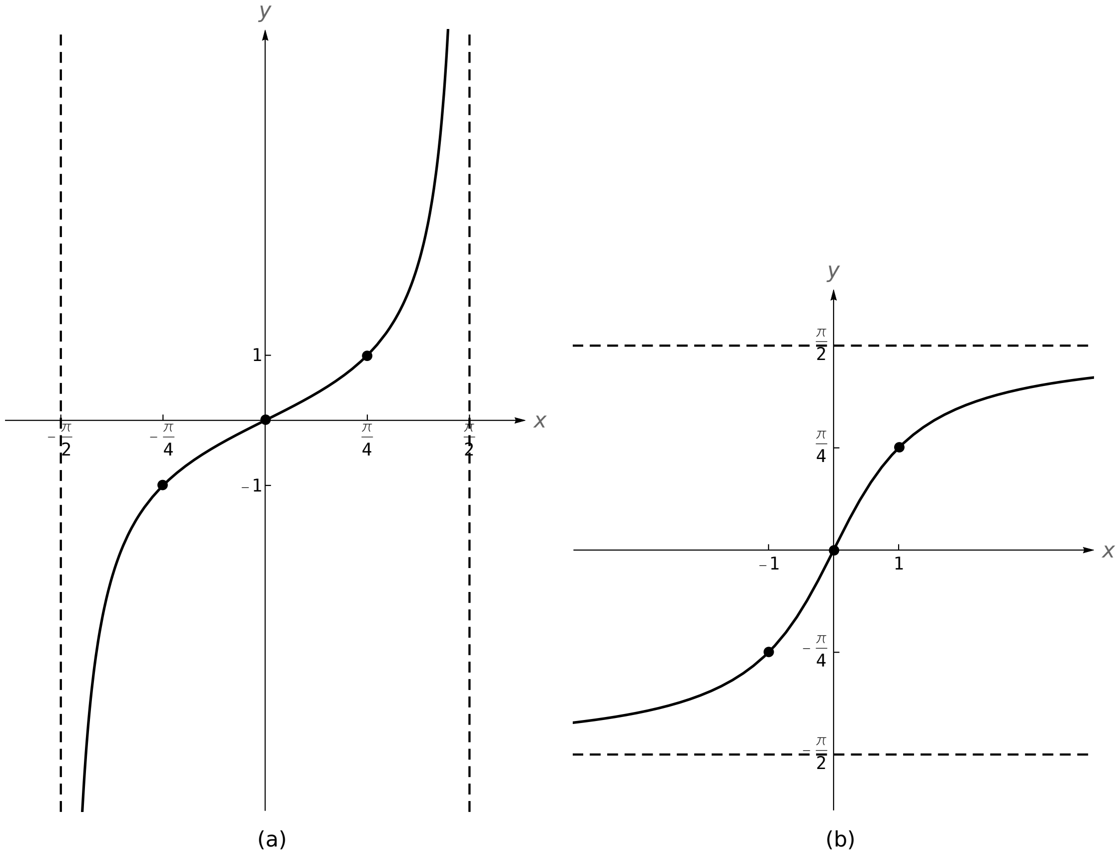


Figure 5.24: One cycle of $y = \tan(x)$ (a) and $y = \arctan(x)$ (b).

Next, we restrict $g(x) = \cot(x)$ to its fundamental cycle on $]\,0,\pi\,[$ to obtain $g^{-1}(x) = \operatorname{arccot}(x)$. Once again, the vertical asymptotes $x = 0$ and $x = \pi$ of the graph of $g(x) = \cot(x)$ become the horizontal asymptotes $y = 0$ and $y = \pi$ of the graph of $g^{-1}(x) = \operatorname{arccot}(x)$ (Figure 5.25). Moreover, as a consequence of the imposed domain restriction, we have

$$\cot\big(\operatorname{arccot}(x)\big) = x\,,$$

for all real numbers x and

$$\operatorname{arccot}(\cot(x)) = x\,,$$

provided $0 < x < \pi$ only. Finally, we also have that

$$\arctan(x) = \operatorname{arccot}\left(\frac{1}{x}\right),$$

for $x > 0$ and likewise

$$\operatorname{arccot}(x) = \arctan\left(\frac{1}{x}\right),$$

for $x > 0$.

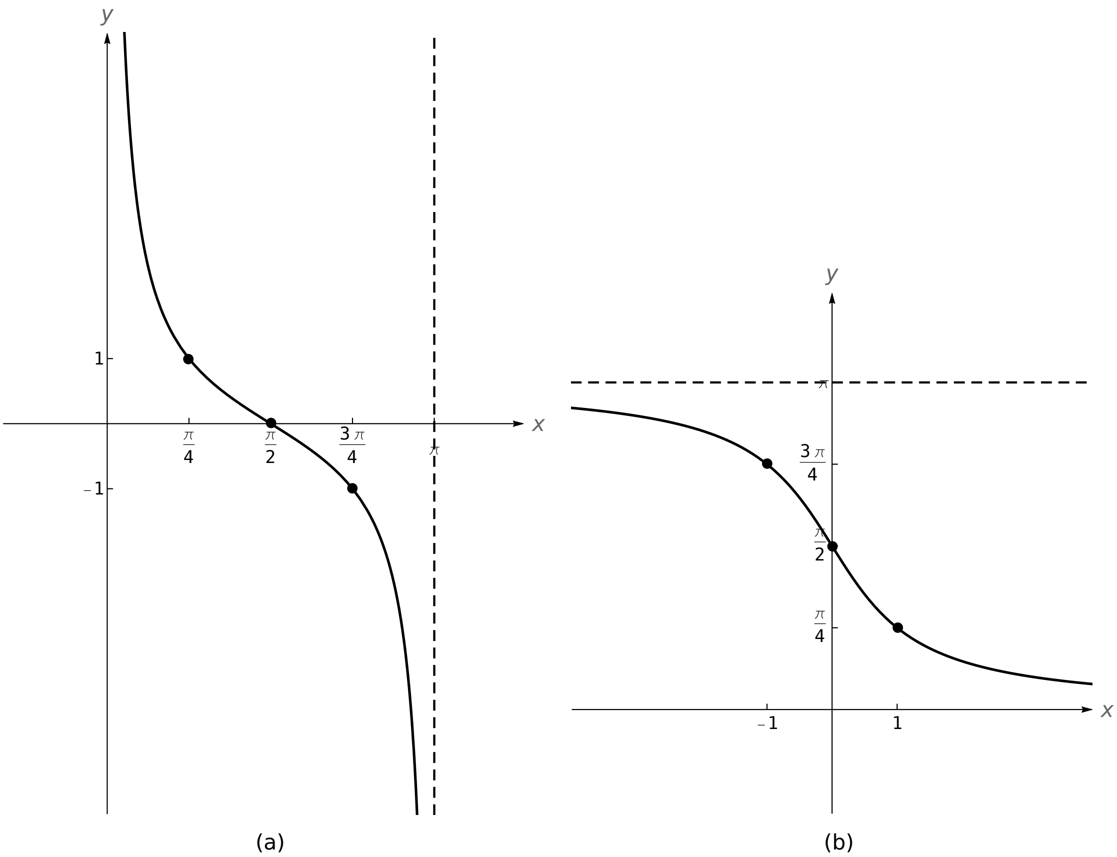


Figure 5.25: One cycle of $y = \cot(x)$ (a) and $y = \operatorname{arccot}(x)$ (b).

Example 5.18

Rewrite the following as algebraic expressions of x and state the domain on which the equivalence is valid.

1. $\tan(2\arctan(x))$
2. $\cos(\operatorname{arccot}(2x))$

Solution

1. If we let $t = \arctan(x)$, then $-\frac{\pi}{2} < t < \frac{\pi}{2}$ and $\tan(t) = x$. We look for a way to express $\tan(2\arctan(x)) = \tan(2t)$ in terms of x. Before we get started using identities, we note that $\tan(2t)$ is undefined when $2t = \frac{\pi}{2} + \pi k$ for integers k. Dividing both sides of this equation by 2 tells us we need to exclude values of t where $t = \frac{\pi}{4} + \frac{\pi}{2}k$, where k is an integer. The only members of this family which lie in $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ are $t = \pm\frac{\pi}{4}$, which means the values of t under consideration are $\left]-\frac{\pi}{2}, -\frac{\pi}{4}\right[\cup \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[\cup \left]\frac{\pi}{4}, \frac{\pi}{2}\right[$. Returning to $\tan(2t)$, we note the double angle identity $\tan(2t) = \frac{2\tan(t)}{1-\tan^2(t)}$, is valid for all the values of t under consideration, hence we get
$$\tan(2\arctan(x)) = \tan(2t) = \frac{2\tan(t)}{1-\tan^2(t)} = \frac{2x}{1-x^2}.$$
To find where this equivalence is valid we check back with our substitution $t = \arctan(x)$. Since the domain of $\arctan(x)$ is all real numbers, the only exclusions come from the values of t we discarded earlier, $t = \pm\frac{\pi}{4}$. Since $x = \tan(t)$, this means we exclude $x = \tan\left(\pm\frac{\pi}{4}\right) = \pm 1$.

Hence, the equivalence $\tan(2\arctan(x)) = \frac{2x}{1-x^2}$ holds for all x in $\mathbb{R} \setminus \{-1, 1\}$.

2. To get started, we let $t = \operatorname{arccot}(2x)$ so that $\cot(t) = 2x$ where $0 < t < \pi$. In terms of t, $\cos(\operatorname{arccot}(2x)) = \cos(t)$, and our goal is to express the latter in terms of x. Since $\cos(t)$ is always defined, there are no additional restrictions on t, so we can begin using identities to relate $\cot(t)$ to $\cos(t)$. The identity $\cot(t) = \frac{\cos(t)}{\sin(t)}$ is valid for t in $]0, \pi[$, so our strategy is to obtain $\sin(t)$ in terms of x, then write $\cos(t) = \cot(t)\sin(t)$. The identity $1 + \cot^2(t) = \csc^2(t)$ holds for all t in $]0, \pi[$ and relates $\cot(t)$ and $\csc(t) = \frac{1}{\sin(t)}$. Substituting $\cot(t) = 2x$, we get $1 + (2x)^2 = \csc^2(t)$, or $\csc(t) = \pm\sqrt{4x^2+1}$. Since t is between 0 and π, $\csc(t) > 0$, so $\csc(t) = \sqrt{4x^2+1}$ which gives $\sin(t) = \left(4x^2+1\right)^{-1/2}$. Hence,

$$\cos(\operatorname{arccot}(2x)) = \cos(t) = \cot(t)\sin(t) = \frac{2x}{\sqrt{4x^2+1}}.$$

Since $\operatorname{arccot}(2x)$ is defined for all real numbers x and we encountered no additional restrictions on t, we have $\cos\left(\operatorname{arccot}(2x)\right) = \frac{2x}{\sqrt{4x^2+1}}$ for all real numbers x.

For comprehensiveness, Table 5.5 lists the domain restrictions and corresponding range of the trigonometric and inverse trigonometric functions that will be used throughout the remainder of this course.

5.4.3 Solving equations involving trigonometric functions

We have already seen how to solve equations like $\tan(x) = -1$ for real numbers x by appealing to the unit circle and relying on the fact that the answers corresponded to a set of common angles listed in Tables 5.1 and 5.2. If, on the other hand, we had been asked to solve $\tan(x) = -2$, we would have been hard-pressed to do so. With the introduction of the inverse trigonometric functions, however, we are now in a position to solve these equations. With our comprehension of the unit circle and hence our understanding between the output of a trigonometric function and its input, we may infer the following strategies for solving basic trigonometric equations.

- To solve $\cos(x) = c$ or $\sin(x) = c$ for $-1 \leq c \leq 1$, first solve for x in the interval $[0, 2\pi[$, thereby not forgetting that opposite and supplementary angles are also solutions of $\cos(x) = c$ and $\sin(x) = c$, respectively. Finally, add integer multiples of the period 2π. If $c < -1$ or of $c > 1$, there are no real solutions.
- To solve $\sec(x) = c$ or $\csc(x) = c$ for $c \leq -1$ or $c \geq 1$, convert to cosine or sine, respectively, and solve as above. If $-1 < c < 1$, there are no real solutions.
- To solve $\tan(x) = c$ for any real number c, first solve for x in the interval $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ and add integer multiples of the period π.
- To solve $\cot(x) = c$ for $c \neq 0$, convert to tangent and solve as above. If $c = 0$, the solution to $\cot(x) = 0$ is $x = \frac{\pi}{2} + \pi k$ for integers k.

The question remains, however, how do we solve something like $\sin(3x) = \frac{1}{2}$? Since this equation has the form $\sin(u) = \frac{1}{2}$, we know the solutions take the form $u = \frac{\pi}{6} + 2\pi k$ or $u = \frac{5\pi}{6} + 2\pi k$ for integers k. Since the argument of sine here is $3x$, we have $3x = \frac{\pi}{6} + 2\pi k$ or $3x = \frac{5\pi}{6} + 2\pi k$ for integers k. To solve for x, we divide both sides of these equations by 3, and obtain $x = \frac{\pi}{18} + \frac{2\pi}{3}k$ or $x = \frac{5\pi}{18} + \frac{2\pi}{3}k$ for integers k. For what concerns equations involving two different trigonometric functions or equations containing the same trigonometric functions but with different arguments, we will need to use the identities we introduced in Section 5.3 identities and some algebra to manipulate the equation up to a point that we reach a solvable one.

Table 5.5: Domains and ranges of the trigonometric and inverse trigonometric functions.

Function	Domain	Range
$\sin(x)$	$\mathbb{R}$	$[-1,1]$
$\cos(x)$	$\mathbb{R}$	$[-1,1]$
$\tan(x)$	$]-\frac{\pi}{2},\frac{\pi}{2}[+k\pi$	$]-\infty,+\infty[$
$\cot(x)$	$]0,\pi[+k\pi$	$]-\infty,+\infty[$
$\csc(x)$	$[-\frac{\pi}{2},0[\cup]0,\frac{\pi}{2}]+k\pi$	$]-\infty,-1]\cup[1,+\infty[$
$\sec(x)$	$[0,\frac{\pi}{2}[\cup]\frac{\pi}{2},\pi]+k\pi$	$]-\infty,-1]\cup[1,+\infty[$
$\arcsin(x)$	$[-1,1]$	$[-\frac{\pi}{2},\frac{\pi}{2}]$
$\arccos(x)$	$[-1,1]$	$[0,\pi]$
$\arctan(x)$	$]-\infty,+\infty[$	$]-\frac{\pi}{2},\frac{\pi}{2}[$
$\operatorname{arccot}(x)$	$]-\infty,+\infty[$	$]0,\pi[$
$\operatorname{arccsc}(x)$	$]-\infty,-1]\cup[1,+\infty]$	$[-\frac{\pi}{2},0[\cup]0,\frac{\pi}{2}]$
$\operatorname{arcsec}(x)$	$]-\infty,-1]\cup[1,+\infty]$	$[0,\frac{\pi}{2}[\cup]\frac{\pi}{2},\pi]$

Example 5.19

Solve the following equations and inequalities and list the solutions which lie in the interval $[0,2\pi[$.

1. $\cos(2x) = -\dfrac{\sqrt{3}}{2}$
2. $\csc\left(\dfrac{1}{3}x-\pi\right) = \sqrt{2}$
3. $\sec^2(x) = 4$
4. $3\sin^3(x) = \sin^2(x)$
5. $\cos(2x) = 3\cos(x)-2$
6. $\sin(2x) > \cos(x)$

Solution

1. From $\cos(2x) = -\frac{\sqrt{3}}{2}$, we immediately infer that the solutions are given by

$$\begin{aligned} & 2x = \tfrac{5\pi}{6}+2\pi k \quad \vee \quad 2x = \tfrac{7\pi}{6}+2\pi k \\ \Leftrightarrow \quad & x = \tfrac{5\pi}{12}+\pi k \quad \vee \quad x = \tfrac{7\pi}{12}+\pi k \end{aligned}$$

for integers k. The solutions that lie in $[0,2\pi[$ are $x = \frac{5\pi}{12}, \frac{7\pi}{12}, \frac{17\pi}{12}$ and $\frac{19\pi}{12}$.

2. Since this equation has the form $\csc(u) = \sqrt{2}$, we rewrite this as $\sin(u) = \frac{\sqrt{2}}{2}$ and we directly find $u = \frac{\pi}{4}+2\pi k$ or $u = \frac{3\pi}{4}+2\pi k$ for integers k. Since the argument of cosecant here is $\left(\frac{1}{3}x-\pi\right)$, we have

$$\begin{aligned} & \tfrac{1}{3}x-\pi = \frac{\pi}{4}+2\pi k \quad \vee \quad \frac{1}{3}x-\pi = \frac{3\pi}{4}+2\pi k \\ \Leftrightarrow \quad & x = \tfrac{3\pi}{4}+6\pi k+3\pi = \frac{15\pi}{4}+6\pi k \quad \vee \quad x = \tfrac{9\pi}{4}+6\pi k+3\pi = \frac{21\pi}{4}+6\pi k \end{aligned}$$

None of the solutions lie in $[0,2\pi[$.

3. We start by extracting square roots to get $\sec(x) = \pm 2$. Converting to cosines, we have $\cos(x) = \pm\frac{1}{2}$. For $\cos(x) = \frac{1}{2}$, we get $x = \frac{\pi}{3}+2\pi k$ or $x = -\frac{\pi}{3}+2\pi k$ for integers k. For

$\cos(x) = -\frac{1}{2}$, we get $x = \frac{2\pi}{3} + 2\pi k$ or $x = -\frac{2\pi}{3} + 2\pi k$ for integers k. These solutions can be combined as $x = \frac{\pi}{3} + \pi k$ and $x = \frac{2\pi}{3} + \pi k$ for integers k. The solutions that lie in $[0, 2\pi[$ are $x = \frac{\pi}{3}$, $\frac{2\pi}{3}$, $\frac{4\pi}{3}$ and $\frac{5\pi}{3}$.

4. We may not divide both sides of $3\sin^3(x) = \sin^2(x)$ by $\sin^2(x)$. Instead we gather all of the terms to one side of the equation and factor.

$$\begin{array}{rrcl} & 3\sin^3(x) & = & \sin^2(x) \\ \Leftrightarrow & 3\sin^3(x) - \sin^2(x) & = & 0 \\ \Leftrightarrow & \sin^2(x)(3\sin(x) - 1) & = & 0 \end{array}$$

So, we get $\sin^2(x) = 0$ or $3\sin(x) - 1 = 0$, from which we deduce that $\sin(x) = 0$ or $\sin(x) = \frac{1}{3}$. The solution to the first equation is $x = \pi k$, with $x = 0$ and $x = \pi$ being the two solutions which lie in $[0, 2\pi[$. To solve $\sin(x) = \frac{1}{3}$, we use the arcsine function to get $x = \arcsin\left(\frac{1}{3}\right) + 2\pi k$ or $x = \pi - \arcsin\left(\frac{1}{3}\right) + 2\pi k$ for integers k. We find the two solutions here which lie in $[0, 2\pi[$, manely $x = \arcsin\left(\frac{1}{3}\right)$ and $x = \pi - \arcsin\left(\frac{1}{3}\right)$.

5. We have cosine on both sides, but the arguments differ. Using the identity $\cos(2x) = 2\cos^2(x) - 1$, we obtain a quadratic in disguise, which can be solved:

$$\begin{array}{rrcll} & \cos(2x) & = & 3\cos(x) - 2 & \\ \Leftrightarrow & 2\cos^2(x) - 1 & = & 3\cos(x) - 2 & \\ \Leftrightarrow & 2\cos^2(x) - 3\cos(x) + 1 & = & 0 & \\ \Leftrightarrow & 2u^2 - 3u + 1 & = & 0 & \text{(Let } u = \cos(x).\text{)} \\ \Leftrightarrow & (2u - 1)(u - 1) & = & 0\,. & \end{array}$$

This gives $u = \frac{1}{2}$ or $u = 1$. Since $u = \cos(x)$, we get $\cos(x) = \frac{1}{2}$ or $\cos(x) = 1$. Solving $\cos(x) = \frac{1}{2}$, we get $x = \frac{\pi}{3} + 2\pi k$ or $x = -\frac{\pi}{3} + 2\pi k$ for integers k. From $\cos(x) = 1$, we get $x = 2\pi k$ for integers k. The answers which lie in $[0, 2\pi[$ are $x = 0$, $\frac{\pi}{3}$, and $\frac{5\pi}{3}$.

6. We first rewrite $\sin(2x) > \cos(x)$ as $\sin(2x) - \cos(x) > 0$ and let $f(x) = \sin(2x) - \cos(x)$. Our original inequality is thus equivalent to $f(x) > 0$. The domain of f is all real numbers, so we can advance to finding the zeros of f. Setting $f(x) = 0$ yields

$$\sin(2x) - \cos(x) = 0 \quad \Leftrightarrow \quad 2\sin(x)\cos(x) - \cos(x) = 0 \quad \Leftrightarrow \quad \cos(x)(2\sin(x) - 1) = 0\,.$$

From $\cos(x) = 0$, we get $x = \frac{\pi}{2} + \pi k$ for integers k of which only $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$ lie in $[0, 2\pi[$. For $\sin(x) = \frac{1}{2}$ we get $x = \frac{\pi}{6} + 2\pi k$ or $x = \frac{5\pi}{6} + 2\pi k$ for integers k. Of those, only $x = \frac{\pi}{6}$ and $x = \frac{5\pi}{6}$ lie in $[0, 2\pi[$. Next, we choose our test values. For $x = 0$ we find $f(0) = -1$; when $x = \frac{\pi}{4}$ we get $f\left(\frac{\pi}{4}\right) = 1 - \frac{\sqrt{2}}{2} = \frac{2-\sqrt{2}}{2}$; for $x = \frac{3\pi}{4}$ we get $f\left(\frac{3\pi}{4}\right) = -1 + \frac{\sqrt{2}}{2} = \frac{\sqrt{2}-2}{2}$; when $x = \pi$ we have $f(\pi) = 1$, and lastly, for $x = \frac{7\pi}{4}$ we get $f\left(\frac{7\pi}{4}\right) = -1 - \frac{\sqrt{2}}{2} = \frac{-2-\sqrt{2}}{2}$. The leads to the following sign diagram:

$\sin(2x) - \cos(x)$	$-$	$\frac{\pi}{6}$	$+$	$\frac{\pi}{2}$	$-$	$\frac{5\pi}{6}$	$+$	$\frac{3\pi}{2}$	$-$

We see $f(x) > 0$ on $]\frac{\pi}{6}, \frac{\pi}{2}[\cup]\frac{5\pi}{6}, \frac{3\pi}{2}[$, so this is our answer.

Our next example puts solving equations and inequalities to good use, namely for finding domains of functions.

Example 5.20

Express the domain of the following functions using extended interval notation.

1. $f(x) = \csc\left(2x + \dfrac{\pi}{3}\right)$
2. $g(x) = \sqrt{1 - \cot(x)}$

Solution

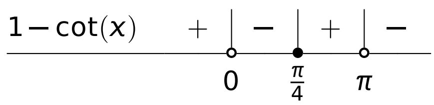
1. We rewrite f in terms of sine as $f(x) = \frac{1}{\sin\left(2x+\frac{\pi}{3}\right)}$. Since the sine function is defined everywhere, our only concern comes from zeros in the denominator. Solving $\sin\left(2x + \frac{\pi}{3}\right) = 0$, we get for integers k:

$$\begin{aligned} 2x + \frac{\pi}{3} &= k\pi \\ \Longleftrightarrow \qquad x &= -\frac{\pi}{6} + \frac{\pi}{2}k\,. \end{aligned}$$

In set-builder notation, our domain is $\left\{x : x \neq -\frac{\pi}{6} + \frac{\pi}{2}k\,, \forall k \in \mathbb{Z}\right\}$. If we now let x_k denote the kth number excluded from the domain, we have $x_k = -\frac{\pi}{6} + \frac{\pi}{2}k = \frac{(3k-1)\pi}{6}$ for integers k. The intervals which comprise the domain are of the form $\left]x_k, x_{k+1}\right[= \left]\frac{(3k-1)\pi}{6}, \frac{(3k+2)\pi}{6}\right[$ as k runs through the integers. Using extended interval notation, we have that the domain is

$$\bigcup_{k=-\infty}^{+\infty} \left]\frac{(3k-1)\pi}{6}, \frac{(3k+2)\pi}{6}\right[\,.$$

2. We first note that, due to the presence of the $\cot(x)$ term, $x \neq \pi k$ for integers k. Next, we recall that for the square root to be defined, we need $1 - \cot(x) \geq 0$. Our strategy is to solve this inequality over $]0,\pi[$, i.e. interval which generates a fundamental cycle of cotangent, and then add integer multiples of the period, in this case, π. We let $r(x) = 1 - \cot(x)$ and set about making a sign diagram for r over the interval $]0,\pi[$ to find where $r(x) \geq 0$. We note that r is undefined for $x = \pi k$ for integers k, in particular, at the endpoints of our interval $x = 0$ and $x = \pi$. Next, we look for the zeros of r. Solving $r(x) = 0$, we get $\cot(x) = 1$ or $x = \frac{\pi}{4} + \pi k$ for integers k and only one of these, $x = \frac{\pi}{4}$, lies in $]0,\pi[$. Choosing the test values $x = \frac{\pi}{6}$ and $x = \frac{\pi}{2}$, we get $r\left(\frac{\pi}{6}\right) = 1 - \sqrt{3}$, and $r\left(\frac{\pi}{2}\right) = 1$, and the following sign diagram.



We find $g(x) \geq 0$ on $\left[\frac{\pi}{4}, \pi\right[$. Adding multiples of the period we get our solution to consist of the intervals

$$\left[\frac{\pi}{4} + \pi k, \pi + \pi k\right[= \left[\frac{(4k+1)\pi}{4}, (k+1)\pi\right[,$$

or more briefly using extended interval notation:

$$\bigcup_{k=-\infty}^{\infty} \left[\frac{(4k+1)\pi}{4}, (k+1)\pi\right[\,.$$

We close this section with an example that demonstrates how to solve equations and inequalities involving inverse trigonometric functions.

Example 5.21

Solve the following equations and inequalities analytically.

1. $\arcsin(2x) = \dfrac{\pi}{3}$
2. $4\arctan^2(x) - 3\pi\arctan(x) - \pi^2 = 0$
3. $\pi^2 - 4\arccos^2(x) < 0$

Solution

1. We first note that $\frac{\pi}{3}$ is in the range of the arcsine function, so a solution exists! Next, we exploit the inverse property of sine and arcsine

$$\begin{aligned} \arcsin(2x) &= \frac{\pi}{3} \\ \Rightarrow \quad \sin\big(\arcsin(2x)\big) &= \sin\left(\frac{\pi}{3}\right) \\ \Leftrightarrow \quad 2x &= \frac{\sqrt{3}}{2} \qquad \text{(Since } \sin(\arcsin(u)) = u.\text{)} \\ \Leftrightarrow \quad x &= \frac{\sqrt{3}}{4}. \end{aligned}$$

2. With the presence of both $\arctan^2(x)$ ($= (\arctan(x))^2$) and $\arctan(x)$, we substitute $u = \arctan(x)$. The equation becomes

$$\begin{aligned} 4u^2 - 3\pi u - \pi^2 = 0 \quad &\Leftrightarrow \quad (4u+\pi)(u-\pi) = 0 \\ &\Leftrightarrow \quad u = \arctan(x) = -\frac{\pi}{4} \quad \vee \quad u = \arctan(x) = \pi. \end{aligned}$$

Since $-\frac{\pi}{4}$ is in the range of arctangent, but π is not, we only get solutions from the first equation. To solve for x, we may write

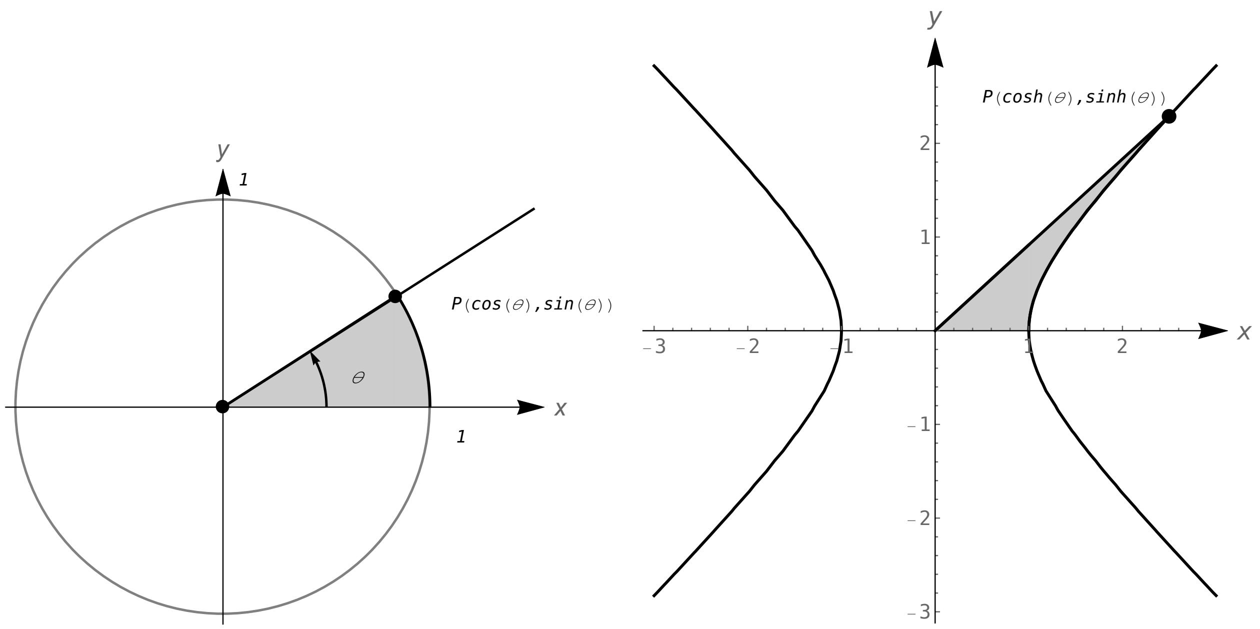
$$\begin{aligned} \arctan(x) &= -\frac{\pi}{4} \\ \Rightarrow \quad \tan(\arctan(x)) &= \tan\left(-\frac{\pi}{4}\right) \\ \Leftrightarrow \quad x &= -1. \qquad \text{(Since } \tan(\arctan(u)) = u.\text{)} \end{aligned}$$

3. Since the inverse trigonometric functions are continuous on their domains, we can solve inequalities featuring these functions using sign diagrams. Since all of the nonzero terms of $\pi^2 - 4\arccos^2(x) < 0$ are on one side of the inequality, we let $f(x) = \pi^2 - 4\arccos^2(x)$ and note that the domain of f is limited by the $\arccos(x)$ to $[-1,1]$. Next, we find the zeros of f by setting $f(x) = \pi^2 - 4\arccos^2(x) = 0$. We get $\arccos(x) = \pm\frac{\pi}{2}$, and since the range of arccosine is $[0,\pi]$, we focus our attention on $\arccos(x) = \frac{\pi}{2}$. Consequently, we get $x = \cos\left(\frac{\pi}{2}\right) = 0$ as our only zero. Hence, we have two test intervals, $[-1,0[$ and $]0,1]$. Choosing test values $x = \pm 1$, we get $f(-1) = -3\pi^2 < 0$ and $f(1) = \pi^2 > 0$, which leads us to conclude that $f(x) < 0$ on $[-1,0[$ and $f(x) > 0$ on $]0,1]$. So, our answer is $[-1,0[$.

5.5 Hyperbolic functions

5.5.1 Hyperbolic cosine and sine

The **hyperbolic functions** (*hyperbolische functies*) are a set of functions that have many applications to mathematics, physics, and engineering. Among many other applications, they are used to describe the formation of satellite rings around planets, to describe the shape of a rope hanging from two points, and they are used in the theory of special relativity. These functions are sometimes referred to as the hyperbolic trigonometric functions as there are many, many connections between them and the standard trigonometric functions. Figure 5.26 demonstrates one such connection. Just as cosine and sine are used to define points on the circle defined by $x^2+y^2=1$, the functions hyperbolic cosine and hyperbolic sine are used to define points on the hyperbola $x^2-y^2=1$. More specifically, for a given θ, the area of region enclosed by the x-axis, the latter hyperbola and the ray connecting the origin and the point P on this hyperbola is nothing else than $\theta/2$.



(a) (b)

Figure 5.26: Using trigonometric functions to define points on a circle (a) and hyperbolic functions to define points on a hyperbola (b).

We start with a formal definition of the hyperbolic cosine and sine.

Definitie 5.3 (Hyperbolic cosine and sine)
For all real values x, **the hyperbolic cosine** (*cosinus hyperbolicus*) and **sine** (*sinus hyperbolicus*) are given by

$$\cosh(x) = \frac{e^x + e^{-x}}{2}, \tag{5.29}$$

and

$$\sinh(x) = \frac{e^x - e^{-x}}{2}, \tag{5.30}$$

respectively.

We can use our knowledge of the graphs of e^x and e^{-x} to sketch the graph of $\cosh(x)$. First, let us calculate the value of $\cosh(0)$. When $x=0$, we have $e^x=1$ and $e^{-x}=1$, so $\cosh(0)=1$. As x gets larger, e^x increases quickly, but e^{-x} decreases. Consequently, the term $\frac{e^{-x}}{2}$ in Equation (5.29) gets very small . Therefore, as x gets larger, $\cosh(x)$ gets closer to $\frac{e^x}{2}$. Yet the graph of $\cosh(x)$ will always stay above the graph of $\frac{e^x}{2}$ because even though $\frac{e^{-x}}{2}$ gets very small, it is always greater than zero. Now, consider negative x. In that case, as x becomes more negative e^{-x} increases quickly, but e^x decreases quickly, so $\cosh(x)$ gets closer and closer to $\frac{e^{-x}}{2}$. Yet again the graph of $\cosh(x)$ will always stay above the one of $\frac{e^{-x}}{2}$. We sketch the resulting graph of $\cosh(x)$ in Figure 5.27(a) together with the graphs of $\frac{e^x}{2}$ and $\frac{e^{-x}}{2}$.

Notice that the graph of $\cosh(x)$ is symmetric about the y-axis because $\cosh(x) = \cosh(-x)$, so $\cosh(x)$ is an even function. Besides, given the properties of the involved exponentials, the domain of this function is $\mathbb{R}$, whereas its range is $[\,1, +\infty\,[$. A similar reasoning can be followed to arrive at the graph of $\sinh(x)$ (Figure 5.27(b)). Notice that the graph of this hyperbolic function is symmetric with respect to the origin, so the function $\sinh(x)$ is an odd function. Besides, both the domain and the range of this function are $\mathbb{R}$.

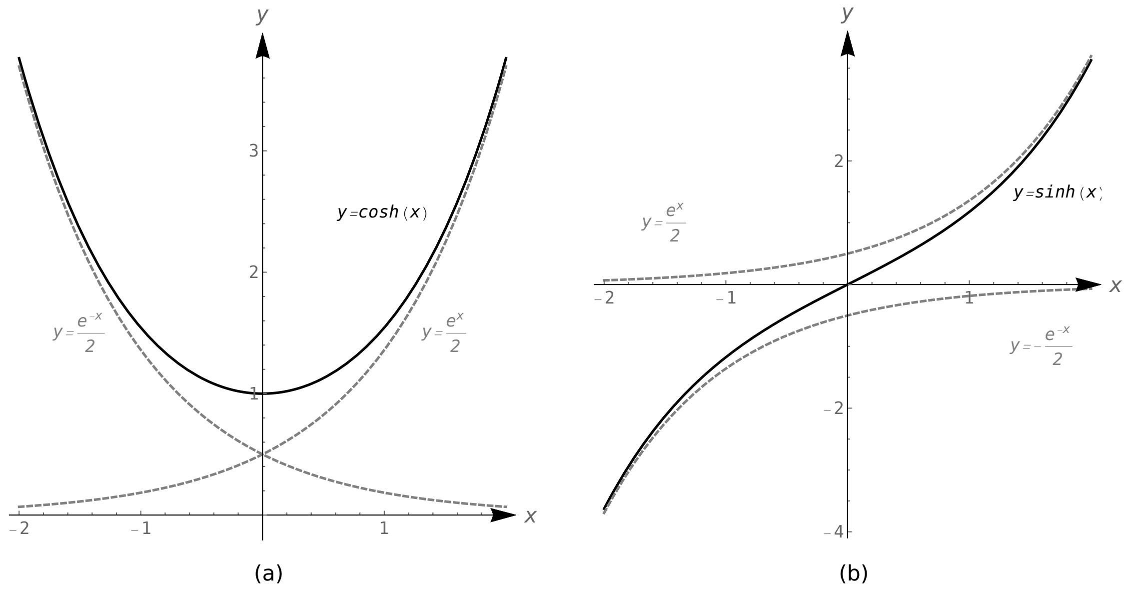


Figure 5.27: The graph of the hyperbolic cosine together with the graphs of $y = \frac{e^x}{2}$ and $\frac{e^{-x}}{2}$ (a) and the graph of the hyperbolic sine together with the graphs of $y = \frac{e^x}{2}$ and $-\frac{e^{-x}}{2}$ (b).

Example 5.22

Use the definition of the hyperbolic cosine and sine to rewrite the following expressions.

1. $\cosh^2(x) - \sinh^2(x)$
2. $2\cosh(x)\sinh(x)$

Solution

1. Direct application of Equations (5.29) and (5.30) yields

$$\begin{aligned}\cosh^2(x) - \sinh^2(x) &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\ &= \frac{e^{2x} + 2e^x e^{-x} + e^{-2x}}{4} - \frac{e^{2x} - 2e^x e^{-x} + e^{-2x}}{4}\end{aligned}$$

$$= \frac{4}{4} = 1.$$

So,

$$\cosh^2(x) - \sinh^2(x) = 1\,.$$

2. Using Equations (5.29) and (5.30), we immediately find

$$\begin{aligned} 2\cosh(x)\sinh(x) &= 2\left(\frac{e^x + e^{-x}}{2}\right)\left(\frac{e^x - e^{-x}}{2}\right) \\ &= 2\cdot\frac{e^{2x} - e^{-2x}}{4} \\ &= \frac{e^{2x} - e^{-2x}}{2} = \sinh(2x). \end{aligned}$$

Thus, $2\cosh(x)\sinh(x) = \sinh(2x)$.

The previous example indicated the remarkable resemblance between the properties of the hyperbolic cosine and sine and their trigonometric counterparts. This resemblance is echoed by **Osborne's rule** (*Regel van Osborne*), which states that $\cos(x)$ appearing in a trigonometric identity should be replaced by $\cosh(x)$, while every $\sin(x)$ should be replaced by $i\sinh(x)$, where i is the imaginary unit. In that way, we arrive at the hyperbolic counterparts of the sum and difference identities for cosine and sine (Theorems 5.7 and 5.9):

$$\cosh(x \pm y) = \cosh(x)\cosh(y) \pm \sinh(x)\sinh(y)\,, \tag{5.31}$$

$$\sinh(x \pm y) = \sinh(x)\cosh(y) \pm \cosh(x)\sinh(y)\,. \tag{5.32}$$

Likewise, we find as hyperbolic counterpart for the double-angle identities (Theorem 5.11):

$$\cosh(2x) = \cosh^2(x) + \sinh^2(x)\,, \tag{5.33}$$

$$\sinh(2x) = 2\sinh(x)\cosh(x)\,, \tag{5.34}$$

and for the power-reduction formulas (Theorem 5.12):

$$\cosh^2(x) = \frac{1 + \cosh(2x)}{2}\,, \tag{5.35}$$

$$\sinh^2(x) = \frac{-1 + \cosh(2x)}{2}\,. \tag{5.36}$$

Of course, there are also hyperbolic product to sum formulas (Theorem 5.13), and vice versa (Theorem 5.14):

$$\cosh(x)\cosh(y) = \frac{1}{2}\left[\cosh(x-y) + \cosh(x+y)\right]\,, \tag{5.37}$$

$$\sinh(x)\sinh(y) = \frac{1}{2}\left[\cosh(x+y) - \cosh(x-y)\right]\,, \tag{5.38}$$

$$\sinh(x)\cosh(y) = \frac{1}{2}\left[\sinh(x-y)+\sinh(x+y)\right], \tag{5.39}$$

$$\cosh(x)+\cosh(y) = 2\cosh\left(\frac{x+y}{2}\right)\cosh\left(\frac{x-y}{2}\right), \tag{5.40}$$

$$\cosh(x)-\cosh(y) = 2\sinh\left(\frac{x+y}{2}\right)\sinh\left(\frac{x-y}{2}\right), \tag{5.41}$$

$$\sinh(x)\pm\sinh(y) = 2\sinh\left(\frac{x\pm y}{2}\right)\cosh\left(\frac{x\mp y}{2}\right). \tag{5.42}$$

5.5.2 The other hyperbolic functions

Just as the cosine and sine functions can be combined to obtain four more trigonometric functions, the hyperbolic cosine and sine can be combined to obtain the **hyperbolic tangent, cotangent, secant and cosecant** (*tangens, cotangens, secans en cosecans hypberbolicus*):

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}, \tag{5.43}$$

$$\coth(x) = \frac{\cosh(x)}{\sinh(x)}, \tag{5.44}$$

$$\operatorname{sech}(x) = \frac{1}{\cosh(x)}, \tag{5.45}$$

$$\operatorname{csch}(x) = \frac{1}{\sinh(x)}. \tag{5.46}$$

The graphs of those four additional hyperbolic functions are given in Figure 5.28. Notice the domains of $\tanh(x)$ and $\operatorname{sech}(x)$ are $\mathbb{R}$, whereas both $\coth(x)$ and $\operatorname{csch}(x)$ have vertical asymptotes at $x=0$. Also note the ranges of these functions, especially $\tanh(x)$: as $x \to +\infty$, both $\sinh(x)$ and $\cosh(x)$ approach $e^x/2$, hence $\tanh(x)$ approaches 1. Table 5.6 summarizes the domains, ranges and symmetries of the six hyperbolic functions.

Table 5.6: Domains, ranges and symmetries of the six hyperbolic functions.

Function	Domain	Range	Symmetry
$\cosh x$	$\mathbb{R}$	$[1,+\infty[$	even
$\sinh x$	$\mathbb{R}$	$\mathbb{R}$	odd
$\tanh x$	$\mathbb{R}$	$]-1,1[$	odd
$\coth x$	$\mathbb{R}_0$	$]-\infty,-1[\,\cup\,]1,+\infty[$	odd
$\operatorname{sech} x$	$\mathbb{R}$	$]0,1]$	even
$\operatorname{csch} x$	$\mathbb{R}_0$	$\mathbb{R}_0$	odd

Once more relying on Osborne's rule, we can formulate the hyperbolic counterparts of the Pythagorean identities given by (Equations (5.8) and (5.9))

$$\tanh^2(x)+\operatorname{sech}^2(x) = 1, \tag{5.47}$$

and

$$\coth^2(x)-\operatorname{csch}^2(x) = 1. \tag{5.48}$$

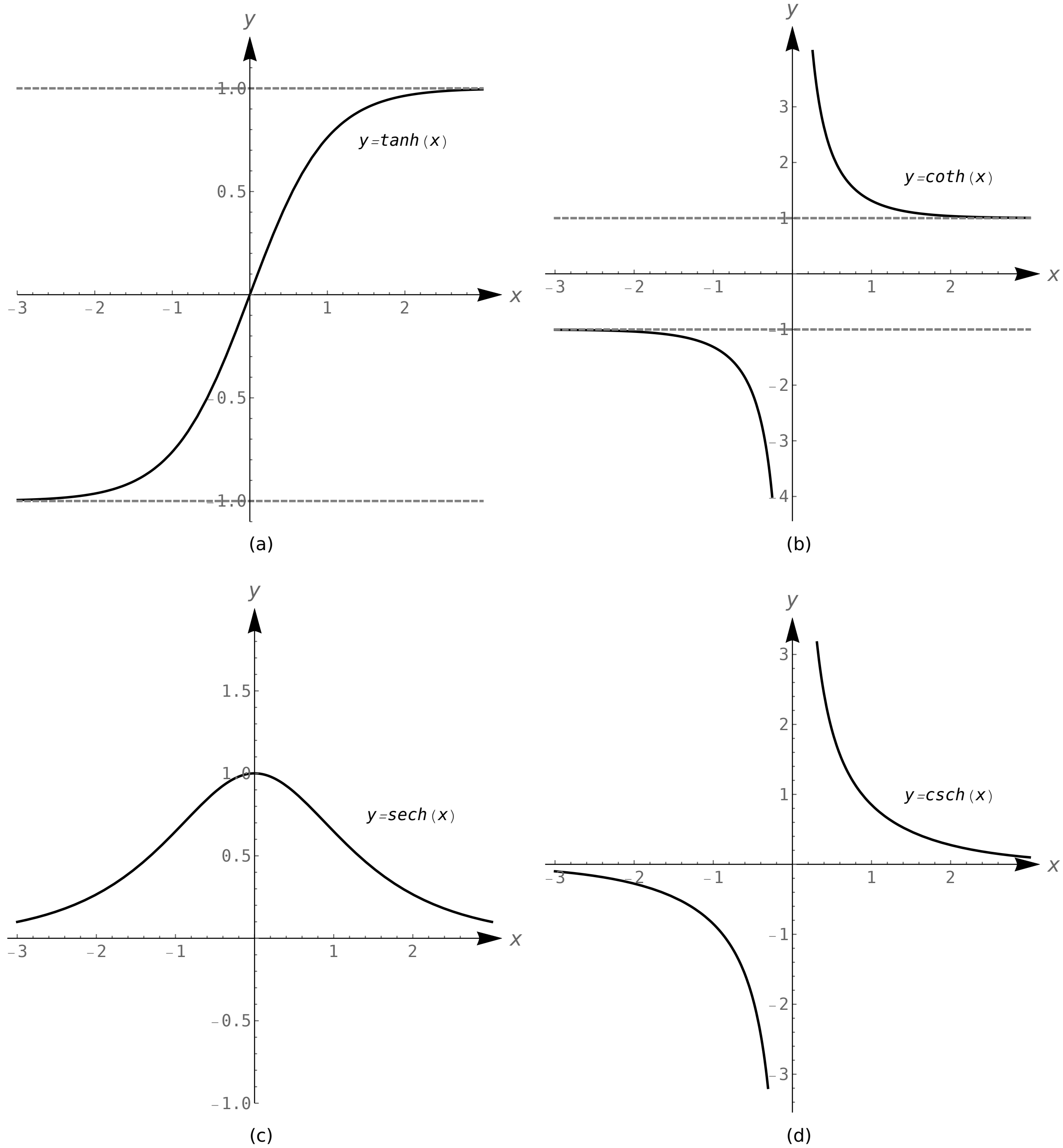


Figure 5.28: The graph of $y = \tanh(x)$ (a), $y = \coth(x)$ (b), $y = \operatorname{sech}(x)$ (c) and $y = \operatorname{csch}(x)$ (d).

Likewise, we can find hyperbolic counterparts to the sum, difference and double-angle identities for tangent:

$$\tanh(x \pm y) = \frac{\tanh(x) \pm \tanh(y)}{1 \pm \tanh(x)\tanh(y)}, \tag{5.49}$$

$$\tanh(2x) = \frac{2\tanh(x)}{1 + \tanh^2(x)}. \tag{5.50}$$

Hyperbolic functions are especially useful to describe, for instance, the shape of the curve a hanging chain or cable assumes under its own weight when supported only at its ends (Figure 5.29(a)). This curve is known in engineering as the **catenary** (*kettinglijn*), but catenaries can as well be found in nature. For instance, the silk on a spider's web forms multiple catenaries (Figure 5.29(b)). Any catenary can be described by the equation

$$y = a\cosh\left(\frac{x}{a}\right),$$

where a is a problem-specific parameter that effectuates a uniform scaling of the curve.

5.5.3 Inverse hyperbolic functions

From Figures 5.27 and 5.28 it is clear that both the hyperbolic cosine and secant function are non-injective on their domains. As such, we proceed in line with the approach we followed to find the inverse of the non-injective trigonometric functions by restricting the domain of the hyperbolic cosine and secant functions. Table 5.7 shows the resulting domains and ranges of the inverse hyperbolic functions. Since for a given value of a hyperbolic function, the corresponding inverse hyperbolic function provides the corresponding hyperbolic angle whose size equals twice the area of the corresponding sector of the unit hyperbola $x^2 - y^2 = 1$, the prefix '-ar' is typically used to denote the inverse hyperbolic functions, e.g. $\operatorname{arcosh}(x)$, $\operatorname{arsinh}(x)$,...

Table 5.7: Domains and ranges of the four inverse hyperbolic functions.

Function	Domain	Range
$\operatorname{arcosh}(x)$	$[\,1,+\infty\,[$	$[\,0,+\infty\,[$
$\operatorname{arsinh}(x)$	$\mathbb{R}$	$\mathbb{R}$
$\operatorname{artanh}(x)$	$]-1,1\,[$	$\mathbb{R}$
$\operatorname{arcoth}(x)$	$]-\infty,-1\,[\,\cup\,]\,1,+\infty\,[$	$\mathbb{R}_0$

The graphs of four inverse hyperbolic functions are shown in Figure 5.30.

Because the hyperbolic functions are defined in terms of exponential functions, their inverses can be expressed in terms of logarithms. For instance, let us try to express the $\operatorname{arcosh}(x)$ in terms of logarithms. So, starting with

$$y = \frac{e^x + e^{-x}}{2},$$

for $x \geq 0$, we first of all swap x and y to get the equation

$$x = \frac{e^y + e^{-y}}{2},$$

which we now should try to solve for $y \geq 0$. Multiplying both sides of this equation by $2e^y$ yields

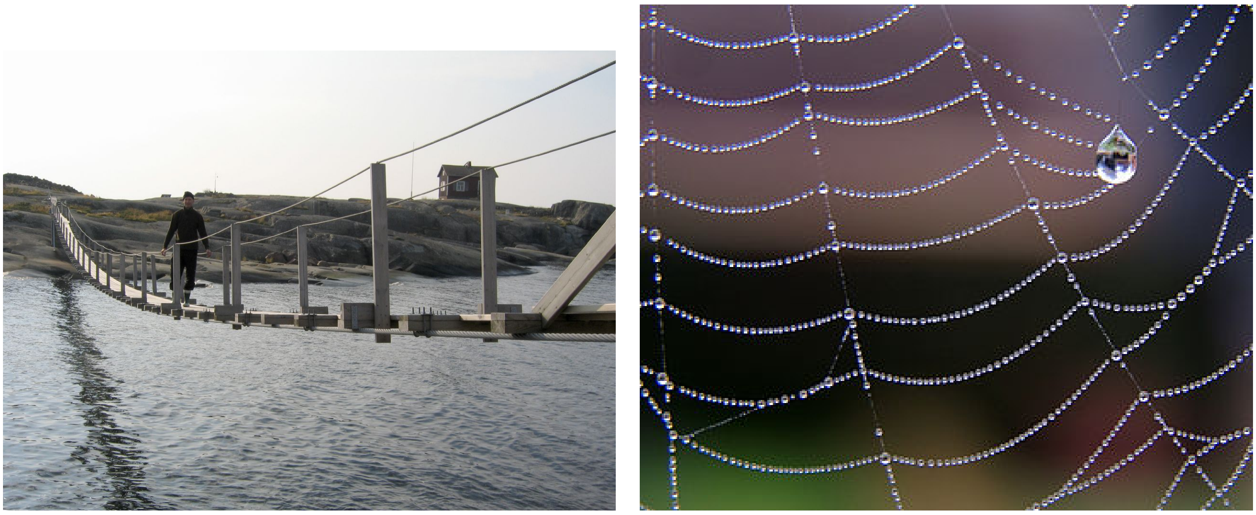
$$\begin{aligned} 2e^y x &= e^{2y} + 1 \\ \Leftrightarrow \quad (e^y)^2 - 2e^y x + 1 &= 0\,. \end{aligned}$$

If we solve the underlying quadratic equation in e^y, we obtain

$$e^y = \frac{2x \pm \sqrt{(2x)^2 - 4}}{2} = x \pm \sqrt{x^2 - 1}\,.$$

The solution $x - \sqrt{x^2-1}$ is, however, extraneous because it is less than 1, whereas e^y is greater than 1 for all $y \geq 0$. Hence, we arrive at

$$y = \ln\left(x + \sqrt{x^2 - 1}\right) = \operatorname{arcosh}(x)\,.$$



(a) (b)

Figure 5.29: Curves from the real world described by a hyperbolic cosine function: suspension bridge (a) and silk threads on a spider's web(b).

Similar arguments apply to each of the inverse hyperbolic functions, so that we get:

$$\operatorname{arcosh}(x) = \ln\left(x+\sqrt{x^2-1}\right), \quad x \geq 1, \tag{5.51}$$

$$\operatorname{arsinh}(x) = \ln\left(x+\sqrt{x^2+1}\right), \tag{5.52}$$

$$\operatorname{artanh}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right), \quad |x| < 1, \tag{5.53}$$

$$\operatorname{arcoth}(x) = \frac{1}{2}\ln\left(\frac{x+1}{x-1}\right), \quad |x| > 1, \tag{5.54}$$

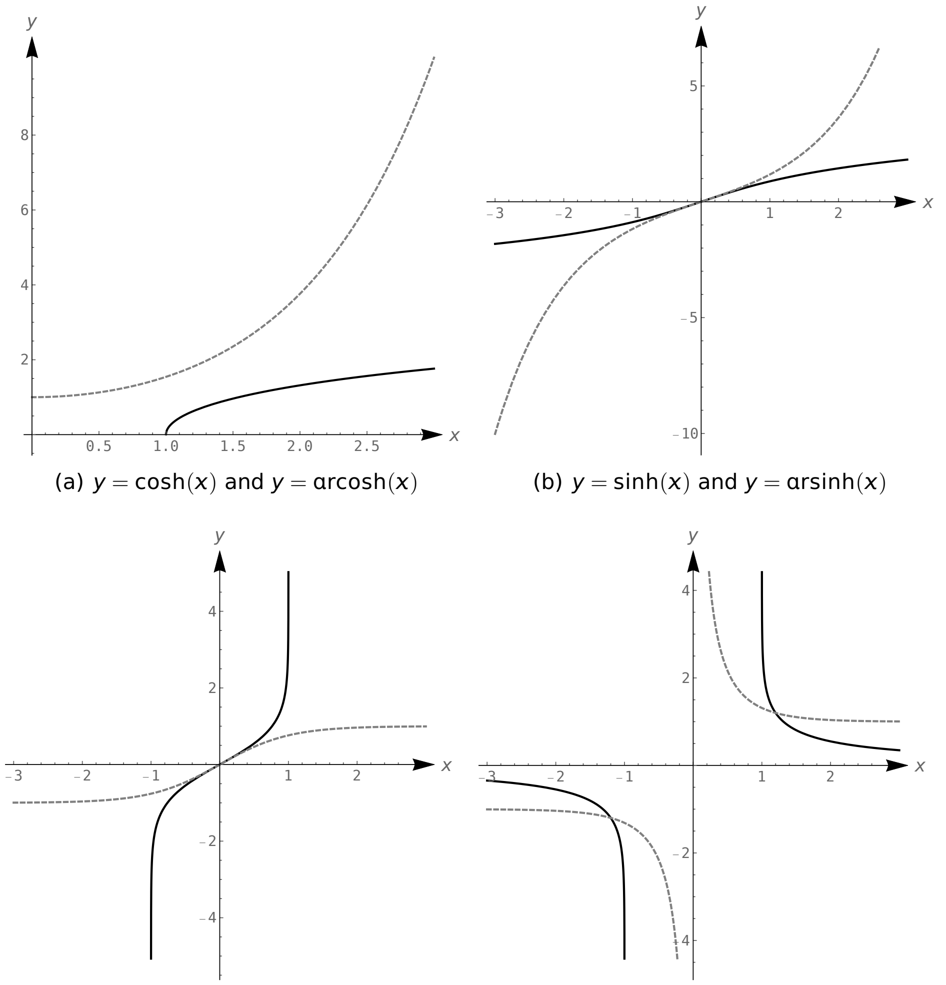
$$\tag{5.55}$$

Earth's albedo

The surface albedo α [–] of planet Earth is defined as the ratio of irradiance reflected to the irradiance received by a surface. The former is not only determined by properties of the surface itself, but also by the spectral and angular distribution of solar radiation reaching the Earth's surface. The higher the albedo, the more irradiance is reflected, which tempers the surface heating. For instance, a typical abedo of icy surfaces is 0.8-0.9. Since temperature affects the area of the Earth's surface covered by ice, it has a direct impact on the Earth's albedo. This effect can be described functionally using the following equation

$$\alpha = 0.5 - 0.2\tanh\left(\frac{T-265}{10}\right),$$

where T is the temperature in degrees Kelvin. This illustrates the value of the hyperbolic tangent function in describing environmental relationships.



(a) $y = \cosh(x)$ and $y = \operatorname{arcosh}(x)$ (b) $y = \sinh(x)$ and $y = \operatorname{arsinh}(x)$

(c) $y = \tanh(x)$ and $y = \operatorname{artanh}(x)$ (d) $y = \coth(x)$ and $y = \operatorname{arcoth}(x)$

Figure 5.30: The graph of the hyperbolic functions (dashed) and their inverses (solid).

5.6 Conic sections continued

5.6.1 Translation of axes

In Section 4.4, we introduced the standard equations of conic sections. The careful reader might have noticed that, for instance, the standard equation of a circle R with radius r and centre in (x_0, y_0) (Equation (4.4)) dictates a shift of the circle

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

with centre in (0,0) x_0 units to the right along the x-axis and y_0 units up along the y-axis. In other words, the studied circle R is just a translated version of a circle with the same radius r but centred at the origin. Besides, upon expanding Equation (4.4), i.e.

$$x^2 - 2x_0x + x_0^2 + y^2 - 2yy_0 + y_0^2 = r^2$$

we see that linear terms in x and y appear in the left-hand side of the equation.

These two facts hint that we might in general be able to get rid of the linear terms in the quadratic equation

$$ax^2 + cy^2 + dx + ey + f = 0 \tag{5.56}$$

by introducing a change of variables; that is by letting

$$\begin{cases} \tilde{x} &= x - x_0 \\ \tilde{y} &= y - y_0 \,, \end{cases}$$

where x_0 and y_0 should be determined in such a way that the squares in Equation (5.56) can be completed, and hence the linear terms vanish. Basically, this change of variables is equivalent to a shift of the original xy-Cartesian coordinate system to an $\tilde{x}\tilde{y}$-Cartesian coordinate system in which the $\tilde{x}$-axis is parallel to the x-axis and x_0 units away, and the $\tilde{y}$- axis is parallel to the y-axis and y_0 units away. This means that the origin $\tilde{o}$ of the new coordinate system has coordinates (x_0, y_0) in the original system (Figure 5.31).

In order to find a suitable change of variables for Equation (5.56), let us first rewrite this equation as

$$a\left(x^2 + \frac{d}{a}x\right) + c\left(y^2 + \frac{e}{c}y\right) + f = 0\,,$$

so that the squares can be completed

$$a\left(\left(x + \frac{d}{2a}\right)^2 - \frac{d^2}{4a^2}\right) + c\left(\left(y + \frac{e}{2c}\right)^2 - \frac{e^2}{4c^2}\right) + f = 0\,.$$

Working out the outer parentheses gives

$$a\left(x + \frac{d}{2a}\right)^2 + c\left(y + \frac{e}{2c}\right)^2 - \frac{d^2}{4a} - \frac{e^2}{4c} + f = 0\,,$$

which, by dividing it by ac and introducing the following change of variables

$$\begin{cases} \tilde{x} &= x + \dfrac{d}{2a} \\ \tilde{y} &= y + \dfrac{e}{2c}\,, \end{cases}$$

can be recast in the standard form of a conic section:

$$\frac{\tilde{x}^2}{c} + \frac{\tilde{y}^2}{a} = \frac{d^2}{4a^2c} + \frac{e^2}{4c^2a} - \frac{f}{ac}\,. \tag{5.57}$$

Example 5.23

Rewrite the following quadratic equation in standard form by eliminating the linear terms in x and y.

$$3x^2 - 7y^2 + 12\sqrt{2}x - 28\sqrt{2}y - 30 = 0$$

Then, draw the corresponding conic section.

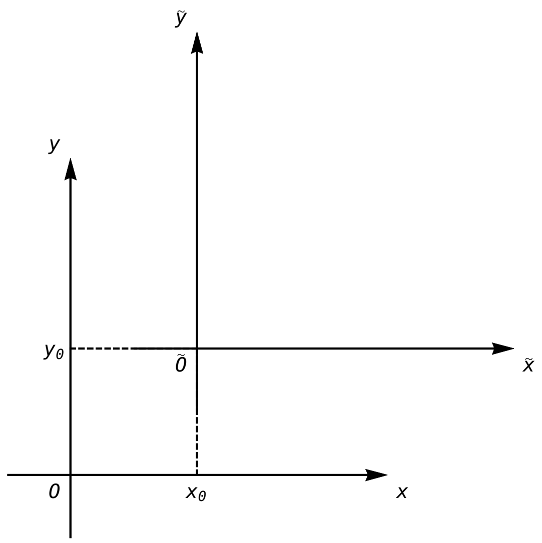


Figure 5.31: Shift of the xy- Cartesian coordinate system to a system $\tilde{x}\tilde{y}$ with origin $\tilde{o}$ with coordinates (x_0, y_0) in the xy-system.

Solution

The linear terms can be eliminated by completing the squares in x and y as follows.

$$\begin{aligned} & 3x^2-7y^2+12\sqrt{2}x-28\sqrt{2}y-30=0 \\ \Leftrightarrow\ & 3(x^2+4\sqrt{2}x)-7(y^2+4\sqrt{2}y)-30=0 \\ \Leftrightarrow\ & 3(x^2+4\sqrt{2}x+8)-7(y^2+4\sqrt{2}y+8)=30+24-56 \\ \Leftrightarrow\ & 3(x+2\sqrt{2})^2-7(y+2\sqrt{2})^2=-2 \end{aligned}$$

Now, let

$$\begin{cases} \tilde{x}=x+2\sqrt{2} \\ \tilde{y}=y+2\sqrt{2} \end{cases}$$

then we get

$$3\tilde{x}^2-7\tilde{y}^2=-2\,,$$

or equivalently

$$-\frac{3\tilde{x}^2}{2}+\frac{7\tilde{y}^2}{2}=1\,.$$

This is the standard equation of a vertical hyperbola centred in $(-2\sqrt{2},-2\sqrt{2})$.

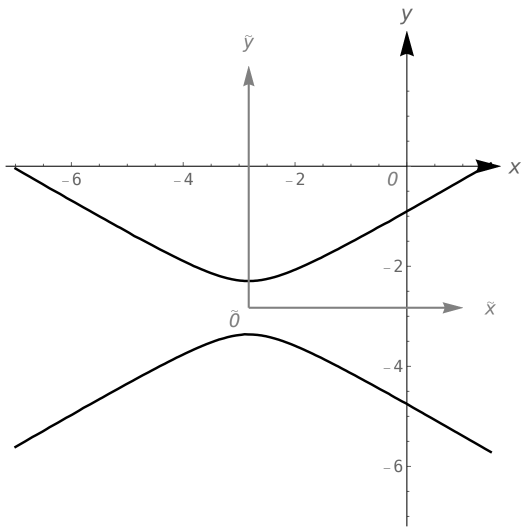


Figure 5.32: Graph of $3x^2 - 7y^2 + 12\sqrt{2}x - 28\sqrt{2}y - 30 = 0$, together with the original and translated coordinate axes.

5.6.2 Rotation of axes

We have just seen that a translation of axes can be used to eliminate linear terms in x and/or y in quadratic, but what about the term in xy that might appear in the general quadratic equation

$$ax^2 + bxy + cy^2 + dx + ey + f = 0\,? \tag{5.58}$$

We shall see that the term bxy can be eliminated by rotating the coordinate axes. For that purpose, we need the trigonometric fuctions that were introduced in Section 5.3.

In Figure 5.33(a) the x- and y-axes have been rotated about the origin through an acute angle θ to produce the $\tilde{x}$- and $\tilde{y}$-axes. Thus, a given point P has coordinates (x, y) in the first coordinate system and $(\tilde{x}, \tilde{y})$ in the new coordinate system. To see how $\tilde{x}$ and $\tilde{y}$ are related to x and y we observe from Figure 5.33(b) that

$$\begin{cases} \tilde{x} &= r\cos(\phi) \\ x &= r\cos(\theta + \phi)\,, \end{cases}$$

and likewise

$$\begin{cases} \tilde{y} &= r\sin(\phi) \\ y &= r\sin(\theta + \phi)\,, \end{cases}$$

The addition formula for the cosine then gives

$$\begin{aligned} x &= r\cos(\theta+\phi) = r\big(\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)\big) \\ &= \big(r\cos(\phi)\big)\cos(\theta) - \big(r\sin(\phi)\big)\sin(\theta) \\ &= \tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\,. \end{aligned}$$

A similar computation gives y in terms of $\tilde{x}$ and $\tilde{y}$, so that we arrive at the following formulas relating x and y and $\tilde{x}$ and $\tilde{y}$:

$$\begin{cases} x &= \tilde{x}\cos(\theta) - \tilde{y}\sin(\theta) \\ y &= \tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\,. \end{cases} \tag{5.59}$$

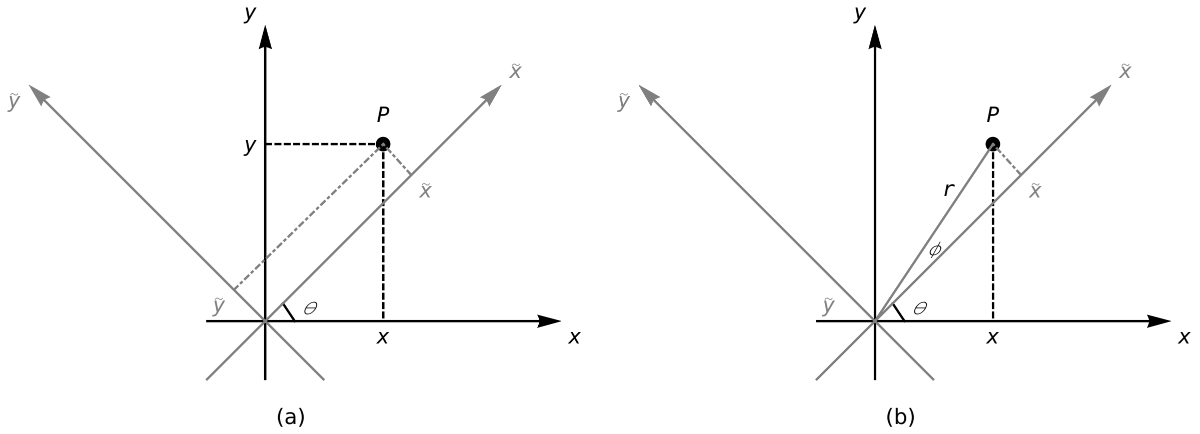


Figure 5.33: Rotation of axes.

Alternatively, these transformation formulas can be written in matrix notation as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix}.$$

By solving these equations for $\tilde{x}$ and $\tilde{y}$, we obtain formulas to determine the coordinates $\tilde{x}$ and $\tilde{y}$ of a point P that has coordinates x and y in the original Cartesian coordinate system:

$$\begin{cases} \tilde{x} &= x\cos(\theta) + y\sin(\theta) \\ \tilde{y} &= -x\sin(\theta) + y\cos(\theta). \end{cases} \tag{5.60}$$

Now let us try to determine an angle θ such that the term bxy in Equation (5.58) disappears when the axes are rotated through the angle θ. If we substitute the expressions from Equation (5.59) in Equation (5.58), we get

$$a\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big)^2 + b\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big)\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big) + c\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big)^2 \\ + d\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big) + e\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big) + f = 0 \tag{5.61}$$

Expanding and collecting terms, we obtain an equation of the form

$$\tilde{a}\tilde{x}^2 + \tilde{b}\tilde{x}\tilde{y} + \tilde{c}\tilde{y}^2 + \tilde{d}\tilde{x} + \tilde{e}\tilde{y} + f = 0,$$

where the coefficient $\tilde{b}$ of $\tilde{x}\tilde{y}$ is

$$\begin{aligned} \tilde{b} &= 2(c-a)\sin(\theta)\cos(\theta) + b\big(\cos^2(\theta) - \sin^2(\theta)\big) \\ &= (c-a)\sin(2\theta) + b\cos(2\theta). \end{aligned}$$

To eliminate the term $\tilde{x}\tilde{y}$ we choose θ in such a way that $\tilde{b} = 0$, i.e.

$$(a-c)\sin(2\theta) = b\cos(2\theta),$$

or

$$\cot(2\theta) = \frac{a-c}{b}. \tag{5.62}$$

Note that, after choosing θ in this way, we end up with an ellipse if $\tilde{a}$ and $\tilde{b}$ have the same sign, whereas we obtain a hyperbola if these coefficients have a different sign. Besides, when one of them is zero, we have a parabola.

Example 5.24

Identify and sketch the graph of

$$73x^2 + 72xy + 52y^2 + 30x - 40y - 75 = 0\,. \tag{5.63}$$

Solution

This equation is in the form of Equation (5.58) with $a = 73$, $b = 72$ and $c = 52$. Consequently, by choosing

$$\cot(2\theta) = \frac{a-c}{b} = \frac{73-52}{72} = \frac{7}{24}$$

we can eliminate the term $\tilde{x}\tilde{y}$ in the rotated counterpart of Equation (5.63). Using the Pythagorean theorem (Figure 5.34), we observe that

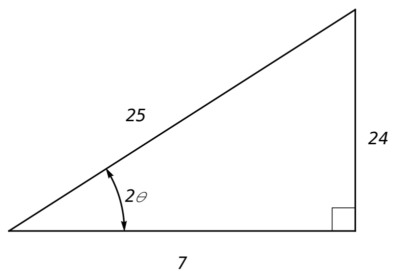


Figure 5.34: Using the Pythagorean theorem to compute 2θ.

$$\cos(2\theta) = \frac{7}{25}\,.$$

This observation allow us to compute $\cos(\theta)$ and $\sin(\theta)$ using the half-angle formulas:

$$\begin{aligned} \cos(\theta) &= \sqrt{\frac{1+\cos(2\theta)}{2}} = \sqrt{\frac{1+\frac{7}{25}}{2}} = \frac{4}{5} \\ \sin(\theta) &= \sqrt{\frac{1-\cos(2\theta)}{2}} = \sqrt{\frac{1-\frac{7}{25}}{2}} = \frac{3}{5}\,. \end{aligned}$$

Consequently, the rotation equations become

$$\begin{cases} x = \frac{4}{5}\tilde{x} - \frac{3}{5}\tilde{y} \\ y = \frac{3}{5}\tilde{x} + \frac{4}{5}\tilde{y}\,, \end{cases}$$

and substituting these in Equation (5.63), yields

$$73\left(\frac{4}{5}\tilde{x} - \frac{3}{5}\tilde{y}\right)^2 + 72\left(\frac{4}{5}\tilde{x} - \frac{3}{5}\tilde{y}\right)\left(\frac{3}{5}\tilde{x} + \frac{4}{5}\tilde{y}\right) + 52\left(\frac{3}{5}\tilde{x} + \frac{4}{5}\tilde{y}\right)^2$$
$$+ 30\left(\frac{4}{5}\tilde{x} - \frac{3}{5}\tilde{y}\right) - 40\left(\frac{3}{5}\tilde{x} + \frac{4}{5}\tilde{y}\right) - 75 = 0\,.$$

This simplifies to

$$4\tilde{x}^2 + \tilde{y}^2 - 2\tilde{y} = 3\,.$$

Completing the square in $\tilde{y}$ leads to:

$$\tilde{x}^2 + \frac{(\tilde{y}-1)^2}{4} = 1\,,$$

and we recognize this as being an ellipse with centre in $(0,1)$ with respect to the rotated coordinate axes $\tilde{x}$ and $\tilde{y}$ (Figure 5.35).

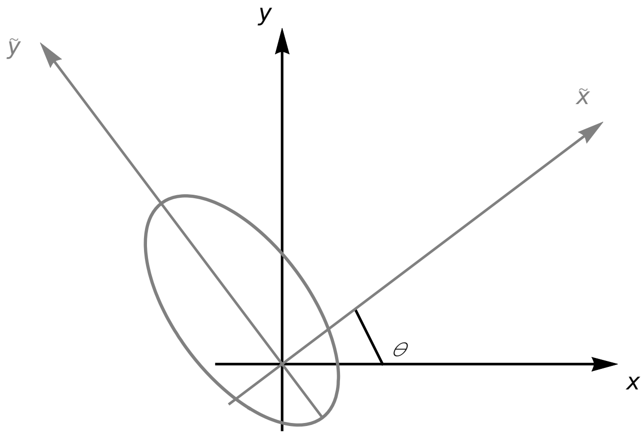


Figure 5.35: Graph of $73x^2 + 72xy + 52y^2 + 30x - 40y - 75 = 0$, together with the original and rotated coordinate axes.

Actually, we even do not have to translate and/or rotate the coordinate axes when we are merely interested in which conic section a given quadratic equations represents by relying on the following theorem.

Theorem 5.15 (Classification of conic sections)
Suppose the equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$ describes a non-degenerate conic section.

- *If $b^2 - 4ac > 0$ then the graph of the equation is a hyperbola.*
- *If $b^2 - 4ac = 0$ then the graph of the equation is a parabola.*
- *If $b^2 - 4ac < 0$ then the graph of the equation is an ellipse or circle.*

As you may expect, the quantity $b^2 - 4ac$ mentioned in Theorem 5.15 is called the **discriminant** (*discriminant*)

5.7 Exercises

Exponential and logarithmic functions

Assignment 5.1 — Simplify the expressions below.

(a) $\log_4\left(\frac{1}{8}\right)$

(b) $\log_{a^2}\left(a^3\right)$

(c) $e^{(3\ln(9))/2}$

(d) $\log_{\frac{1}{3}}\left(3^{2x}\right)$

(e) $\log_x\left(x\left(\log_y\left(y^2\right)\right)\right)$

(f) $\log_{15}(75)+\log_{15}(3)$

(g) $2\log_3(12)-4\log_3(6)$

(h) $2\ln(x)+5\ln(x-2)$

Assignment 5.2 — Prove the equalities below.

(a) $\log_a(b)\cdot\log_b(c)\cdot\log_c(a)=1$

(b) $\log_b(a)\cdot\log_c(b)=\log_c(a)$

(c) $\dfrac{1}{\log_a(x)}+\dfrac{1}{\log_b(x)}=\dfrac{1}{\log_{ab}(x)}$

(d) $\log_{a^n}(b^n)\cdot\log_{b^m}(a^m)=1$

(e) $\log_{b^n}(a^m)=\dfrac{m}{n}\log_b(a)$

Assignment 5.3 — Solve the equations below

(a) $3^{2x+1}=\sqrt[3]{3}$

(b) $4^{\frac{1}{x}}\cdot 16^{\frac{1}{x+2}}=64^{\frac{1}{x+1}}$

(c) $\log_4(x+4)-2\log_4(x+1)=\dfrac{1}{2}$

(d) $2\log_3(x)+\log_9(x)=10$

(e) $\dfrac{1}{2^x}=\dfrac{5}{8^{x+3}}$

(f) $(\ln x)^2+\ln(x^2)=\ln(x)$

(g) $4^{x-2}\cdot 8^{3x-1}=\sqrt{2}$

(h) $5^{2x-1}-3\cdot 5^{x+2}+6250=0$

(i) $\ln(x)-5\ln(2)=\ln(3x+84)+\ln(12)$

(j) $8^x+4^x=5\cdot 2^{x-4}$

(k) $5^{2x}+5^3=5^{x+2}+5^{x+1}$

(l) $4\cdot 8^{x-1}+1=4^x+2^{x-1}$

(m) $\dfrac{\log_4(x)}{\log_4(3)}\cdot\log_3\left(\dfrac{x}{9}\right)=3\cdot 27^{\log_3(2)}$

(n) $\log_3(x+4)+\log_3(x-2)=2\log_3(x)$

(o) $\log_2(3)\cdot\log_3(5)\cdot\log_5(7)\cdot\log_7(x)=4$

(p) $25\sqrt{0.1^x}=\dfrac{1}{2}10^{x+1}\sqrt{2.5}$

(q) $2(3^x+2^{x-1})=3(2^x-3^{x-1})$

(r) $2^{2x+1}-9\cdot 2^{x-1}+1=0$

(s) $4^{x-1}+2^{x-3}=2^{2x-3}+7$

(t) $\dfrac{1}{\log_{x-1}(5x-7)}+\dfrac{1}{\log_{x+1}(5x-7)}=\dfrac{1}{\log_x(x)}$

(u) $2\log_4(x)+\dfrac{1}{\log_{x-4}(2)}=2+\dfrac{1}{2\log_{x-3}(\sqrt{2})}$

Assignment 5.4 — Solve the inequalities below.

(a) $\ln(x^2-2) \leq \ln(x)$

(b) $9^{-x} < \dfrac{2+3^{x+1}}{3^x}$

(c) $\log\left(\dfrac{x(10-x)}{16}\right) < 0$

(d) $\log_{\frac{1}{3}}(4x) < \log_{\frac{1}{3}}(x-1)-2$

(e) $x\ln(x)-x > 0$

(f) $2^{3x-x^2} > \left(\dfrac{1}{8}\right)^{1-x}$

Assignment 5.5 — Solve the following systems of equations in $\mathbb{R}^2$.

(a) $\begin{cases} 9^x.3^y = 81 \\ \dfrac{2^x}{8^y} = \dfrac{1}{32} \end{cases}$

(b) $\begin{cases} 5^x = 4y \\ 2^{2x} = 5y \end{cases}$

(c) $\begin{cases} \log_{10}(x) + 3\log_{10^3}(y) = 2 \\ y^2 - 300 = 4x^2 \end{cases}$

Assignment 5.6 — Determine the domain and intersections with the x- and y-axis of the (logarithmic) functions below.

(a) $f(x) = \log\left(\dfrac{x+2}{x^2-1}\right)$

(b) $f(x) = \ln(4x-20) + \ln\left(x^2+9x+18\right)$

(c) $f(x) = \ln\left(\sqrt{x-4}-3\right)$

(d) $f(x) = \dfrac{\sqrt{-1-x}}{\log_{\frac{1}{2}}(x)}$

(e) $f(x) = \log_2\left(\dfrac{1-x}{1+x}\right)$

(f) $f(x) = \log_6\left(\log_{\frac{1}{5}}(x)\right)$

Assignment 5.7 — A bacterial culture grows at a rate proportional to the number of cells present. Suppose the culture initially contains 500 cells, while 800 cells are present still after 24 hours. How many more cells will there be another 12 hours later?

Assignment 5.8 — A radioactive material has a half-life of 1200 years. What percentage of the initial radioactivity is left after 10 years? How many years does it take to decrease the level of radioactivity by 10%?

Assignment 5.9 — A scientist inoculates a bacterial culture with 20 bacteria. The number of bacteria increases by a constant factor every hour such that after 1 day the number of bacteria equals 220.

(a) Determine how many bacteria there will be after 1 week.

(b) How long does it take to reach 5 million bacteria?

Assignment 5.10 — A thermometer is removed from an oven at 72°C and placed in a room at 20°C . After 1 minute, the thermometer gives 48°C. What temperature will the thermometer indicate after 5 minutes?

Assignment 5.11 — An object is placed in a freezer of $-5°C$. In 40 minutes, the object cools from $45°C$ to $20°C$. How much longer is needed for the object to cool to $0°C$?

Assignment 5.12 — Consider the function

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

(a) Show that the inverse function of $f(x)$ equals $f^{-1}(x) = 10^{\frac{x}{x+1}}$.

(b) For which $x \in \mathbb{R}$ does $\left(f^{-1} \circ f\right)(x) = x$? For which $x \in \mathbb{R}$ does $\left(f \circ f^{-1}\right)(x) = x$?

(c) Define the image of f by determining the domain of f^{-1}.

(d) Assume $g(x) = \dfrac{x}{1-x}$ and $h(x) = \log(x)$. Prove that $f = g \circ h$ and that $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$.

Trigonometric functions

Assignment 5.13 — Determine the trigonometric numbers below.

(a) $\cos\left(\dfrac{3\pi}{4}\right)$

(b) $\tan\left(-\dfrac{3\pi}{4}\right)$

(c) $\sin\left(\dfrac{2\pi}{3}\right)$

(d) $\sin\left(\dfrac{7\pi}{12}\right)$

(e) $\cos\left(\dfrac{5\pi}{12}\right)$

(f) $\sin\left(\dfrac{11\pi}{12}\right)$

(g) $\cot\left(\dfrac{13\pi}{12}\right)$

(h) $\sec\left(-\dfrac{\pi}{12}\right)$

(i) $\csc\left(-\dfrac{\pi}{3}\right)$

(j) $\cot\left(\dfrac{7\pi}{6}\right)$

Assignment 5.14 — Below, a list of trigonometric numbers corresponding to an angle θ is given. Determine the other trigonometric numbers.

(a) $\sin(\theta) = \dfrac{3}{5}$ met $\theta \in \left[\dfrac{\pi}{2}, \pi\right]$

(b) $\tan(\theta) = 2$ met $\theta \in \left[0, \dfrac{\pi}{2}\right[$

(c) $\sec(\theta) = 3$ met $\theta \in \left]-\dfrac{\pi}{2}, 0\right]$

(d) $\cos(\theta) = -\dfrac{5}{13}$ met $\theta \in \left[\dfrac{\pi}{2}, \pi\right]$

(e) $\csc(\theta) = -2$ met $\theta \in \left]\pi, \dfrac{3\pi}{2}\right]$

(f) $\tan(\theta) = \dfrac{1}{2}$ met $\theta \in \left[\pi, \dfrac{3\pi}{2}\right[$

Assignment 5.15 — Prove the identities below.

(a) $\cos^4(x) - \sin^4(x) = \cos(2x)$

(b) $\dfrac{1-\cos(x)}{\sin(x)} = \dfrac{\sin(x)}{1+\cos(x)} = \tan\left(\dfrac{x}{2}\right)$

(c) $\dfrac{1-\cos(x)}{1+\cos(x)} = \tan^2\left(\dfrac{x}{2}\right)$

(d) $\dfrac{\cos(x)-\sin(x)}{\cos(x)+\sin(x)} = \sec(2x)-\tan(2x)$

(e) $\dfrac{\tan^2(x)-\sin^2(x)}{1-\sin^2(x)} = \tan^4(x)$

(f) $\sin(x+y)\sin(x-y) = \cos^2(y)-\cos^2(x)$

(g) $\cos(x+y)\cos(x)+\sin(x+y)\sin(x) = \cos(y)$

(h) $\dfrac{\sin(x+y)}{\sin(x-y)} = \dfrac{\tan(x)+\tan(y)}{\tan(x)-\tan(y)}$

(i) $\sin^2(2x)-\sin^2(x) = \sin(3x)\sin(x)$

Assignment 5.16 — Consider the general sinusoids below. Determine for each them the amplitude, period, phase shift, vertical shift, domain, image, and zeros.

(a) $f(x) = 3\sin\left(\dfrac{x}{2\pi}\right)$

(b) $f(x) = \dfrac{2}{3}\sin\left(\dfrac{2}{3}\left(x-\dfrac{\pi}{4}\right)\right)-11$

(c) $f(x) = \sin\left(10\pi\left(x+\dfrac{1}{2}\right)\right)+3$

(d) $f(x) = 2\sin(3x-2)+1$

Assignment 5.17 — Solve the trigonometric equations below.

(a) $\cos(3x) = \sin(7x)$

(b) $\sin(x)-\sin^3(x) = -\cos^2(x)$

(c) $\sin^2(x)+3\sin(x)\cos(x) = 1$

(d) $\dfrac{1+\sin(x)}{1-\sin(x)} = 2$

(e) $\sin(x)(\tan(x)+\cot(x)) = 2$

(f) $\tan(2x) = 5\tan(x)$

(g) $\cos(3x) = \cos(2x)-\cos(x)$

(h) $2\tan^2(x)+6 = 5\sec^2(x)$

(i) $2\cos(x)\cos(3x) = -1$

(j) $\sin^3(x)+\cos^3(x)$
$= \sin(x)\cos(x)(\sin(x)+\cos(x))$

Assignment 5.18 — Solve the trigonometric equations below.

(a) $\tan(2x) < \dfrac{1}{3}$

(b) $2\sin\left(2x-\dfrac{\pi}{3}\right)+1 < 0$

(c) $4\cos\left(3\left(x+\dfrac{\pi}{5}\right)\right)+2 > 4$

(d) $\tan\left(2x+\dfrac{\pi}{6}\right) < \sqrt{3}$

(e) $2\cos^2(2x)+(\sqrt{3}+2)\cos(2x)+\sqrt{3} < 0$

(f) $\dfrac{\sin(x)}{2\sin(x)-1} > \dfrac{1-\sin(x)}{4\sin^2(x)-1}$

Inverse trigonometric functions

Assignment 5.19 — Determine the domain of the functions below.

(a) $\arctan(x^3+1)$

(b) $\arcsin\left(\frac{1}{x}\right)$

(c) $\arccos(x^2-3)$

(d) $\arccos\left(\frac{6}{\pi}\arcsin(x)\right)$

(e) $\arcsin\left(\frac{4}{\pi}\arccos(2x)\right)$

Assignment 5.20 — Calculate the values below.

(a) $\arcsin\left(-\frac{\sqrt{2}}{2}\right)$

(b) $\arctan\left(\frac{\sqrt{3}}{3}\right)$

(c) $\operatorname{arccot}\left(\sqrt{3}\right)$

(d) $\sin\left(\arccos\left(\frac{\sqrt{3}}{2}\right)\right)$

(e) $\cot\left(\arctan(-1)\right)$

(f) $\sin\left(2\arctan\left(\frac{1}{\sqrt{5}}\right)\right)$

(g) $\cos\left(2\operatorname{arccot}\left(\sqrt{7}\right)\right)$

(h) $\tan\left(4\arcsin\left(\frac{1}{3}\right)\right)$

(i) $\cot\left(3\arccos\left(-\frac{1}{\sqrt{2}}\right)\right)$

(j) $\cos\left(\arcsin\left(\frac{\sqrt{2}}{2}\right)\right)$

(k) $\tan\left(\arccos\left(-\frac{\sqrt{2}}{2}\right)\right)$

(l) $\cos\left(2\arcsin\left(\frac{2}{5}\right)\right)$

(m) $\arccos\left(\sin\left(\arctan(-1)\right)\right)$

Assignment 5.21 — Rewrite the expressions below as a function of x and give in each case the domain for which these equivalences hold true.

(a) $\tan\left(\arccos(x)\right)$

(b) $\cot\left(\arcsin(x)\right)$

(c) $\sin\left(2\arctan(x)\right)$

(d) $\cos\left(2\operatorname{arccot}(x)\right)$

(e) $\cos\left(2\arcsin(x)\right)$

Assignment 5.22 — Prove the identities below.

(a) $\arcsin(x)+\arcsin(-x)=0$

(b) $\arccos(x)+\arccos(-x)=\pi$

(c) $\arctan(x)+\arctan(-x)=0$

(d) $\operatorname{arccot}(x)+\operatorname{arccot}(-x)=0$

Assignment 5.23 — Prove the expressions below.

(a) $\sin\left(2\arcsin(x)\right)=2x\sqrt{1-x^2}$

(b) $\cos\big(3\arcsin(x)\big) = \big(1-4x^2\big)\sqrt{1-x^2}$

(c) $\arctan\left(\frac{1}{2}\right) + \arctan\left(\frac{1}{3}\right) = \frac{\pi}{4}$

(d) $\forall x \in]-3,1[: \quad \arcsin\left(\frac{x+1}{2}\right) = \arctan\left(\frac{x+1}{\sqrt{3-2x-x^2}}\right)$

Assignment 5.24 — Prove that $\arcsin\left(\frac{3}{5}\right) + \arcsin\left(\frac{4}{5}\right) = \frac{\pi}{2}$.

Hyperbolic functions

Assignment 5.25 — Prove Equations (5.31) and (5.32).

Assignment 5.26 — Simplify

$$\frac{\cosh\big(\ln(x)\big) + \sinh\big(\ln(x)\big)}{\cosh\big(\ln(x)\big) - \sinh\big(\ln(x)\big)}.$$

Assignment 5.27 — Prove that for every $x \in \mathbb{R}$ the following holds.

(a) $\dfrac{1}{\cosh^2(x)} = 1 - \tanh^2(x)$

(b) $\dfrac{1}{\sinh^2(x)} = \coth^2(x) - 1$

(c) $\sinh(x) = \dfrac{\tanh(x)}{\sqrt{1-\tanh^2(x)}}$

(d) $\cosh(x) = \dfrac{1}{\sqrt{1-\tanh^2(x)}}$

(e) $\sinh(x) = \dfrac{2\tanh\left(\frac{x}{2}\right)}{1-\tanh^2\left(\frac{x}{2}\right)}$

(f) $\cosh(x) = \dfrac{1+\tanh^2\left(\frac{x}{2}\right)}{1-\tanh^2\left(\frac{x}{2}\right)}$

(g) $\sinh(3x) = 3\sinh(x) + 4\sinh^3(x)$

(h) $\cosh(3x) = 4\cosh^3(x) - 3\cosh(x)$

(i) $\cosh(x) + \cosh(y) = 2\cosh\left(\frac{x+y}{2}\right)\cosh\left(\frac{x-y}{2}\right)$

(j) $\cosh(x) - \cosh(y) = 2\sinh\left(\frac{x+y}{2}\right)\sinh\left(\frac{x-y}{2}\right)$

(k) $\cosh\big(\operatorname{arsinh}(x)\big) = \sqrt{x^2+1}$

(l) $\cosh\left[\ln\left(ax+\sqrt{(ax)^2+1}\right)\right] = \sqrt{(ax)^2+1}$

(m) $\tanh\left[\ln\left(ax+\sqrt{(ax)^2+1}\right)\right] = \dfrac{ax}{\sqrt{(ax)^2+1}}$

Assignment 5.28 — Prove that for every $x \in \mathbb{R}_0^+$ the following holds:

(a) $\sinh(\ln(x)) = \dfrac{x^2-1}{2x}$

(b) $\cosh(\ln(x)) = \dfrac{x^2+1}{2x}$

(c) $\sinh\left(2\operatorname{arsinh}\left(\dfrac{x-1}{2\sqrt{x}}\right)\right) = \dfrac{x^2-1}{2x}$

Assignment 5.29 — A cord whose ends are attached to two points on a horizontal line, hangs according to the catenary with equation

$$y = 80\cosh\left(\frac{x}{80}\right).$$

If the center of this cord is 20 meter below the horizontal line connecting the suspension points, what is the distance between these suspension points?

Conic sections continued

Assignment 5.30 — Investigate the nature of the following conic sections and reduce the equation to a standard form.

(a) $9y^2+12y+4=0$

(b) $3x^2-y^2-12x-2y+6=0$

Q: What do you get when you cross a mosquito with a mountain climber?
A: Nothing. You can't cross a vector and a scalar.

6 Vector math

For many applications, real numbers suffice, but are other times, these do not suffice. Perhaps it is important to know, for instance, how close the nearest cuckoo's nest is as well as the direction in which it lies. To answer questions like these which involve both a quantitative answer, or magnitude, along with a direction, we use the mathematical objects called **vectors** (*vector*).

6.1 Definition and representation

A vector is represented geometrically as a directed line segment where the magnitude of the vector is taken to be the length of the line segment and the direction is made clear with the use of an arrow at one endpoint of the segment. When referring to vectors in this text, we shall adopt the bold-faced arrow notation, so the symbol $\vec{\mathbf{v}}$ is read as the vector v.

In Figure 6.1(a) is a typical vector $\vec{\mathbf{v}}$ with endpoints $P(1,2)$ and $Q(4,6)$. The point P is called the **initial point** or **head** (*aangrijpingspunt*) of $\vec{\mathbf{v}}$ and the point Q is called the **terminal point** or **tail** of $\vec{\mathbf{v}}$. Since we can reconstruct $\vec{\mathbf{v}}$ completely from P and Q, we write $\vec{\mathbf{v}} = \overrightarrow{PQ}$, where the order of points P (initial point) and Q (terminal point) is important.

While it is true that P and Q completely determine $\vec{\mathbf{v}}$, it is important to note that since vectors are defined in terms of their two characteristics, **magnitude** (*grootte*) and **direction** (*richting*), any directed line segment with the same length and direction as $\vec{\mathbf{v}}$ is considered to be the same vector as $\vec{\mathbf{v}}$, regardless of its initial point. In the case of our vector $\vec{\mathbf{v}}$ above, any vector which moves three units to the right and four up from its initial point to arrive at its terminal point is considered the same vector as $\vec{\mathbf{v}}$. The notation we use to capture this idea is the **component form** of the vector, $\vec{\mathbf{v}} = (3,4)$, where the first number, 3, is called the ***x*-component** (*x-component*) of $\vec{\mathbf{v}}$ and the second number, 4, is called the ***y*-component** (*y-component*) of $\vec{\mathbf{v}}$. If we wanted to reconstruct $\vec{\mathbf{v}} = (3,4)$ with initial point $\widetilde{P}(-2,3)$, then we would find the terminal point of $\vec{\mathbf{v}}$ by adding 3 to the x-coordinate and adding 4 to the y-coordinate to obtain the terminal point $\widetilde{Q}(1,7)$ (Figure 6.2).

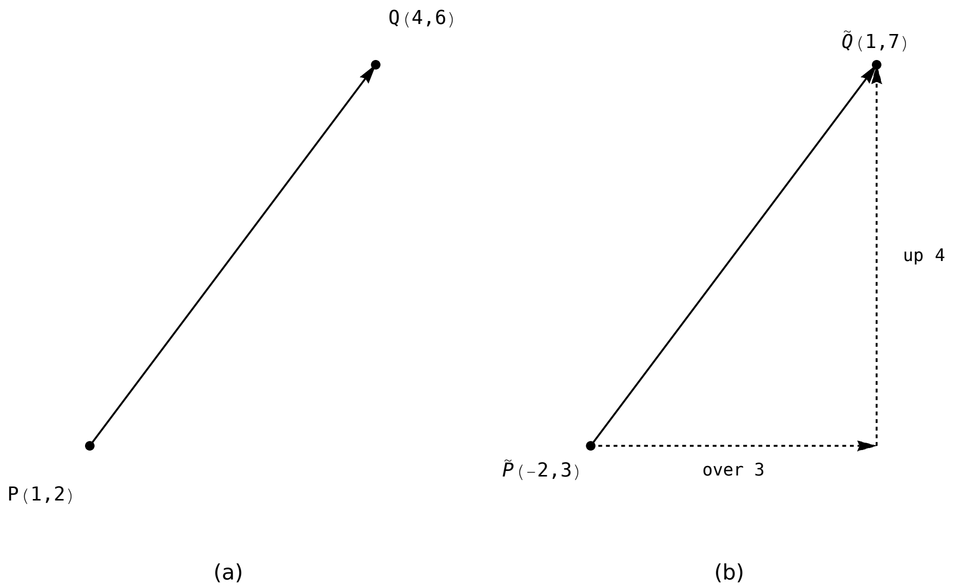


Figure 6.1: $\vec{v} = (3,4)$ with initial point $P(1,2)$ (a) and $\widetilde{P}(-2,3)$ (b).

This idea is formalized in the following definition.

Definitie 6.1 (Component form of a vector)
Suppose $\vec{v}$ is represented by a directed line segment with initial point $P(x_0, y_0)$ and terminal point $Q(x_1, y_1)$. The component form of $\vec{v}$ is given by

$$\vec{v} = \overrightarrow{PQ} = (x_1 - x_0, y_1 - y_0).$$

Using the language of components, we have that two vectors are equal if and only if their corresponding components are equal. That is, $(v_1, v_2) = (\widetilde{v}_1, \widetilde{v}_2)$ if and only if $v_1 = \widetilde{v}_1$ and $v_2 = \widetilde{v}_2$.

6.2 Vector arithmetic

6.2.1 Addition

We are now set to define operations on vectors. Suppose we are given two vectors $\vec{v}$ and $\vec{w}$. The sum $\vec{v} + \vec{w}$ is obtained as illustrated in Figure 6.2. First, plot $\vec{v}$. Next, plot $\vec{w}$ so that its initial point is the terminal point of $\vec{v}$. To plot the vector $\vec{v} + \vec{w}$ we begin at the initial point of $\vec{v}$ and end at the terminal point of $\vec{w}$. It is helpful to think of the vector $\vec{v} + \vec{w}$ as the net result of moving along $\vec{v}$ then moving along $\vec{w}$.

Example 6.1

A plane leaves an airport with an airspeed of 175 kilometres per hour at a bearing of N40°E. A 35 kilometres per hour wind is blowing at a bearing of S60°E. Find the speed and bearing of the plane.

Solution

For both the plane and the wind, we are given their speeds and their directions. Coupling speed

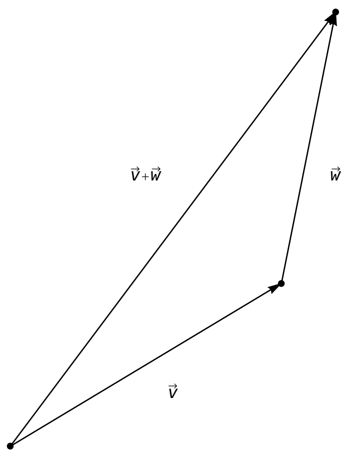


Figure 6.2: The vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ and their sum $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$.

(as a magnitude) with direction is the concept of velocity. We let $\vec{\boldsymbol{v}}$ denote the plane's velocity and $\vec{\boldsymbol{w}}$ denote the wind's velocity in the diagram below. The true speed and bearing is found by analysing the vector $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$ (Figure 6.3(a)). From the vector diagram, we get a triangle, the lengths of whose sides are the magnitude of $\vec{\boldsymbol{v}}$, which is 175, the magnitude of $\vec{\boldsymbol{w}}$, which is 35, and the magnitude of $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$, which we call c (Figure 6.3(b)). From the given bearing information, we go through the usual geometry to determine that the angle between the sides of length 35 and 175 measures 100°.

From the law of cosines, we determine $c=\sqrt{31850-12250\cos(100°)} \approx 184$, which means the true speed of the plane is (approximately) 184 kilometres per hour. To determine the true bearing of the plane, we need to determine the angle α. Using the law of cosines once more, we find $\cos(\alpha)=\frac{c^2+29400}{350c}$ so that $\alpha \approx 11°$. Given the geometry of the situation, we add α to the given 40° and find the true bearing of the plane to be (approximately) N51°E.

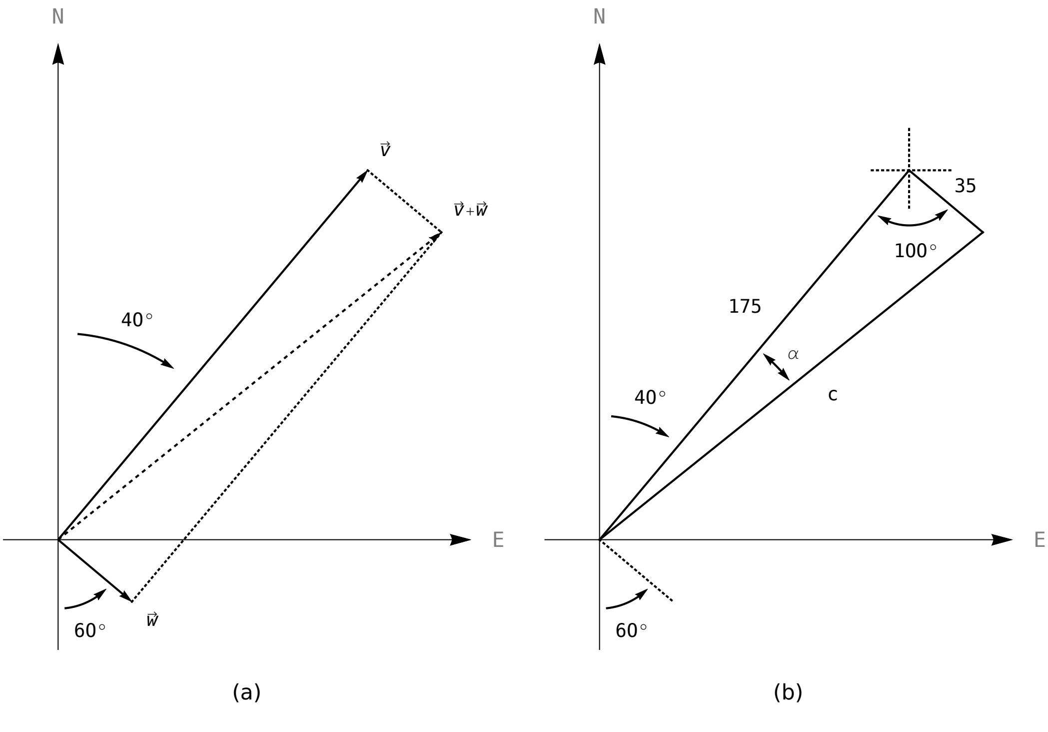


Figure 6.3: Finding the speed and bearing of a plane travelling at 175 kilometres per hour at a bearing of N40°E under a 35 kilometre per hour wind is blowing at a bearing of S60°E.

Having now a geometric understanding of the addition of vectors, we are ready to give its algebraic counterpart.

Definitie 6.2 (Vector addition)
Suppose $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$. The vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ is defined by

$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = (v_1 + w_1, v_2 + w_2).$$

In order for vector addition to enjoy the same kinds of properties as real number addition, it is necessary to extend our definition of vectors to include a **zero vector** (*nulvector*) , $\vec{\boldsymbol{0}} = (0,0)$. Geometrically, $\vec{\boldsymbol{0}}$ represents a point, which we can think of as a directed line segment with the same initial and terminal points. The direction of $\vec{\boldsymbol{0}}$ is in fact undefined. Having introduced the vector counterpart of the real number 0, we can go ahead and list the properties of vector addition, which are completely in line with those for real numbers. Essentially, for all vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ we have

- **Commutative property:**
$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}},$$
- **Associative property:**
$$(\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}) + \vec{\boldsymbol{w}} = \vec{\boldsymbol{u}} + (\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}),$$
- **Identity property:**
$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{0}} = \vec{\boldsymbol{0}} + \vec{\boldsymbol{v}} = \vec{\boldsymbol{v}},$$
- **Inverse property:** for every vector $\vec{\boldsymbol{v}}$, there is a vector $-\vec{\boldsymbol{v}}$ so that
$$\vec{\boldsymbol{v}} + (-\vec{\boldsymbol{v}}) = (-\vec{\boldsymbol{v}}) + \vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}.$$

Proof These properties are easily verified using the definition of vector addition. For instance, for the commutative property, we note that if $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$ then

$$\begin{aligned}\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} &= (v_1, v_2) + (w_1, w_2) \\ &= (v_1 + w_1, v_2 + w_2) \\ &= (w_1 + v_1, w_2 + v_2) \\ &= \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}}.\end{aligned}$$

Geometrically, we can see the commutative property by realizing that the sums $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ and $\vec{\boldsymbol{w}} + \vec{\boldsymbol{v}}$ are the same directed diagonal determined by the parallelogram in Figure 6.4. The proofs of the associative and identity properties proceed similarly. □

For what concerns the additive inverse $-\vec{\boldsymbol{v}} = (-v_1, -v_2)$ of a vector $\vec{\boldsymbol{v}} = (v_1, v_2)$, it is clear that both vectors have the same length, but opposite directions. As a result, when adding the vectors geometrically, the sum $\vec{\boldsymbol{v}} + (-\vec{\boldsymbol{v}})$ results in starting at the initial point of $\vec{\boldsymbol{v}}$ and ending back at the initial point of $\vec{\boldsymbol{v}}$, or in other words, the net result of moving $\vec{\boldsymbol{v}}$ then $-\vec{\boldsymbol{v}}$ is not moving at all. Using the additive inverse of a vector, we can define the difference of two vectors, $\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}} = \vec{\boldsymbol{v}} + (-\vec{\boldsymbol{w}})$. If $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$ then

$$\begin{aligned}\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}} &= \vec{\boldsymbol{v}} + (-\vec{\boldsymbol{w}}) \\ &= (v_1, v_2) + (-w_1, -w_2) \\ &= \big(v_1 + (-w_1), v_2 + (-w_2)\big) \\ &= (v_1 - w_1, v_2 - w_2).\end{aligned}$$

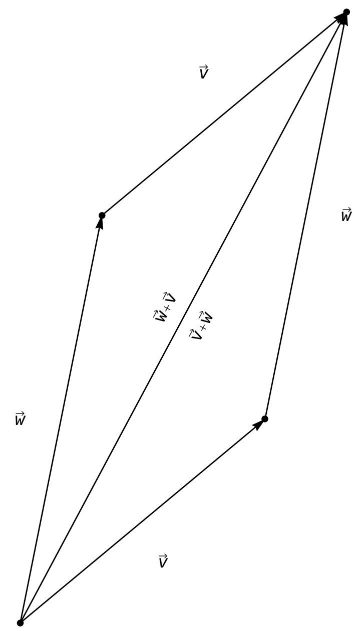


Figure 6.4: Proving the commutative property of vector addition.

In other words, like vector addition, vector subtraction works component-wise. From the diagram in Figure 6.5, we see that $\vec{v}-\vec{w}$ may be interpreted as the vector whose initial point is the terminal point of $\vec{w}$ and whose terminal point is the terminal point of $\vec{v}$ as depicted below. It is also worth mentioning that in the parallelogram determined by the vectors $\vec{v}$ and $\vec{w}$, the vector $\vec{v}-\vec{w}$ is one of the diagonals – the other being $\vec{v}+\vec{w}$.

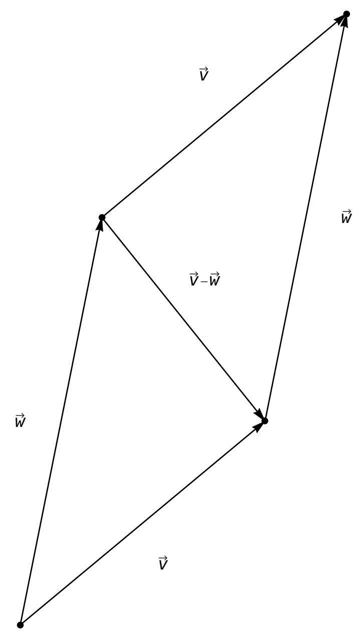


Figure 6.5: The vectors $\vec{v}$ and $\vec{w}$ and their difference $\vec{v}-\vec{w}$.

6.2.2 Scalar multiplication

Next, we discuss **scalar multiplication** (*scalaire vermenigvuldiging*) – that is, taking a real number times a vector.

Definitie 6.3 (Scalar multiplication)
If k is a real number and $\vec{\boldsymbol{v}} = (v_1, v_2)$, we define $k\vec{\boldsymbol{v}}$ by

$$k\vec{\boldsymbol{v}} = k\,(v_1, v_2) = (kv_1, kv_2)\,.$$

Scalar multiplication by k in vectors can be understood geometrically as scaling the vector (if $k > 0$) or scaling the vector and reversing its direction (if $k < 0$) as demonstrated in Figure 6.6.

The properties of scalar multiplication are summarized below for vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ and scalars k and r.

- **Associative property:**
$$(kr)\vec{\boldsymbol{v}} = k(r\vec{\boldsymbol{v}})\,,$$
- **Identity property:**
$$1\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}\,,$$
- **Additive inverse property:**
$$-\vec{\boldsymbol{v}} = (-1)\vec{\boldsymbol{v}}\,,$$
- **Distributive property of scalar multiplication over scalar addition:**
$$(k+r)\vec{\boldsymbol{v}} = k\vec{\boldsymbol{v}} + r\vec{\boldsymbol{v}}\,,$$
- **Distributive property of scalar multiplication over vector addition:**
$$k(\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}) = k\vec{\boldsymbol{v}} + k\vec{\boldsymbol{w}}\,,$$
- **Zero product property:**
$$k\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}} \quad \Leftrightarrow \quad k = 0 \quad \vee \quad \vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}\,.$$

Proof The proof of these properties, ultimately boils down to the definition of scalar multiplication and properties of real numbers. For example, to prove the associative property, we let $\vec{\boldsymbol{v}} = (v_1, v_2)$. If k and r are scalars then

$$\begin{aligned}
(kr)\vec{\boldsymbol{v}} &= (kr)\,(v_1, v_2) \\
&= \big((kr)v_1, (kr)v_2\big) && \text{(Definition of scalar multiplication.)} \\
&= \big(k(rv_1), k(rv_2)\big) && \text{(Associative property of real number multiplication.)} \\
&= k\,(rv_1, rv_2) && \text{(Definition of scalar multiplication.)} \\
&= k\big(r\,(v_1, v_2)\big) && \text{(Definition of scalar multiplication.)} \\
&= k(r\vec{\boldsymbol{v}}).
\end{aligned}$$

□

Our next example demonstrates how Definition 6.3 allows us to do the same kind of algebraic manipulations with vectors as we do with variables – multiplication and division of vectors notwithstanding. If the pedantry seems familiar, it should. We spell out the solution of the following example in excruciating detail to encourage the reader to think carefully about why each step is justified.

Example 6.2

Solve $5\vec{\boldsymbol{v}} - 2\big(\vec{\boldsymbol{v}} + (1,-2)\big) = \vec{\boldsymbol{0}}$ for $\vec{\boldsymbol{v}}$.

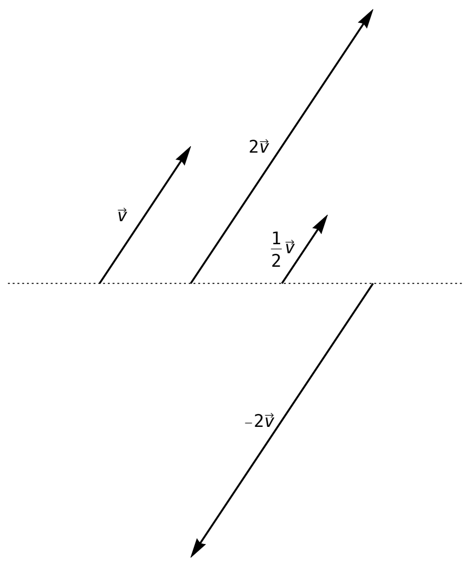


Figure 6.6: The scalar multiplication of a vector $\vec{\boldsymbol{v}}$.

Solution

$$\begin{aligned}
5\vec{\boldsymbol{v}} - 2\left(\vec{\boldsymbol{v}} + (1,-2)\right) &= \vec{\mathbf{0}} \\
5\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{v}} - 2(1,-2) &= \vec{\mathbf{0}} \\
(5-2)\vec{\boldsymbol{v}} + \left((-2)1, (-2)(-2)\right) &= \vec{\mathbf{0}} \\
3\vec{\boldsymbol{v}} + (-2,4) &= \vec{\mathbf{0}} \\
3\vec{\boldsymbol{v}} &= \vec{\mathbf{0}} - (-2,4) \\
\vec{\boldsymbol{v}} &= \frac{1}{3}(2,-4) \\
\vec{\boldsymbol{v}} &= \left(\frac{2}{3}, \frac{-4}{3}\right)
\end{aligned}$$

6.2.3 Magnitude and direction

A vector whose initial point is $(0,0)$ is said to be in **standard position** (*standaardvoorstelling*). If $\vec{\boldsymbol{v}} = (v_1, v_2)$ is plotted in standard position, then its terminal point is necessarily (v_1, v_2) (Figure 6.7).

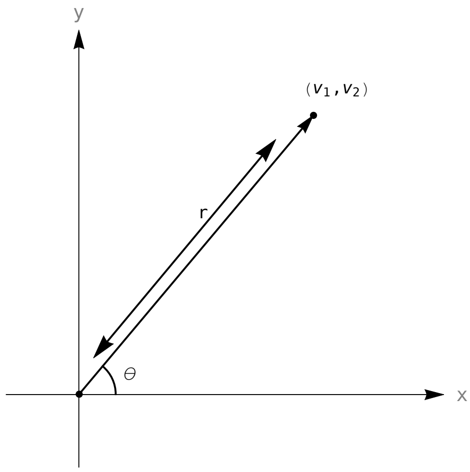


Figure 6.7: The $\vec{\mathbf{v}} = (v_1, v_2)$ in standard position.

Plotting a vector in standard position enables us to more easily quantify the concepts of magnitude and direction of the vector. We can convert the point (v_1, v_2) in rectangular coordinates to a pair (r, θ) in polar coordinates where $r \geq 0$. The magnitude of $\vec{\mathbf{v}}$, which we said earlier was the length of the directed line segment, is $r = \sqrt{v_1^2 + v_2^2}$ and is denoted by $\|\vec{\mathbf{v}}\|$. From Theorem 5.5, we know $v_1 = r\cos(\theta) = \|\vec{\mathbf{v}}\|\cos(\theta)$ and $v_2 = r\sin(\theta) = \|\vec{\mathbf{v}}\|\sin(\theta)$. From the definition of scalar multiplication and vector equality, we get

$$\begin{aligned} \vec{\mathbf{v}} &= (v_1, v_2) \\ &= \big(\|\vec{\mathbf{v}}\|\cos(\theta), \|\vec{\mathbf{v}}\|\sin(\theta)\big) \\ &= \|\vec{\mathbf{v}}\|\big(\cos(\theta), \sin(\theta)\big). \end{aligned}$$

This motivates the following definition.

Definitie 6.4 (Magnitude and direction of a vector)
Suppose $\vec{\mathbf{v}}$ is a vector with component form $\vec{\mathbf{v}} = (v_1, v_2)$. Let (r, θ) be a polar representation of the point with rectangular coordinates (v_1, v_2) with $r \geq 0$.

- The **magnitude** (*grootte*) of $\vec{\mathbf{v}}$, denoted $\|\vec{\mathbf{v}}\|$, is given by

$$\|\vec{\mathbf{v}}\| = r = \sqrt{v_1^2 + v_2^2}\,.$$

- If $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$, the **(vector) direction** (*richting*) of $\vec{\mathbf{v}}$, denoted $\hat{\mathbf{v}}$ is given by

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

Both magnitude and direction of a vector come along with a few important properties.

- It holds that $\|\vec{\mathbf{v}}\| \geq 0$ and $\|\vec{\mathbf{v}}\| = 0$ if and only if $\vec{\mathbf{v}} = \vec{\mathbf{0}}$.
- For all scalars k, it holds that
$$\|k\,\vec{\mathbf{v}}\| = |k|\,\|\vec{\mathbf{v}}\|\,.$$
- If $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$ then $\vec{\mathbf{v}} = \|\vec{\mathbf{v}}\|\,\hat{\mathbf{v}}$, so that
$$\hat{\mathbf{v}} = \left(\frac{1}{\|\vec{\mathbf{v}}\|}\right)\vec{\mathbf{v}}\,. \tag{6.1}$$

Example 6.3

1. Find the polar representation of the vector $\vec{\boldsymbol{v}} = \left(3, -3\sqrt{3}\right)$, assuming that $0 \le \theta < 2\pi$.
2. For the vectors $\vec{\boldsymbol{v}} = (3,4)$ and $\vec{\boldsymbol{w}} = (1,-2)$, find the following.

(a) $\hat{\boldsymbol{v}}$ (b) $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\|$ (c) $\|\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}\|$ (d) $\|\hat{\boldsymbol{w}}\|$

Solution

1. For $\vec{\boldsymbol{v}} = \left(3, -3\sqrt{3}\right)$, we get $\|\vec{\boldsymbol{v}}\| = \sqrt{(3)^2 + (-3\sqrt{3})^2} = 6$. We can find the θ we are after by converting the point with rectangular coordinates $(3, -3\sqrt{3})$ to polar form (r, θ) where $r = \|\vec{\boldsymbol{v}}\| > 0$. This leads to $\tan(\theta) = -3\sqrt{3}/3 = -\sqrt{3}$. Since $(3, -3\sqrt{3})$ is a point in Quadrant IV, θ is a Quadrant IV angle. Hence, we pick $\theta = \frac{5\pi}{3}$.
2. (a) Since we are given the component form of $\vec{\boldsymbol{v}}$, we will use Equation (6.1). For $\vec{\boldsymbol{v}} = (3,4)$, we have $\|\vec{\boldsymbol{v}}\| = \sqrt{3^2+4^2} = \sqrt{25} = 5$. Hence, $\hat{\boldsymbol{v}} = \frac{1}{5}(3,4) = \left(\frac{3}{5}, \frac{4}{5}\right)$.

   (b) We already know that $\|\vec{\boldsymbol{v}}\| = 5$, so to find $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\|$, we need only find $\|\vec{\boldsymbol{w}}\|$. Since $\vec{\boldsymbol{w}} = (1,-2)$, we get $\|\vec{\boldsymbol{w}}\| = \sqrt{1^2 + (-2)^2} = \sqrt{5}$. Hence, $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\| = 5 - 2\sqrt{5}$.

   (c) Our first step is to find the component form of the vector $\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}$. As such, we get $\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}} = (3,4) - 2(1,-2) = (1,8)$. Hence, $\|\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}\| = \|(1,8)\| = \sqrt{1^2+8^2} = \sqrt{65}$.

   (d) To find $\|\hat{\boldsymbol{w}}\|$, we first need $\hat{\boldsymbol{w}}$. Using Equation (6.1) along with $\|\vec{\boldsymbol{w}}\| = \sqrt{5}$, we get

$$\hat{\boldsymbol{w}} = \frac{1}{\sqrt{5}}(1,-2) = \left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right) = \left(\frac{\sqrt{5}}{5}, -\frac{2\sqrt{5}}{5}\right).$$

   Hence,

$$\|\hat{\boldsymbol{w}}\| = \sqrt{\left(\frac{\sqrt{5}}{5}\right)^2 + \left(-\frac{2\sqrt{5}}{5}\right)^2} = \sqrt{\frac{5}{25} + \frac{20}{25}} = \sqrt{1} = 1\,.$$

The process exemplified in Example 6.3 by which we take information about the magnitude and direction of a vector and find the component form of a vector is called **resolving** a vector into its components.

6.3 Unit vectors

Vectors with length 1 are called unit vectors and are very important in algebra.

Definitie 6.5 (Unit vector)
Let $\vec{\boldsymbol{v}}$ be a vector. If $\|\vec{\boldsymbol{v}}\| = 1$, we say that $\vec{\boldsymbol{v}}$ is a **unit vector** (*eenheidsvector*).

If $\vec{\boldsymbol{v}}$ is a unit vector, then necessarily,

$$\vec{\boldsymbol{v}} = \|\vec{\boldsymbol{v}}\|\hat{\boldsymbol{v}} = 1 \cdot \hat{\boldsymbol{v}} = \hat{\boldsymbol{v}}.$$

Conversely, it can be shown that

$$\hat{\boldsymbol{v}} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\right)\vec{\boldsymbol{v}}$$

is a unit vector for any nonzero vector $\vec{\boldsymbol{v}}$. The process of multiplying a nonzero vector by the factor $\|\vec{\boldsymbol{v}}\|^{-1}$ to produce a unit vector is called **normalizing** (*normeren*) the vector and the resulting vector $\hat{\boldsymbol{v}}$ is called the 'unit vector in the direction of $\vec{\boldsymbol{v}}$'. The terminal points of unit vectors, when plotted in standard position, lie on the unit circle. As a result, we may visualize normalizing a nonzero vector $\vec{\boldsymbol{v}}$ as shrinking its terminal point, when plotted in standard position, back to the unit circle. In practice, if $\vec{\boldsymbol{v}}$ is a unit vector we write it as $\hat{\boldsymbol{v}}$ (Figure 6.8).

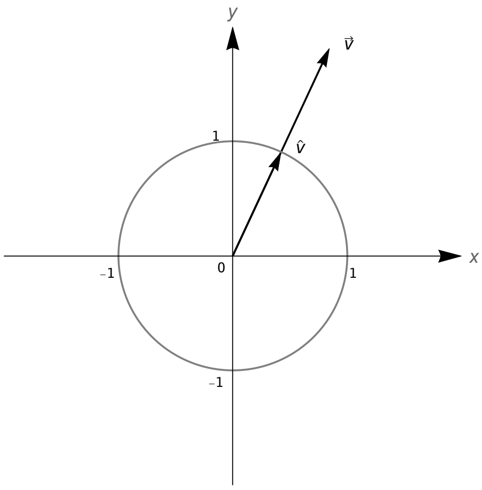


Figure 6.8: Vector normalisation $\hat{\boldsymbol{v}} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\right)\vec{\boldsymbol{v}}$.

Of all of the unit vectors, the so-called **principal unit vectors** (*basis eenheidsvector*) deserve special attention. In two dimensions, they are given by

- The vector $\hat{\boldsymbol{\imath}} = (1, 0)$,
- The vector $\hat{\boldsymbol{\jmath}} = (0, 1)$.

We may think of the vector $\hat{\boldsymbol{\imath}}$ as representing the positive x-direction, while $\hat{\boldsymbol{\jmath}}$ represents the positive y-direction. Consequently, the coordinate axes x and y are the axes in the direction of $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$, respectively. Together, $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$ make up the so-called **standard basis** (*standaardbasis*) for the Euclidean plane. Having introduced principal unit vectors, we are now ready to get up to the following decomposition theorem.

Theorem 6.1 (Principal vector decomposition theorem)
Let $\vec{\boldsymbol{v}}$ be a vector with component form $\vec{\boldsymbol{v}} = (v_1, v_2)$. Then $\vec{\boldsymbol{v}} = v_1\hat{\boldsymbol{\imath}} + v_2\hat{\boldsymbol{\jmath}}$.

Proof The proof of this theorem is straightforward. Since $\hat{\boldsymbol{\imath}} = (1, 0)$ and $\hat{\boldsymbol{\jmath}} = (0, 1)$, we have from the definition of scalar multiplication and vector addition that

$$v_1\hat{\boldsymbol{\imath}} + v_2\hat{\boldsymbol{\jmath}} = v_1(1, 0) + v_2(0, 1) = (v_1, 0) + (0, v_2) = (v_1, v_2) = \vec{\boldsymbol{v}}$$

Geometrically, the situation looks like the diagram in Figure 6.9. □

We conclude this section with a classic example that demonstrates how vectors are used to model forces. A force is defined as a push or a pull. The intensity of the push or pull is the magnitude of the force, and is measured in Newton (N).

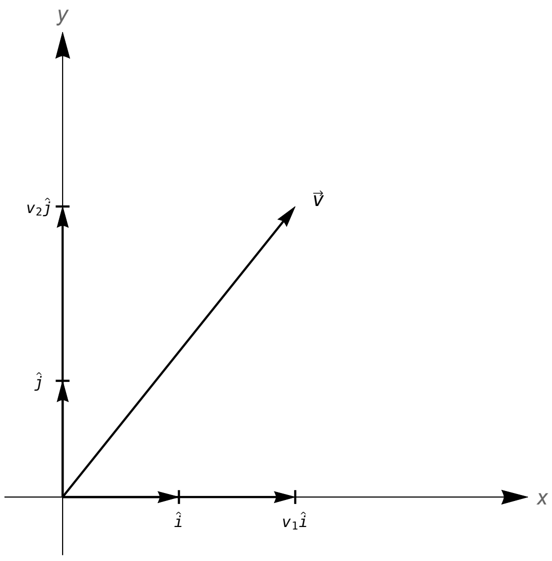


Figure 6.9: $\vec{\boldsymbol{v}} = (v_1, v_2) = v_1\hat{\boldsymbol{\imath}} + v_2\hat{\boldsymbol{\jmath}}$.

Example 6.4

A speaker exerting a fore by gravity of 50 Newton is suspended from the ceiling by two support braces. If one of them makes a 60° angle with the ceiling and the other makes a 30° angle with the ceiling, what are the tensions on each of the supports?

Solution

The problem is sketched schematically in Figure 6.10(a) and the corresponding vector diagram is given in Figure 6.10(b).

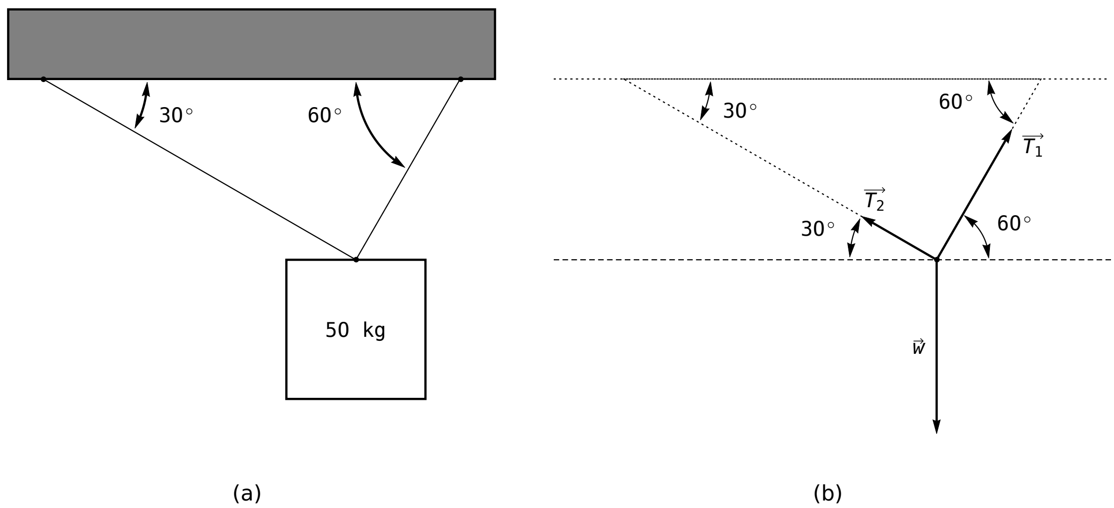


Figure 6.10: Schematic representation and corresponding vector diagram for the problem in Example 6.4.

We have three forces acting on the speaker: the weight of the speaker, which we will call $\vec{\boldsymbol{w}}$, pulling the speaker directly downward, and the forces on the support rods, which we will call $\overrightarrow{\mathbf{T_1}}$ and $\overrightarrow{\mathbf{T_2}}$ (for tensions) acting upward at angles 60° and 30°, respectively. We are looking for the tensions on the support, which are the magnitudes $\|\overrightarrow{\mathbf{T_1}}\|$ and $\|\overrightarrow{\mathbf{T_2}}\|$. In order for the speaker to remain stationary, we require $\vec{\boldsymbol{w}} + \overrightarrow{\mathbf{T_1}} + \overrightarrow{\mathbf{T_2}} = \vec{\mathbf{0}}$. Viewing the common initial point of these vectors as the origin and the dashed line as the x-axis, we get the component representations for the three vectors involved. We can model the weight of the speaker as a vector pointing directly downwards with a magnitude of 50 Newton. That is, $\|\vec{w}\| = 50$ and $\hat{w} = -\hat{j} = (0,-1)$. Hence, $\vec{w} = 50\,(0,-1) = (0,-50)$. For the force in the first support, we get

$$\begin{aligned}\overrightarrow{\mathbf{T_1}} &= \|\overrightarrow{\mathbf{T_1}}\|\big(\cos(60°),\sin(60°)\big)\\ &= \left(\frac{\|\overrightarrow{\mathbf{T_1}}\|}{2},\frac{\|\overrightarrow{\mathbf{T_1}}\|\sqrt{3}}{2}\right).\end{aligned}$$

For the second support, we note that the angle 30° is measured from the negative x-axis, so the angle needed to write $\overrightarrow{\mathbf{T_2}}$ in component form is 150°. Hence

$$\begin{aligned}\overrightarrow{\mathbf{T_2}} &= \|\overrightarrow{\mathbf{T_2}}\|\big(\cos(150°),\sin(150°)\big)\\ &= \left(-\frac{\|\overrightarrow{\mathbf{T_2}}\|\sqrt{3}}{2},\frac{\|\overrightarrow{\mathbf{T_2}}\|}{2}\right).\end{aligned}$$

The requirement $\vec{w} + \overrightarrow{\mathbf{T_1}} + \overrightarrow{\mathbf{T_2}} = \vec{0}$ gives us the following vector equation.

$$\begin{aligned}\vec{w} + \overrightarrow{\mathbf{T_1}} + \overrightarrow{\mathbf{T_2}} &= \vec{0}\\ (0,-50) + \left(\frac{\|\overrightarrow{\mathbf{T_1}}\|}{2},\frac{\|\overrightarrow{\mathbf{T_1}}\|\sqrt{3}}{2}\right) + \left(-\frac{\|\overrightarrow{\mathbf{T_2}}\|\sqrt{3}}{2},\frac{\|\overrightarrow{\mathbf{T_2}}\|}{2}\right) &= (0,0)\\ \left(\frac{\|\overrightarrow{\mathbf{T_1}}\|}{2} - \frac{\|\overrightarrow{\mathbf{T_2}}\|\sqrt{3}}{2},\frac{\|\overrightarrow{\mathbf{T_1}}\|\sqrt{3}}{2} + \frac{\|\overrightarrow{\mathbf{T_2}}\|}{2} - 50\right) &= (0,0)\end{aligned}$$

Equating the corresponding components of the vectors on each side, we get a system of linear equations in the variables $\|\overrightarrow{\mathbf{T_1}}\|$ and $\|\overrightarrow{\mathbf{T_2}}\|$.

$$\frac{\|\overrightarrow{\mathbf{T_1}}\|}{2} - \frac{\|\overrightarrow{\mathbf{T_2}}\|\sqrt{3}}{2} = 0 \tag{6.2}$$

$$\frac{\|\overrightarrow{\mathbf{T_1}}\|\sqrt{3}}{2} + \frac{\|\overrightarrow{\mathbf{T_2}}\|}{2} - 50 = 0 \tag{6.3}$$

From Equation (6.2), we get $\|\overrightarrow{\mathbf{T_1}}\| = \|\overrightarrow{\mathbf{T_2}}\|\sqrt{3}$. Substituting that into Equation (6.3) gives

$$\frac{(\|\overrightarrow{\mathbf{T_2}}\|\sqrt{3})\sqrt{3}}{2} + \frac{\|\overrightarrow{\mathbf{T_2}}\|}{2} - 50 = 0$$

which yields $2\|\overrightarrow{\mathbf{T_2}}\| - 50 = 0$. Hence, $\|\overrightarrow{\mathbf{T_2}}\| = 25$ Newton and $\|\overrightarrow{\mathbf{T_1}}\| = \|\overrightarrow{\mathbf{T_2}}\|\sqrt{3} = 25\sqrt{3}$ Newton.

6.4 The dot product

In Section 6.2, we learned how add and subtract vectors and how to multiply vectors by scalars. Here, we will see how to multiply vectors. We begin with the following definition.

Definitie 6.6 (Dot product)
Suppose $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ are vectors whose component forms are $\vec{\mathbf{v}} = (v_1, v_2)$ and $\vec{\mathbf{w}} = (w_1, w_2)$. The **dot product** (*scalair product*) of $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ is given by

$$\vec{\mathbf{v}} \cdot \vec{\mathbf{w}} = (v_1, v_2) \cdot (w_1, w_2) = v_1 w_1 + v_2 w_2.$$

.

For example, let $\vec{\mathbf{v}} = (3,4)$ and $\vec{\mathbf{w}} = (1,-2)$. Then $\vec{\mathbf{v}} \cdot \vec{\mathbf{w}} = (3,4) \cdot (1,-2) = (3)(1) + (4)(-2) = -5$. Note that the dot product takes two vectors and produces a scalar. The dot product enjoys the following properties for all vectors $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ and scalar k.

- **Commutative property:**
$$\vec{\mathbf{v}} \cdot \vec{\mathbf{w}} = \vec{\mathbf{w}} \cdot \vec{\mathbf{v}}\,,$$
- **Distributive property:**
$$\vec{\mathbf{u}} \cdot (\vec{\mathbf{v}} + \vec{\mathbf{w}}) = \vec{\mathbf{u}} \cdot \vec{\mathbf{v}} + \vec{\mathbf{u}} \cdot \vec{\mathbf{w}}\,,$$
- **Scalar property:**
$$(k\vec{\mathbf{v}}) \cdot \vec{\mathbf{w}} = k(\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}) = \vec{\mathbf{v}} \cdot (k\vec{\mathbf{w}})\,,$$
- **Relation to magnitude:**
$$\vec{\mathbf{v}} \cdot \vec{\mathbf{v}} = \|\vec{\mathbf{v}}\|^2\,.$$

Like most of the theorems involving vectors, the proof of these properties amounts to using the definition of the dot product and properties of real number arithmetic.

Proof To show the commutative property for instance, let $\vec{\mathbf{v}} = (v_1, v_2)$ and $\vec{\mathbf{w}} = (w_1, w_2)$. Then

$$\begin{array}{rcll}
\vec{\mathbf{v}} \cdot \vec{\mathbf{w}} & = & (v_1, v_2) \cdot (w_1, w_2) & \\
 & = & v_1 w_1 + v_2 w_2 & \text{(Definition of dot product.)} \\
 & = & w_1 v_1 + w_2 v_2 & \text{(Commutativity of real number multiplication.)} \\
 & = & (w_1, w_2) \cdot (v_1, v_2) & \text{(Definition of dot product.)} \\
 & = & \vec{\mathbf{w}} \cdot \vec{\mathbf{v}} &
\end{array}$$

The distributive property is proved similarly.

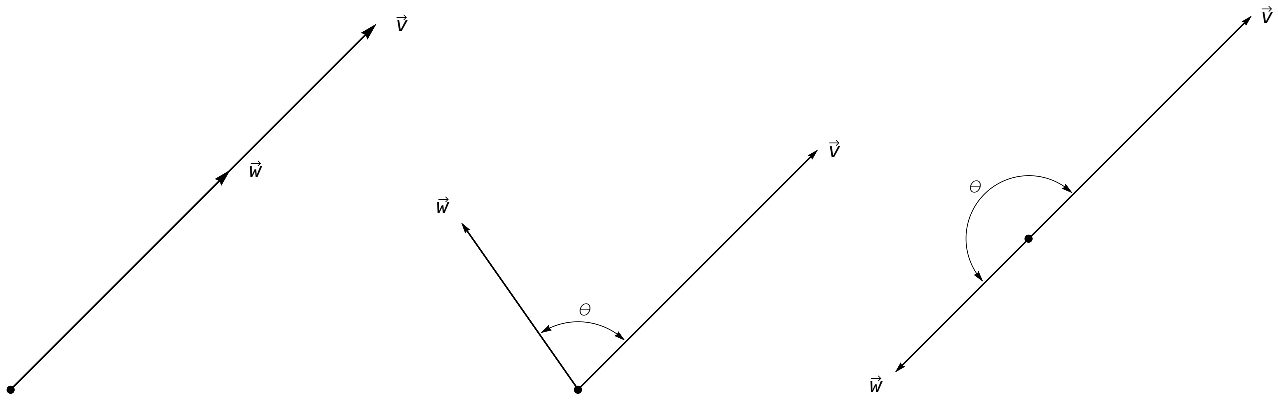
For the scalar property, assume that $\vec{\mathbf{v}} = (v_1, v_2)$ and $\vec{\mathbf{w}} = (w_1, w_2)$ and k is a scalar. Then

$$\begin{array}{rcll}
(k\vec{\mathbf{v}}) \cdot \vec{\mathbf{w}} & = & \big(k\,(v_1, v_2)\big) \cdot (w_1, w_2) & \\
 & = & (kv_1, kv_2) \cdot (w_1, w_2) & \text{(Definition of scalar multiplication.)} \\
 & = & (kv_1)(w_1) + (kv_2)(w_2) & \text{(Definition of dot product.)} \\
 & = & k(v_1 w_1) + k(v_2 w_2) & \text{(Associativity of real number multiplication.)} \\
 & = & k(v_1 w_1 + v_2 w_2) & \text{(Distributive law of real numbers.)} \\
 & = & k\,(v_1, v_2) \cdot (w_1, w_2) & \text{(Definition of dot product.)} \\
 & = & k(\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}) &
\end{array}$$

The property $k(\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}) = \vec{\boldsymbol{v}} \cdot (k\vec{\boldsymbol{w}})$ can be proved similarly

For the last property, we note that if $\vec{\boldsymbol{v}} = (v_1, v_2)$, then $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}} = (v_1, v_2) \cdot (v_1, v_2) = v_1^2 + v_2^2 = \|\vec{\boldsymbol{v}}\|^2$ (Definition 6.4). □

We now explore the geometric properties of the dot product. Suppose for that purpose $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are two nonzero vectors. If we draw $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ with the same initial point, we define the angle between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ to be the angle θ determined by the rays containing the vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$, as illustrated in Figure 6.11. We require $0 \leq \theta \leq \pi$.



(a) $\theta = 0$ (b) $0 < \theta < \pi$ (c) $\theta = \pi$

Figure 6.11: Two vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ with the same initial point and the angle θ between them.

The following theorem gives us some insight into the geometric role the dot product plays.

Theorem 6.2 (Geometric interpretation of dot product)
If $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are nonzero vectors then

$$\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = \|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\| \cos(\theta), \tag{6.4}$$

where θ is the angle between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$.

This theorem states that taking a dot product of two vectors boils down to projecting one vector onto the direction of the second vector and subsequently scaling it with the magnitude of the latter.

An immediate consequence of Theorem 6.2 is the following.

Theorem 6.3 (Angle between vectors)
Let $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ be nonzero vectors and let θ the angle between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$. Then

$$\theta = \arccos\left(\frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|}\right) = \arccos(\hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}}). \tag{6.5}$$

Proof We arrive at Equation (6.5) by solving Equation (6.4) for θ. Since $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are nonzero, so are $\|\vec{\boldsymbol{v}}\|$ and $\|\vec{\boldsymbol{w}}\|$. Hence, we may divide both sides of $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = \|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\| \cos(\theta)$ by $\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|$. Since $0 \leq \theta \leq \pi$ by definition, the values of θ exactly match the range of the arccosine function, so we get Equation (6.5). We can rewrite the argument of the arccosine function as

$$\frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\vec{\boldsymbol{v}}\right) \cdot \left(\frac{1}{\|\vec{\boldsymbol{w}}\|}\vec{\boldsymbol{w}}\right) = \hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}},$$

giving us the alternative formula $\theta = \arccos(\hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}})$. □

Example 6.5

Find the angle between the following pairs of vectors.

1. $\vec{v} = \left(3, -3\sqrt{3}\right)$ and $\vec{w} = \left(-\sqrt{3}, 1\right)$
2. $\vec{v} = (2, 2)$ and $\vec{w} = (5, -5)$

Solution

We use Equation (6.5) in each case below.

1. We have $\vec{v} \cdot \vec{w} = \left(3, -3\sqrt{3}\right) \cdot \left(-\sqrt{3}, 1\right) = -3\sqrt{3} - 3\sqrt{3} = -6\sqrt{3}$. As $\|\vec{v}\| = \sqrt{3^2 + (-3\sqrt{3})^2} = 6$ and $\|\vec{w}\| = \sqrt{(-\sqrt{3})^2 + 1^2} = \sqrt{4} = 2$, we find that
$$\theta = \arccos\left(\frac{-6\sqrt{3}}{12}\right) = \arccos\left(-\frac{\sqrt{3}}{2}\right) = \frac{5\pi}{6}.$$
2. For $\vec{v} = (2, 2)$ and $\vec{w} = (5, -5)$, we find $\vec{v} \cdot \vec{w} = (2, 2) \cdot (5, -5) = 10 - 10 = 0$. Hence, it does not matter what $\|\vec{v}\|$ and $\|\vec{w}\|$ are, and
$$\theta = \arccos\left(\frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}\right) = \arccos(0) = \frac{\pi}{2}.$$

Note that the vectors $\vec{v} = (2, 2)$, and $\vec{w} = (5, -5)$ in Example 6.5 are called **orthogonal** (*orthogonaal*) and we write $\vec{v} \perp \vec{w}$, because the angle between them is $\frac{\pi}{2}$ radians $= 90°$. Geometrically this means that when orthogonal vectors are sketched with the same initial point, the lines containing the vectors are perpendicular. We state the relationship between orthogonal vectors and their dot product in the following theorem.

Theorem 6.4 (Orthogonality of vectors)
Let $\vec{v}$ and $\vec{w}$ be nonzero vectors. Then $\vec{v} \perp \vec{w}$ if and only if $\vec{v} \cdot \vec{w} = 0$.

Proof To prove this theorem, we first assume $\vec{v}$ and $\vec{w}$ are nonzero vectors with $\vec{v} \perp \vec{w}$. By definition, the angle between $\vec{v}$ and $\vec{w}$ is $\frac{\pi}{2}$. By Theorem 6.2, it holds that $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\left(\frac{\pi}{2}\right) = 0$. Conversely, if $\vec{v}$ and $\vec{w}$ are nonzero vectors and $\vec{v} \cdot \vec{w} = 0$, then Theorem 6.3 gives
$$\theta = \arccos\left(\frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}\right) = \arccos\left(\frac{0}{\|\vec{v}\|\|\vec{w}\|}\right) = \arccos(0) = \frac{\pi}{2},$$
so $\vec{v} \perp \vec{w}$. □

6.5 Vectors in n-dimensional space

Vectors are of course not limited to the plane as most processes involving forces happen in three-dimensional space. For that reason, we will move our mathematics out of the plane and into space. That is, we begin to think mathematically not only in two dimensions, but in three, and even more. With this foundation, we can explore vectors in space.

6.5.1 Cartesian coordinates and distance in space

Up to this point we have considered mathematics in a two–dimensional world. We have plotted graphs on the xy-plane using rectangular and polar coordinates and found the area of regions in the plane. Here, we introduce Cartesian coordinates in space.

Each point P in space can be represented with an ordered triple, $P = (x, y, z)$, where x, y and z represent the relative position of P along the x-, y- and z-axes, respectively. Each axis is perpendicular to the other two. For plotting shapes in space a first convention is that the axes must conform to the **right hand rule** (*rechterhandregel*). This rule states that when the index finger of the right hand is extended in the direction of the positive x-axis, and the middle finger (bent inward so it is perpendicular to the palm) points along the positive y-axis, then the extended thumb will point in the direction of the positive z-axis (Figure 6.12).

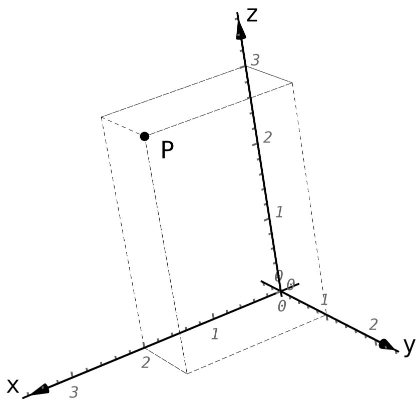


Figure 6.12: Plotting the point $P = (2, 1, 3)$ in space.

As long as the coordinate axes are positioned so that they follow this rule, it does not matter how the axes are drawn on paper. Throughout this text we will adhere to the convention with xy-plane as being a horizontal plane, where the positive z-axis is pointing up. This point of view is preferred by most mathematicians. Just as the x- and y-axes divide the plane into four quadrants, the x-, y-, and z-coordinate planes divide space into eight **octants** (*octant*). The octant in which x, y, and z are positive is called the first octant. The Euclidean distance between two points $P(x, y, z)$ and $Q(\tilde{x}, \tilde{y}, \tilde{z})$ in such a three-dimensional space is given by

$$d(P, Q) = \sqrt{(x-\tilde{x})^2 + (y-\tilde{y})^2 + (z-\tilde{z})^2}\,. \tag{6.6}$$

Even though it becomes more difficult to present graphically, it is mathematically of course possible to extend the notion of space to n dimensions. Essentially, a point P in n-dimensional space can then be represented in rectangular coordinates by the n-tuple $(x_1, x_2, \ldots, x_n)$ using n mutually perpendicular axes. The Euclidean distance between two points $P(x_1, x_2, \ldots, x_n)$ and $Q(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ in n-dimensional space is given by

$$d(P, Q) = \sqrt{(x_1-\tilde{x}_1)^2 + (x_2-\tilde{x}_2)^2 + \ldots + (x_n-\tilde{x}_n)^2}\,. \tag{6.7}$$

Higher-dimensional mathematics in art

The concept of four-dimensional space and the difficulties involved in trying to visualize it helped inspire many modern artists in the first half of the twentieth century, such as Picasso and Weber. The former got introduced to the topic through Poincaré's *Elementary Treatise on the Geometry of Four Dimensions*. He incorporated some aspects of four-dimensional mathematics in his painting *Portrait of Daniel-Henry Kahnweiler* (Figure 6.13).

Figure 6.13: Picasso's portrait of Daniel-Henry Kahnweiler.

6.5.2 Vector representation and operations in space

Essentially, all of the definitions and vector operations we introduced in Sections 6.1 and 6.2 may be generalized intuitively to three or more dimensions.

In n dimensions the component form of a vector is defined as follosws.

Definitie 6.7 (Component form of a vector in *n* dimensions)
Suppose $\vec{\boldsymbol{v}}$ is represented by a directed line segment with initial point $P\left(x_1, x_2, \ldots, x_n\right)$ and terminal point $Q\left(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\right)$. The component form of $\vec{\boldsymbol{v}}$ is given by

$$\vec{\boldsymbol{v}} = \overrightarrow{PQ} = \left(\tilde{x}_1 - x_1, \tilde{x}_2 - x_2, \ldots, \tilde{x}_n - x_n\right).$$

Likewise, vector addition, scalar multiplication and the dot product, as well as their corresponding properties, generalize naturally to n dimensions.

Definitie 6.8 (Vector addition in *n* dimensions)
Suppose $\vec{\boldsymbol{v}} = (v_1, v_2, \ldots, v_n)$ and $\vec{\boldsymbol{w}} = (w_1, w_2, \ldots, w_n)$. The vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ is defined by

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

Definitie 6.9 (Scalar multiplication in *n* dimensions)
If k is a real number and $\vec{\boldsymbol{v}} = (v_1, v_2, \ldots, v_n)$, we define $k\vec{\boldsymbol{v}}$ by

$$k\vec{\boldsymbol{v}} = k\,(v_1, v_2, \ldots, v_n) = (kv_1, kv_2, \ldots, kv_n)\,.$$

Definitie 6.10 (Dot product in *n* dimensions)
Suppose $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are vectors whose component forms are $\vec{\boldsymbol{v}} = (v_1, v_2, \ldots, v_n)$ and $\vec{\boldsymbol{w}} = (w_1, w_2, \ldots, w_n)$. The dot product of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ is given by

$$\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = (v_1, v_2, \ldots, v_n) \cdot (w_1, w_2, \ldots, w_n) = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

Besides, in n dimensions, we can define n principal unit vectors $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$, …, $\hat{\boldsymbol{e}}_n$, which make up a standard basis. All components of the i-th such principal unit vector are zero, expect the i-th component, which is 1. For instance, $\hat{\boldsymbol{e}}_2 = (0, 1, 0, \ldots, 0)$ or $\hat{\boldsymbol{e}}_n = (0, 0, 0, \ldots, 1)$.

Example 6.6

Given $\vec{\boldsymbol{v}} = (1, 2, 2)$ and $\vec{\boldsymbol{w}} = (-1, 0, 3)$, find

1. The unit vector in the direction of $\vec{\boldsymbol{v}}$.
2. $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}$.

Solution

1. We find $\|\vec{\boldsymbol{v}}\| = 3$, so the unit vector $\hat{\boldsymbol{z}}$ in the direction of $\vec{\boldsymbol{v}}$ is

$$\hat{\boldsymbol{z}} = \frac{1}{3}\vec{\boldsymbol{v}} = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right).$$

2. Using Definition 6.10 we immediately find

$$\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = 1(-1) + 2(0) + 2(3) = 5.$$

Example 6.7

Let $\vec{\boldsymbol{u}} = (1, 1, 1)$, $\vec{\boldsymbol{v}} = (-1, 3, -2)$ and $\vec{\boldsymbol{w}} = (-5, 1, 4)$. Find the angle between each pair of vectors.

Solution

For this purpose, we use Equation (6.5).

1. Between $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$:

$$\begin{aligned} \theta &= \arccos\left(\frac{\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}}{\|\vec{\boldsymbol{u}}\|\,\|\vec{\boldsymbol{v}}\|}\right) \\ &= \arccos\left(\frac{0}{\sqrt{3}\sqrt{14}}\right) \\ &= \frac{\pi}{2}. \end{aligned}$$

2. Between $\vec{u}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{u}\cdot\vec{w}}{\|\vec{u}\|\,\|\vec{w}\|}\right)\\ &= \arccos\left(\frac{0}{\sqrt{3}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

3. Between $\vec{v}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\,\|\vec{w}\|}\right)\\ &= \arccos\left(\frac{0}{\sqrt{14}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

For n-dimensional vectors we have the following important theorem.

Theorem 6.5 (Cauchy-Schwartz inequality)
If $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^n$, then

$$|\vec{u}\cdot\vec{v}| \leq \|\vec{u}\|\,\|\vec{v}\|$$

We conclude this section with an example.

Example 6.8
For any x, y $z \in \mathbb{R}^+$ prove that

$$\sqrt{x(3x+y)}+\sqrt{y(3y+z)}+\sqrt{z(3z+x)} \leq 2(x+y+z).$$

Solution

By Cauchy-Schwartz we have that

$$\begin{aligned}&\sqrt{x(3x+y)}+\sqrt{y(3y+z)}+\sqrt{z(3z+x)}\\ \leq\ &\sqrt{\left((\sqrt{x})^2+(\sqrt{y})^2+(\sqrt{z})^2\right)\left((\sqrt{3x+y})^2+(\sqrt{3y+z})^2+(\sqrt{3z+x})^2\right)}\\ =\ &\sqrt{4(x+y+z)^2}\\ =\ &2(x+y+z)\end{aligned}$$

6.6 The cross product

6.6.1 Definition and properties

Orthogonality is immensely important. A quick scan of your current environment will undoubtedly reveal numerous surfaces and edges that are perpendicular to each other. The dot product provides a quick test for orthogonality: vectors $\vec{u}$ and $\vec{v}$ are perpendicular if and only if, $\vec{u}\cdot\vec{v}=0$.

Given two non–parallel, nonzero vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ in space, it is very useful to find a vector $\vec{\boldsymbol{w}}$ that is perpendicular to both $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$. There is a operation, called the cross product, that creates such a vector.

Definitie 6.11 (Cross product)
Let $\vec{\boldsymbol{u}} = (u_1, u_2, u_3)$ and $\vec{\boldsymbol{v}} = (v_1, v_2, v_3)$ be vectors in $\mathbb{R}^3$. The **cross product** (*kruisproduct of vectorieel product*) of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$, denoted $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$, is the vector

$$\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = \big(u_2v_3 - u_3v_2, -(u_1v_3 - u_3v_1), u_1v_2 - u_2v_1\big).$$

The cross product can also be expressed as the formal determinant:

$$\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = \begin{vmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

which makes it easier to remember. Using cofactor expansion along the first row, it expands to

$$\begin{aligned} \vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \hat{\boldsymbol{i}} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \hat{\boldsymbol{j}} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \hat{\boldsymbol{k}} \\ &= (u_2v_3 - u_3v_2)\hat{\boldsymbol{i}} - (u_1v_3 - u_3v_1)\hat{\boldsymbol{j}} + (u_1v_2 - u_2v_1)\hat{\boldsymbol{k}}, \end{aligned}$$

which gives the components of the resulting vector directly.

Example 6.9

Let $\vec{\boldsymbol{u}} = (2, -1, 4)$ and $\vec{\boldsymbol{v}} = (3, 2, 5)$. Find $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$, and verify that it is orthogonal to both $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

Solution

Using Definition 6.11, we have

$$\begin{aligned} \vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} &= \begin{vmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix} \\ &= \Big(\big((-1)5 - (4)2\big)\hat{\boldsymbol{i}}, -\big((2)5 - (4)3\big)\hat{\boldsymbol{j}}, \big((2)2 - (-1)3\big)\hat{\boldsymbol{k}}\Big) \\ &= \Big(-13\hat{\boldsymbol{i}}, 2\hat{\boldsymbol{j}}, 7\hat{\boldsymbol{k}}\Big) \end{aligned}$$

We now test whether or not $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ is orthogonal to $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ using the dot product:

$$\big(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}\big) \cdot \vec{\boldsymbol{u}} = (-13, 2, 7) \cdot (2, -1, 4) = 0,$$

$$\big(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}\big) \cdot \vec{\boldsymbol{v}} = (-13, 2, 7) \cdot (3, 2, 5) = 0.$$

Since both dot products are zero, $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ is indeed orthogonal to both $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

Given the vectors $\vec{\boldsymbol{u}}$, $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ in $\mathbb{R}^3$ and the scalar c, the following properties hold for the cross product, each of which can be verified by referring to the definition.

- **Anticommutative property:**
$$\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = -(\vec{\boldsymbol{v}} \times \vec{\boldsymbol{u}}),$$
- **Distributive properties:**
$$(\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}) \times \vec{\boldsymbol{w}} \quad = \quad \vec{\boldsymbol{u}} \times \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}},$$

$$\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w},$$

- **Compatibility with scalar multiplication:**

$$c(\vec{u} \times \vec{v}) = (c\vec{u}) \times \vec{v} = \vec{u} \times (c\vec{v}),$$

- **Orthogonality properties:**

$$\begin{aligned}(\vec{u} \times \vec{v}) \cdot \vec{u} &= 0,\\ (\vec{u} \times \vec{v}) \cdot \vec{v} &= 0,\end{aligned}$$

- **Zero cross product:**

$$\vec{u} \times \vec{0} = \vec{0},$$

- **Self cross product:**

$$\vec{u} \times \vec{u} = \vec{0},$$

- **Triple Scalar Product**:

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}.$$

Another wonderful property of the cross product, as defined, is that it follows the right hand rule. Given $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ with the same initial point, point the index finger of your right hand in the direction of $\vec{u}$ and let your middle finger point in the direction of $\vec{v}$. Your thumb will naturally extend in the direction of $\vec{u} \times \vec{v}$. If you switch, and point the index finger in the direction of $\vec{v}$ and the middle finger in the direction of $\vec{u}$, your thumb will now point in the opposite direction, allowing you to visualize the anticommutative property of the cross product.

The property $\vec{u} \times \vec{u} = \vec{0}$ reveals something more interesting than the cross product of a vector with itself is $\vec{0}$. Let $\vec{u}$ and $\vec{v}$ be parallel vectors; that is, let there be a scalar c such that $\vec{v} = c\vec{u}$. Consider their cross product:

$$\begin{aligned}\vec{u} \times \vec{v} &= \vec{u} \times (c\vec{u})\\ &= c(\vec{u} \times \vec{u})\\ &= \vec{0}.\end{aligned}$$

This hints at something deeper. Theorem 6.2 related the angle between two vectors and their dot product; there is a similar relationship relating the cross product of two vectors and the angle between them, given by the following theorem.

Theorem 6.6 (The cross product and angles)
Let $\vec{u}$ and $\vec{v}$ be nonzero vectors in $\mathbb{R}^3$. Then

$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\,\|\vec{v}\| \sin(\theta),$$

where θ, $0 \le \theta \le \pi$, is the angle between $\vec{u}$ and $\vec{v}$.

Note that this theorem makes a statement about the magnitude of the cross product. When the angle between $\vec{u}$ and $\vec{v}$ is 0 or π (i.e., the vectors are parallel), the magnitude of the cross product is 0. The only vector with a magnitude of 0 is $\vec{0}$, hence the cross product of parallel vectors is $\vec{0}$.

We demonstrate the truth of this theorem in the following example.

Example 6.10

Let $\vec{u} = (1, 3, 6)$ and $\vec{v} = (-1, 2, 1)$. Find the angle between $\vec{u}$ and $\vec{v}$, and the magnitude of $\vec{u} \times \vec{v}$.

Solution

We use Theorem 6.2 to find the angle between $\vec{u}$ and $\vec{v}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\,\|\vec{v}\|}\right)\\ &= \arccos\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\\ &\approx 0.8471 = 48.54°.\end{aligned}$$

Since $\vec{u} \times \vec{v} = (-9, -7, 5)$ we have $\|\vec{u} \times \vec{v}\| = \sqrt{155}$. The question now is whether or not $\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\,\|\vec{v}\| \sin(\theta)$? Using numerical approximations, we find:

$$\begin{aligned}\|\vec{u} \times \vec{v}\| &= \sqrt{155} & \qquad \|\vec{u}\|\,\|\vec{v}\|\sin(\theta) &= \sqrt{46}\sqrt{6}\sin(0.8471)\\ &\approx 12.45, & &\approx 12.45.\end{aligned}$$

Numerically, they seem equal. Using a right triangle, one can show that

$$\sin\left(\arccos\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\right) = \frac{\sqrt{155}}{\sqrt{46}\sqrt{6}},$$

which allows us to verify Theorem 6.6 exactly.

It is a standard geometry fact that the area of a parallelogram is $A = bh$, where b is the length of the base and h is the height of the parallelogram. As shown when defining vector addition (Figure 6.2), two vectors $\vec{u}$ and $\vec{v}$ define a parallelogram when drawn from the same initial point, as illustrated in Figure 6.14. Trigonometry tells us that $h = \|\vec{u}\| \sin(\theta)$, hence the area of the parallelogram is

$$A = \|\vec{u}\|\,\|\vec{v}\|\sin(\theta) = \|\vec{u} \times \vec{v}\|, \tag{6.8}$$

where the second equality comes from Theorem 6.6. We illustrate using Equation (6.8) in the following examples.

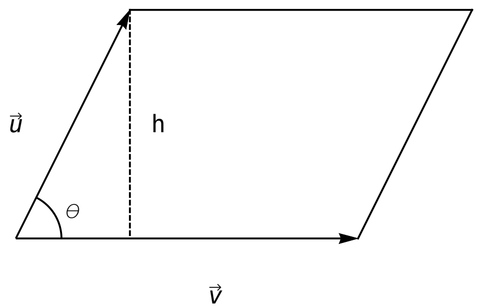


Figure 6.14: Using the cross product to find the area of a parallelogram.

Example 6.11

1. Find the area of the parallelogram defined by the vectors $\vec{u} = (2, 1)$ and $\vec{v} = (1, 3)$.

2. Find the area of the triangle with vertices $A = (1, 2)$, $B = (2, 3)$ and $C = (3, 1)$.

Solution

1. We have a slight problem in that our vectors exist in $\mathbb{R}^2$, not $\mathbb{R}^3$, and the cross product is only defined on vectors in $\mathbb{R}^3$. We skirt this issue by viewing $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ as vectors in the xy-plane of $\mathbb{R}^3$, and rewrite them as $\vec{\boldsymbol{u}} = (2,1,0)$ and $\vec{\boldsymbol{v}} = (1,3,0)$. We can now compute the cross product. It is easy to show that $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = (0,0,5)$; therefore the area of the parallelogram is $A = \|\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}\| = 5$.

2. We can choose any two sides of the triangle to use to form vectors; we choose $\overrightarrow{AB} = (1,1)$ and $\overrightarrow{AC} = (2,-1)$. As in the previous example, we will rewrite these vectors with a third component of 0 so that we can apply the cross product. The area of the triangle is

$$\frac{1}{2}\left\|\overrightarrow{AB} \times \overrightarrow{AC}\right\| = \frac{1}{2}\left\|(1,1,0) \times (2,-1,0)\right\| = \frac{1}{2}\left\|(0,0,-3)\right\| = \frac{3}{2}.$$

Computational geometry

In computational geometry of the plane, the cross product is used to determine the sign of the acute angle defined by three points $P_1 = (x_1,y_1), P_2 = (x_2,y_2)$, and $P_3 = (x_3,y_3)$. It corresponds to the direction (upward or downward) of the cross product of the two vectors defined by the two pairs of points (P_1,P_2) and (P_1,P_3). The sign of the acute angle is the sign of the expression

$$(x_2-x_1)(y_3-y_1)-(y_2-y_1)(x_3-x_1),$$

which is the signed length of the cross product of the two vectors. If the result is 0, the points are collinear; if it is positive, the three points constitute a positive angle of rotation around P_1 from P_2 to P_3, otherwise a negative angle.

6.7 Exercises

Vector arithmetic

Assignment 6.1 — Draw the vectors

$$\vec{\boldsymbol{a}}=(3,-2),\quad \vec{\boldsymbol{b}}=\left(-\frac{4}{3},0\right),\quad \vec{\boldsymbol{c}}=(0,2)\quad \text{en}\quad \vec{\boldsymbol{d}}=(1,-1),$$

and determine their size and direction.

Assignment 6.2 — Solve the equations below to $\vec{\boldsymbol{x}}$. What are the coordinates for $\vec{\boldsymbol{x}}$ if $A=(3,2)$ and $B=(1,4)$?

(a) $\vec{\boldsymbol{a}}+\vec{\boldsymbol{x}}+\vec{\boldsymbol{b}}=\vec{\boldsymbol{0}}$

(b) $\vec{\boldsymbol{a}}-\vec{\boldsymbol{b}}=2\vec{\boldsymbol{b}}+\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}}$

(c) $3(\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}})=\vec{\boldsymbol{x}}-\vec{\boldsymbol{b}}$

(d) $2(\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}})=3\left(\vec{\boldsymbol{x}}-\vec{\boldsymbol{b}}\right)$

Assignment 6.3 — Consider M as the center of $[AB]$. Prove that for any point X in the plane, the following holds.

(a) $\left(\overrightarrow{AX}\right)^2+\left(\overrightarrow{BX}\right)^2=2\left(\overrightarrow{MX}\right)^2+2\left(\overrightarrow{MB}\right)^2$

(b) $\left(\overrightarrow{AX}\right)^2-\left(\overrightarrow{BX}\right)^2=2\overrightarrow{AB}\,\overrightarrow{MX}$

Assignment 6.4 — Show that the triangle with vertices $A=(2,1)$, $B=(6,4)$ and $C=(5,-3)$ is isosceles.

Assignment 6.5 — Show that the triangle with vertices $A=(0,0)$, $B=\left(1,\sqrt{3}\right)$ and $C=(2,0)$ is equilateral.

Assignment 6.6 — A wind vane mounted on top of a car traveling north at 50 km/h indicates that the wind is coming from the west. If the car is traveling twice as fast, the wind vane indicates that the wind is coming from the northwest. From which direction is the wind coming and what is its speed?

Unit vectors

Assignment 6.7 — Consider the following points: $A=(-1,2)$, $B=(2,0)$, $C=(1,-3)$ and $D=(0,4)$. Express the vectors below using the basis vectors $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$.

(a) $\overrightarrow{AB}$

(b) $\overrightarrow{BA}$

(c) $\overrightarrow{AC}$

(d) $\overrightarrow{AB}-\overrightarrow{BC}$

(e) $\overrightarrow{AC}-2\overrightarrow{AB}+3\overrightarrow{CD}$

(f) $\dfrac{\overrightarrow{AB}+\overrightarrow{AC}+\overrightarrow{AD}}{3}$

The dot product

Assignment 6.8 — Consider the vectors: $\vec{\boldsymbol{a}} = (3,2)$ and $\vec{\boldsymbol{b}} = (x, 2-x)$. Determine x such that $\vec{\boldsymbol{a}} \perp \vec{\boldsymbol{b}}$.

Assignment 6.9 — Determine the dot product and the angle between the vectors $\vec{\boldsymbol{a}} = (2,1)$ and $\vec{\boldsymbol{b}} = (1,2)$.

Assignment 6.10 — Consider the vectors $\vec{\boldsymbol{a}} = (a_1, a_2)$, $\vec{\boldsymbol{b}} = (b_1, b_2)$ and $\vec{\boldsymbol{c}} = (c_1, c_2)$. Calculate the coordinates of $\left(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}\right) \vec{\boldsymbol{c}}$ and $\vec{\boldsymbol{a}} \left(\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{c}}\right)$.

Assignment 6.11 — Verify that

$$(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}})^2 = \|\vec{\boldsymbol{x}}\|^2 + \|\vec{\boldsymbol{y}}\|^2 + 2\|\vec{\boldsymbol{x}}\|\|\vec{\boldsymbol{y}}\| \cos(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}).$$

Assignment 6.12 — Prove the following.

(a) $\forall\, \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \neq \vec{\boldsymbol{0}};\ r, s \in \mathbb{R}_0: \quad \cos(r\vec{\boldsymbol{x}}, s\vec{\boldsymbol{y}}) = \cos(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}) \neq 0 \quad \Leftrightarrow \quad rs > 0$

(b) $\forall\, \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \neq \vec{\boldsymbol{0}};\ r, s \in \mathbb{R}_0: \quad \cos(r\vec{\boldsymbol{x}}, s\vec{\boldsymbol{y}}) = -\cos(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}) \neq 0 \quad \Leftrightarrow \quad rs < 0$

Assignment 6.13 — Verify that

$$\forall\, \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}: \quad \|\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}\| = \|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}\| \quad \Leftrightarrow \quad \vec{\boldsymbol{x}} \perp \vec{\boldsymbol{y}}.$$

Assignment 6.14 — Prove that the points $A = (2,-1)$, $B = (1,3)$ and $C = (-3,2)$ are the vertices of a square and determine the fourth vertex.

Vectors in n-dimensional space

Assignment 6.15 — Determine the parameter h such that the given vectors are orthogonal.

(a) $\vec{\boldsymbol{v}} = (-1,2,3)$ en $\vec{\boldsymbol{w}} = (1,2,h)$

(b) $\vec{\boldsymbol{a}} = (\sqrt{3}, h, 8)$ en $\vec{\boldsymbol{b}} = (h, -4, 2)$

Assignment 6.16 — Consider the vectors

$$\vec{\boldsymbol{u}} = \frac{3}{5}\hat{\boldsymbol{\imath}} + \frac{4}{5}\hat{\boldsymbol{\jmath}}, \quad \vec{\boldsymbol{v}} = \frac{4}{5}\hat{\boldsymbol{\imath}} - \frac{3}{5}\hat{\boldsymbol{\jmath}} \quad \text{en} \quad \vec{\boldsymbol{w}} = \hat{\boldsymbol{k}}.$$

(a) Prove that $\|\vec{\boldsymbol{u}}\| = \|\vec{\boldsymbol{v}}\| = \|\vec{\boldsymbol{w}}\| = 1$ and that $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = \vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{w}} = \vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = 0$.

(b) Consider $\vec{\boldsymbol{r}} = x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}$. Prove that

$$\vec{\boldsymbol{r}} = (\vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{u}})\vec{\boldsymbol{u}} + (\vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{v}})\vec{\boldsymbol{v}} + (\vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{w}})\vec{\boldsymbol{w}}.$$

Assignment 6.17 — Prove the so-called parallelogram rule

$$\|\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}}\|^2+\|\vec{\boldsymbol{u}}-\vec{\boldsymbol{v}}\|^2=2\|\vec{\boldsymbol{u}}\|^2+2\|\vec{\boldsymbol{v}}\|^2$$

for vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ in $\mathbb{R}^n$.

The cross product

Assignment 6.18 — Determine $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ with

(a) $\vec{\boldsymbol{u}}=\hat{\boldsymbol{\imath}}-2\hat{\boldsymbol{\jmath}}+3\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}}=3\hat{\boldsymbol{\imath}}+\hat{\boldsymbol{\jmath}}-4\hat{\boldsymbol{k}}$,

(b) $\vec{\boldsymbol{u}}=\hat{\boldsymbol{\jmath}}+2\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}}=-\hat{\boldsymbol{\imath}}-\hat{\boldsymbol{\jmath}}+\hat{\boldsymbol{k}}$.

Assignment 6.19 — Determine $\vec{\boldsymbol{u}} \times (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}})$ and $(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}) \times \vec{\boldsymbol{w}}$ with $\vec{\boldsymbol{u}}=\hat{\boldsymbol{\imath}}+2\hat{\boldsymbol{\jmath}}+3\hat{\boldsymbol{k}}$, $\vec{\boldsymbol{v}}=2\hat{\boldsymbol{\imath}}-3\hat{\boldsymbol{\jmath}}$ and $\vec{\boldsymbol{w}}=\hat{\boldsymbol{\jmath}}-\hat{\boldsymbol{k}}$. Give an explanation as to why both results are not equal to each other.

Assignment 6.20 — Prove that $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}=\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}=\vec{\boldsymbol{w}} \times \vec{\boldsymbol{u}}$ if it is true that $\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}=\vec{\boldsymbol{0}}$.

Review exercises

Assignment 6.21 — Consider the vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$:

(a) $\vec{\boldsymbol{u}}=\hat{\boldsymbol{\imath}}-\hat{\boldsymbol{\jmath}}$ and $\vec{\boldsymbol{v}}=\hat{\boldsymbol{\jmath}}+2\hat{\boldsymbol{k}}$,

(b) $\vec{\boldsymbol{u}}=3\hat{\boldsymbol{\imath}}+4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}}=3\hat{\boldsymbol{\imath}}-4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}$.

Determine the following for the given vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

(a) $\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}}$, $\vec{\boldsymbol{u}}-\vec{\boldsymbol{v}}$, $2\vec{\boldsymbol{u}}-3\vec{\boldsymbol{v}}$

(b) $\|\vec{\boldsymbol{u}}\|$ and $\|\vec{\boldsymbol{v}}\|$

(c) $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$

(d) $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}$

(e) the angle between $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$

Q: Why did the chicken cross the road?
A: The answer is trivial and is left as an exercise for the reader.

7

Three-dimensional analytical geometry

In Chapter 6 we already introduced Cartesian coordinates in space when discussing vectors in space. Of course, we can make use of that framework to extend also analytical geometry to three dimensions. We start our brief discussion with lines and planes, after which we turn to more involved objects such as quadratic surfaces.

7.1 Lines

7.1.1 Definition

To find the equation of **a line** (*rechte*) in the xy-plane, we need a point on the line and the direction of the line. This also holds true for lines in space. Let P be a point in space, let $\vec{\boldsymbol{p}}$ be the vector with initial point at the origin and terminal point at P (i.e., $\vec{\boldsymbol{p}}$ points to P), and let $\vec{\boldsymbol{d}}$ be a vector. Consider the points on the line through P in the direction of $\vec{\boldsymbol{d}}$. Clearly one point on the line is P; we can say that the vector $\vec{\boldsymbol{p}}$ lies at this point on the line. To find another point on the line, we can start at $\vec{\boldsymbol{p}}$ and move in a direction parallel to $\vec{\boldsymbol{d}}$. For instance, starting at $\vec{\boldsymbol{p}}$ and travelling one length of $\vec{\boldsymbol{d}}$ places one at another point on the line. Consider Figure 7.1 where certain points along the line are indicated. The figure illustrates how every point on the line can be obtained by starting with $\vec{\boldsymbol{p}}$ and moving a certain distance in the direction of $\vec{\boldsymbol{d}}$. That is, we can define the line as a function of t using a **vector equation** (*vectorvergelijking van een rechte*):

$$\vec{\boldsymbol{\ell}}(t) = \vec{\boldsymbol{p}} + t\,\vec{\boldsymbol{d}}. \tag{7.1}$$

In many ways, this is not a new concept. Compare Equation (7.1) to the familiar $y = mx + b$ equation of a line:

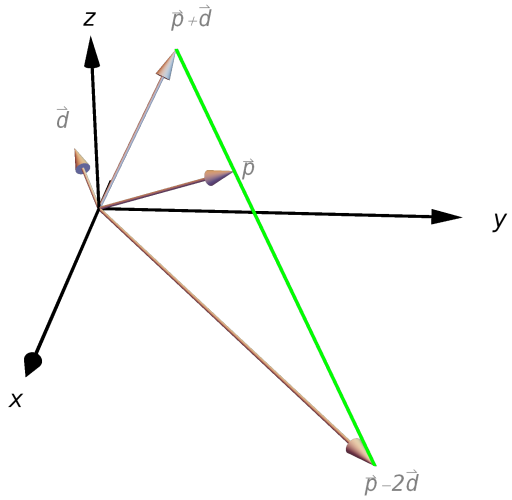


Figure 7.1: Defining a line in space.

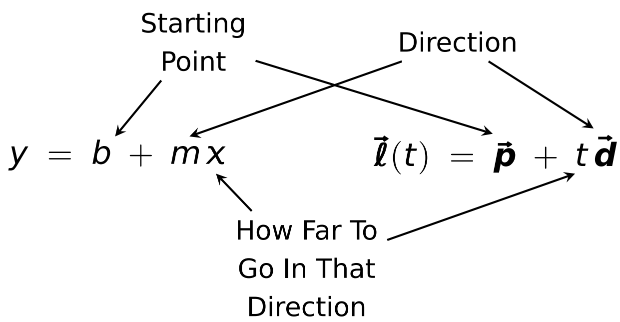


Figure 7.2: Understanding the vector equation of a line.

The equations exhibit the same structure: they give a starting point, define a direction, and state how far in that direction to travel.

Equation (7.1) is an example of a **vector–valued function** (*vectorfunctie*); the input of the function is a real number and the output is a vector. We will cover vector–valued functions extensively in Chapter 15.

There are other ways to represent a line. Let $\vec{p} = (x_0, y_0, z_0)$ and let $\vec{d} = (a, b, c)$. Then the equation of the line through $\vec{p}$ in the direction of $\vec{d}$ is:

$$\begin{aligned}\vec{\ell}(t) &= \vec{p} + t\vec{d}\\ &= (x_0, y_0, z_0) + t(a, b, c)\\ &= (x_0 + at, y_0 + bt, z_0 + ct).\end{aligned}$$

The last line states that the x-values of the line are given by $x = x_0 + at$, the y-values are given by $y = y_0 + bt$, and the z-values are given by $z = z_0 + ct$. These three equations, taken together, are the **parametric equations of the line** (*parametervoorstelling van een rechte*) through $\vec{p}$ in the direction of $\vec{d}$.

Finally, each of the equations for x, y and z above contain the variable t:

$$\begin{cases} x &= x_0 + at,\\ y &= y_0 + bt,\\ z &= z_0 + ct.\end{cases}$$

We can solve for t in each equation to obtain:

$$\begin{cases} t &= \dfrac{x-x_0}{a}, \\ t &= \dfrac{y-y_0}{b}, \\ t &= \dfrac{z-z_0}{c}. \end{cases}$$

assuming $a, b, c \neq 0$. Since t is equal to each expression on the right, we can set these equal to each other, forming the **Cartesian equations of the line** (*cartesische vergelijkingen van een rechte*) through $\vec{\boldsymbol{p}}$ in the direction of $\vec{\boldsymbol{d}}$:

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}.$$

Each representation has its own advantages, depending on the context. We summarize these three forms in the following definition, then give examples of their use.

Definitie 7.1 (Equations of lines in space)
Consider the **line** in space that passes through $\vec{\boldsymbol{p}} = (x_0, y_0, z_0)$ in the direction of $\vec{\boldsymbol{d}} = (a, b, c)$.

1. The **vector equation** of the line is

$$\vec{\boldsymbol{\ell}}(t) = \vec{\boldsymbol{p}} + t\vec{\boldsymbol{d}}.$$

2. The **parametric equations** of the line are

$$\begin{cases} x &= x_0 + at, \\ y &= y_0 + bt, \\ z &= z_0 + ct. \end{cases}$$

3. The **symmetric equations** of the line are

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}.$$

Example 7.1

Give all three equations, as given in Definition 7.1, of the line through $P = (2, 3, 1)$ in the direction of $\vec{\boldsymbol{d}} = (-1, 1, 2)$.

Solution

We identify the point $P = (2, 3, 1)$ with the vector $\vec{\boldsymbol{p}} = (2, 3, 1)$. Following the definition, we have

- the vector equation of the line is $\vec{\boldsymbol{\ell}}(t) = (2, 3, 1) + t(-1, 1, 2)$;
- the parametric equations of the line are

$$\begin{cases} x &= 2 - t, \\ y &= 3 + t, \\ z &= 1 + 2t; \end{cases}$$

and

- the symmetric equations of the line are

$$\frac{x-2}{-1} = \frac{y-3}{1} = \frac{z-1}{2}.$$

The resulting line is graphed in Figure 7.3.

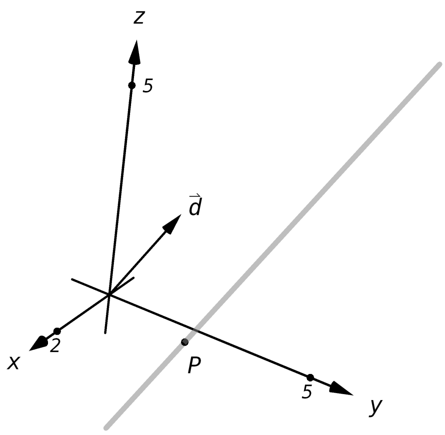


Figure 7.3: Graphing the line from Example 7.1.

The first two equations of the line are useful when a t value is given: one can immediately find the corresponding point on the line. These forms are good when calculating with a computer; most software programs easily handle equations in these formats.

7.1.2 Relative position of lines

In the plane, two distinct lines can either be parallel or they will intersect at exactly one point. In space, given the equations of two lines, it can sometimes be difficult to tell whether the lines are distinct or not (i.e., the same line can be represented in different ways). Given lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$, we have four possibilities: $\vec{\ell}_1$ and $\vec{\ell}_2$ are

the **same line** (*samenvallend*)	they share all points;
intersecting (*snijdend*) lines	share only 1 point;
parallel (*evenwijdig*) lines	$\vec{d}_1 \parallel \vec{d}_2$, no points in common;
skew (*kruisend*) lines	$\vec{d}_1 \nparallel \vec{d}_2$, no points in common.

Example 7.2

Consider lines ℓ_1 and ℓ_2, given in parametric equation form

$$\ell_1 : \begin{cases} x &= 1+3t \\ y &= 2-t \\ z &= t \end{cases} \quad \text{and} \quad \ell_2 : \begin{cases} x &= -2+4s \\ y &= 3+s \\ z &= 5+2s\,. \end{cases}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution

We start by looking at the directions of each line. Line ℓ_1 has the direction given by $\vec{d}_1 = (3, -1, 1)$ and line ℓ_2 has the direction given by $\vec{d}_2 = (4, 1, 2)$. It should be clear that $\vec{d}_1$ and $\vec{d}_2$ are not parallel, hence ℓ_1 and ℓ_2 are not the same line, nor are they parallel. Figure 7.4 verifies this fact. It shows the points and directions indicated by the equations of each line are identified.

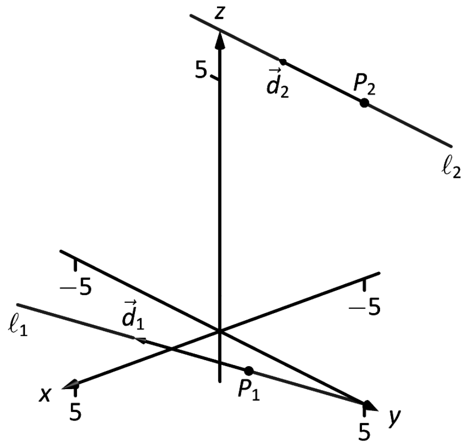


Figure 7.4: Sketching the lines from Example 7.2.

We next check to see if they intersect (if they do not, they are skew lines). To find if they intersect, we look for t and s values such that the respective x, y and z values are the same. That is, we want s and t such that:

$$\begin{cases} 1+3t &= -2+4s \\ 2-t &= 3+s \\ t &= 5+2s\,. \end{cases}$$

This is a relatively simple system of linear equations. Since the last equation is already solved for t, we substitute that value of t into the equation above it:

$$2-(5+2s) = 3+s \quad \Rightarrow \quad s = -2, \quad t = 1\,.$$

A key to remember is that we have three equations; we need to check if $s = -2$, $t = 1$ satisfies the first equation as well:

$$1+3(1) \neq -2+4(-2).$$

It does not. Therefore, we conclude that the lines ℓ_1 and ℓ_2 are skew.

One of the key points to understand from this section is this: to describe a line, we need a point and a direction. Lines are one of two fundamental objects of study in space. The other fundamental object is the plane, which we study in detail in the next section. Many complex three dimensional objects are studied by approximating their surfaces with lines and planes.

7.2 Planes

7.2.1 Definition

Any flat surface, such as a wall, table top or stiff piece of cardboard can be thought of as representing part of a **plane** (*vlak*). Consider a piece of cardboard with a point P marked on it. One can take a nail and stick it into the cardboard at P such that the nail is perpendicular to the cardboard. This nail provides a handle for the cardboard. Moving the cardboard around moves P to different locations in space. Tilting the nail but keeping P fixed tilts the cardboard. Both moving and tilting the cardboard defines a different plane in space. In fact, we can define a plane by: 1) the location of P in space, and 2) the direction of the nail.

The previous section showed that one can define a line given a point on the line and the direction of the line. One can make a similar statement about planes: we can define a plane in space given a point on the plane and the direction the plane faces. Once again, the direction information will be supplied by a vector, called a **normal vector** (*normaalvector*), that is orthogonal to the plane.

What exactly does orthogonal to the plane mean? Choose any two points P and Q in the plane, and consider the vector $\overrightarrow{PQ}$. We say a vector $\vec{\boldsymbol{n}}$ is orthogonal to the plane if $\vec{\boldsymbol{n}}$ is perpendicular to $\overrightarrow{PQ}$ for all choices of P and Q; that is, if $\vec{\boldsymbol{n}} \cdot \overrightarrow{PQ} = 0$ for all P and Q. This gives us way of writing an equation describing the plane. Let $P = (x_0, y_0, z_0)$ be a point in the plane and let $\vec{\boldsymbol{n}} = (a, b, c)$ be a normal vector to the plane. A point $Q = (x, y, z)$ lies in the plane defined by P and $\vec{\boldsymbol{n}}$ if and only if, $\overrightarrow{PQ}$ is orthogonal to $\vec{\boldsymbol{n}}$. Knowing $\overrightarrow{PQ} = (x - x_0, y - y_0, z - z_0)$, consider:

$$\begin{aligned} & \overrightarrow{PQ} \cdot \vec{\boldsymbol{n}} = 0 \\ \Leftrightarrow \quad & (x - x_0, y - y_0, z - z_0) \cdot (a, b, c) = 0 \\ \Leftrightarrow \quad & a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \end{aligned} \tag{7.2}$$

Equation (7.2) defines an implicit function describing the plane. More algebra produces:

$$ax + by + cz = ax_0 + by_0 + cz_0.$$

The right hand side is just a number, so we replace it with d:

$$ax + by + cz = d. \tag{7.3}$$

As long as $c \neq 0$, we can solve for z:

$$z = \frac{1}{c}(d - ax - by). \tag{7.4}$$

Equation (7.4) is especially useful as many computer programs can graph functions in this form. Equations (7.2) and (7.3) have specific names, given next.

Definitie 7.2 (Equations of a plane)
The **plane** passing through the point $P = (x_0, y_0, z_0)$ with normal vector $\vec{\boldsymbol{n}} = (a, b, c)$ can be described by an **equation with standard form**

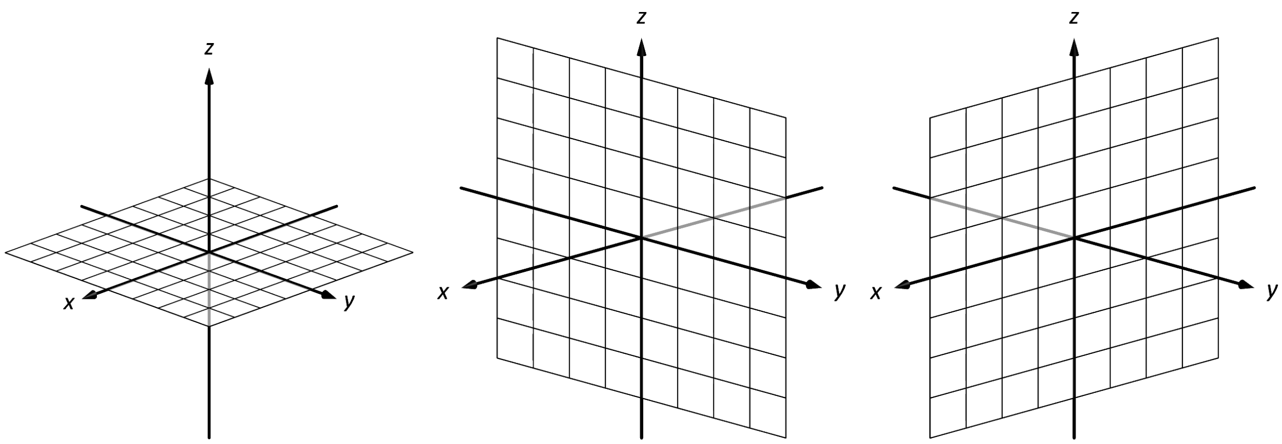
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0;$$

the **equation's general form** is

$$ax + by + cz = d.$$

Clearly, the coordinate axes naturally define three planes (shown in Figure 7.5), the **coordinate planes** (*coördinaatvlak*): the xy-plane, the yz-plane and the xz-plane. The xy-plane is characterized as the set of all points in space where the z-value is 0. This, in fact, gives us an equation that describes this plane: $z = 0$. Likewise, the xz-plane is all points where the y-value is 0, characterized by $y = 0$. Furthermore, the equation $x = 2$ describes all points in space where the x-value is 2.

A key to remember throughout this section is this: to find the equation of a plane, we need a point and a normal vector. We will give several examples of finding the equation of a plane, and in each one different types of information are given. In each case, we need to use the given information to find a point on the plane and a normal vector.



(a) The xy-plane. (b) The yz-plane. (c) The xz-plane.

Figure 7.5: The coordinate planes.

Example 7.3

Write the equation of the plane that passes through the points $P = (1, 1, 0)$, $Q = (1, 2, -1)$ and $R = (0, 1, 2)$ in standard form.

Solution

We need a vector $\vec{n}$ that is orthogonal to the plane. Since P, Q and R are in the plane, so are the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$; $\overrightarrow{PQ} \times \overrightarrow{PR}$ is orthogonal to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ and hence the plane itself.

It is straightforward to compute $\vec{n} = \overrightarrow{PQ} \times \overrightarrow{PR} = (2, 1, 1)$. We can use any point we wish in the plane (any of P, Q or R will do) and we arbitrarily choose P. Following Definition 7.2, the equation of the plane in standard form is

$$2(x-1) + (y-1) + z = 0.$$

The plane is sketched in Figure 7.6.

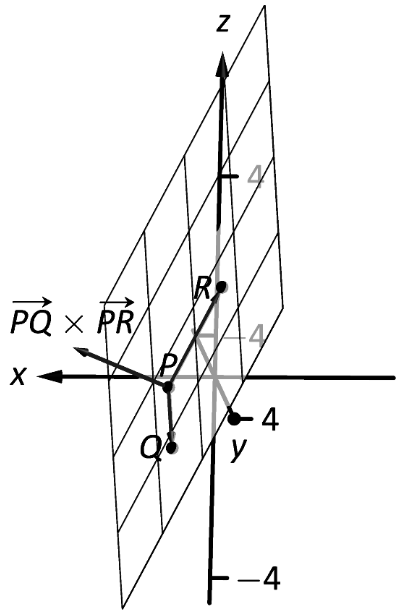


Figure 7.6: Sketching the plane in Example 7.3.

We have just demonstrated the fact that any three non-collinear points define a plane. This is why a three-legged stool does not rock; its three feet always lie in a plane. A four-legged stool will rock unless all four feet lie in the same plane.

Example 7.4

Verify that lines ℓ_1 and ℓ_2, whose parametric equations are given below, intersect, then give the equation of the plane that contains these two lines in general form.

$$\ell_1 : \begin{cases} x &= -5+2s \\ y &= 1+s \\ z &= -4+2s \end{cases} \qquad \ell_2 : \begin{cases} x &= 2+3t \\ y &= 1-2t \\ z &= 1+t \end{cases}$$

Solution

The lines clearly are not parallel. If they do not intersect, they are skew, meaning there is not a plane that contains them both. If they do intersect, there is such a plane.

To find their point of intersection, we set the x, y and z equations equal to each other and solve for s and t:

$$\begin{cases} -5+2s &= 2+3t \\ 1+s &= 1-2t \\ -4+2s &= 1+t \end{cases} \quad \Rightarrow \quad s=2, \quad t=-1.$$

When $s=2$ and $t=-1$, the lines intersect at the point $P=(-1,3,0)$.

Let $\vec{\boldsymbol{d}}_1 = (2,1,2)$ and $\vec{\boldsymbol{d}}_2 = (3,-2,1)$ be the directions of lines ℓ_1 and ℓ_2, respectively. A normal vector to the plane containing these the two lines will also be orthogonal to $\vec{\boldsymbol{d}}_1$ and $\vec{\boldsymbol{d}}_2$. Thus we find a normal vector $\vec{\boldsymbol{n}}$ by computing $\vec{\boldsymbol{n}} = \vec{\boldsymbol{d}}_1 \times \vec{\boldsymbol{d}}_2 = (5,4-7)$.

We can pick any point in the plane with which to write our equation; each line gives us infinite choices of points. We choose P, the point of intersection. We follow Definition 7.2 to write the plane's equation in general form:

$$\begin{aligned} 5(x+1)+4(y-3)-7z &= 0 \\ \Leftrightarrow \quad 5x+5+4y-12-7z &= 0 \\ \Leftrightarrow \quad 5x+4y-7z &= 7. \end{aligned}$$

The plane is $5x+4y-7z=7$; it is sketched in Figure 7.7.

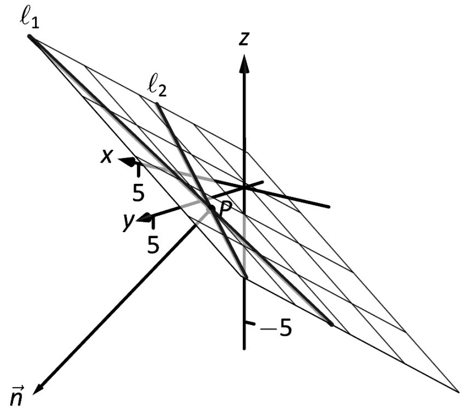


Figure 7.7: Sketching the plane in Example 7.4.

Having now defined lines and planes in space, it makes of course sense to look for the intersection between planes or between a plane and a line.

Example 7.5

Give the parametric equations of the line that is the intersection of the following planes:

$$\begin{aligned} p_1: &\quad x-(y-2)+(z-1)=0, \\ p_2: &\quad -2(x-2)+(y+1)+(z-3)=0. \end{aligned}$$

Solution

To find an equation of a line, we need a point on the line and the direction of the line. We can find a point on the line by solving each equation of the planes for z:

$$\begin{aligned} p_1: &\quad z=-x+y-1 \\ p_2: &\quad z=2x-y-2. \end{aligned}$$

We can now set these two equations equal to each other to find values of x and y where the planes have the same z value:

$$\begin{aligned} -x+y-1 &= 2x-y-2 \\ \Leftrightarrow \quad 2y &= 3x-1 \\ \Leftrightarrow \quad y &= \frac{1}{2}(3x-1). \end{aligned}$$

We can choose any value for x; we choose $x=1$. This determines that $y=1$. We can now use the equations of either plane to find z: when $x=1$ and $y=1$, $z=-1$ on both planes. We have found a point P on the line: $P=(1,1,-1)$.

We now need the direction of the line. Since the line lies in each plane, its direction is orthogonal to a normal vector for each plane. Considering the equations for p_1 and p_2, we can quickly determine their normal vectors. For p_1, $\vec{n}_1 = (1,-1,1)$ and for p_2, $\vec{n}_2 = (-2,1,1)$. A direction orthogonal to both of these directions is their cross product: $\vec{d} = \vec{n}_1 \times \vec{n}_2 = (-2,-3,-1)$.

The parametric equations of the line through $P = (1,1,-1)$ in the direction of $d = (-2,-3,-1)$ is:

$$\ell : \begin{cases} x &= 1-2t \\ y &= 1-3t \\ z &= -1-t. \end{cases}$$

The planes and line are graphed in Figure 7.8.

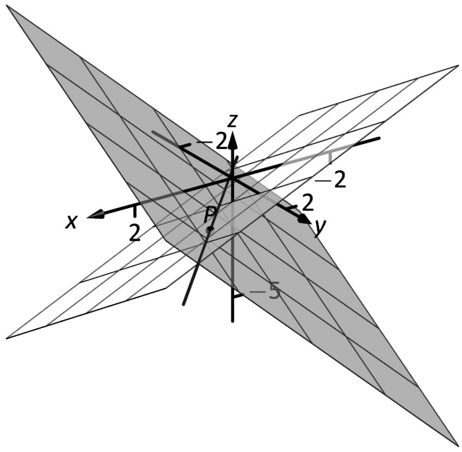


Figure 7.8: Graphing the planes and their line of intersection in Example 7.5.

Example 7.6

Find the point of intersection, if any, of the line $\ell(t) = (3,-3,-1) + t(-1,2,1)$ and the plane with equation in general form $2x + y + z = 4$.

Solution

The equation of the plane shows that the vector $\vec{n} = (2,1,1)$ is a normal vector to the plane, and the equation of the line shows that the line moves parallel to $\vec{d} = (-1,2,1)$. Since these are not orthogonal, we know there is a point of intersection.

To find the point of intersection, we need to find a t-value such that $\ell(t)$ satisfies the equation of the plane. Rewriting the equation of the line with parametric equations will help:

$$\ell(t) : \begin{cases} x = 3-t \\ y = -3+2t \\ z = -1+t. \end{cases}$$

Replacing x, y and z in the equation of the plane with the expressions containing t found in the equation of the line allows us to determine a t value that indicates the point of intersection:

$$\begin{aligned} 2x+y+z &= 4 \\ \Leftrightarrow \quad 2(3-t)+(-3+2t)+(-1+t) &= 4 \\ \Leftrightarrow \quad t &= 2\,. \end{aligned}$$

When $t = 2$, the point on the line satisfies the equation of the plane; that point is $\ell(2) = (1,1,1)$.

Thus the point $(1,1,1)$ is the point of intersection between the plane and the line, illustrated in Figure 7.9.

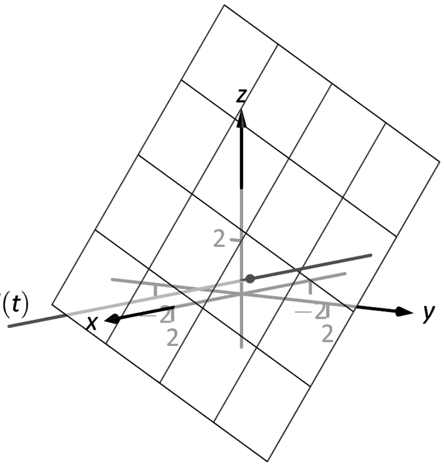


Figure 7.9: Illustrating the intersection of a line and a plane in Example 7.6.

Finite element method

One of the most popular numerical method for solving problems of engineering that relies on an approximation of surfaces by means of small planes is the finite element method. It proceeds by dividing the domain of the problem into a collection of subdomains (i.e. small planes), with each subdomain represented by a set of element equations to the original problem. For instance, the stresses on the surface of a human knee joint can be assessed by means of this method.

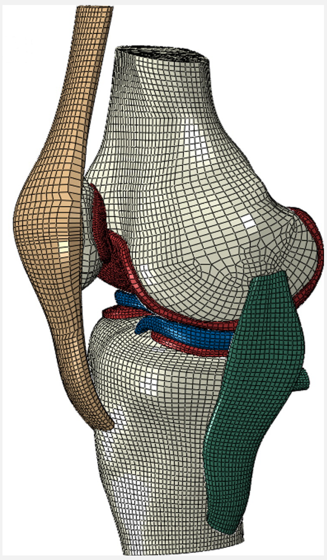


Figure 7.10: Illustrating finding the distance from a point to a plane.

In the final section of this chapter we will investigate more complex three-dimensional objects.

7.3 Three-dimensional objects

7.3.1 Spheres and cylinders

Just as a circle is the set of all points in the plane equidistant from its centre, a sphere is the set of all points in space that are equidistant from a given point. Equation (6.6) allows us to write an equation of the sphere.

We start with a point $C = (a, b, c)$ which is to be the centre of a sphere with radius r. If a point $P = (x, y, z)$ lies on the sphere, then P is r units from C; that is,

$$||\overrightarrow{PC}|| = \sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2} = r.$$

Squaring both sides, we get the standard equation of a sphere in space with center at $C = (a, b, c)$ with radius r, as given in the following definition.

Definitie 7.3 (Standard equation of a sphere)
The standard equation of the **sphere** (*bol*) with radius r, centred at $C = (a, b, c)$, is

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2.$$

The equation of a sphere is an example of an implicit function defining a surface in space. In the case of a sphere, the variables x, y and z are all used. We now consider a situation where surfaces are defined where one or two of these variables are absent, in addition to the coordinate planes that we encountered before.

The equation $x = 1$ obviously lacks the y and z variables, meaning it defines points where the y and z coordinates can take on any value. Now consider the equation $x^2 + y^2 = 1$ in space. In the plane, this equation describes a circle of radius 1, centred at the origin. In space, the z coordinate is not specified, meaning it can take on any value. In Figure 7.11(a), we show part of the graph of the equation $x^2 + y^2 = 1$ by sketching 3 circles: the bottom one has a constant z-value of -1.5, the middle one has a z-value of 0 and the top circle has a z-value of 1. By plotting all possible z-values, we get the surface shown in Figure 7.11(b). This surface is a cylinder.

Definitie 7.4 (Cylinder)
Let C be a curve in a plane and let L be a line not parallel to C. A **cylinder** (*cilinder*) is the set of all lines parallel to L that pass through C. The curve C is the **directrix** (*richtkromme*) of the cylinder, and the lines are the **rulings** (*beschrijvenden*).

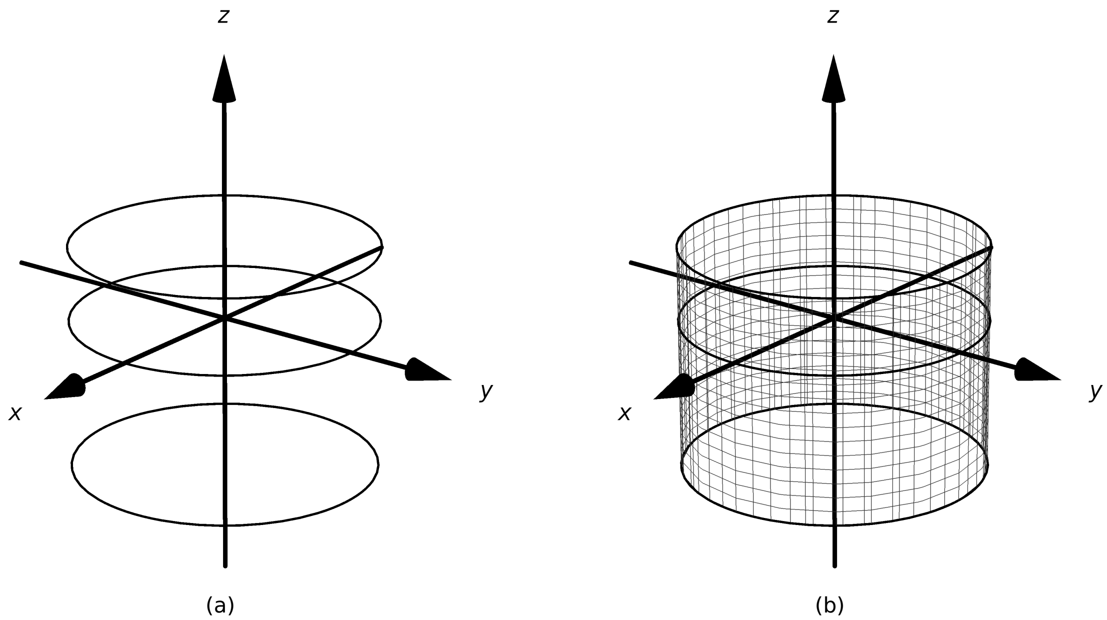


Figure 7.11: Sketching $x^2 + y^2 = 1$.

In this text, we consider curves C that lie in planes parallel to one of the coordinate planes, and lines L that are perpendicular to these planes, forming **right cylinders** (*rechte cilinder*). Thus the directrix can be defined using equations involving 2 variables, and the rulings will be parallel to the axis of the third variable.

In the example preceding the definition, the curve $x^2 + y^2 = 1$ in the xy-plane is the directrix and the rulings are lines parallel to the z-axis. Actually, any circle shown in Figure 7.11(a) can be considered a directrix; we simply choose the one where $z = 0$.

Example 7.7

Graph the following cylinders.

1. $z = y^2$
2. $x = \sin(z)$

Solution

1. We can view the equation $z = y^2$ as a parabola in the yz-plane, as illustrated in Figure 7.12(a). As x does not appear in the equation, the rulings are lines through this parabola parallel to the x-axis, shown in Figure 7.12(b). These rulings give an idea as to what the surface looks like, drawn in Figure 7.12(c).

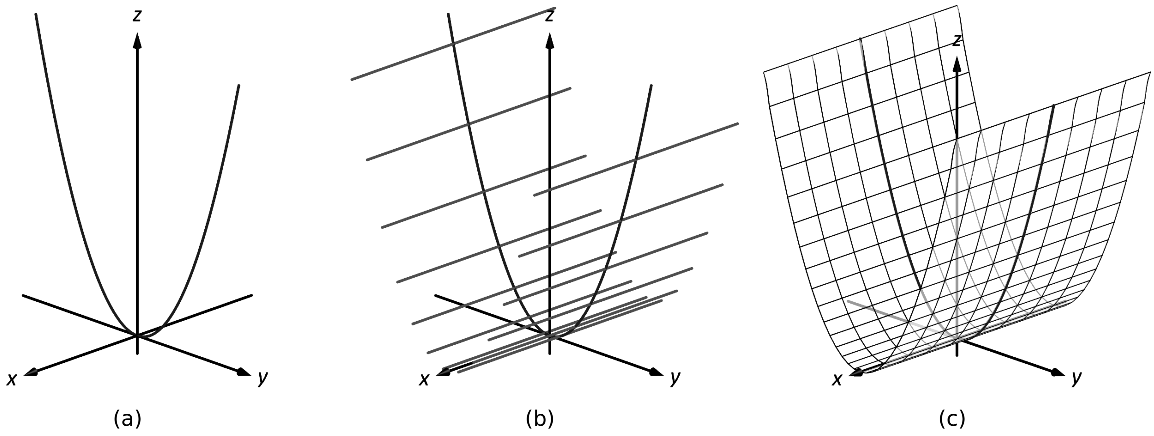


Figure 7.12: Sketching the parabolic cylinder defined by $z = y^2$.

2. We can view the equation $x = \sin(z)$ as a sine curve that exists in the xz-plane, as shown in Figure 7.13(a). The rules are parallel to the y-axis as the variable y does not appear in the equation $x = \sin(z)$; some of these are shown in Figure 7.13(b). The surface is shown in Figure 7.13(c).

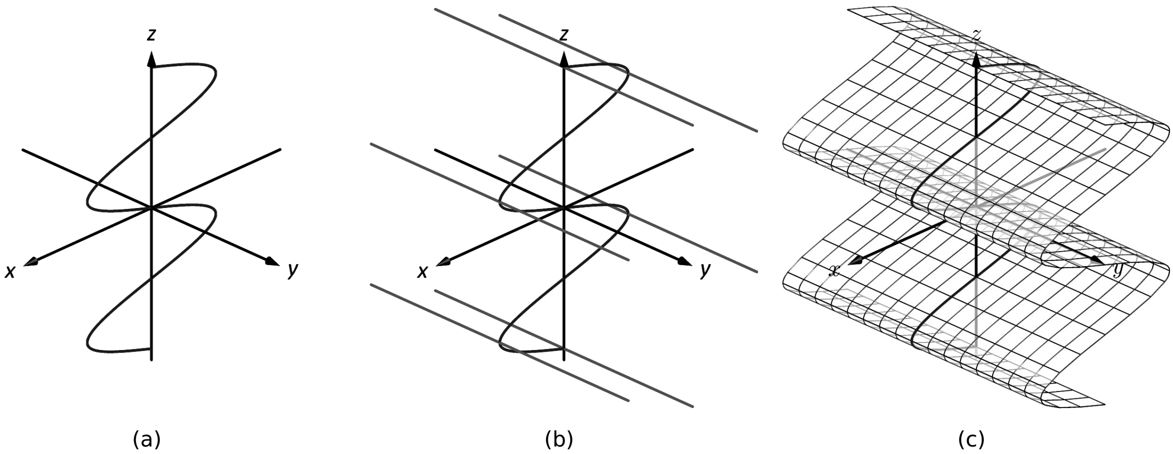


Figure 7.13: Sketching the cylinder defined by $x = \sin(z)$.

7.3.2 Surface of revolution

For the sake of later chapters, we now consider how to find the equation of the surface of a solid formed by revolving a curve about a horizontal or vertical axis. These surfaces are called **surfaces of revoluting** (*omwentelingslichaam*).

Consider the surface formed by revolving $y = \sqrt{x}$ about the x-axis. Cross–sections of this surface parallel to the yz-plane are circles, as shown in Figure 7.14(a). Each circle has equation of the form $y^2 + z^2 = r^2$ for some radius r. The radius is a function of x; in fact, it is $r(x) = \sqrt{x}$. Thus the equation of the surface shown in Figure 7.14(b) is $y^2 + z^2 = (\sqrt{x})^2$.

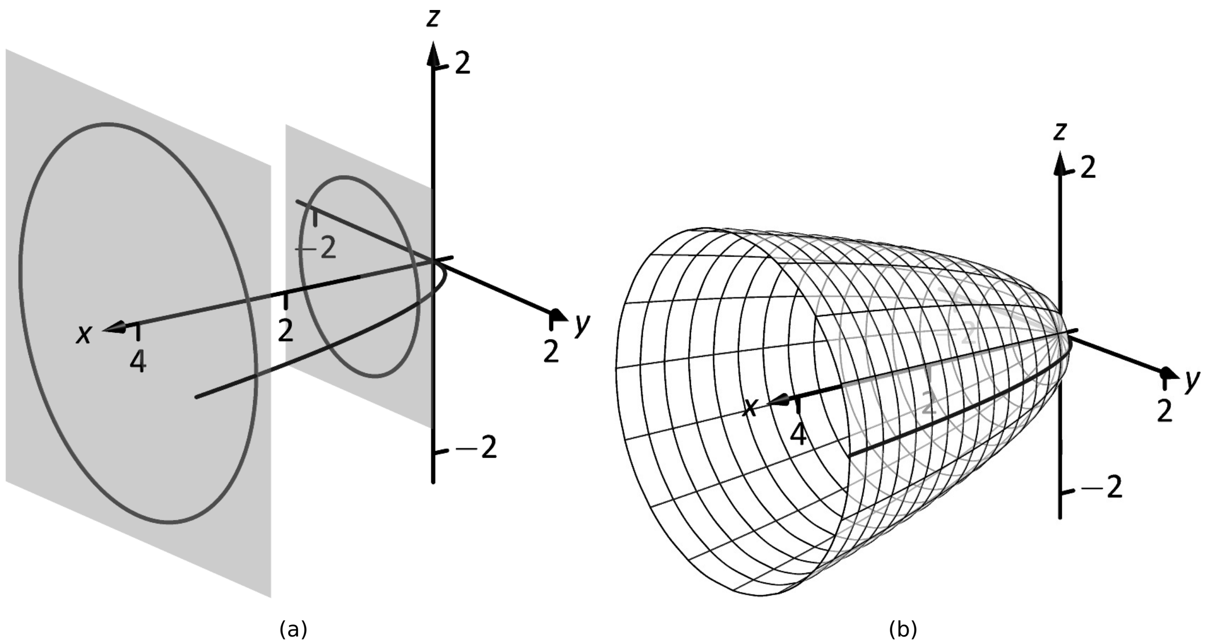


Figure 7.14: Introducing surfaces of revolution.

We generalize the above principles to give the equations of surfaces formed by revolving curves about the coordinate axes.

Definitie 7.5 (Surfaces of revolution, Part 1)
Let r be a radius function.

1. The equation of the surface formed by revolving $y = r(x)$ or $z = r(x)$ about the x-axis is $y^2 + z^2 = r(x)^2$.
2. The equation of the surface formed by revolving $x = r(y)$ or $z = r(y)$ about the y-axis is $x^2 + z^2 = r(y)^2$.
3. The equation of the surface formed by revolving $x = r(z)$ or $y = r(z)$ about the z-axis is $x^2 + y^2 = r(z)^2$.

Example 7.8

Let $y = \sin(z)$ on $[0, \pi]$. Find the equation of the surface of revolution formed by revolving $y = \sin(z)$ about the z-axis.

Solution

Using Definition 7.5, we find the surface has equation $x^2 + y^2 = \sin^2(z)$. The curve is sketched in Figure 7.15(a) and the surface is drawn in Figure 7.16(b).

Note how the surface and hence the resulting equation is the same if we began with the curve $x = \sin(z)$, which is also drawn in Figure 7.15(a).

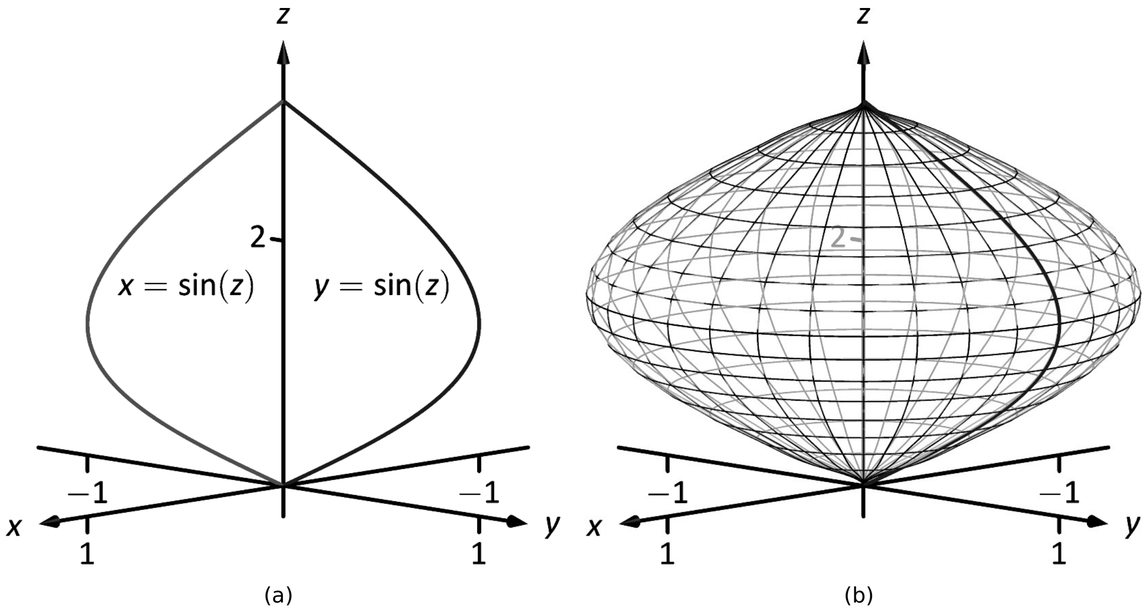


Figure 7.15: Revolving $y = \sin(z)$ about the z-axis in Example 7.8.

This particular method of creating surfaces of revolution is limited. Our current method of forming surfaces can only rotate, for instance, $y = \sin(x)$ about the x-axis. Trying to rewrite $y = \sin(x)$ as a function of y is not trivial, as simply writing $x = \arcsin(y)$ only gives part of the region we desire. What we desire is a way of writing the surface of revolution formed by rotating $y = f(x)$ about the y-axis. We start by first recognizing this surface is the same as revolving $z = f(x)$ about the z-axis. This will give us a more natural way of viewing the surface.

A value of x is a measurement of distance from the z-axis. At the distance r, we plot a z-height of $f(r)$. When rotating $f(x)$ about the z-axis, we want all points a distance of r from the z-axis in the xy-plane to have a z-height of $f(r)$. All such points satisfy the equation $r^2 = x^2 + y^2$; hence $r = \sqrt{x^2 + y^2}$. Replacing r with $\sqrt{x^2 + y^2}$ in $f(r)$ gives $z = f(\sqrt{x^2 + y^2})$. This is the equation of the surface, and is clearly stated in the following definition.

Definitie 7.6 (Surfaces of revolution, Part 2)
Let $z = f(x)$, $x \geq 0$, be a curve in the xz-plane. The surface formed by revolving this curve about the z-axis has equation

$$z = f\big(\sqrt{x^2 + y^2}\big).$$

Example 7.9

Find the equation of the surface found by revolving $z = \sin(x)$ about the z-axis.

Solution

Using Definition 7.6, the surface has equation $z = \sin\big(\sqrt{x^2 + y^2}\big)$. The curve and surface are graphed in Figure 7.16(b).

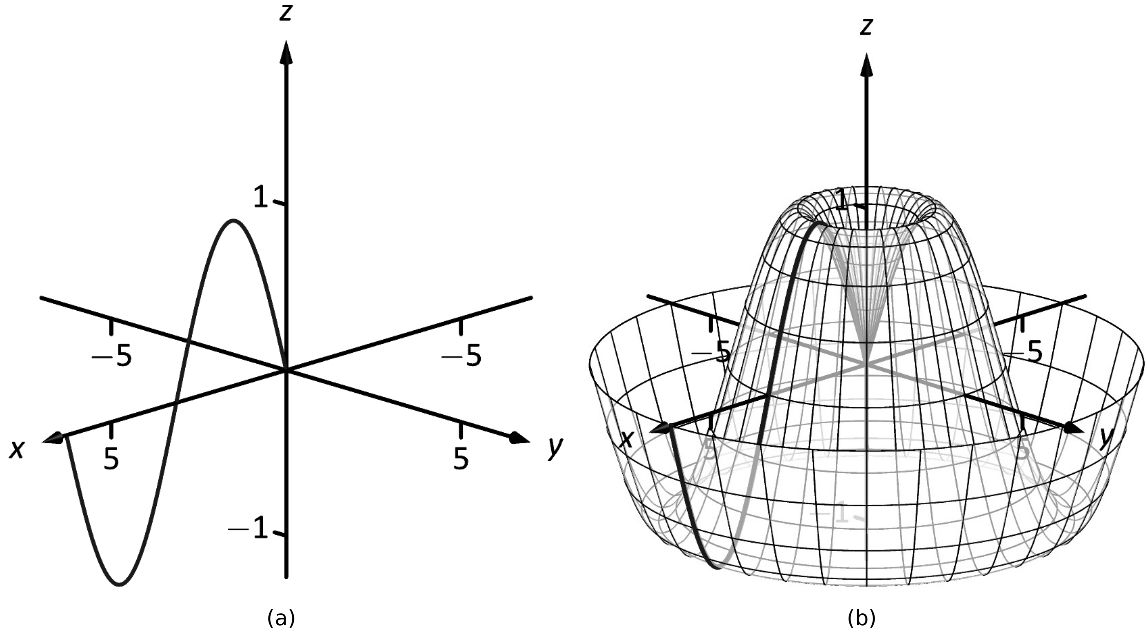


Figure 7.16: Revolving $z = \sin(x)$ about the z-axis in Example 7.9.

7.3.3 Quadratic surfaces

Spheres, planes and cylinders are important surfaces to understand. We now consider one last type of surface, a quadric surface. Essentially, they are the three-dimensional extension of the conic sections we discussed in Section 4.4. Their definition is given below.

Definitie 7.7 (Quadric surface)
A **quadric surface** (*kwadratisch oppervlak*) is the graph of the general second–degree equation in three variables:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0.$$

When the coefficients d, e or f are not zero, the basic shapes of the quadric surfaces are rotated in space. We will focus on quadric surfaces where these coefficients are 0; we will not consider rotations. There are six basic quadric surfaces: the elliptic paraboloid, elliptic cone, ellipsoid, hyperboloid of one sheet, hyperboloid of two sheets, and the hyperbolic paraboloid.

We study each shape by considering **traces**, that is, intersections of each surface with a plane parallel to a coordinate plane. For instance, consider the elliptic paraboloid $z = x^2/4 + y^2$, shown in Figure 7.17. If we intersect this shape with the plane $z = d$ (i.e., replace z with d), we have the equation:

$$d = \frac{x^2}{4} + y^2.$$

Divide both sides by d:

$$1 = \frac{x^2}{4d} + \frac{y^2}{d}.$$

This describes an ellipse – so cross sections parallel to the x-y coordinate plane are ellipses. This ellipse is drawn in the figure. Now consider cross sections parallel to the xz-plane. For instance, letting $y = 0$

gives the equation $z = x^2/4$, clearly a parabola. Intersecting with the plane $x = 0$ gives a cross section defined by $z = y^2$, another parabola. These parabolas are also sketched in the figure.

Thus we see where the elliptic paraboloid gets its name: some cross sections are ellipses, and others are parabolas.

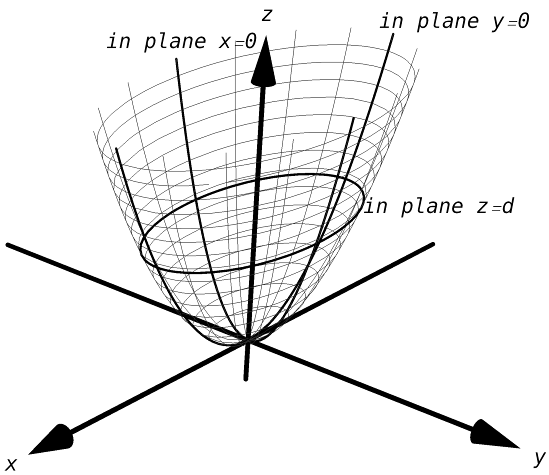
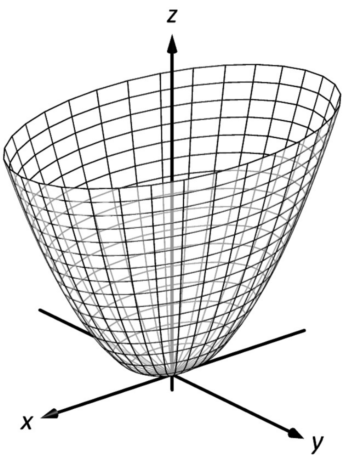


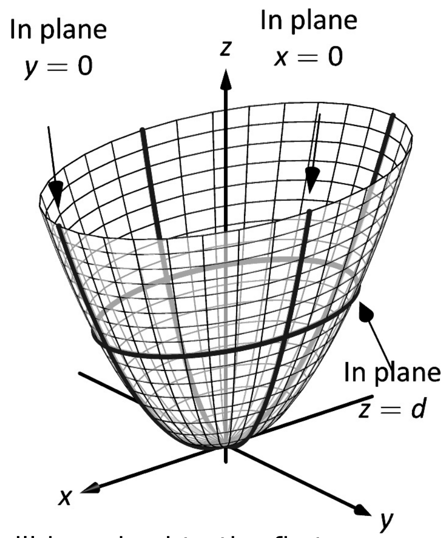
Figure 7.17: The elliptic paraboloid $z = x^2/4 + y^2$.

Such an analysis can be made with each of the quadric surfaces. We give a sample equation of each, provide a sketch with representative traces, and describe these traces.

Elliptic paraboloid (*elliptische paraboloïde*): $z = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$



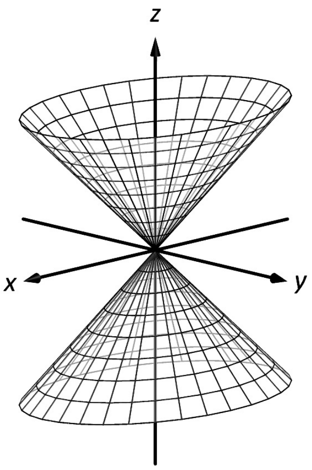
Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Ellipse



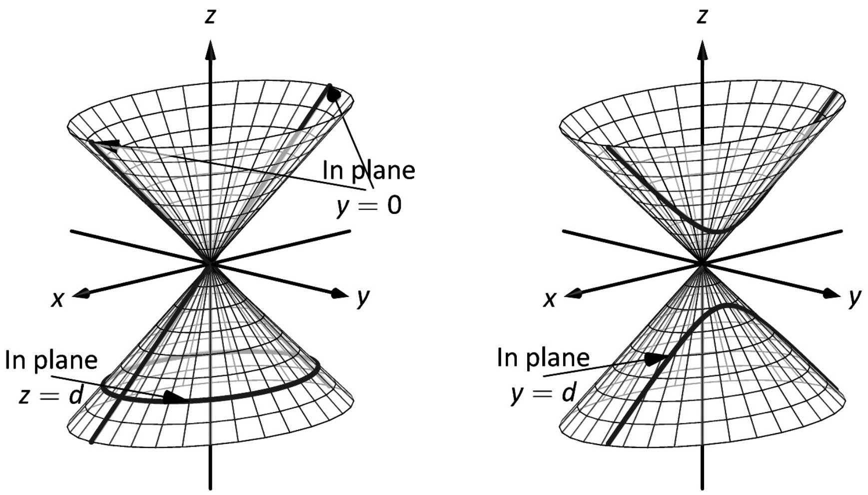
One variable in the equation of the elliptic paraboloid will be raised to the first power; above, this is the z variable. The paraboloid will open in the direction of this variable's axis. Thus $x = y^2/a^2 + z^2/b^2$ is an elliptic paraboloid that opens along the x-axis.

Multiplying the right hand side by (-1) defines an elliptic paraboloid that opens in the opposite direction.

Elliptic cone (*elliptische kegel*): $z^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$



Plane	Trace
$x = 0$	Crossed Lines
$y = 0$	Crossed Lines
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

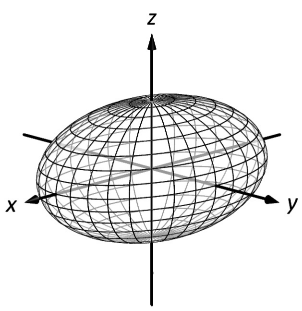


One can rewrite the governing equation as

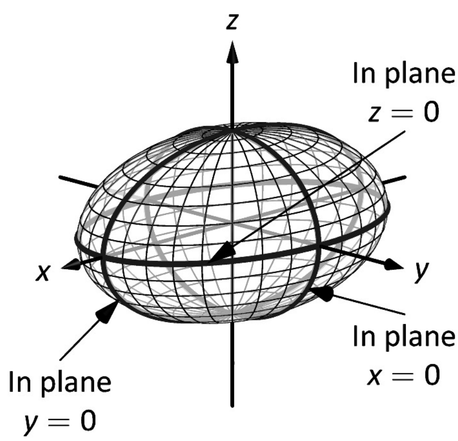
$$z^2 - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$$

The one variable with a positive coefficient corresponds to the axis that the cones open along.

Ellipsoid (*ellipsoïde*): $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$



Plane	Trace
$x = d$	Ellipse
$y = d$	Ellipse
$z = d$	Ellipse



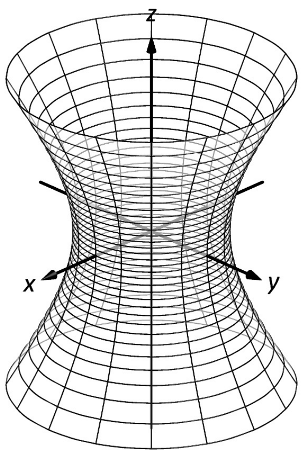
If $a = b = c \neq 0$, the ellipsoid is a sphere with radius a.

Earth ellipsoid

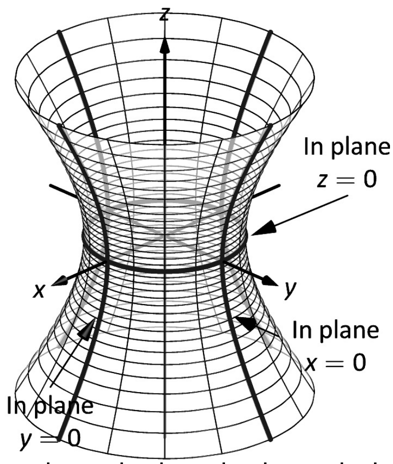
An Earth ellipsoid is a mathematical figure approximating the Earth's form, used as a reference frame for computations in the geosciences. Various different ellipsoids have been used as approximations. It is an ellipsoid of revolution whose minor axis, which connects the geographical North Pole and South Pole, is approximately aligned with the Earth's axis of rotation.

Many methods exist for determination of the axes of an Earth ellipsoid, but several ellipsoids are of special importance, such as the Bessel ellipsoid of 1841 and (for GPS positioning) the WGS84 ellipsoid. In the latter, the semi major axis measures 6 378 137.0 m, while its semi minor axis measures approximately 6 356 752.314 245 m.

Hyperboloid of one sheet (*eenbladige hyperboloïde*): $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$

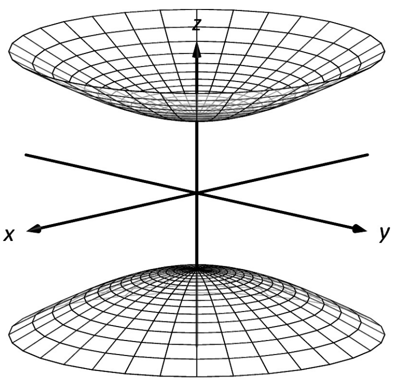


Plane	**Trace**
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

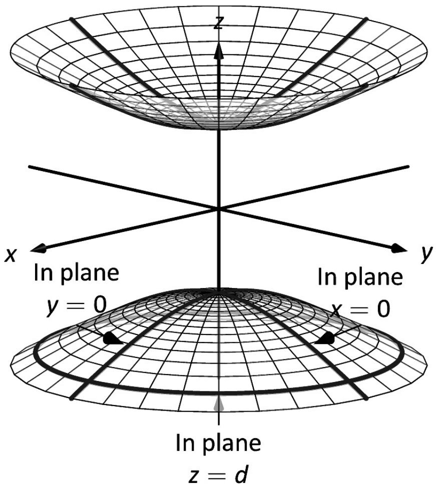


The one variable with a negative coefficient corresponds to the axis that the hyperboloid opens along.

Hyperboloid of two sheets (*tweebladige hyperboloïde*): $\dfrac{z^2}{c^2} - \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$

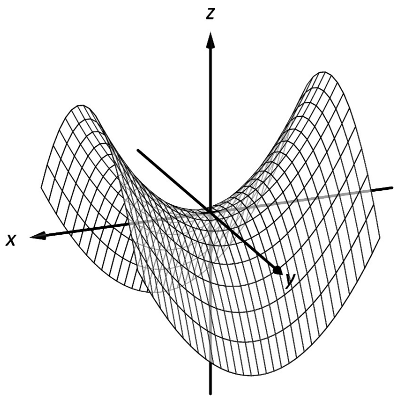


Plane	**Trace**
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

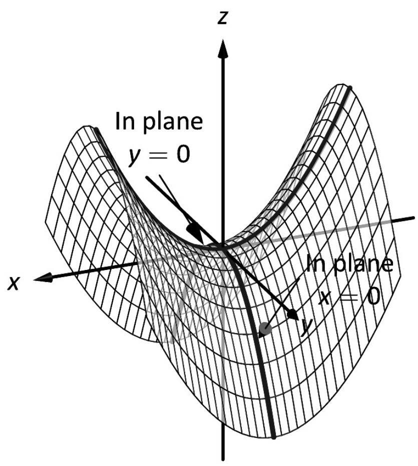


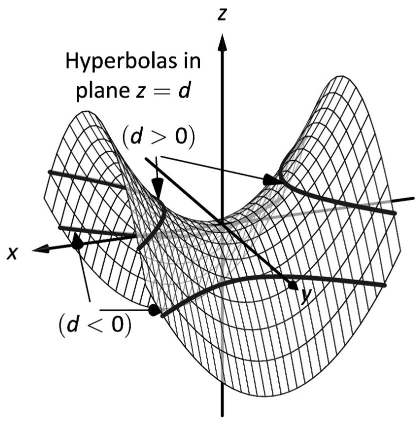
The one variable with a positive coefficient corresponds to the axis that the hyperboloid opens along. In the case illustrated, when $|d| < |c|$, there is no trace.

Hyperbolic Paraboloid (*hyperbolische paraboloïde*): $z = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$



Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Hyperbola





The parabolic traces will open along the axis of the one variable that is raised to the first power.

Example 7.10

Sketch the quadratic surface defined by the given equation.

1. $y = \dfrac{x^2}{4} + \dfrac{z^2}{16}$ 2. $x^2 + \dfrac{y^2}{9} + \dfrac{z^2}{4} = 1$ 3. $z = y^2 - x^2$.

Solution

1. We first identify the quadratic by pattern–matching with the equations given previously. Only two surfaces have equations where one variable is raised to the first power, the elliptic paraboloid and the hyperbolic paraboloid. In the latter case, the other variables have different signs, so we conclude that this describes an elliptic paraboloid. As the variable with the first power is y, we note the paraboloid opens along the y-axis.

   To make a decent sketch by hand, we need only draw a few traces. In this case, the traces $x = 0$ and $z = 0$ form parabolas that outline the shape.

   $x = 0$: The trace is the parabola $y = z^2/16$.

   $z = 0$: The trace is the parabola $y = x^2/4$.

   Graphing each trace in the respective plane creates a sketch as shown in Figure 7.18(a). This is enough to give an idea of what the paraboloid looks like. The surface is filled in

Figure 7.18(b).

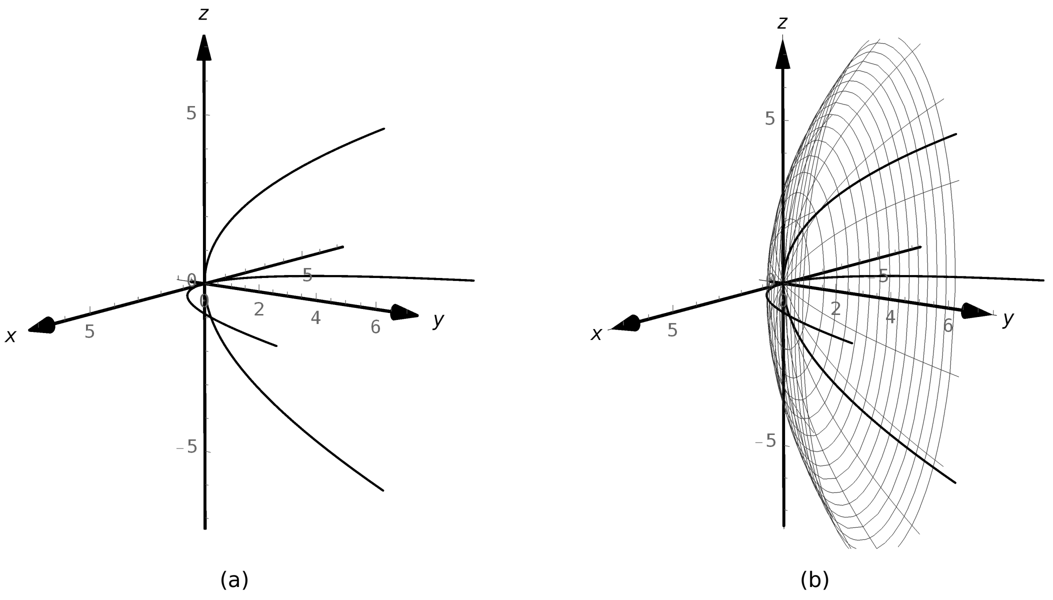


Figure 7.18: Sketching the elliptic paraboloid from Example 7.10.1.

2. This is an ellipsoid. We can get a good idea of its shape by drawing the traces in the coordinate planes.

$x = 0$: The trace is the ellipse $y^2/9 + z^2/4 = 1$. The major axis is along the y–axis with length 6 the minor axis is along the z-axis with length 4.

$y = 0$: The trace is the ellipse $x^2 + z^2/4 = 1$. The major axis is along the z-axis, and the minor axis has length 2 along the x-axis.

$z = 0$: The trace is the ellipse $x^2 + y^2/9 = 1$, with major axis along the y-axis.

Graphing each trace in the respective plane creates a sketch as shown in Figure 7.19(a). Filling in the surface gives Figure 7.19(b).

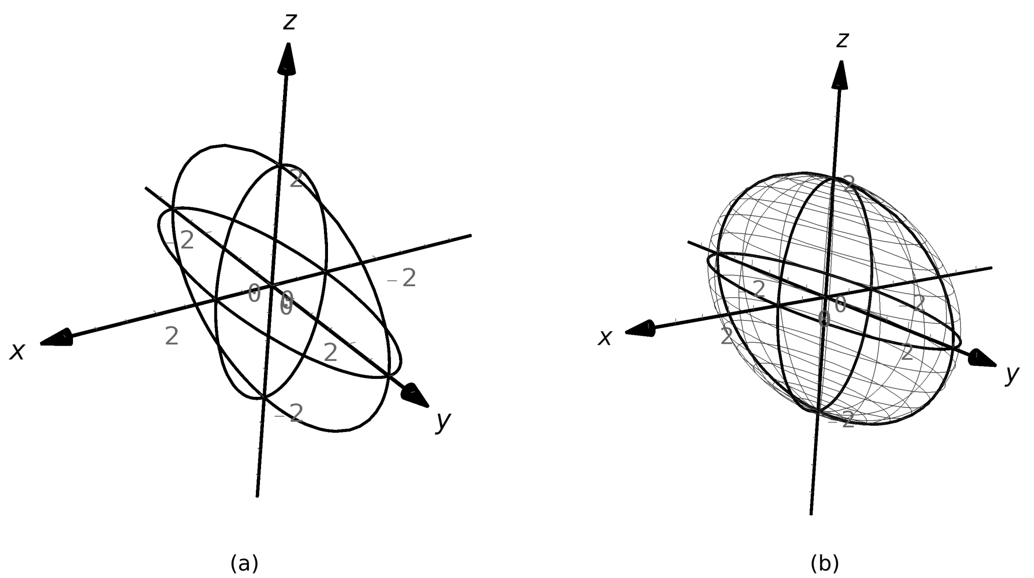


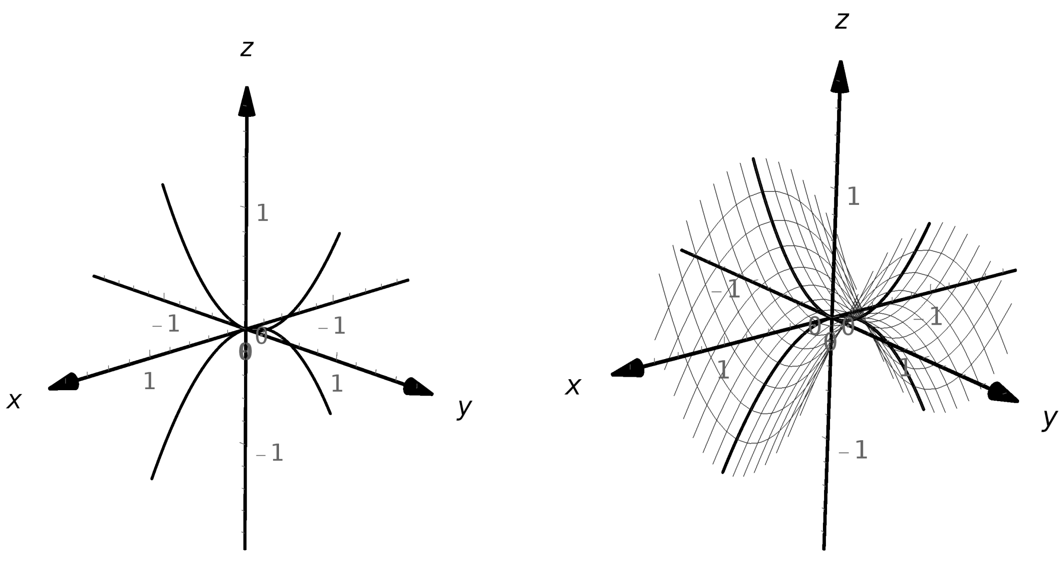
Figure 7.19: Sketching the ellipsoid from Example 7.10.2.

3. This defines a hyperbolic paraboloid. Consider the traces in the yz- and xz-planes:

$x = 0$: The trace is $z = y^2$, a parabola opening up in the yz-plane.

$y = 0$: The trace is $z = -x^2$, a parabola opening down in the xz-plane.

Sketching these two parabolas gives a sketch like that in Figure 7.20(a), and filling in the surface gives a sketch like in Figure 7.20(b).



(a) (b)

Figure 7.20: Sketching the hyperbolic paraboloid from Example 7.10.3.

7.4 Exercises

Lines

Assignment 7.1 — Prove that the points $A = (1,2,3)$, $B = (2,3,4)$ and $C = (3,4,5)$ are on the same line.

Assignment 7.2 — Determine the parametric representation of the line that is the intersection of the following planes:

$$\begin{aligned} p_1 : &\quad x - 2y + 3z = 0, \\ p_2 : &\quad 2x + 3y - 4z = 4. \end{aligned}$$

Assignment 7.3 — Determine the cartesian equations, parameter equations, and vector equation of the line ℓ_1 in each of the following cases.

(a) through $O = (0,0,0)$ and in the direction of vector $\vec{\boldsymbol{d}} = (1,2,3)$

(b) through $A = (3,4,1)$ and $B = (1,4,0)$

(c) through $A=(1,2,0)$ and parallel to

$$\ell_2: \frac{2x+2}{3}=\frac{y-1}{2}=\frac{2z+3}{1}$$

(d) through $A=(1,2,3)$ and parallel to $\vec{\boldsymbol{d}}=(2,-3,-4)$

(e) through $A=(-1,0,1)$ and perpendicular to $p: 2x-y+7z=12$

(f) through $O=(0,0,0)$ and parallel to the intersection of $p_1: x+2y-z=2$ and $p_2: 2x-y+4z=5$

(g) through $A=(2,-1,-1)$ and parallel with $p_1: x+y=0$ and $p_2: x-y+2z=0$.

Assignment 7.4 — Determine for all cases below if the lines ℓ_1 and ℓ_2 are parallel, (perpendicular) intersecting or skew. If they are intersecting, determine the coordinates of their intersection.

(a) $\ell_1: \dfrac{x-2}{3}=\dfrac{y-2}{4}=\dfrac{z}{2}$ and $\ell_2: \dfrac{x}{6}=\dfrac{3y+2}{24}=\dfrac{3z+4}{12}$

(b) $\ell_1: \begin{cases} 2x-3y+z=0 \\ x+y+z=0 \end{cases}$ and $\ell_2: \begin{cases} 2x-y=1 \\ x+2y-z=0 \end{cases}$

(c) $\ell_1: \begin{cases} 2x+3y-z=4 \\ x-y+3z=1 \end{cases}$ and $\ell_2: \dfrac{x-2}{8}=\dfrac{y-1}{-7}=\dfrac{-z+5}{5}$

(d) $\ell_1: \begin{cases} x+y+z=1 \\ x+2z=0 \end{cases}$ and $\ell_2: \begin{cases} x+y=4 \\ z=0 \end{cases}$

Planes

Assignment 7.5 — Determine the cartesian equation of the plane p in each of the following cases.

(a) through $P=(1,4,2)$ and with normal vector $\vec{\boldsymbol{n}}=(3,1,-4)$

(b) through $P_1=(1,-2,1)$, $P_2=(2,0,3)$ and $P_3=(0,1,-1)$

(c) through $P=(1,2,-3)$ and perpendicular to

$$\ell: \begin{cases} x=t \\ y=-2-2t \\ z=1+3t \end{cases},$$

with parameter $t\in\mathbb{R}$.

(d) through $P=(0,0,1)$ and perpendicular to

$$\ell: \frac{2x+2}{1}=\frac{y-1}{3}=\frac{z+1}{-2}$$

(e) through the lines

$$\ell_1 : \frac{x+1}{2} = \frac{y-2}{3} = \frac{z-1}{1} \quad \text{and} \quad \ell_2 : \frac{x+1}{1} = \frac{y-2}{-1} = \frac{z-1}{2}$$

(f) through $P_1 = (1,1,1)$ and $P_2 = (2,0,3)$ and perpendicular to $p : x+2y-3z=0$

(g) through the cross-section of $p_1 : 2x+3y-z=0$ and $p_2 : x-4y+2z=-5$ and through $P=(-2,0,-1)$

Assignment 7.6 — We consider the planes $p_1 : 3x+2y-z+3=0$, $p_2 : -x+2y+z-3=0$ and $p_3 : 6x+4y-2z-3=0$.

(a) Show that p_1 en p_2 are perpendicular.

(b) Show that p_1 en p_3 are parallel.

(c) What can you conclude about the location of p_2 and p_3 relative to each other? Solve this question without doing any math.

Assignment 7.7 — Determine the angle between $p_1 : 3x-2y+z-4=0$ and $p_2 : x+4y-3z-2=0$.

Three-dimensional objects

Assignment 7.8 — Determine an equation for the following surfaces of revolution created by:

(a) revolving the curve $4x^2+9y^2=36$ around the y-axis.

(b) revolving the curve $x=2z^2$ around the x-axis.

(c) revolving the curve $(y-z)^2+z^2=1$ around the z-axis.

Assignment 7.9 — Name and draw the surfaces below.

(a) $3x^2-2y^2+z^2+3=0$

(b) $-4x^2+2z^2-3=0$

(c) $3x^2-y^2=z^2$

(d) $(x-1)^2+(y-2)^2=(z-4)^2$

(e) $2x^2+3z^2=1$

(f) $y^2+2z^2=x$

(g) $x^2+4y^2+9z^2=36$

(h) $\frac{25}{9}x^2-25y^2+z^2=25$

Review exercises

Assignment 7.10 — Determine the equation of the line through $P=(2,-1,5)$ and perpendicular to $p: 3x+2y-2z-7=0$.

Assignment 7.11 — Determine the positions of the lines and the planes below relative to each other. If they intersect, determine whether this is perpendicularly.

(a) $\ell: 1-x=y=z$ and $p: 3x+2y+z-3=0$

(b) $\ell: \frac{x-5}{2}=\frac{y-2}{3}=\frac{z-3}{8}$ and $p: x+2y-z-2=0$

(c) $\ell: \begin{cases} 2x+y+z=1 \\ 3x+y-z=4 \end{cases}$ and $p: 5x+2y+2z=8$

(d) $\ell: x=y=z$ and $p: x+y+z=3$

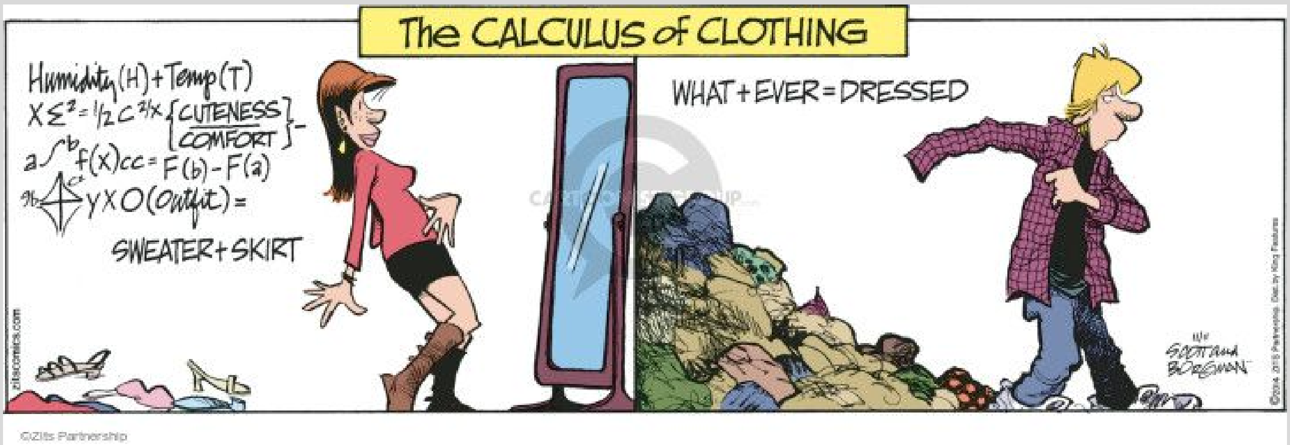
(e) $\ell: \begin{cases} 3x+2y=8 \\ x+3z=4 \end{cases}$ and $p: x+2y-6z=2$

(f) $\ell: \begin{cases} 2x+y+z=1 \\ x-y+3z=0 \end{cases}$ and $p: x+2y-2z=1$

(g) $\ell: \frac{x-1}{2}=\frac{y+1}{3}=\frac{z-1}{-2}$ and $p: 3x-y+2z-5=0$

PART II

SINGLE VARIABLE CALCULUS



Mathematics as an expression of the human mind reflects the active will, the contemplative reason, and the desire for aesthetic perfection. Its basic elements are logic and intuition, analysis and construction, generality and individuality.

— Richard Courant —

8

Limits and continuity

Calculus means a method of calculation or reasoning. When one computes the sales tax on a purchase, one employs a simple calculus. Proving a theorem in geometry employs another calculus. Despite the wonderful advances in mathematics that had taken place into the first half of the 17^{th} century, mathematicians and scientists were keenly aware of what they could not do. In particular, two important concepts eluded mastery by the great thinkers of that time: area and rates of change.

Area seems innocuous enough; areas of circles, rectangles, parallelograms, etc., are standard topics of study for students today just as they were then. However, the areas of arbitrary shapes could not be computed, even if the boundary of the shape could be described exactly. Rates of change were also important. When an object moves at a constant rate of change, then distance = rate × time. But what if the rate is not constant – can distance still be computed?

It turns out that these two concepts were related. Two mathematicians, Sir Isaac Newton and Gottfried Leibniz, are credited with independently formulating a system of computing that solved the above problems and showed how they were connected. Their system of reasoning was "a" calculus. However, as the power and importance of their discovery took hold, it became known to many as "the" calculus. Today, we generally shorten this to discuss calculus. The foundation of the calculus is the **limit** (*limiet*). It is a tool to describe a particular behaviour of a function. This chapter begins our study of the limit by approximating its value graphically and numerically. After a formal definition of the limit, properties are established that make finding limits tractable. Once the limit is understood, then the problems of area and rates of change can be approached.

Leibniz versus Newton

As stated today, both Newton and Leibniz are credited for developing calculus independently. Yet, the question "Who was the first?" was a major intellectual controversy back in the beginning of the Eighteenth century. It is known as the Leibniz–Newton calculus controversy and divided the mathematical community into two bickering groups for years after.

8.1 An intuitive introduction

8.1.1 Limits and their approximation

Consider the function

$$y = \frac{\sin(x)}{x}. \tag{8.1}$$

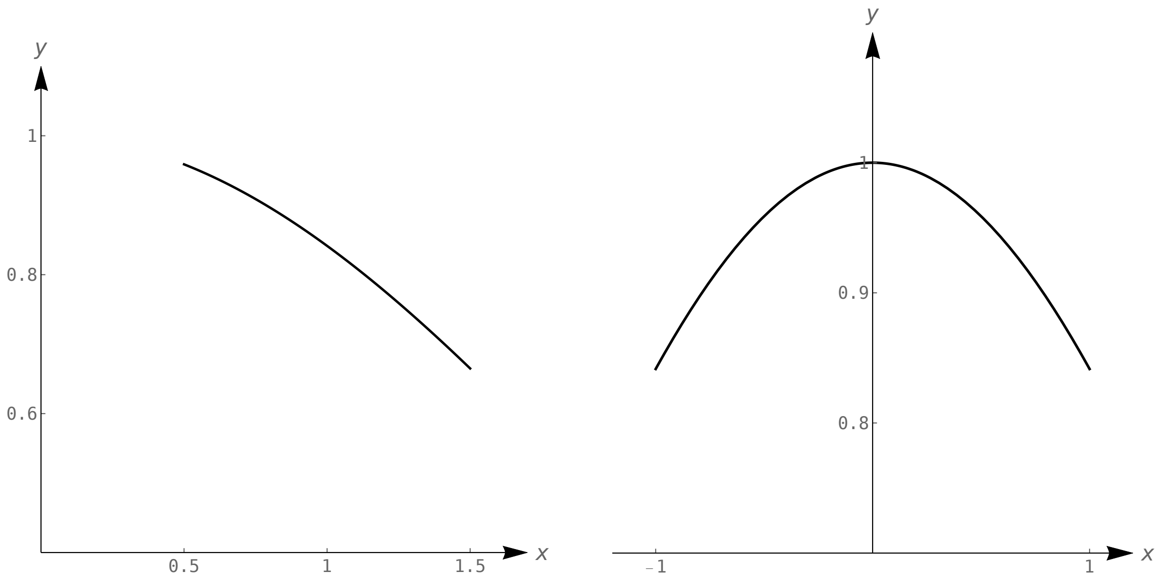
We could ask ourselves the question: When x is near the value 1, what value (if any) is y near?

While our question is not precisely formed (what constitutes "near the value 1"?), the answer does not seem difficult to find. One might think first to look at a graph of this function to approximate the appropriate y values. Consider Figure 8.1(a), where this function is graphed. For values of x near 1, it seems that y takes on values near 0.85. In fact, when $x = 1$, then $y = \sin(1)/1 \approx 0.84$, so it makes sense that when x is near 1, y will be near 0.84.

Consider this again at a different value for x. When x is near 0, what value (if any) is y near? By considering Figure 8.1(b), one can see that it seems that y takes on values near 1. But what happens when $x = 0$? We have

$$y \to \frac{\sin(0)}{0} \to \frac{0}{0}.$$

The expression $0/0$ has no value; it is **indeterminate** (*onbepaald*). Such an expression gives no information about what is going on with the function nearby. We cannot find out how y behaves near $x = 0$ for this function simply by letting $x = 0$.



(a) (b)

Figure 8.1: The graph of $y = \frac{\sin(x)}{x}$ near $x = 1$ (a) and $x = 0$ (b).

Finding a limit entails understanding how a function behaves near a particular value of x. Before continuing, it will be useful to establish some notation. Let $y = f(x)$; that is, let y be a function of x for some function f. The expression "the limit of y as x approaches 1" describes a number, often referred to as L, that y nears as x nears 1. We write all this as

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

This is not a formal definition, but an intuitive one to settle the mind. It allows us to approximate limits both graphically and numerically.

For what concerns the function defined by Equation (8.1), we approximated graphically that

$$\lim_{x\to 1}\frac{\sin(x)}{x}\approx 0.84 \quad \text{and} \quad \lim_{x\to 0}\frac{\sin(x)}{x}\approx 1.$$

8.1.2 Existence of limits

A function may not have a limit for all values of x. That is, we cannot say $\lim_{x\to c} f(x) = L$ for some numbers L for all values of c, for there may not be a number that $f(x)$ is approaching. There are three common ways in which a limit may fail to exist.

1. The function $f(x)$ may approach different values on either side of c.
2. The function may grow without upper or lower bound as x approaches c.
3. The function may oscillate as x approaches c without approaching a specific value.

Each of these cases is illustrated in the following examples

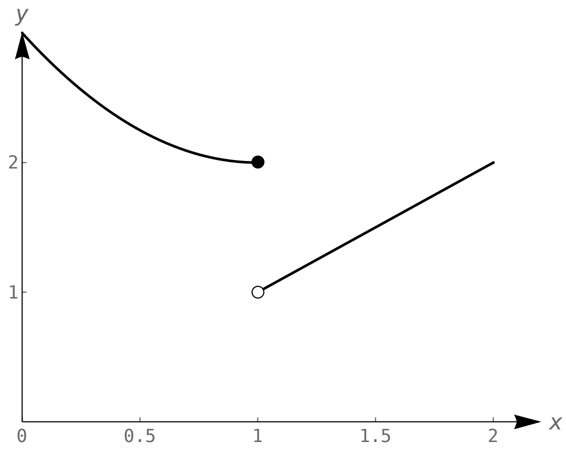
Example 8.1

Let us consider the function

$$f(x) = \begin{cases} x^2 - 2x + 3, & \text{if} \quad x \le 1, \\ x, & \text{if} \quad x > 1, \end{cases} \tag{8.2}$$

and try to determine $\lim_{x\to 1} f(x)$.

A graph of $f(x)$ around $x = 1$ and a corresponding table are given in Figure 8.2(a) and 8.2(b), respectively. It is clear that as x approaches 1, $f(x)$ does not seem to approach a single number. Instead, it seems as though $f(x)$ approaches two different numbers. When considering values of x less than 1, so approaching 1 from the left, it seems that $f(x)$ is approaching 2; when considering values of x greater than 1 ,so approaching 1 from the right, it seems that $f(x)$ is approaching 1. Consequently, the limit does not exist since $f(x)$ is not approaching one value as x approaches 1.



(a)

x	$f(x)$
0.9	2.01
0.99	2.0001
0.999	2.000001
1.001	1.001
1.01	1.01
1.1	1.1

(b)

Figure 8.2: Graphically (a) and numerically (b) approximating $\lim_{x\to 1} f(x)$ for f given by Equation (8.2).

Example 8.2

Let us now have a closer look at

$$\lim_{x\to 1} \frac{1}{(x-1)^2}.$$

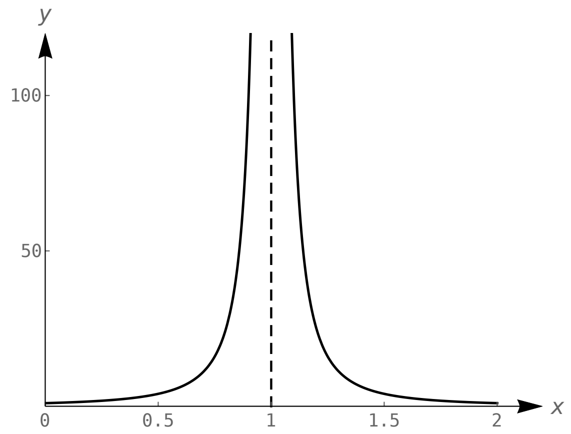
A graph and table of $f(x) = 1/(x-1)^2$ are given in Figure 8.3(a) and 8.3(b), respectively. Both show that as x approaches 1, $f(x)$ grows larger and larger. Indeed, if x is near 1, then $(x-1)^2$ is very small, and:

$$\frac{1}{\text{very small number}} = \text{very large number}.$$

Since $f(x)$ is not approaching a single number, we conclude that

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

does not exist.



(a)

x	$f(x)$
0.9	100.
0.99	10000.
0.999	$1. \times 10^6$
1.001	$1. \times 10^6$
1.01	10000.
1.1	100.

(b)

Figure 8.3: Graphically (a) and numerically (b) approximating $\lim_{x\to 1} \frac{1}{(x-1)^2}$.

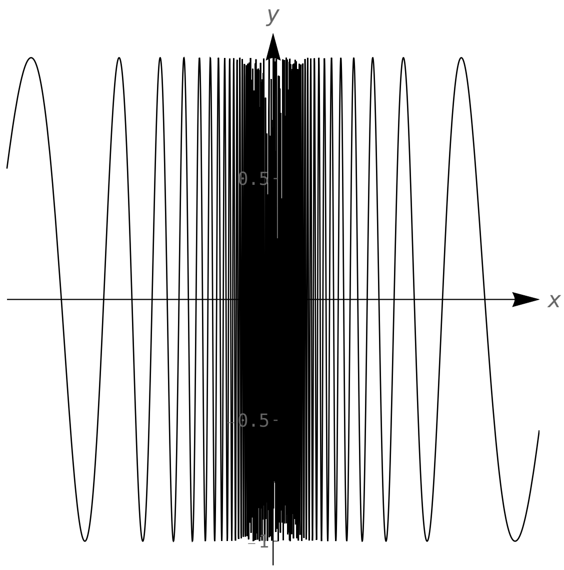
Example 8.3

Let us finally explore why

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

does not exist.

For that purpose, Figure 8.4(a) shows $f(x)$ on the interval $[-0.1, 0.1]$; notice how $f(x)$ clearly seems to oscillate near $x = 0$. This is confirmed in Table 8.4(b), where we see $\sin(1/x)$ evaluated for values of x near 0. As x approaches 0, $f(x)$ does not appear to approach any value. It can be shown that in reality, as x approaches 0, $\sin(1/x)$ takes on all values between -1 and 1 infinitely many times. Because of this oscillation, so the considered limit does not exist.



(a)

x	$\sin(1/x)$
0.1	−0.544021
0.01	−0.506366
0.001	0.82688
0.0001	−0.305614
$1. \times 10^{-5}$	0.0357488
$1. \times 10^{-6}$	−0.349994
$1. \times 10^{-7}$	0.420548

(b)

Figure 8.4: Graphically (a) and numerically (b) approximating $\lim_{x\to 0} \sin\left(\frac{1}{x}\right)$.

8.1.3 Limits of difference quotients

Let $f(x)$ represent the position function, in metres, of some particle that is moving in a straight line, where x is measured in seconds. Let us say that when $x = 1$, the particle is at position 10 m, and when $x = 5$, the particle is at 20 m. Another way of expressing this is to say

$$f(1) = 10 \quad \text{and} \quad f(5) = 20.$$

Since the particle travelled 10 metres in 4 seconds, we can say the particle's average velocity was 2.5 m/s. We write this using a quotient of differences, or, a **difference quotient** (*differentiequotiënt*):

$$\frac{f(5) - f(1)}{5 - 1} = \frac{10}{4} = 2.5\text{m/s}.$$

In fact, we are finding in this way the slope of the **secant line** (*snijlijn*) through those two points.

Now consider finding the average speed on another time interval. We again start at $x = 1$, but consider the position of the particle h seconds later. That is, consider the positions of the particle when $x = 1$ and when $x = 1 + h$. The difference quotient is now

$$\frac{f(1+h) - f(1)}{(1+h) - 1} = \frac{f(1+h) - f(1)}{h}. \tag{8.3}$$

Let now

$$f(x) = -1.5x^2 + 11.5x;$$

for which it holds that $f(1) = 10$ and $f(5) = 20$. We can compute this difference quotient for all values of h except $h = 0$, for then we get 0/0, an indeterminate form. For all values $h \neq 0$, the difference quotient computes the average velocity of the particle over an interval of time of length h starting at $x = 1$. For small values of h, i.e., values of h close to 0, we get average velocities over very short time periods and compute secant lines over small intervals (Figure 8.5). This leads us to wonder what the limit of the difference quotient is as h approaches 0. That is,

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

As we do not yet have a true definition of a limit nor an exact method for computing it, we settle for approximating the value in Table 8.1. This table gives us reason to assume the value of the limit is about 8.5.

Table 8.1: The difference quotient given by Equation (8.3) evaluated at values of h near 0.

h	$\frac{f(1+h)-f(1)}{h}$
−0.5	9.25
−0.1	8.65
−0.01	8.515
0.01	8.485
0.1	8.35
0.5	7.75

Proper understanding of limits is key to understanding calculus. With limits, we can accomplish seemingly impossible mathematical things, like adding up an infinite number of numbers (and not get infinity) and finding the slope of a line between two points, where the two points are actually the same point. These are not just mathematical curiosities; they allow us to link position, velocity and acceleration together, connect cross-sectional areas to volume, find the work done by a variable force, and much more.

In the next section we give the formal definition of the limit and begin our study of finding limits analytically.

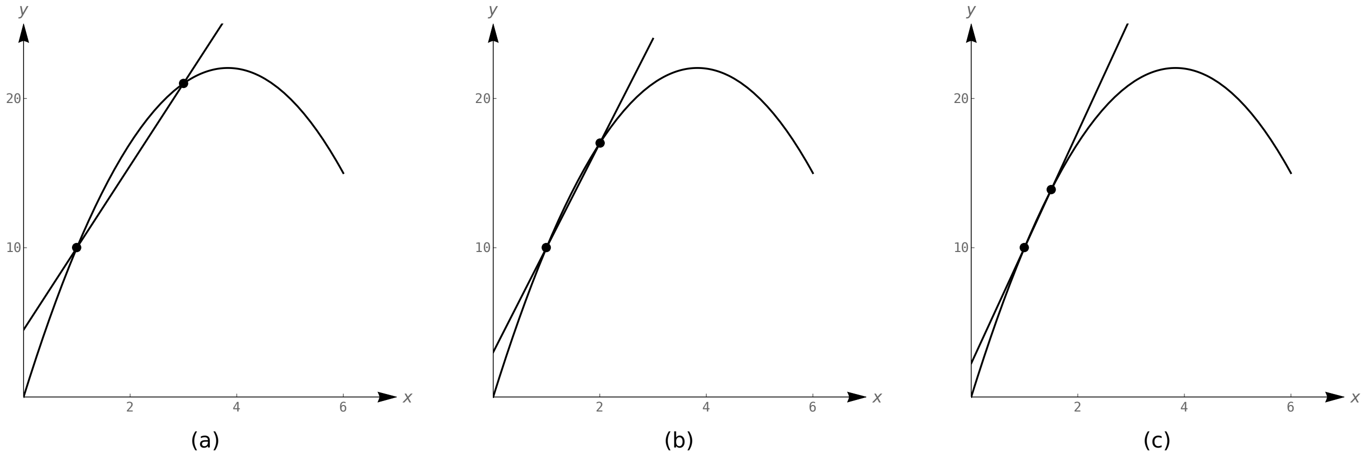


Figure 8.5: Secant lines of $f(x)$ at $x=1$ and $x=1+h$, for shrinking values of h.

8.2 Epsilon-delta definition of a limit

This section introduces the formal definition of a limit, typically called the **epsilon-delta definition** (*epsilon-delta definitie*), referring to the letters ε and δ of the Greek alphabet.

Definitie 8.1 (The limit of a function f)
Let I be an open interval containing c, and let f be a function defined on I, except possibly at c. The limit of $f(x)$, as x approaches c, is L, denoted by

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

and means that given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x-c| < \delta$, then $|f(x)-L| < \varepsilon$.

Note the order in which ε and δ are given. In the definition, the y-tolerance ε is given first and then the limit will exist if we can find an x-tolerance δ that works. Note also that Definition 8.1 basically requires that the point c is a limit point (see Definition 2.5). The (ε, δ) definition is illustrated in Figure 8.6.

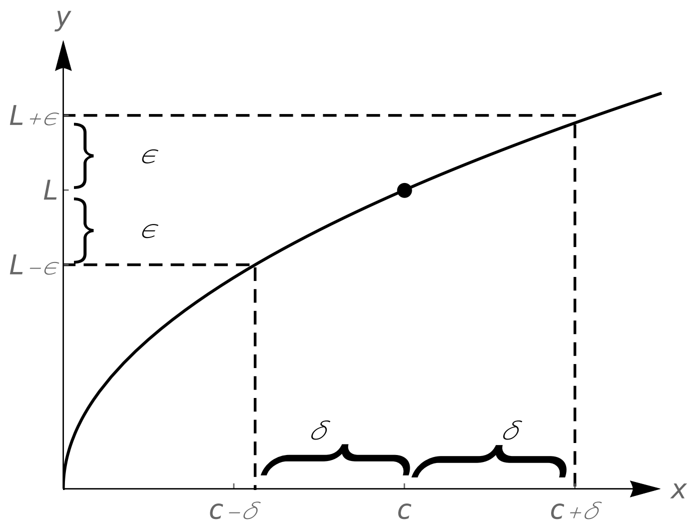


Figure 8.6: Illustrating the (ε, δ) definition.

Using logic operators only, Definition 8.1 becomes

$$\lim_{x\to c} f(x) = L \quad \Leftrightarrow \quad \forall\,\varepsilon > 0, \exists\,\delta(\varepsilon) > 0: \big(\forall x \in I \setminus \{c\} \,:\, 0 < |x-c| < \delta \,\Rightarrow\, |f(x)-L| < \varepsilon\big)\,.$$

With the epsilon-delta definition of a limit in mind, we can easily show that there can be at most one limit of a function, as x approaches c.

Theorem 8.1
Let I be an open interval containing c, and let f be a function defined on I, except possibly at c, then f has at most one limit in c.

Proof To prove this theorem, let us assume that f has two limits in c; that is

$$\lim_{x\to c} f(x) = L_1 \text{ and } \lim_{x\to c} f(x) = L_2.$$

Take $\varepsilon > 0$. Then, there exist $\delta_1(\varepsilon)$ and $\delta_2(\varepsilon)$ for which it holds that

$$\left|f(x) - L_1\right| < \varepsilon_1 \text{ en } \left|f(x) - L_2\right| < \varepsilon_2$$

for $x \in I$ if $0 < |x - c| < \delta_1(\varepsilon)$ and $0 < |x - c| < \delta_2(\varepsilon)$. Hence, for every such x the following holds:

$$\begin{aligned} |L_1 - L_2| &= \left|L_1 - f(x) + f(x) - L_2\right| \\ &\leq \left|L_1 - f(x)\right| + \left|f(x) - L_2\right| < \varepsilon_1 + \varepsilon_2 = \varepsilon. \end{aligned}$$

Consequently, since $\varepsilon > 0$ was chosen arbitrarily, it follows that $L_1 - L_2 = 0$. □

An example will help us understand this definition.

Example 8.4

Show that $\lim\limits_{x \to 4} \sqrt{x} = 2$.

Solution

We start by assuming $y = \sqrt{x}$ is within ε units of 2:

$$\begin{array}{rrcccll} & & & |y-2| & < & \varepsilon & \\ \Leftrightarrow & -\varepsilon & < & y-2 & < & \varepsilon & \text{(Definition of absolute value.)} \\ \Leftrightarrow & -\varepsilon & < & \sqrt{x}-2 & < & \varepsilon & (y = \sqrt{x}.) \\ \Leftrightarrow & 2-\varepsilon & < & \sqrt{x} & < & 2+\varepsilon & \text{(Add 2.)} \\ \Leftrightarrow & (2-\varepsilon)^2 & < & x & < & (2+\varepsilon)^2 & \text{(Square all.)} \\ \Leftrightarrow & 4-4\varepsilon+\varepsilon^2 & < & x & < & 4+4\varepsilon+\varepsilon^2 & \text{(Expand.)} \\ \Leftrightarrow & 4-(4\varepsilon-\varepsilon^2) & < & x & < & 4+(4\varepsilon+\varepsilon^2). & \text{(Rewrite in the desired form.)} \end{array}$$

The desired form in the last step is "$4 - \textit{something} < x < 4 + \textit{something}$." Since we want this last interval to describe an x tolerance around 4, we have that either $\delta < 4\varepsilon - \varepsilon^2$ or $\delta < 4\varepsilon + \varepsilon^2$, whichever is smaller:

$$\delta < \min\{4\varepsilon - \varepsilon^2, 4\varepsilon + \varepsilon^2\}.$$

Since $\varepsilon > 0$, the minimum is $\delta < 4\varepsilon - \varepsilon^2$. So, given an ε, we must set $\delta < 4\varepsilon - \varepsilon^2$.

So given any $\varepsilon > 0$, set $\delta < 4\varepsilon - \varepsilon^2$. Then if $|x - 4| < \delta$ (and $x \neq 4$), it holds that $|f(x) - 2| < \varepsilon$, clearly satisfying the definition of the limit. We have shown formally that $\lim\limits_{x \to 4} \sqrt{x} = 2$. We can check this for $\varepsilon = 0.5$. In that case, the formula gives $\delta < 4(0.5) - (0.5)^2 = 1.75$.

Make note of the general pattern exhibited in the last example. In some sense, each starts out backwards. That is, while we want to

1. start with $|x - c| < \delta$ and conclude that
2. $|f(x) - L| < \varepsilon$,

we actually start by assuming

1. $|f(x) - L| < \varepsilon$, then perform some algebraic manipulations to give an inequality of the form
2. $|x - c| <$ something.

When we have properly done this, the something on the greater than side of the last inequality becomes our δ we are looking for. Once we have such a δ, we can formally start with $|x - c| < \delta$ and use algebraic manipulations to conclude that $|f(x) - L| < \varepsilon$. This is once more illustrated in an example.

Example 8.5

Prove that $\lim_{x\to 0} e^x = 1$.

Solution

Symbolically, we want to take the equation $|e^x - 1| < \varepsilon$ and unravel it to the form $|x - 0| < \delta$. Hence:

$$\begin{array}{rrcll} & |e^x-1| & < & \varepsilon & \\ \Leftrightarrow \quad -\varepsilon < & e^x - 1 & < & \varepsilon & \\ \Leftrightarrow \quad 1-\varepsilon < & e^x & < & 1+\varepsilon & \text{(Add 1.)} \\ \Leftrightarrow \quad \ln(1-\varepsilon) < & x & < & \ln(1+\varepsilon) & \text{(Take natural logs.)} \end{array}$$

There is a caveat here. If it happens that $\varepsilon \geq 1$, then $\ln(1-\varepsilon)$ would be undefined! The way to work around this is to simply define a new epsilon, denoted ϵ_1, that is guaranteed to be smaller than the original epsilon and less than 1, e.g. $\epsilon_1 = \min\{\varepsilon, 1/2\}$. Anyhow, let us continue now under the assumption that $\varepsilon < 1$, so that $\ln(1-\varepsilon)$ is defined.

Moreover, since $\ln(1-\varepsilon) < 0$, we consider its absolute value, and consequently, we can then set δ to be the minimum of $|\ln(1-\varepsilon)|$ and $\ln(1+\varepsilon)$; i.e.,

$$\delta = \min\{|\ln(1-\varepsilon)|, \ln(1+\varepsilon)\} = \ln(1+\varepsilon).$$

Now, we work through the actual the proof:

$$\begin{array}{rrcll} & |x-0| & < & \delta & \\ \Leftrightarrow \quad -\delta < & x & < & \delta & \\ \Leftrightarrow \quad -\ln(1+\varepsilon) < & x & < & \ln(1+\varepsilon). & \\ \Leftrightarrow \quad \ln(1-\varepsilon) < & x & < & \ln(1+\varepsilon). & \text{(Since } \ln(1-\varepsilon) < -\ln(1+\varepsilon)\text{).} \end{array}$$

The above line is true by our choice of δ and by the fact that since $|\ln(1-\varepsilon)| > \ln(1+\varepsilon)$ and $\ln(1-\varepsilon) < 0$, we know $\ln(1-\varepsilon) < -\ln(1+\varepsilon)$. That is; $\ln(1-\varepsilon)$ decreases much more rapidly to $-\infty$ for $0 < \varepsilon < 1$ than $\ln(1+\varepsilon)$ increases to $+\infty$ (see Section 5.2).

$$\begin{array}{rrcll} 1-\varepsilon < & e^x & < & 1+\varepsilon & \text{(Exponentiate.)} \\ \Leftrightarrow \quad -\varepsilon < & e^x - 1 & < & \varepsilon & \text{(Subtract 1.)} \end{array}$$

In summary, given $\varepsilon > 0$, let $\delta = \ln(1+\varepsilon)$. Then $|x-0| < \delta$ implies $|e^x - 1| < \varepsilon$ as desired. We have shown that $\lim_{x\to 0} e^x = 1$.

We may as well use the epsilon-delta definition to show that a limit does not exist.

Example 8.6

Prove that

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

does not exist.

Solution

Suppose that $\lim_{x\to 0} \cos\left(x^{-1}\right) = L$ and choose $\varepsilon = 1/2$. Then, according to the epsilon-delta definition there exists a number $\delta > 0$ such that

$$0 < |x| < \delta \quad \Rightarrow \quad \left|\cos\left(\frac{1}{x}\right) - L\right| < \frac{1}{2}.$$

for all $x \in \mathbb{R}_0$. Now, consider $n \in \mathbb{N}$ such that $\frac{1}{2n\pi} < \delta$. Substituting $x = \frac{1}{2n\pi}$ in the last inequality yields

$$|\cos(2n\pi) - L| = |1 - L| < \frac{1}{2} \quad \Rightarrow \quad \frac{1}{2} < L < \frac{3}{2},$$

whereas choosing $x = \frac{1}{(2n+1)\pi}$, we arrive at

$$|\cos\big((2n+1)\pi\big) - L| = |-1 - L| < \frac{1}{2} \quad \Rightarrow \quad -\frac{3}{2} < L < -\frac{1}{2}.$$

This last two inequalities are contradicting each other, so conclude that the limit does not exist.

In the light of the examples we examined, it should be clear that (ε, δ)-proofs are long and difficult to do. Luckily, in the next section we will learn some theorems that allow us to evaluate limits analytically, that is, without using the (ε, δ)-definition explicitly.

8.3 Finding limits analytically

Recognizing that (ε, δ)-proofs are cumbersome, this section gives a series of theorems which allow us to find limits much more quickly and intuitively.

8.3.1 Properties of limits

The following properties of limits indicate that already established limits do behave nicely. For that purpose, let b, c, L_1 and L_2 be real numbers, let n be a positive integer, and let f and g be functions defined on an open interval I containing c with the following limits:

$$\lim_{x \to c} f(x) = L_1 \qquad \text{and} \qquad \lim_{x \to c} g(x) = L_2.$$

Then, the following limits hold.

1. **Constants**: $\lim_{x \to c} b = b$
2. **Identity**: $\lim_{x \to c} x = c$
3. **Sums/Differences**: $\lim_{x \to c} (f(x) \pm g(x)) = L_1 \pm L_2$
4. **Scalar multiples**: $\lim_{x \to c} b f(x) = bL_1$
5. **Products**: $\lim_{x \to c} f(x)g(x) = L_1 L_2$
6. **Quotients**: $\lim_{x \to c} \dfrac{f(x)}{g(x)} = \dfrac{L_1}{L_2}$, $(L_2 \neq 0)$
7. **Powers**: $\lim_{x \to c} f(x)^n = L_1^n$
8. **Roots**: $\lim_{x \to c} \sqrt[n]{f(x)} = \sqrt[n]{L_1}$, where $f(x) \geq 0$ on I if n is even.

For what concerns function composition, we get

$$\lim_{x\to c} g(f(x)) = N\,,$$

provided

$$\lim_{x\to c} f(x) = M, \qquad \lim_{x\to M} g(x) = N \qquad \text{and} \qquad g(M) = N.$$

By relying on the epsilon-delta definition of a limit (Definition 8.1), these properties can be shown rather easily. We illustrate this for

$$\lim_{x\to c}(f(x)+g(x)) = L_1 + L_2.$$

Let us assume that

$$\lim_{x\to x_0} f(x) = L_1 \qquad \text{and} \qquad \lim_{x\to x_0} g(x) = L_2.$$

From the epsilon-delta definition of a limit, we then have that

$$\forall\,\varepsilon > 0,\ \exists\,\delta_1(\varepsilon) > 0: \quad 0 < |x - x_0| < \delta_1 \quad \Rightarrow \quad |f(x) - L_1| < \varepsilon$$

$$\forall\,\varepsilon > 0,\ \exists\,\delta_2(\varepsilon) > 0: \quad 0 < |x - x_0| < \delta_2 \quad \Rightarrow \quad |g(x) - L_2| < \varepsilon.$$

Moreover, it holds that

$$|f(x) + g(x) - (L_1 + L_2)| \le |f(x) - L_1| + |g(x) - L_2| < 2\varepsilon.$$

Let $\varepsilon' = 2\varepsilon$ and $\delta'(\varepsilon) = \min\{\delta_1(\varepsilon), \delta_2(\varepsilon)\}$, then it holds that

$$\forall\,\varepsilon' > 0,\ \exists\,\delta'(\varepsilon') > 0: \quad 0 < |x - x_0| < \delta' \quad \Rightarrow \quad |f(x) + g(x) - (L_1 + L_2)| < \varepsilon',$$

from which

$$\lim_{x\to x_0}[f(x) + g(x)] = L_1 + L_2 = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x)\,.$$

We use these properties in the following example.

Example 8.7

Let

$$\lim_{x\to 2} f(x) = 2, \quad \lim_{x\to 2} g(x) = 3 \quad \text{and} \quad p(x) = 3x^2 - 5x + 7.$$

Find the following limits:

1. $\lim_{x\to 2}\big(f(x) + g(x)\big)$
2. $\lim_{x\to 2}\big(5f(x) + g(x)^2\big)$
3. $\lim_{x\to 2} p(x)$.

Solution

1. Using the sum/difference rule, we know that

$$\lim_{x\to 2}\big(f(x) + g(x)\big) = 2 + 3 = 5.$$

2. Using the scalar multiple and sum/difference rules, we find that

$$\lim_{x\to 2}\big(5f(x) + g(x)^2\big) = 5\cdot 2 + 3^2 = 19.$$

3. Here we combine the power, scalar multiple, sum/difference and constant rules:

$$\begin{aligned}\lim_{x\to 2} p(x) &= \lim_{x\to 2}(3x^2 - 5x + 7)\\ &= \lim_{x\to 2} 3x^2 - \lim_{x\to 2} 5x + \lim_{x\to 2} 7\\ &= 3\cdot 2^2 - 5\cdot 2 + 7\\ &= 9.\end{aligned}$$

We can also verify this result with Mathematica, using the built-in command **Limit** as follows.

```
In[9]:= Limit[3*x^2-5*x+7,x→2]

Out[9]= 9
```

Part 3 of Example 8.7 demonstrates how the limit of a quadratic polynomial can be determined using the properties of limits. Not only that, recognize that

$$\lim_{x\to 2} p(x) = 9 = p(2);$$

i.e., the limit at 2 was found just by plugging 2 into the function. This holds true for all polynomials, and also for rational functions , as stated in the following theorem.

Theorem 8.2 (Limits of polynomials and rational functions)
Let $p(x)$ and $q(x)$ be polynomials and c a real number. Then:

1. $\lim_{x\to c} p(x) = p(c)$
2. $\lim_{x\to c} \frac{p(x)}{q(x)} = \frac{p(c)}{q(c)}$, *where* $q(c) \neq 0$.

Likewise, for what concerns irrational functions we have the following theorem.

Theorem 8.3 (Limits of irrational functions)
Let f be an irrational function and c a real number. Then:

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

provided $c \in \operatorname{dom} f$.

It was likely frustrating in Section 8.2 to do a lot of work to prove that

$$\lim_{x\to 2} x^2 = 4,$$

as this seemed fairly obvious. Theorem 8.2 shows, however, that polynomial and rational functions behave in an obvious fashion.

Polynomial and rational functions are not the only functions to behave in such a predictable way. The same holds true for the power, exponential, logarithmic, trigonometric and hyperbolic functions we studied in Chapters 4 and 5.

If the limit of a function f exists in a point c, and hence is a finite number, it is intuitively clear that this function should be bounded in a neighbourhood containing c. This insight is formalized in the following theorem.

Theorem 8.4 (Limit and boundedness of a function)
Let I be an open interval containing c, let f be a function defined on I, except possibly at c, and let $\lim_{x\to c} f(x) = L$, then f is bounded in the neighbourhood of c.

Proof In order to prove this theorem, let us take $\varepsilon = 1$, so that there exists $\delta > 0$ for which it holds that $\left|f(x) - L\right| < 1$ for $x \in I$ if $0 < |x - c| < \delta$. Then, for every such x we have

$$\left|f(x)\right| - |L| \le \left|f(x) - L\right| < 1$$

and consequently

$$\left|f(x)\right| \le 1 + |L|.$$

Let us now consider the case that $c \notin I$. Then we can take $M = 1 + |L|$. On the other hand, if $c \in I$, we take $M = \text{maximum}\ (\left|f(c)\right|,\ 1 + |L|)$. So, for every x satisfying $x \in I$ and $0 < |x - c| < \delta$, it holds that $\left|f(x)\right| \le M$. This implies that f is bounded on the δ neighbourhood of c. □

8.3.2 The squeeze theorem

By knowing certain limits of functions, we can find limits involving sums, products, powers, etc., of these functions. We further the development of such comparative tools with the squeeze theorem, a clever and intuitive way to find the value of some limits.

Before stating this theorem formally, suppose we have functions f, g and h where g always takes on values between f and h; that is, for all x in an interval,

$$f(x) \le g(x) \le h(x).$$

If f and h have the same limit at c, and g is always squeezed between them, then g must have the same limit as well. That is essentially what the **squeeze theorem** (*insluitstelling*) states.

Theorem 8.5 (Squeeze theorem)
Let f, g and h be functions on an open interval I containing c such that for all x in I,

$$f(x) \le g(x) \le h(x).$$

If

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

then

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

Proof In order to prove the squeeze theorem, we resort to the (ε, δ)-definition of a limit. More precisely,

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

means that

$$\forall \varepsilon > 0, \exists\ \delta_1 > 0 : \left(\forall x \in I \setminus \{c\}\ :\ 0 < |x - c| < \delta_1\ \Rightarrow\ |f(x) - L| < \varepsilon\right). \tag{8.4}$$

and

$$\lim_{x\to c} h(x) = L$$

means that

$$\forall \varepsilon > 0, \exists\ \delta_2 > 0 : \left(\forall x\ \in I \setminus \{c\} : 0 < |x - c| < \delta_2\ \Rightarrow\ |h(x) - L| < \varepsilon\right). \tag{8.5}$$

Moreover,

$$f(x) \leq g(x) \leq h(x)$$

implies

$$f(x) - L \leq g(x) - L \leq h(x) - L.$$

Now, let us choose δ to be the minimum of δ_1 and δ_2. Then, if $|x - c| < \delta$, combining Statements (8.4) and (8.5), we have

$$-\varepsilon < f(x) - L \leq g(x) - L \leq h(x) - L\ < \varepsilon.$$

□

It can take some work to figure out appropriate functions by which to squeeze a given function. However, that is generally the only place where work is necessary; the theorem makes the evaluating the limit part very simple. We use this theorem in the following example.

Example 8.8

Show that

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

Solution

We begin by considering the unit circle (Section 5.3). Remember that each point on the unit circle has coordinates $(\cos(\theta), \sin(\theta))$ for some angle θ (Figure 8.7). Using similar triangles, we can extend the line from the origin through the point to the point $(1, \tan(\theta))$, as shown. Here we are assuming that $0 \leq \theta \leq \pi/2$. Later we will show that we can also consider $\theta \leq 0$.

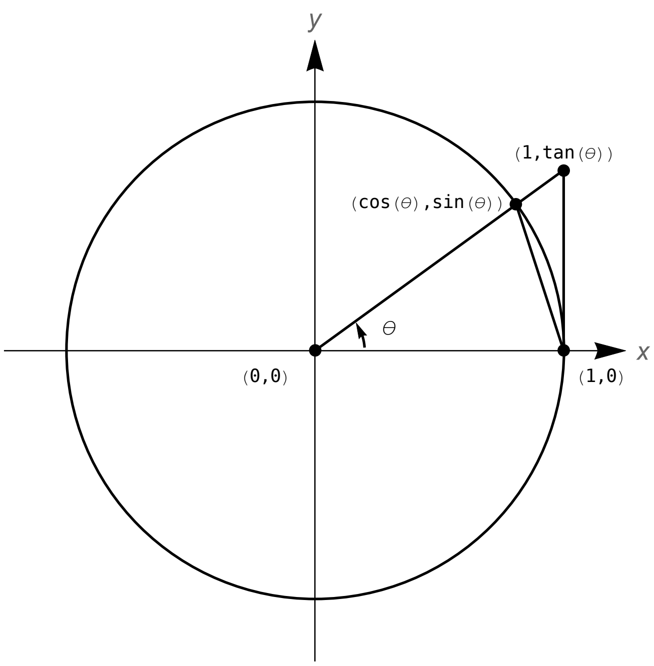
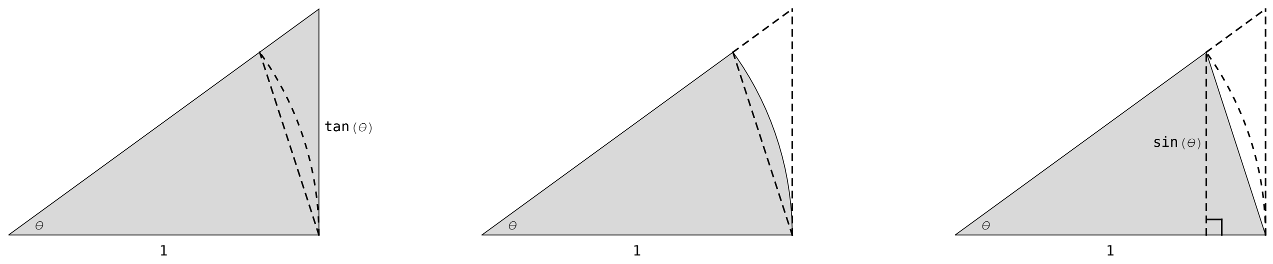


Figure 8.7: The unit circle and related triangles.

Figure 8.7 shows three regions have been constructed in the first quadrant, two triangles and a sector of a circle, which are also drawn below. The area of the large triangle is $\tan(\theta)/2$; the area of the sector is $\theta/2$; the area of the triangle contained inside the sector is $\sin(\theta)/2$ (Figure 5.10). It is then clear from the diagram that



$$\frac{\tan(\theta)}{2} \geq \frac{\theta}{2} \geq \frac{\sin(\theta)}{2}$$

Multiplying all terms in this inequality by $2\sin^{-1}(\theta)$, yields

$$\frac{1}{\cos(\theta)} \geq \frac{\theta}{\sin(\theta)} \geq 1.$$

Taking reciprocals reverses the inequalities, giving

$$\cos(\theta) \leq \frac{\sin(\theta)}{\theta} \leq 1.$$

Not that these inequalities hold for all values of θ near 0, even negative values, since $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$.

Now take limits for $\theta \to 0$.

$$\begin{array}{rcccl} \lim\limits_{\theta\to0} \cos(\theta) & \leq & \lim\limits_{\theta\to0} \dfrac{\sin(\theta)}{\theta} & \leq & \lim\limits_{\theta\to0} 1 \\ \cos(0) & \leq & \lim\limits_{\theta\to0} \dfrac{\sin(\theta)}{\theta} & \leq & 1 \\ 1 & \leq & \lim\limits_{\theta\to0} \dfrac{\sin(\theta)}{\theta} & \leq & 1 \end{array}$$

Clearly, Theorem 8.5 guarantees that

$$\lim_{\theta\to0} \frac{\sin(\theta)}{\theta} = 1.$$

Actually, this limit tells us more than just that as x approaches 0, $\sin(x)/x$ approaches 1. Both x and $\sin(x)$ are approaching 0, but the ratio of x and $\sin(x)$ approaches 1, meaning that they are approaching 0 in essentially the same way. So for small x, the functions $y = x$ and $y = \sin(x)$ are essentially indistinguishable.

We include this special limit, along with three others, which can be determined in a similar way using Theorem 8.5, in the following theorem.

Theorem 8.6 (Special limits)

1. $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$
2. $\lim_{x\to 0} \frac{\cos(x)-1}{x} = 0$
3. $\lim_{x\to 0} (1+x)^{\frac{1}{x}} = e$
4. $\lim_{x\to 0} \frac{e^x-1}{x} = 1$

A short word on how to interpret the latter three limits in Theorem 8.6. We know that as x goes to 0, $\cos(x)$ goes to 1. So, in the second limit, both the numerator and denominator are approaching 0. However, since the limit is 0, we can interpret this as saying that $\cos(x)$ is approaching 1 faster than x is approaching 0.

In the third limit in Theorem 8.6, inside the parentheses we have an expression that is approaching 1 (though never equalling 1), and we know that 1 raised to any power is still 1. At the same time, the power is growing toward infinity. What happens to a number near 1 raised to a very large power? In this particular case, the result approaches **Euler's number** (*Eulergetal*), e, approximately 2.718. Upon an appropriate change of variables, we can also write this as

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

In the fourth limit in Theorem 8.6, we see that as $x \to 0$, e^x approaches 1 just as fast as $x \to 0$, resulting in a limit of 1.

Euler's number and interests

Although the symbol e was introduced by Leonhard Euler around 1727, it was Jacob Bernoulli who already discovered this constant in 1683 through the third limit in Theorem 8.6. He came across this special limit while studying a question about compound interest:

A bank account starts with $\$1.00$ *and pays 100 percent interest per year. If the interest is credited once, at the end of the year, the value of the account at year-end will be* $\$2.00$. *What happens if the interest is computed and credited more frequently during the year?*

If there are n compounding intervals, the interest for each interval will be $100\%/n$ and the value at the end of the year will be $\$1.00\,(1+1/n)^n$.

8.3.3 Limits of functions equal at all but one point

Consider the following limit:

$$\lim_{x\to 1} \frac{x^2-1}{x-1}.$$

We begin by attempting to apply Theorem 8.2 and substituting 1 for x in the quotient. This, however, gives:

$$\lim_{x\to 1} \frac{x^2-1}{x-1} = \frac{1^2-1}{1-1} = \frac{0}{0},$$

an indeterminate form. We cannot apply the theorem.

By graphing the function $y = \frac{x^2-1}{x-1}$ (Figure 8.8), we see that the function seems to be linear, implying

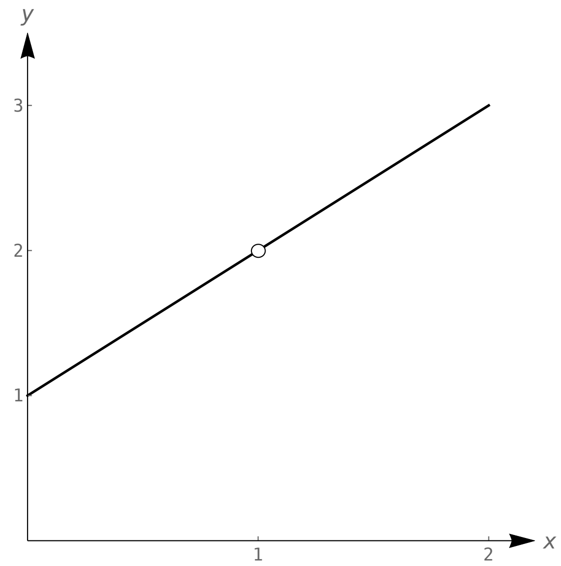


Figure 8.8: The graph of $y = \frac{x^2-1}{x-1}$.

that the limit should be easy to evaluate. Recognize that the numerator of our quotient can be factored:

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

The function is not defined when $x = 1$, but for all other x,

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

Clearly $\lim\limits_{x\to 1}(x+1) = 2$. Recall that when considering limits, we are not concerned with the value of the function at 1, only the value the function approaches as x approaches 1. Since $(x^2-1)/(x-1)$ and $x+1$ are the same at all points except $x = 1$, they both approach the same value as x approaches 1. Therefore we may conclude that

$$\lim_{x\to 1}\frac{x^2-1}{x-1} = 2.$$

This finding is formalized in the following theorem

Theorem 8.7 (Equality of functions and limits)
Let $g(x) = f(x)$ for all x in an open interval, except possibly at c, and let $\lim\limits_{x\to c} g(x) = L$ for some real number L. Then

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

So, when dealing with a rational function of the form $g(x)/f(x)$ and directly evaluating the limit

$$\lim_{x\to c}\frac{g(x)}{f(x)}$$

returns 0/0, the fundamental theorem of algebra tells us that $(x-c)$ is a factor of both $g(x)$ and $f(x)$. One can then use algebra to factor this term out, cancel, and then apply Theorem 8.7.

We end this section by revisiting a limit first seen in Section 8.1, a limit of a difference quotient. Let

$f(x) = -1.5x^2 + 11.5x$; we approximated the limit

$$\lim_{h \to 0} \frac{f(1+h) - f(1)}{h} \approx 8.5.$$

We now formally evaluate this limit in the following example.

Example 8.9

Let $f(x) = -1.5x^2 + 11.5x$, then find

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

Solution

Since f is a polynomial, our first attempt should be to employ Theorem 8.2 and substitute 0 for h. However, we see that this gives us $0/0$.

Knowing that we have a rational function hints that some algebra will help. Consider the following:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(1+h)-f(1)}{h} &= \lim_{h\to 0} \frac{-1.5(1+h)^2 + 11.5(1+h) - \left(-1.5(1)^2 + 11.5(1)\right)}{h} \\
&= \lim_{h\to 0} \frac{-1.5(1+2h+h^2) + 11.5 + 11.5h - 10}{h} \\
&= \lim_{h\to 0} \frac{-1.5h^2 + 8.5h}{h} \\
&= \lim_{h\to 0} \frac{h(-1.5h + 8.5)}{h} \\
&= 8.5.
\end{aligned}$$

This matches our previous approximation (see Table 8.1).

This section contains several valuable tools for evaluating limits. One of the main results is that many functions behave in a very nice, predictable way. In Section 8.5 we give a name to this nice behaviour; we label such functions as continuous. Defining that term will require us to look again at what a limit is and what causes limits to not exist.

For the sake of comprehensiveness, We list below the steps that should be taken when evaluating the limit $\lim\limits_{x\to a} f(x)$:

1. Compute $f(a)$.
2. You arrive at one of the following cases:
   - $f(a) \in \mathbb{R}$: the limit is computed.
   - $f(a) = \left(\frac{0}{0}\right)$: try to get $x-a$ as a common factor in the nominator and denominator, possibly after multiplying with its conjugate binomial, then simplify and return to Step 1.
   - $f(a) = \left(\frac{c}{0}\right) = \pm\infty\ (c \neq 0) : x = a$ is a vertical asymptote of the function f.

8.4 One-sided limits

Remember from Section 8.1 that one of the ways in which limits of functions fail to exist is when the function approaches different values from the left and right. To explore in depth the concepts underlying this we introduce in this section the **one-sided limit** (*eenzijdige limiet*). We begin with formal definitions that are very similar to Definition 8.1, but the notation is slightly different and $x \neq c$ is replaced with either $x < c$ or $x > c$.

Definitie 8.2 (One-sided limits)
Left-hand Limit (*linkerlimiet*)

Let f be a function defined on $]a, c[$ for some $a < c$ and let L be a real number.

The limit of $f(x)$, as x approaches c from the left, is L, or, the left-hand limit of f at c is L, denoted by

$$\lim_{\substack{x \to c \\ <}} f(x) = L,$$

means given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $a < x < c$, if $|x - c| < \delta$, then $|f(x) - L| < \varepsilon$.

Right-hand Limit (*rechterlimiet*)

Let f be a function defined on $]c, b[$ for some $b > c$ and let L be a real number.

The limit of $f(x)$, as x approaches c from the right, is L, or, the right-hand limit of f at c is L, denoted by

$$\lim_{\substack{x \to c \\ >}} f(x) = L,$$

means given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $c < x < b$, if $|x - c| < \delta$, then $|f(x) - L| < \varepsilon$.

Practically speaking, when evaluating a left-hand limit, we consider only values of x to the left of c, i.e., where $x < c$. The notation $x \underset{<}{\to} c$ is used to imply that we look at values of x to the left of c. A similar statement holds for evaluating right-hand limits; there we consider only values of x to the right of c, i.e., $x > c$. We can use the theorems from previous sections to help us evaluate these limits; we just restrict our view to one side of c.

We practice evaluating left- and right-hand limits through a series of examples.

Example 8.10

Let the function f_1 be defined by

$$f_1(x) = \begin{cases} x, & \text{if} \quad 0 \leq x \leq 1, \\ 3 - x, & \text{if} \quad 1 < x < 2. \end{cases}$$

Its graph is shown in Figure 8.9. Find each of the following:

1. $\lim_{\substack{x \to 1 \\ <}} f_1(x)$
2. $\lim_{\substack{x \to 1 \\ >}} f_1(x)$
3. $\lim_{x \to 1} f_1(x)$
4. $f_1(1)$
5. $\lim_{\substack{x \to 0 \\ >}} f_1(x)$
6. $f_1(0)$
7. $\lim_{\substack{x \to 2 \\ <}} f_1(x)$
8. $f_1(2)$

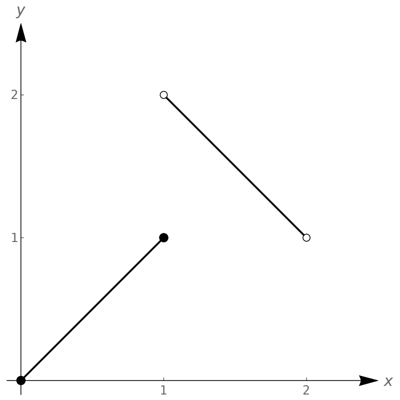


Figure 8.9: The graph of f_1 in Example 8.10.

Solution

For these problems, the visual aid of the graph is likely more effective in evaluating the limits than using f_1 itself.

1. As x goes to 1 from the left, we see that $f_1(x)$ is approaching the value of 1. Therefore $\lim\limits_{\substack{x\to 1\\<}} f_1(x) = 1$.
2. As x goes to 1 from the right, we see that $f_1(x)$ is approaching the value of 2. Recall that it does not matter that there is an open circle there; we are evaluating a limit, not the value of the function. Therefore $\lim\limits_{\substack{x\to 1\\>}} f_1(x) = 2$.
3. The limit of f_1 as x approaches 1 does not exist. The function does not approach one particular value, but two different values from the left and the right.
4. Using the definition and by looking at the graph we see that $f_1(1) = 1$.
5. As x goes to 0 from the right, we see that $f_1(x)$ is also approaching 0. Therefore $\lim\limits_{\substack{x\to 0\\>}} f_1(x) = 0$.

   Note we cannot consider a left-hand limit at 0 as f_1 is not defined for values of $x < 0$.
6. Using the definition and the graph, $f_1(0) = 0$.
7. As x goes to 2 from the left, we see that $f_1(x)$ is approaching the value of 1. Therefore $\lim\limits_{\substack{x\to 2\\<}} f_1(x) = 1$.
8. The graph and the definition of the function show that $f_1(2)$ is not defined.

Alternatively, we could again make use of Mathematica to determine the one-sided limits. The option **`Direction`** specifies which one-sided limit should be computed ("FromBelow" and "FromAbove" for left-hand and right-hand limits, respectively). For example, we can compute the limit in 7 as follows:

```
In[10]:= Limit[Piecewise[{{x, 0≤x≤1}, {3-x, 1<x<2}}], x → 2, Direction → "FromBelow"]

Out[10]= 1
```

Note how the left and right-hand limits in the previous examples were different at $x = 1$. This, of course, causes the limit to not exist. The following theorem states what is fairly intuitive: the limit exists precisely when the left and right-hand limits are equal.

Theorem 8.8 (Limits and one sided limits)
Let f be a function defined on an open interval I containing c. Then

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

if and only if,

$$\lim_{\substack{x\to c\\ <}} f(x) = L \quad \textit{and} \quad \lim_{\substack{x\to c\\ >}} f(x) = L.$$

Throughout these examples pay attention to the fact that the value of the function may/may not be equal to the value(s) of its left/right-hand limits, even when these limits agree.

Example 8.11

Let

$$f_2(x) = \begin{cases} 2-x, & \text{if} \quad 0 < x < 1, \\ (x-2)^2, & \text{if} \quad 1 < x < 2. \end{cases}$$

A graph of this function is shown in Figure 8.10. Evaluate the following.

1. $\lim_{\substack{x\to 1\\ <}} f_2(x)$
2. $\lim_{\substack{x\to 1\\ >}} f_2(x)$
3. $\lim_{x\to 1} f_2(x)$
4. $f_2(1)$
5. $\lim_{\substack{x\to 0\\ >}} f_2(x)$
6. $f_2(0)$
7. $\lim_{\substack{x\to 2\\ <}} f_2(x)$
8. $f_2(2)$

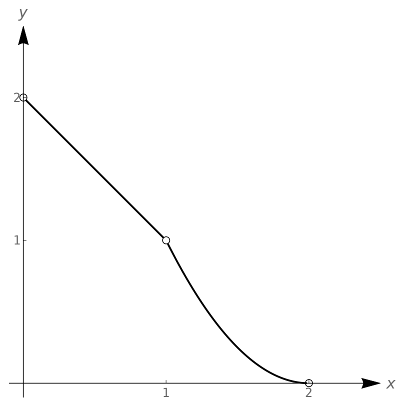


Figure 8.10: A graph of f_2 in Example 8.11.

Solution

We will evaluate each using both the definition of f_2 and its graph.

1. As x approaches 1 from the left, we see that $f_2(x)$ approaches 1. Therefore $\lim_{\substack{x\to 1\\ <}} f_2(x) = 1$.
2. As x approaches 1 from the right, we see that $f_2(x)$ approaches 1. Therefore $\lim_{\substack{x\to 1\\ >}} f_2(x) = 1$.
3. The limit of f_2 as x approaches 1 exists and is 1, as f_2 approaches 1 from both the right and left. Therefore $\lim_{x\to 1} f_2(x) = 1$.
4. $f_2(1)$ is not defined. Note that 1 is not in the domain of f_2 as defined by the problem, which is indicated on the graph by an open circle when $x = 1$.

5. As x goes to 0 from the right, $f_2(x)$ approaches 2. So $\lim\limits_{x \underset{>}{\to} 0} f_2(x) = 2$.

6. $f_2(0)$ is not defined as 0 is not in the domain of f_2.

7. As x goes to 2 from the left, $f_2(x)$ approaches 0. So $\lim\limits_{x \underset{<}{\to} 2} f_2(x) = 0$.

8. $f_2(2)$ is not defined as 2 is not in the domain of f_2.

Example 8.12

Consider the following piecewise-defined functions:

1.
$$f_3(x) = \begin{cases} (x-1)^2, & \text{if} \quad 0 \le x \le 2, x \ne 1, \\ 1, & \text{if} \quad x = 1, \end{cases}$$

2.
$$f_4(x) = \begin{cases} x^2, & \text{if} \quad 0 \le x \le 1, \\ 2-x, & \text{if} \quad 1 < x \le 2. \end{cases}$$

Their graphs are shown in Figure 8.11(a) and 8.11(b), respectively. Evaluate for both functions $\lim\limits_{x \underset{<}{\to} 1} f_i(x)$, $\lim\limits_{x \underset{>}{\to} 1} f_i(x)$, $\lim\limits_{x \to 1} f_i(x)$ and $f_i(1)$.

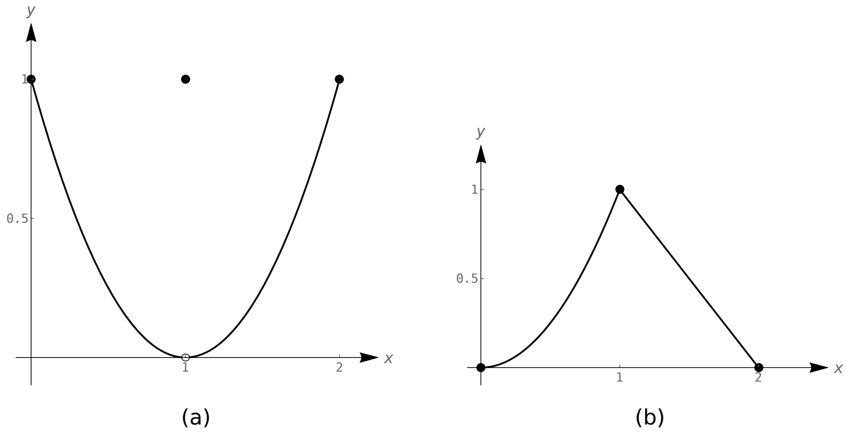


Figure 8.11: A graph of f_3 (a) and f_4 (b) in Example 8.12.

Solution

1. It is clear by looking at the graph that both the left and right-hand limits of f_3, as x approaches 1, are 0. Thus it is also clear that the limit is 0; i.e., $\lim\limits_{x \to 1} f_3(x) = 0$. It is also clearly stated that $f_3(1) = 1$.

2. It is clear from the definition of the function and its graph that all of the following are equal:

$$\lim_{x \underset{<}{\to} 1} f_4(x) = \lim_{x \underset{>}{\to} 1} f_4(x) = \lim_{x \to 1} f_4(x) = f_4(1) = 1.$$

In Examples 8.10 – 8.12 we were asked to find both $\lim\limits_{x \to 1} f_i(x)$ and $f_i(1)$. Consider the following table:

	$\lim_{x\to 1} f_i(x)$	$f_i(1)$
Example 8.10	does not exist	1
Example 8.11	1	not defined
Example 8.12.1	0	1
Example 8.12.2	1	1

Only in Example 8.12.2 do both the function and the limit exist and agree. This seems nice; in fact, it seems normal. This is in fact an important situation which we explore in the next section and refers to a function being continuous. In short, a continuous function is one in which when a function approaches a value as $x \to c$ (i.e., when $\lim_{x\to c} f(x) = L$), it actually attains that value at c. Such functions behave nicely as they are very predictable.

8.5 Continuity

8.5.1 Definition

As we have studied limits, we have gained the intuition that limits measure where a function is heading. That is, if $\lim_{x\to 1} f(x) = 3$, then as x is close to 1, $f(x)$ is close to 3. We have seen, though, that this is not necessarily a good indicator of what $f(1)$ actually is. This can be problematic; functions can tend to one value but attain another. This section focuses on functions that do not exhibit such behaviour.

Definitie 8.3 (Continuous function)
Let f be a function defined on an open interval I containing c.

1. f is **continuous** (*continu*) at c if $\lim_{x\to c} f(x) = f(c)$.
2. f is continuous on I if f is continuous at c for all values of c in I. If f is continuous on $\mathbb{R}$, we say f is continuous everywhere.

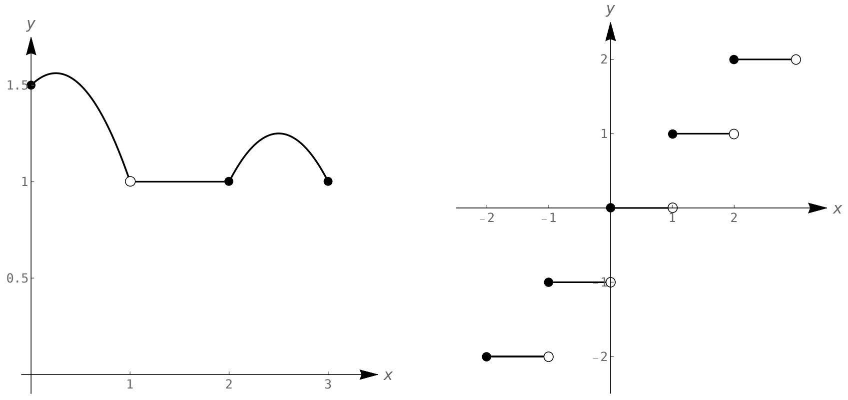
Note that this definition of continuity (currently) only applies to open intervals.

To establish whether or not a function f is continuous at c one should verify:

1. $\lim_{x\to c} f(x)$ exists,
2. $f(c)$ is defined, and
3. $\lim_{x\to c} f(x) = f(c)$.

Example 8.13

Let f (a) and g (b) be defined as shown in Figures 8.12(a) and 8.12(b), respectively. Give the interval(s) on which these functions are continuous.



(a) (b)

Figure 8.12: The graph of f (a) and g (b) in Example 8.13.

Solution

a. We proceed by examining the three criteria for continuity.

   (a) The limits $\lim_{x\to c} f(x)$ exists for all c between 0 and 3.

   (b) $f(c)$ is defined for all c between 0 and 3, except for $c = 1$. We know immediately that f cannot be continuous at $x = 1$.

   (c) The limit $\lim_{x\to c} f(x) = f(c)$ for all c between 0 and 3, except, of course, for $c = 1$.

   We conclude that f is continuous at every point of $]0,3[$ except at $x = 1$. Therefore f is continuous on $]0,1[$ and $]1,3[$.

b. We examine the three criteria for continuity.

   (a) The limits $\lim_{x\to c} g(x)$ do not exist at the jumps from one step to the next, which occur at all integer values of c. Therefore the limits exist for all c except when c is an integer.

   (b) The function is defined for all values of c.

   (c) The limit $\lim_{x\to c} g(x) = g(c)$ for all values of c where the limit exist, since each step consists of just a line.

   We conclude that g is continuous everywhere except at integer values of c. So the intervals on which g is continuous are

$$\ldots,]-2,-1[,]-1,0[,]0,1[,]1,2[, \ldots .$$

Our definition of continuity on an interval specifies the interval is an open interval. We can extend the definition of continuity to closed intervals by considering the appropriate one-sided limits at the endpoints.

Definitie 8.4 (Continuity on closed intervals)
Let f be defined on the closed interval $[a,b]$ for some real numbers $a < b$. f is continuous on $[a,b]$ if:

1. f is continuous on $]a,b[$,

2. $\lim\limits_{\substack{x \to a \\ >}} f(x) = f(a)$ and

3. $\lim\limits_{\substack{x \to b \\ <}} f(x) = f(b)$.

We can of course make the appropriate adjustments to talk about continuity on half–open intervals such as $[a, b[$ or $]a, b]$ if necessary. Also note that we call the function f **right-continuous** (*rechtscontinu*) at a if

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

and **left-continuous** (*linkscontinu*) at b if

$$\lim_{\substack{x \to b \\ <}} f(x) = f(b),$$

where we of course assumed that the respective one-sided limits exist.

Using this new definition, we can adjust our answer in Example 8.13 by stating that f is continuous on $[0,1[$ and $]1,3]$. Likewise, the function g in Example 8.13 is continuous on the following half–open intervals

$$\ldots, [-2,-1[\,, [-1,0[\,, [0,1[\,, [1,2[\,, \ldots.$$

Moreover, bearing in mind Theorem 8.4, which guarantees that a function is bounded in a δ neighbourhood of c if the limit at c exists, the next theorem should not come as a surprise.

Theorem 8.9 (Continuity and boundedness of a function)
Let f be continuous on the closed interval $[a, b]$, then f is bounded on $[a, b]$.

Most of the functions you have likely seen in the past are continuous on their domains. This is demonstrated in the following example where we examine the intervals of continuity of a variety of common functions.

Example 8.14

For each of the following functions, give the domain of the function and the interval(s) on which it is continuous.

1. $f(x) = \dfrac{1}{x}$
2. $f(x) = \sqrt{x}$
3. $f(x) = \sqrt{1 - x^2}$
4. $f(x) = |x|$

Solution

We examine each in turn.

1. The domain of $f(x) = 1/x$ is $\mathbb{R}_0$. As it is a rational function, we apply Theorem 8.2 together with Definition 8.3 to recognize that f is continuous on all of its domain.

2. The domain of $f(x) = \sqrt{x}$ is $\mathbb{R}^+$. It follows that $f(x) = \sqrt{x}$ is continuous on its domain of $\mathbb{R}^+$.

3. The domain of $f(x) = \sqrt{1 - x^2}$ is $[-1, 1]$. Using properties of limits shows that f is continuous on all of its domain, $[-1, 1]$.

4. The domain of $f(x) = |x|$ is $\mathbb{R}$. We can define the absolute value function as

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

Each piece of this piecewise defined function is continuous on all of its domain, giving that f is continuous on $]-\infty, 0[$ and $[0, +\infty[$. We cannot assume this implies that f is continuous on $\mathbb{R}$; we need to check that $\lim\limits_{x\to 0} f(x) = f(0)$, as $x = 0$ is the point where f transitions from one piece of its definition to the other. It is easy to verify that this is indeed true, hence we conclude that $f(x) = |x|$ is continuous everywhere.

The continuity of the function $f(x) = |x|$ demonstrated in the previous example is nothing but an illustration of the following theorem.

Theorem 8.10 (Continuity of absolute value functions)
Let f be continuous on an interval I, then also the function $|f|$, defined as $y = \left|f(x)\right|$, where $x \in I$, is continuous on I.

Continuous functions can be combined to form other continuous functions, which is an immediate consequence of the properties of limits. So, if we let f and g be continuous functions on an interval I, c be a real number and n be a positive integer, then the following functions are continuous on I.

1. **Sums/Differences**: $f \pm g$
2. **Constant Multiples**: $c \cdot f$
3. **Products**: $f \cdot g$
4. **Quotients**: f/g (As long as $g \neq 0$ on I.)
5. **Powers**: f^n
6. **Roots**: $\sqrt[n]{f}$ (If n is even then require $f(x) \geq 0$ on I.)

For what concerns function compositions, we consider a function f which is continuous on I, whose range on I is J, and a function g which is continuous on J. Then $g \circ f$, i.e., $g(f(x))$, is continuous on I.

A function f that is not continuous at c is called **discontinuous** (*discontinu*) at c, i.e. there is a **discontinuity** (*discontinuïteit*) at c.

There exist several types of discontinuities. For a so-called **removable discontinuity** (*ophefbare discontinuïteit*), it holds that

$$L^- = \lim_{\substack{x \to c \\ <}} f(x)$$

and

$$L^+ = \lim_{\substack{x \to c \\ >}} f(x)$$

at c both exist, are finite, and are equal, i.e. $L = L^- = L^+$, but at the same time that the actual value of $f(c)$ is not equal to L. This discontinuity can be removed to make f continuous at c, or more precisely, the function

$$g(x) = \begin{cases} f(x), & x \neq c \\ L, & x = c \end{cases}$$

is continuous at $x = c$.

In the case a single limit does not exist because the one-sided limits, L^- and L^+, exist and are finite, but are not equal, c is called a **jump discontinuity** (*sprong-discontinuïteit*). For this type of discontinuity, the function f may have any value at c. Finally, there exists also a so-called **essential discontinuity** (*essentiële discontinuïteit*), for which it holds that only one of the two one-sided limits exists or is infinite.

Example 8.15

Identify and classify the discontinuity occurring for each of the following functions.

1. $f(x) = \begin{cases} x^2 - 2x + 3, & \text{if} \quad x \leq 1, \\ x, & \text{if} \quad x > 1. \end{cases}$
2. $f(x) = \begin{cases} (x-1)^2, & \text{if} \quad 0 \leq x \leq 2, x \neq 1, \\ 1, & \text{if} \quad x = 1. \end{cases}$
3. $f(x) = \begin{cases} \dfrac{1}{x}, & \text{if} \quad x \neq 0, \\ 1, & \text{if} \quad x = 0. \end{cases}$

Solution

1. A graph of this function is shown in Figure 8.2(a). Using the tools we have introduced in the preceding sections, we infer that $\lim\limits_{\substack{x \to 1 \\ <}} f(x) = 2$ and $\lim\limits_{\substack{x \to 1 \\ >}} f(x) = 1$. Clearly, the left and right-hand limits of f as x approaches 1 exist, but they are different, so the point $x = 1$ has a jump discontinuity for this function.
2. This is the piecewise-defined function that we already encountered in Example 8.12 (Figure 8.11(a)). From our analysis in that example we have that the discontinuity occurs at $x = 1$ and that both the left and right-hand limits of f, as x approaches 1, are 0. Thus it is also clear that the limit is 0; i.e., $\lim\limits_{x \to 1} f(x) = 0$. At the same time, it holds that f has the value 1 in $x = 1$. Consequently, the point $x = 1$ is a removable discontinuity, which can be removed by defining
$$g(x) = \begin{cases} f(x), & x \neq 1 \\ 0, & x = 1 \end{cases}.$$
3. For this function, we can see that at 0 we have some problem of continuity, so we look at what happens with the continuity of the function at this point. We infer that $\lim\limits_{\substack{x \to 0 \\ <}} f(x) = -\infty$ and $\lim\limits_{\substack{x \to 0 \\ >}} f(x) = \infty$, while $f(0) = 1$. Hence, we see ourselves confronted with an essential discontinuity. Note that the function $f(x) = 1/x$ does not present an essential discontinuity since the function is continuous. We would have the discontinuity at the point $x = 0$, but it does not belong to the domain of the function, so it is not possible to define the discontinuity.

8.5.2 Intermediate value theorem

A common way of thinking of a continuous function is that its graph can be sketched without lifting your pencil. That is, its graph forms a continuous curve, without holes, breaks or jumps. This pseudo–definition glosses, however, over some of the finer points of continuity. Very strange functions are continuous that one would be hard pressed to actually sketch by hand, an example of this being for instance the Weierstrass function (Figure 8.13).

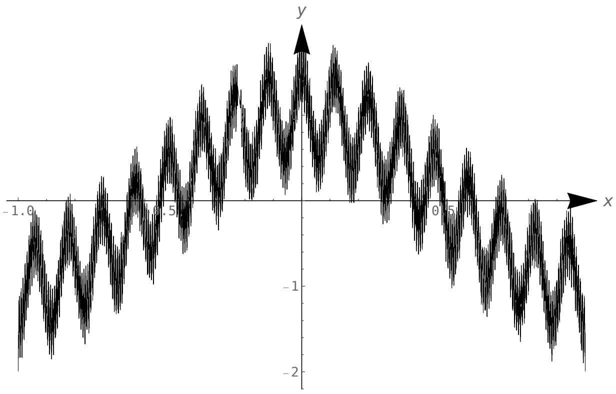


Figure 8.13: The Weierstrass function.

This intuitive notion of continuity does nonetheless help us understand another important concept as follows. Suppose f is defined on $[1,2]$ and $f(1) = -10$ and $f(2) = 5$. If f is continuous on $[1,2]$ (i.e., its graph can be sketched as a continuous curve from $(1,-10)$ to $(2,5)$) then we know intuitively that somewhere on $[1,2]$ f must be equal to -9, and -8, and -7, -6, $\ldots$, 0, $1/2$, etc. In short, f takes on all intermediate values between -10 and 5. It may take on more values; f may actually equal 6 at some time, for instance, but we are guaranteed all values between -10 and 5 will be covered.

This notion seems intuitive and its importance will turn out to be profound. Therefore the concept is stated in the form of a theorem, the so-called **intermediate value theorem** (*tussenwaardestelling*).

> **Theorem 8.11 (Intermediate value theorem)**
> *Let f be a continuous function on $[a,b]$ and, without loss of generality, let $f(a) < f(b)$. Then for every value u, where $f(a) < u < f(b)$, there is at least one value c in $]a,b[$ such that $f(c) = u$.*

There are several ways to prove the intermediate value theorem, but one way proceeds as follows.

Proof Let S be the set of all $x \in [a,b]$ such that $f(x) \leq u$. Then S is non-empty since a must be an element of S. Moreover, S is bounded above by b. Hence, by completeness, the supremum $c = \sup S$ exists. That is, c is the lowest number that is greater than or equal to every member of S. We now claim that $f(c) = u$.

Fix some $\varepsilon > 0$. Since f is continuous, we have according to Definition 8.1 that there is a $\delta > 0$ such that $\left|f(x) - f(c)\right| < \varepsilon$ whenever $|x - c| < \delta$. This means that

$$f(x) - \varepsilon < f(c) < f(x) + \varepsilon$$

on all $x \in \,]c-\delta, c+\delta[$ (Figure 8.14(a)). By the properties of the supremum, there exist a $a^* \in \,]c-\delta, c]$ that is contained in S (Figure 8.14(b)), so that for that a^*

$$f(c) < f(a^*) + \varepsilon \leq u + \varepsilon.$$

Choose now a $a^{**} \in [c, c+\delta[$ that will obviously not be contained in S (Figure 8.14(c)), so we have

$$f(c) > f(a^{**}) - \varepsilon \geq u - \varepsilon.$$

Both inequalities

$$u - \varepsilon < f(c) < u + \varepsilon$$

are valid for all $\varepsilon > 0$, from which we deduce $f(c) = u$ as the only possible value, as stated. □

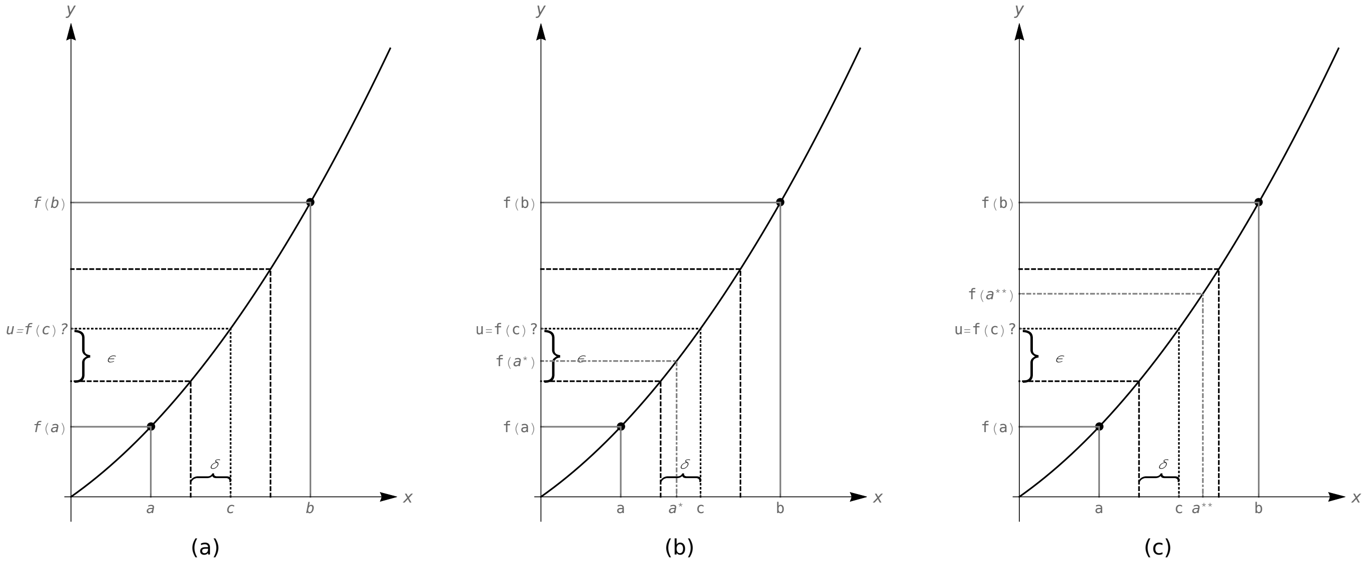


Figure 8.14: Proving the intermediate value theorem.

One important application of the intermediate value theorem is root finding. Given a function f, we are often interested in finding values of x where $f(x) = 0$. These roots may be very difficult to find exactly. Good approximations can be found through successive applications of this theorem. Suppose through direct computation we find that $f(a) < 0$ and $f(b) > 0$, where $a < b$. The intermediate value theorem states that there is at least one c in $]a,b[$ such that $f(c) = 0$. The theorem does not give us any clue as to where to find such a value in the interval $]a,b[$, just that at least one such value exists.

There is a technique that produces a good approximation of c. Let d be the midpoint of the interval $[a,b]$ and consider $f(d)$. There are three possibilities:

1. $f(d) = 0$: We got lucky and stumbled on the actual value. We stop as we found a root.
2. $f(d) < 0$: Then we know there is a root of f on the interval $[d,b]$ – we have halved the size of our interval, hence are closer to a good approximation of the root.
3. $f(d) > 0$: Then we know there is a root of f on the interval $[a,d]$ – again,we have halved the size of our interval, hence are closer to a good approximation of the root.

Successively applying this technique is called the **bisection method** (*halveringsmethode*) of root finding. We continue until the interval is sufficiently small. We demonstrate this in the following example.

Example 8.16

Approximate the root of $f(x) = x - \cos(x)$, accurate to three places after the decimal.

Solution

Consider the graph of $f(x) = x - \cos(x)$, shown in Figure 8.15(a). It is clear that the graph crosses the x-axis somewhere near $x = 0.8$. To start the bisection method, pick an interval that contains 0.8. We choose $[0.7, 0.9]$. Note that all we care about are signs of $f(x)$, not their actual value, so this is all we display.

Iteration 1: $f(0.7) < 0$, $f(0.9) > 0$, and $f(0.8) > 0$. So replace 0.9 with 0.8 and repeat.

Iteration 2: $f(0.7) < 0$, $f(0.8) > 0$, and at the midpoint, 0.75, we have $f(0.75) > 0$. So replace 0.8 with 0.75 and repeat. Note that we do not need to continue to check the endpoints, just the midpoint. Thus we put the rest of the iterations in Table 8.15(b).

Notice that in the 12^{th} iteration we have the endpoints of the interval each starting with 0.739. Thus we have narrowed the zero down to an accuracy of the first three places after the decimal. Using a computer, we have

$$f(0.7390) = -0.00014, \quad f(0.7391) = 0.000024.$$

Either endpoint of the interval gives a good approximation of where f is 0. The intermediate value theorem states that the actual zero is still within this interval. While we do not know its exact value, we know it starts with 0.739.

This type of exercise is rarely done by hand. Rather, it is simple to program a computer to run such an algorithm and stop when the endpoints differ by a preset small amount.

It is a simple matter to extend the bisection method. For instance, we can find x, where $f(x) = 1$. It actually works very well to define a new function g where $g(x) = f(x) - 1$. Then use the bisection method to solve $g(x) = 0$. Similarly, given two functions f and g, we can use this method to solve $f(x) = g(x)$. Once again, create a new function h where $h(x) = f(x) - g(x)$ and solve $h(x) = 0$.

8.6 Limits involving infinity

In Definition 8.1 we stated that in the equation $\lim_{x\to c} f(x) = L$, both c and L were real numbers. In this section we relax that definition a bit by considering situations when it makes sense to let c and/or L be infinity. Essentially, we allow c and/or L to be in the set of extended real number $\overline{\mathbb{R}}$ (Section 2.2).

As a motivating example, consider $f(x) = \frac{1}{x^2}$, as shown in Figure 8.16. Note how, as x approaches 0, $f(x)$ grows very, very large – in fact, it grows without bound. It seems appropriate, and descriptive, to state that

$$\lim_{x\to 0} x^{-2} = +\infty.$$

Also note that as x gets very large, $f(x)$ gets very, very small. We could represent this concept with notation such as

$$\lim_{x\to +\infty} \frac{1}{x^2} = 0.$$

We explore both types of use of infinity in turn.

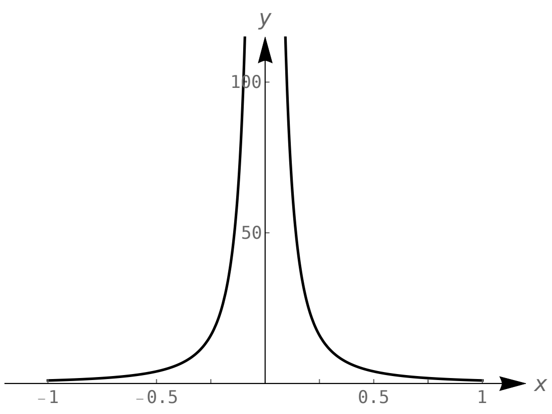
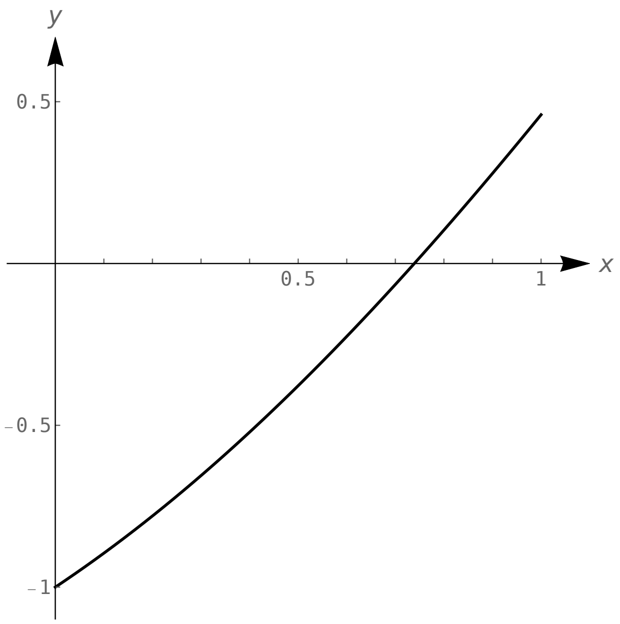


Figure 8.16: Graphing $f(x) = \frac{1}{x^2}$ for values of x near 0.



(a) Graph of $f(x) = x - \cos x$ near a root.

Iteration #	Interval	Midpoint Sign
1	$[0.7, 0.9]$	$f(0.8) > 0$
2	$[0.7, 0.8]$	$f(0.75) > 0$
3	$[0.7, 0.75]$	$f(0.725) < 0$
4	$[0.725, 0.75]$	$f(0.7375) < 0$
5	$[0.7375, 0.75]$	$f(0.7438) > 0$
6	$[0.7375, 0.7438]$	$f(0.7407) > 0$
7	$[0.7375, 0.7407]$	$f(0.7391) > 0$
8	$[0.7375, 0.7391]$	$f(0.7383) < 0$
9	$[0.7383, 0.7391]$	$f(0.7387) < 0$
10	$[0.7387, 0.7391]$	$f(0.7389) < 0$
11	$[0.7389, 0.7391]$	$f(0.7390) < 0$
12	$[0.7390, 0.7391]$	

(b)

Figure 8.15: Finding a root of $f(x) = x - \cos x$.

8.6.1 Limits of infinity and vertical asymptotes

Definitie 8.5 (Limit of infinity)
Let I be an open interval containing c, and let f be a function defined on I, except possibly at c.

- The limit of $f(x)$, as x approaches c, is positive infinity, denoted by

  $$\lim_{x \to c} f(x) = +\infty,$$

  means that given any $N > 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x - c| < \delta$, then $f(x) > N$.

- The limit of $f(x)$, as x approaches c, is negative infinity, denoted by

  $$\lim_{x \to c} f(x) = -\infty,$$

  means that given any $N < 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x - c| < \delta$, then $f(x) < N$.

The first definition is similar to the (ε, δ)-definition (Definition 8.1). In that definition, given any (small) value ε, if we let x get close enough to c (within δ units of c) then $f(x)$ is guaranteed to be within ε of L. Here, given any (large) value N, if we let x get close enough to c (within δ units of c), then $f(x)$ will be at least as large as N. In other words, if we get close enough to c, then we can make $f(x)$ as large as we want.

Of course, we may easily extend the squeeze theorem to limits of infinity, as shown in the following theorem.

Theorem 8.12 (Squeeze theorem involving limits of infinity)
Let f and g be functions on an open interval I containing c such that for all x in I

$$f(x) \leq g(x).$$

Then, the following hold:

1. *if $\lim_{x\to c} f(x) = +\infty$ then $\lim_{x\to c} g(x) = +\infty$;*
2. *if $\lim_{x\to c} g(x) = -\infty$ then $\lim_{x\to c} f(x) = -\infty$.*

We define one-sided limits that approach infinity in a similar way.

Definitie 8.6 (One-sided limits of infinity)

- Let f be a function defined on $]a,c[$ for some $a < c$.

  The limit of $f(x)$, as x approaches c from the left, is infinity, or, the left-hand limit of f at c is positive infinity, denoted by
  $$\lim_{\substack{x\to c\\ <}} f(x) = +\infty,$$
  means given any $N > 0$, there exists $\delta > 0$ such that for all $a < x < c$, if $|x-c| < \delta$, then $f(x) > N$.

- Let f be a function defined on $]c,b[$ for some $b > c$.

  The limit of $f(x)$, as x approaches c from the right, is positive infinity, or, the right-hand limit of f at c is infinity , denoted by
  $$\lim_{\substack{x\to c\\ >}} f(x) = +\infty,$$
  means given any $N > 0$, there exists $\delta > 0$ such that for all $c < x < b$, if $|x-c| < \delta$, then $f(x) > N$.

- The left- (or, right-) hand limit of f at c is negative infinity is defined as in Definition 8.5.

Example 8.17

Find

1. $\lim_{x\to 1} \dfrac{1}{(x-1)^2}$,
2. $\lim_{x\to 0} \dfrac{1}{x}$,
3. $\lim_{\substack{x\to -2\\ >}} \dfrac{x-1}{\sqrt{x+2}}$,
4. $\lim_{x\to 1} \dfrac{\sqrt{x^2+3}-2x}{x-1}$.

Solution

1. In Example 8.2, by inspecting values of x close to 1 we concluded that this limit does not exist. That is, it cannot equal any real number. But the limit could be infinite. And in fact, we see that the function does appear to be growing larger and larger, as $f(0.99) = 10^4$, $f(0.999) = 10^6$, $f(0.9999) = 10^8$. A similar thing happens on the other side of 1. In general,

let a large value N be given. Let $\delta = 1/\sqrt{N}$. If x is within δ of 1, i.e., if $|x-1| < 1/\sqrt{N}$, then:

$$\begin{aligned} |x-1| &< \frac{1}{\sqrt{N}} \\ \Leftrightarrow \quad (x-1)^2 &< \frac{1}{N} \\ \Leftrightarrow \quad \frac{1}{(x-1)^2} &> N, \end{aligned}$$

which is what we wanted to show. So we may say

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

2. It is easy to see that the function grows without bound near 0, but it does so in different ways on different sides of 0. Since its behaviour is not consistent, we cannot say that

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

However, we can make a statement about one–sided limits. We can state that

$$\lim_{\substack{x\to 0 \\ >}} \frac{1}{x} = +\infty, \qquad \text{and,} \qquad \lim_{\substack{x\to 0 \\ <}} \frac{1}{x} = -\infty.$$

3. Here, we are confronted with the composition of an rational and irrational function. The point $x = -2$ does not belong to the corresponding function's domain, as it is a zero of the denominator only, so we expect to find positive or minus infinity. Indeed, we easily find

$$\lim_{\substack{x\to -2 \\ >}} \frac{x-1}{\sqrt{x+2}} = \frac{-3}{0} = -\infty.$$

4. In this case $x = 1$ is a zero of both the numerator and denominator, but we can simplify the expression as follows.

$$\begin{aligned} \lim_{x\to 1} \frac{\sqrt{x^2+3}-2x}{x-1} &= \lim_{x\to 1} \frac{x^2+3-4x^2}{(x-1)\left(\sqrt{x^2+3}+2x\right)} \\ &= \lim_{x\to 1} \frac{-3(x^2-1)}{(x-1)\left(\sqrt{x^2+3}+2x\right)} \\ &= \lim_{x\to 1} \frac{-3(x+1)(x-1)}{(x-1)\left(\sqrt{x^2+3}+2x\right)} = -\frac{3}{2} \end{aligned}$$

In Mathematica, these limits are computed as any other (one-sided) limit. For example, to compute

$$\lim_{\substack{x\to 0 \\ >}} \frac{1}{x},$$

we write the following.

```
In[11]:= Limit[1/x, x → 0, Direction → "FromAbove"]

Out[11]= ∞
```

If a function f has a limit (or, left- or right-hand limit) of infinity at $x = c$, then the graph of f looks similar to a vertical line near $x = c$. This observation leads to a definition.

Definitie 8.7 (Vertical asymptote)
Let I be an interval that either contains c or has c as an endpoint, and let f be a function defined on I, except possibly at c.

If the limit of $f(x)$ as x approaches c from either the left or right (or both) is $+\infty$ or $-\infty$, then the line $x = c$ is a **vertical asymptote** (*verticale asymptoot*) of f.

Example 8.18

Find the vertical asymptotes of

$$f(x) = \frac{3x}{x^2 - 4}.$$

Solution

Vertical asymptotes occur where the function grows without bound; this can occur at values of c where the denominator is 0. When x is near c, the denominator is small, which in turn can make the function take on large values. In the case of the given function, the denominator is 0 at $x = \pm 2$. Substituting in values of x close to 2 and -2 seems to indicate that the function tends toward ∞ or $-\infty$ at those points. We can graphically confirm this by looking at Figure 8.17. Thus the vertical asymptotes are at $x = \pm 2$.

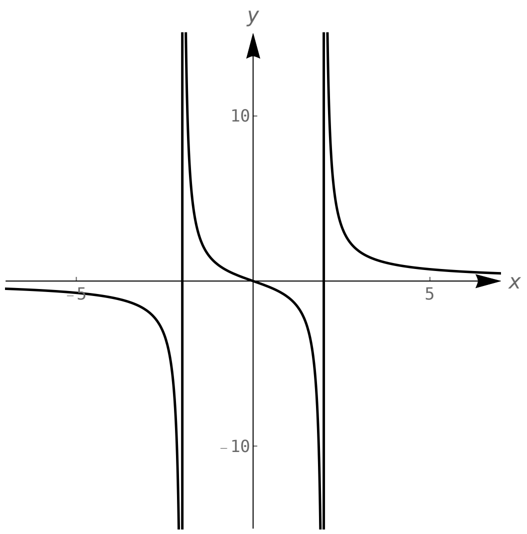


Figure 8.17: Graphing $f(x) = \frac{3x}{x^2-4}$.

If the denominator of a rational function is 0 at a certain point but the numerator is not, then there will usually be a vertical asymptote at that point. On the other hand, if the numerator and denominator are both zero at that point, then there may or may not be a vertical asymptote at that point. This case where the numerator and denominator are both zero returns us to an important topic.

8.6.2 Indeterminate forms

We have seen how the limits

$$\lim_{x\to 0} \frac{\sin(x)}{x} \quad \text{and} \quad \lim_{x\to 1} \frac{x^2-1}{x-1}$$

each return the indeterminate form $0/0$ when we blindly plug in $x = 0$ and $x = 1$, respectively. However, $0/0$ is not a valid arithmetical expression. With a little cleverness, one can come up with $0/0$ expressions which have a limit of ∞, 0, or any other real number. That is why this expression is called **indeterminate** (*onbepaald*).

A key concept to understand is that such limits do not really return $0/0$. Rather, keep in mind that we are taking limits. What is really happening is that the numerator is shrinking to 0 while the denominator is also shrinking to 0. The respective rates at which they do this are very important and determine the actual value of the limit.

An indeterminate form indicates that one needs to do more work in order to compute the limit. That work may be algebraic (such as factoring and cancelling) or it may require a tool such as the squeeze theorem. In Chapter 9 we will learn a technique called l'Hôpital's Rule that provides another way to handle indeterminate forms.

Some other common indeterminate forms are $+\infty - \infty$, $\infty \cdot 0$, ∞/∞, 0^0, ∞^0 and 1^∞. Again, keep in mind that the expression $\infty - \infty$ does not really mean subtract infinity from infinity. Rather, it means one quantity is subtracted from the other, but both are growing without bound. What is the result? It is possible to get every value between $-\infty$ and $+\infty$.

8.6.3 Limits at infinity and horizontal asymptotes

In Figure 8.16 we briefly considered what happens to $f(x) = x^{-2}$ as x grew very large. Graphically, it concerns the behaviour of the function to the far right of the graph. We make this notion more explicit in the following definition.

Definitie 8.8 (Limits at infinity and horizontal asymptotes)
Let L be a real number.

1. Let f be a function defined on $]a, +\infty[$ for some number a. The limit of f at infinity is L, or $\lim_{x\to+\infty} f(x) = L$, means for every $\varepsilon > 0$ there exists $M > a$ such that if $x > M$, then $|f(x) - L| < \varepsilon$.
2. Let f be a function defined on $]-\infty, b[$ for some number b. The limit of f at negative infinity is L, or $\lim_{x\to-\infty} f(x) = L$, means for every $\varepsilon > 0$ there exists $M < b$ such that if $x < M$, then $|f(x) - L| < \varepsilon$.
3. If $\lim_{x\to+\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$, we say the line $y = L$ is a **horizontal asymptote** (*horizontale asymptoot*) of f.

Horizontal asymptotes can take on a variety of forms. Figure 8.18(a) shows that $f(x) = x/(x^2+1)$ has a horizontal asymptote of $y = 0$, where 0 is approached from both above and below. On the other hand, Figure 8.18(b) shows that $f(x) = x/\sqrt{x^2+1}$ has two horizontal asymptotes; one at $y = 1$ and the other at $y = -1$. Figure 8.18(c) shows that $f(x) = (\sin(x))/x$ has even more interesting behaviour than at just $x = 0$; as x approaches $\pm\infty$, $f(x)$ approaches 0, but oscillates as it does this.

We can analytically evaluate limits at infinity for rational functions once we understand $\lim_{x\to+\infty} 1/x$. As x gets larger and larger, $1/x$ gets smaller and smaller, approaching 0. We can, in fact, make $1/x$ as small as we want by choosing a large enough value of x. Given ε, we can make $1/x < \varepsilon$ by choosing

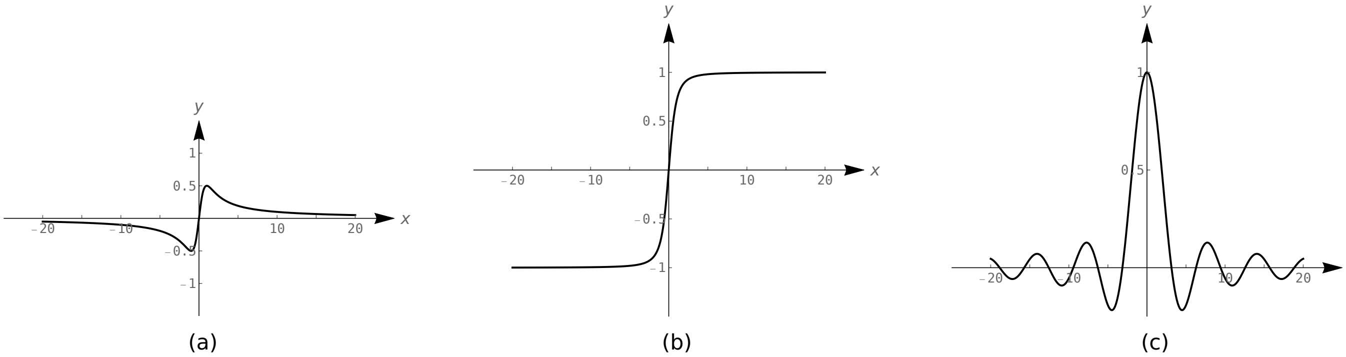


Figure 8.18: A graph of $f(x) = x/(x^2+1)$ (a), $f(x) = x/\sqrt{x^2+1}$ (b) and $f(x) = (\sin(x))/x$.

$x > 1/\varepsilon$. Thus we have $\lim_{x\to+\infty} 1/x = 0$. It is now not much of a jump to conclude the following:

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

Now suppose we need to compute the following limit:

$$\lim_{x\to+\infty} \frac{x^3+2x+1}{4x^3-2x^2+9}.$$

A good way of approaching this is to divide through the numerator and denominator by x^3, which is the largest power of x to appear in the function. Doing this, we get

$$\lim_{x\to+\infty} \frac{x^3+2x+1}{4x^3-2x^2+9} = \lim_{x\to+\infty} \frac{x^3\left(1+2/x^2+1/x^3\right)}{x^3\left(4-2/x+9/x^3\right)}.$$

Then using the rules for limits (which also hold for limits at infinity), as well as the fact about limits of x^{-n}, we see that the limit becomes

$$\frac{1+0+0}{4-0+0} = \frac{1}{4}.$$

This procedure works for any rational function. In fact, it gives us the following theorem.

Theorem 8.13 (Limits of rational functions at infinity)
Let $f(x)$ be a rational function of the following form:

$$f(x) = \frac{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0},$$

where any of the coefficients may be 0 except for a_n and b_m.

1. *If $n = m$, then $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = \frac{a_n}{b_m}$.*
2. *If $n < m$, then $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$.*
3. *If $n > m$, then $\lim_{x\to+\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$ are both infinite.*

Intuitively, as x gets very large, all the terms in the numerator are small in comparison to $a_n x^n$, and likewise all the terms in the denominator are small compared to $b_n x^m$. If $n = m$, looking only at these two important terms, we have $(a_n x^n)/(b_n x^m)$. This reduces to a_n/b_m. If $n < m$, the function behaves like $a_n/(b_m x^{m-n})$, which tends toward 0. If $n > m$, the function behaves like $a_n x^{n-m}/b_m$, which will tend to either $+\infty$ or $-\infty$ depending on the values of n, m, a_n, b_m and whether you are looking for $\lim\limits_{x\to+\infty} f(x)$ or $\lim\limits_{x\to-\infty} f(x)$.

With care, we can quickly evaluate limits at infinity for a large number of functions by considering the largest powers of x. This is, for instance, the case for irrational functions. As an example, consider again

$$\lim_{x\to\pm\infty} \frac{x}{\sqrt{x^2+1}},$$

graphed in Figure 8.18(b). When x is very large, $x^2+1 \approx x^2$. Thus

$$\sqrt{x^2+1} \approx \sqrt{x^2} = |x|, \quad \text{and} \quad \frac{x}{\sqrt{x^2+1}} \approx \frac{x}{|x|}.$$

This expression is 1 when x is positive and -1 when x is negative. Hence we get asymptotes of $y=1$ and $y=-1$, respectively. In general, when evaluation limits at infinity involving irrational functions we should bear in mind that

$$\begin{aligned}
\sqrt{x^2} &= x, && \forall x \in \mathbb{R}^+, \\
\sqrt[3]{x^3} &= x, && \forall x \in \mathbb{R}, \\
\sqrt{x^2} &= -x, && \forall x \in \mathbb{R}^-, \\
\sqrt[3]{(-x)^3} &= -x, && \forall x \in \mathbb{R}.
\end{aligned}$$

Example 8.19

Evaluate each of the following limits.

1. $\lim\limits_{x\to-\infty} \dfrac{x^2+2x-1}{x^3+1}$
2. $\lim\limits_{x\to+\infty} \dfrac{x^2-1}{3-x}$
3. $\lim\limits_{x\to-\infty} \dfrac{\sqrt{x^2+5}+7x}{2x-3}$
4. $\lim\limits_{x\to-\infty} \left(\sqrt{x^2+8x}+x\right)$
5. $\lim\limits_{x\to+\infty} \left(\cosh(x)-\sinh(x)\right)$

Solution

1. The highest power of x is in the denominator. Therefore, the limit is 0.
2. We see that

$$\begin{aligned}
\lim_{x\to+\infty} \frac{x^2-1}{3-x} &= \lim_{x\to+\infty} \frac{x^2\left(1-1/x^2\right)}{x\,(3/x-1)} \\
&= \lim_{x\to+\infty} \frac{x\left(1-1/x^2\right)}{3/x-1} = -\infty.
\end{aligned}$$

3. We first should realize that the highest power of x in both the numerator and denominator is the same. Hence, we proceed as follows:

$$\lim_{x\to-\infty}\frac{\sqrt{x^2+5}+7x}{2x-3} = \lim_{x\to-\infty}\frac{|x|\sqrt{1+\frac{5}{x^2}}+7x}{x\left(2-\frac{3}{x}\right)}$$
$$= \lim_{x\to-\infty}\frac{x\left(-\sqrt{1+\frac{5}{x^2}}+7\right)}{x\left(2-\frac{3}{x}\right)}$$
$$= \frac{-1+7}{2} = 3.$$

4. At first sight, we would say that this limit leads to the indeterminate form $\infty-\infty$, but this can be overcome by multiplying both numerator and denominator by the conjugate expression.

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

5. At first sight, this limit returns the indeterminate form $\infty-\infty$, but we can try to work around it by multiplying the nominator and denominator with the conjugate binomial $\cosh(x)+\sinh(x)$. This leads to

$$\lim_{x\to+\infty}\left(\cosh(x)-\sinh(x)\right) = \lim_{x\to+\infty}\frac{\cosh^2(x)-\sinh^2(x)}{\cosh(x)+\sinh(x)} \tag{8.6}$$
$$= \lim_{x\to+\infty}\frac{1}{\cosh(x)+\sinh(x)} \tag{8.7}$$
$$= 0\,. \tag{8.8}$$

In Mathematica, these limits are again computed as any other limit. To specify that the limit is at (minus) infinity, we simply write `(-)`**`Infinity`**. For example,

$$\lim_{x\to-\infty}\frac{x^2+2x-1}{x^3+1},$$

is computed as follows.

```
In[12]:= Limit[(x^2 + 2 x - 1) (x^3 + 1), x → -Infinity]

Out[12]= 0
```

For the sake of comprehensiveness, we list below the steps that should be taken when evaluating the limit $\lim_{x\to\pm\infty} f(x)$:

1. Compute $f(\pm\infty)$.
2. You arrive at one of the following cases:
   - $f(\pm\infty)=\pm\infty$: the function values approaches ∞ as $x\to\pm\infty$

- $f(\pm\infty) = b \in \mathbb{R}$: $y = b$ is a horizontal asymptote of the function f.
- $f(\pm\infty) = \left(\frac{\infty}{\infty}\right)$: factor out the highest-degree term in both the nominator and denominator, then simplify and return to Step 1.
- $f(\pm\infty) = (\infty - \infty)$: multiply with the conjugate binomial, then factor out the highest-degree term and return to Step 1.

8.6.4 Slant asymptotes

In addition to vertical and horizontal asymptotes, we can also define **slant or oblique asymptotes** (*schuine asymptoot*). These are diagonal lines such that the difference between the curve and the line approaches 0 as x tends to $+\infty$ or $-\infty$.

Definitie 8.9 (Slant asymptotes)
The line $y = ax + b$ is a **slant asymptote** for the function f if and only if

$$\lim_{x\to+\infty}\left[f(x) - (ax+b)\right] = 0,$$

or/and

$$\lim_{x\to-\infty}\left[f(x) - (ax+b)\right] = 0\,.$$

From this definition, it follows that

$$\lim_{x\to\pm\infty}\left[\frac{f(x)}{x} - \left(a + \frac{b}{x}\right)\right] = 0,$$

so

$$a = \lim_{x\to\pm\infty}\frac{f(x)}{x} \qquad \text{and} \qquad b = \lim_{x\to\pm\infty}(f(x) - ax)\,.$$

Example 8.20

Determine the asymptotes, if any, of the following functions.

1. $f(x) = \dfrac{x^3+2}{x^2-9}$
2. $f(x) = \sqrt{x^2-4x+3}$

Solution

1. (a) Vertical asymptotes

Since $f(x)$ tends to infinity as x approaches 3 or -3, the vertical asymptotes are $x = -3$ and $x = 3$. More precisely, we find that

$$\lim_{\substack{x\to-3\\<}} f(x) = -\infty \qquad \lim_{\substack{x\to 3\\<}} f(x) = -\infty$$
$$\lim_{\substack{x\to-3\\>}} f(x) = +\infty \qquad \lim_{\substack{x\to 3\\>}} f(x) = +\infty$$

This allows us to conclude that $f(x)$ tends towards $-\infty$ as x is approaching -3 from the left, whereas $f(x)$ tends towards ∞ when approaching -3 from the right, and likewise for what concerns the vertical asymptote at $x = 3$.

(b) Horizontal asymptotes

There are none because

$$\lim_{x\to\pm\infty} \frac{x^3+2}{x^2-9} = \pm\infty.$$

(c) Slant asymptotes

We verify that

$$a = \lim_{x\to\pm\infty} \frac{f(x)}{x} = \lim_{x\to\pm\infty} \frac{x^3+2}{x^3-9x} = 1,$$

$$b = \lim_{x\to\pm\infty} \big(f(x) - ax\big) = \lim_{x\to\pm\infty} \left(\frac{x^3+2}{x^2-9} - x\right) = \lim_{x\to\pm\infty} \left(\frac{2+9x}{x^2-9}\right) = 0.$$

Consequently, the function has slant asymptote $y = x$ for $x \to \pm\infty$. The position of this asymptote with respect to the graph of the function f can be found be determining the sign of

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

The sign diagram of the function g is:

x		-3		$-\frac{2}{9}$		3	
$g(x)$	$-$	|	$+$	0	$-$	|	$+$

Consequently, we may conclude that the graph of f lies above the slant asymptote for $x \to +\infty$ because then $g(x) > 0$, whereas the graph of f lies below the slant asymptote for $x \to -\infty$ because then $g(x) < 0$. This is confirmed by the graph of the function f shown in Figure 8.19(a).

2. $f(x) = \sqrt{x^2-4x+3}$

(a) Vertical asymptotes

The function $f(x)$ only tends to infinity if x does so, so there are no vertical asymptotes.

(b) Horizontal asymptotes

There are none because

$$\lim_{x\to\pm\infty} \sqrt{x^2-4x+3} = +\infty.$$

(c) Slant asymptotes

We verify that

- $x \to +\infty$

$$a = \lim_{x\to+\infty} \frac{f(x)}{x} = \lim_{x\to+\infty} \frac{\sqrt{x^2-4x+3}}{x} = 1,$$

$$b = \lim_{x\to+\infty} \big(f(x) - ax\big) = \lim_{x\to+\infty} \left(\sqrt{x^2-4x+3} - x\right)$$

$$= \lim_{x\to+\infty}\left(\frac{-4x+3}{\sqrt{x^2-4x+3}+x}\right) = -2.$$

- $x \to -\infty$

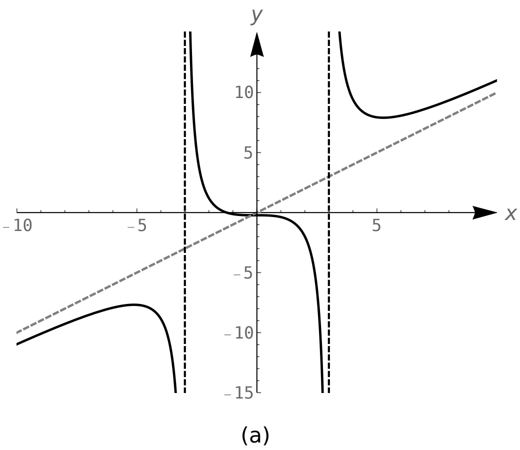
$$a = \lim_{x\to-\infty}\frac{f(x)}{x} = \lim_{x\to-\infty}\frac{\sqrt{x^2-4x+3}}{x} = -1,$$

$$\begin{aligned} b &= \lim_{x\to-\infty}\big(f(x)-ax\big) = \lim_{x\to-\infty}\left(\sqrt{x^2-4x+3}+x\right) \\ &= \lim_{x\to-\infty}\left(\frac{-4x+3}{\sqrt{x^2-4x+3}-x}\right) = 2. \end{aligned}$$

Hence, $y = x-2$ is a slant asymptote of f for $x \to +\infty$, while $y = -x+2$ is a slant asymptote of f for $x \to -\infty$. Again, the position of the slant asymptote with respect to the graph of f can be determined by verifying the sign of

$$g(x) = \sqrt{x^2-4x+3}-(x-2) \qquad \text{and} \qquad h(x) = \sqrt{x^2-4x+3}-(-x+2).$$

It follows that as $x \to +\infty$, then we have that $g(x) < 0$, so the graph of f lies below the slant asymptote $y = x-2$. Similarly, as $x \to -\infty$, then we have that $h(x) < 0$, so the graph of f lies below the slant asymptote $y = -x+2$. This is confirmed by the graph of the function f shown in Figure 8.19(b).



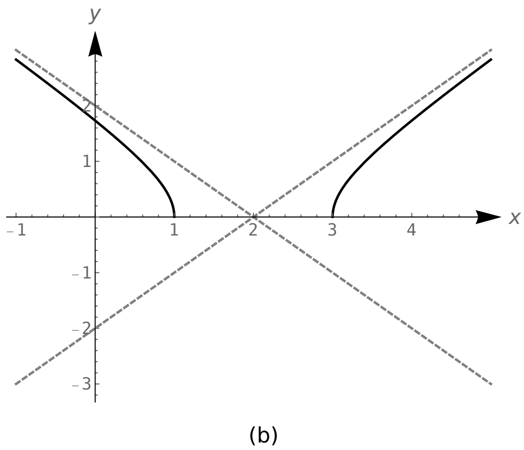


Figure 8.19: A graph of $f(x) = \frac{x^3+2}{x^2-9}$ (a) and $f(x) = \sqrt{x^2-4x+3}$ (b).

8.7 Exercises

Epsilon-delta definition of a limit

Assignment 8.1 — Consider

$$f(x) = \sqrt{2x+3}, \quad c = 3, \quad L = 3, \quad \varepsilon = 0,01.$$

Find a $\delta > 0$ such that $|x - c| < \delta$, $|f(x) - L|$ is smaller than the given ε.

Assignment 8.2 — Prove the given limit by using the $(\varepsilon,\ \delta)$-definition (Definition 8.1).

(a) $\lim_{x\to 5}(3-x) = -2$ (b) $\lim_{x\to 4}\left(x^2+x-5\right) = 15$ (c) $\lim_{x\to 0}\sin(x) = 0$

Finding limits analytically

Assignment 8.3 — Prove the product rule for limits. If

$$\lim_{x\to c} f(x) = L_1 \quad \text{and} \quad \lim_{x\to c} g(x) = L_2,$$

than

$$\lim_{x\to c} f(x)g(x) = L_1 L_2$$

or:

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

Assignment 8.4 — Calculate the following limits.

(a) $\displaystyle \lim_{x \underset{>}{\to} 1} \frac{\sqrt{(x-1)^2}}{x-1}$

(b) $\displaystyle \lim_{x \underset{>}{\to} -4} \frac{2x+8}{|x+4|}$

(c) $\displaystyle \lim_{x \to 1} \frac{(x-1)^3}{x^2-4x+3}$

(d) $\displaystyle \lim_{x \to -1} \frac{x^3+2x^2+x}{x^8-2x^4+1}$

(e) $\displaystyle \lim_{x \to 1} \frac{x-\sqrt{x}}{2-\sqrt{x+3}}$

(f) $\displaystyle \lim_{x \to 5} \frac{\sqrt{x+4}+x-8}{x^2-8x+15}$

(g) $\displaystyle \lim_{x \to 1} \frac{\sqrt{8x+1}+\sqrt{2x-1}-4}{x-1}$

(h) $\displaystyle \lim_{x \to 0} \frac{|3x-1|-|3x+1|}{x}$

(i) $\displaystyle \lim_{x \to 3} \frac{x^2-6x+9}{x^2-9}$

(j) $\displaystyle \lim_{h \to 0} \frac{\sqrt{4+h}-2}{h}$

(k) $\displaystyle \lim_{x \to 1} \frac{x^2-1}{\sqrt{x+3}-2}$

(l) $\displaystyle \lim_{x \to 2} \frac{x^4-16}{x^3-8}$

(m) $\displaystyle \lim_{x \to 8} \frac{x^{2/3}-4}{x^{1/3}-2}$

(n) $\displaystyle \lim_{x \to 2} \left(\frac{1}{x-2}-\frac{1}{x^2-4}\right)$

(o) $\displaystyle \lim_{x \to 64} \frac{x^{1/3}-4}{x^{1/2}-8}$

(p) $\displaystyle \lim_{x \to 1} \frac{\sqrt{3+x}-2}{\sqrt[3]{7+x}-2}$

Assignment 8.5 — Determine $\displaystyle \lim_{x \to 0} \left(x^2 \sin\left(\frac{1}{x}\right)\right)$ by using the squeeze theorem.

Assignment 8.6 — Assume $|f(x)| \leq g(x)$ for all x. What can be concluded for $\lim_{x \to c} f(x)$ if (a) $\lim_{x \to c} g(x) = 0$ and (b) $\lim_{x \to c} g(x) = 3$?

One-sided limits

Assignment 8.7 — Consider the function $y = f(x)$ as given in Figure 8.20 and calculate the limits.

(a) $\lim_{x \underset{>}{\to} 0} f(x)$

(b) $\lim_{x \to 1} f(x)$

(c) $\lim_{x \underset{>}{\to} 2} f(x)$

(d) $\lim_{x \underset{<}{\to} 2} f(x)$

(e) $\lim_{x \underset{<}{\to} 3} f(x)$

(f) $\lim_{x \underset{>}{\to} 3} f(x)$

(g) $\lim_{x \underset{>}{\to} 4} f(x)$

(h) $\lim_{x \underset{<}{\to} 4} f(x)$

(i) $\lim_{x \underset{<}{\to} 5} f(x)$

(j) $\lim_{x \underset{>}{\to} 5} f(x)$

(k) $\lim_{x \to +\infty} f(x)$

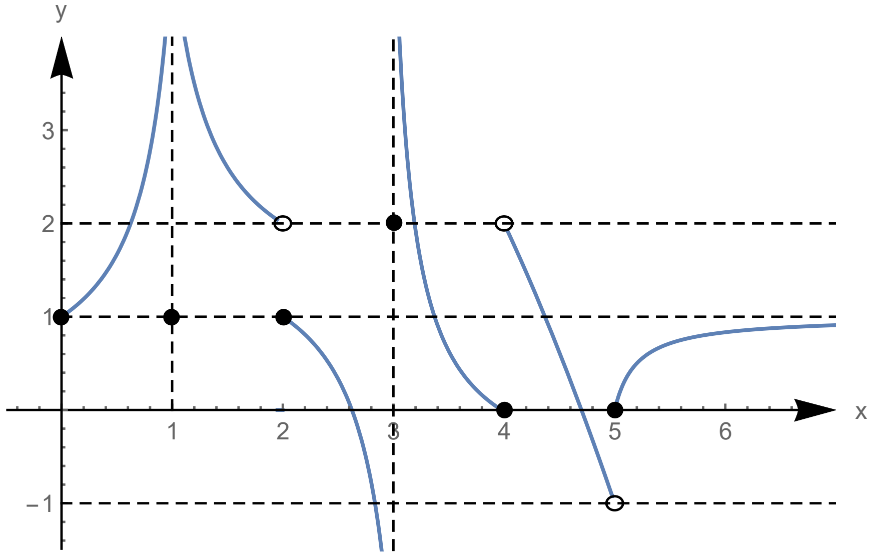


Figure 8.20: The function $y = f(x)$ from Exercise 8.7 and 8.10.

Assignment 8.8 — Consider the function

$$f(x) = \begin{cases} x-1, & \text{if } x \leq -1, \\ x^2+1, & \text{if } -1 < x \leq 0, \\ (x+\pi)^2, & \text{if } x > 0. \end{cases}$$

Compute the following limits

(a) $\lim\limits_{\substack{x \to -1 \\ <}} f(x)$

(b) $\lim\limits_{\substack{x \to -1 \\ >}} f(x)$

(c) $\lim\limits_{\substack{x \to 0 \\ >}} f(x)$

(d) $\lim\limits_{\substack{x \to 0 \\ <}} f(x)$

Assignment 8.9 — If $\lim\limits_{\substack{x \to 0 \\ >}} f(x) = A$ and $\lim\limits_{\substack{x \to 0 \\ <}} f(x) = B$, determine

(a) $\lim\limits_{\substack{x \to 0 \\ <}} f\left(x^2 - x^4\right)$,

(b) $\lim\limits_{\substack{x \to 0 \\ >}} f\left(x^2 - x^4\right)$.

Continuity

Assignment 8.10 — Consider Figure 8.20 once more. In which points is $f(x)$ discontinuous? In which points is $f(x)$ left or right continuous? Can the function $f(x)$ be defined in $x = 1$ such that $f(x)$ becomes continuous in this point?

Assignment 8.11 — Determine the value of a and/or b such that the functions below are continuous for all x.

(a) $f(x) = \begin{cases} x-a, & \text{if } x < 3, \\ 1-ax, & \text{if } x \geq 3, \end{cases}$

(b) $f(x) = \begin{cases} \dfrac{a}{x-2}, & \text{if } x \leq 0, \\ 2x-b, & \text{if } 0 < x < 2, \\ 6, & \text{if } x \geq 2, \end{cases}$

(c) $f(x) = \begin{cases} x^2, & \text{if } 1 \leq x \leq 2, \\ ax+b, & \text{if } 2 < x, \\ ax-b, & \text{if } x < 1, \end{cases}$

(d) $f(x) = \begin{cases} x^3, & \text{if } x \leq 2, \\ ax^2, & \text{if } x > 2. \end{cases}$

Assignment 8.12 — Identify and classify the discontinuities of the functions below.

(a) $f(x) = \dfrac{x}{(x+3)^3}$

(b) $f(x) = \dfrac{x^2+5x+6}{x+2}$

Assignment 8.13 — Investigate whether the given functions in the given interval I contains at least one zero. If so, determine the zero(s) using the bisection method.

(a) $f(x) = x^5 - 5x + 1, \quad I = [0,1]$

(b) $f(x) = x^3 + 8x^2 - 3, \quad I = [-10,-1]$

(c) $f(x) = \dfrac{|x|}{x}, \quad I = [-1,1]$

(d) $f(x) = 1 - x + \sin(x), \quad I = [0,\pi]$

(e) $f(x) = x^3 - 15x + 1, \quad I = [-4,4]$

Assignment 8.14 — Prove that $\cos(x) = x$ has at least one solution.

Assignment 8.15 — Consider the function $f(x) = \dfrac{x^2+x}{x-1}$ with $x \in [5/2, 4]$. Assume c such that $f(c) = 6$. Verify that the intermediate value theorem is applicable and find a value of c that satisfies the stated condition.

Limits involving infinity

Assignment 8.16 — Determine the following limits:

(a) $\lim\limits_{x\to\pm\infty} \dfrac{6x^3+4x^2-x-1}{x^4+5}$

(b) $\lim\limits_{x\to\pm\infty} \dfrac{\sqrt{x^2-1}+2x}{x+7}$

(c) $\lim\limits_{x\to+\infty} \dfrac{2x-3}{\sqrt{x^2+4}-\sqrt{2x}}$

(d) $\lim\limits_{x\to\pm\infty} \dfrac{\sqrt{x^4+x^3+8}}{1-x^2}$

(e) $\lim\limits_{x\to\pm\infty} \left(\sqrt{4x^2+3}-2x\right)$

(f) $\lim\limits_{x\to\pm\infty} \left(2x-1-\sqrt{4x^2+x}\right)$

(g) $\lim\limits_{x\to+\infty} \dfrac{3x^3-5x^2+7}{8+2x-5x^3}$

(h) $\lim\limits_{x\to\pm\infty} \dfrac{2x-1}{\sqrt{3x^2+x+1}}$

(i) $\lim\limits_{x\to-\infty} \dfrac{2x-5}{|3x+2|}$

(j) $\lim\limits_{x\to\pm\infty} \left(x+\sqrt{x^2-4x+1}\right)$

(k) $\lim\limits_{x\to+\infty} \dfrac{x\sqrt{x+1}\left(1-\sqrt{2x+3}\right)}{7-6x+4x^2}$

(l) $\lim\limits_{x\to\pm\infty} \dfrac{1}{\sqrt{x^2-2x}-x}$

Assignment 8.17 — Determine the asymptotes of the graph of the functions below. Also, determine how these graphs approach the asymptotes.

(a) $f(x) = \dfrac{1-x}{2x-1}$

(b) $f(x) = \dfrac{x^2-x-2}{x-3}$

(c) $f(x) = \dfrac{3x^2-5}{x^2+7}$

(d) $f(x) = \dfrac{2x^3+x}{x^2+1}$

(e) $f(x) = x - \dfrac{7}{x}$

(f) $f(x) = \dfrac{x^4-1}{3x^3-12x}$

(g) $f(x) = \sqrt{x^2-4}-3$

(h) $f(x) = \sqrt{x^2+2x-3}$

(i) $f(x) = x-\sqrt{x^2-4}$

(j) $f(x) = \dfrac{x^2+x-6}{\sqrt{x^2-x-6}}$

Review Exercises

Assignment 8.18 — Calculate the following limits.

(a) $\lim\limits_{x\to+\infty} \dfrac{\sin(x)}{x}$

(b) $\lim\limits_{x\to-\infty} (x+\cos(x))$

(c) $\lim\limits_{x\to+\infty} \dfrac{1+\cos(x)}{x^2}$

(d) $\lim\limits_{x\to 0} \dfrac{\sin(3x)}{x}$

(e) $\lim\limits_{x\to 0} \dfrac{\tan(x)}{x}$

(f) $\lim\limits_{x\to 0} \dfrac{\sin(2x)}{\tan(5x)}$

(g) $\lim\limits_{x\to 0} \dfrac{\sin^2(x)}{x.\tan(2x)}$

(h) $\lim\limits_{x\to 0} \dfrac{\sin^2(7x)}{\tan^2(2x)}$

(i) $\lim\limits_{x\to 0} \dfrac{1-\cos(4x)}{x^2}$

(j) $\lim\limits_{x\to 0} \dfrac{1-\sqrt{1-\sin(x)}}{x}$

Assignment 8.19 — Calculate the following limits.

(a) $\displaystyle\lim_{x\to+\infty}\left(\frac{2x+1}{2x}\right)^x$

(b) $\displaystyle\lim_{x\to 0}(1+3x)^{1/x}$

(c) $\displaystyle\lim_{x\to 0}\frac{e^{-x}-1}{x}$

(d) $\displaystyle\lim_{x\to 0}\frac{e^{ax}-1}{x}$

(e) $\displaystyle\lim_{x\to a}\frac{e^x-e^a}{x-a}$

(f) $\displaystyle\lim_{m\to+\infty}\left(\frac{m^2-1}{m^2+1}\right)^{m^2}$

(g) $\displaystyle\lim_{m\to+\infty}\left(1-\frac{1}{m}\right)^m$

Assignment 8.20 — Investigate the continuity of the following functions within their domain.

(a) $f(x)=\sin(4x)$

(b) $f(x)=\dfrac{x+2}{x^2-9}$

(c) $f(x)=\sqrt{x^2+5}$

(d) $f(x)=\sqrt{x^2-1}$

(e) $f(x)=\begin{cases} x^2, & \text{als } x\in\mathbb{R}^+,\\ -x, & \text{als } x\in\mathbb{R}^-,\end{cases}$

(f) $f(x)=\dfrac{|x|}{x}$

(g) $f(x)=\begin{cases} \dfrac{|4-x^2|}{2-x}, & \text{als } x\neq 2,\\ 4, & \text{als } x=2,\end{cases}$

(h) $f(x)=\begin{cases} x+\dfrac{\sqrt{x^2}}{x}, & \text{als } x\neq 0\\ 1, & \text{als } x=0.\end{cases}$

Assignment 8.21 — Determine whether the graphs of the rational functions below have one or more vertical asymptotes and/or perforations.

(a) $f(x)=\dfrac{x^2+5x+6}{x+3}$

(b) $f(x)=\dfrac{x^2+3x-4}{x^2+x-6}$

(c) $f(x)=\dfrac{x^2-9}{x^2-2x-3}$

Q: What is the first derivative of a cow?
A: Prime Rib!

— Queen Elizabeth II —

9

Derivatives and their applications

The previous chapter introduced the most fundamental of calculus topics: the limit. This chapter introduces the second most fundamental of calculus topics: the derivative. Limits describe where a function is going; derivatives describe how fast the function is going.

9.1 Definition

9.1.1 Intuitive introduction

A common amusement park ride lifts riders to a height then allows them to freefall a certain distance before safely stopping them. Suppose such a ride drops riders from a height of 150 metres. From physics we know that the height (in metres) of the riders, t seconds after freefall (and ignoring air resistance, etc.) can be accurately modelled by $f(t) = -16t^2 + 150$. It allows us to verify that, without intervention, the riders will hit the ground at $t = 2.5\sqrt{1.5} \approx 3.06$ seconds, but how fast will the riders be travelling after two seconds?

We have been given a position function, but what we want to compute is a velocity at a specific point in time, i.e., we want an instantaneous velocity. We do not currently know how to calculate this. However, we do know how to calculate an average velocity using the difference quotient introduced in Section 8.1. More specifically, we have

$$\frac{\text{change in distance}}{\text{change in time}} = \frac{\text{rise}}{\text{run}} = \text{average velocity.}$$

We can approximate the instantaneous velocity at $t = 2$ by considering the average velocity over some time period containing $t = 2$. If we make the time interval small, we will get a good approximation. For instance, consider the interval from $t = 2$ to $t = 3$. On that interval, the average velocity is

$$\frac{f(3)-f(2)}{3-2} = \frac{f(3)-f(2)}{1} = -80 \text{ m/s},$$

where the minus sign indicates that the riders are moving down. By narrowing the considered interval, we get a better approximation of the instantaneous velocity. On $[2, 2.5]$ we have

$$\frac{f(2.5)-f(2)}{2.5-2} = \frac{f(2.5)-f(2)}{0.5} = -72 \text{ m/s}.$$

We can do this for smaller and smaller intervals of time. For instance, over a time span of 1/10$^{\text{th}}$ of a second, i.e., on $[2, 2.1]$, we have

$$\frac{f(2.1)-f(2)}{2.1-2} = \frac{f(2.1)-f(2)}{0.1} = -65.6 \text{ m/s}.$$

Likewise, over a time span of 1/100$^{\text{th}}$ of a second, on $[2, 2.01]$, the average velocity is

$$\frac{f(2.01)-f(2)}{2.01-2} = \frac{f(2.01)-f(2)}{0.01} = -64.16 \text{ m/s}.$$

Essentially, we are computing the average velocity on the interval $[2, 2+h]$ for small values of h. That is, we are computing

$$\frac{f(2+h)-f(2)}{h},$$

where h is small. Still, we really want to use $h = 0$, but this, of course, returns the indeterminate form $0/0$. Computing this limit directly gives

$$\begin{aligned}
\lim_{h\to 0} \frac{f(2+h)-f(2)}{h} &= \lim_{h\to 0} \frac{-16(2+h)^2+150-(-16(2)^2+150)}{h} \\
&= \lim_{h\to 0} \frac{-64h-16h^2}{h} \\
&= \lim_{h\to 0} (-64-16h) \\
&= -64.
\end{aligned}$$

Graphically, we can view the average velocities we computed numerically as the slopes of secant lines on the graph of f going through the points $(2, f(2))$ and $(2+h, f(2+h))$. In Figure 9.1(a), the secant line corresponding to $h = 1$ is shown. Notice how well it approximates f between those two points – it is a common practice to approximate functions with straight lines. As $h \to 0$, these secant lines approach the **tangent line** (*raaklijn*), a line that goes through the point $(2, f(2))$ with the special slope of -64 (Figure 9.1(b)). It is clear that this tangent line approximates the function f even better than the secant line.

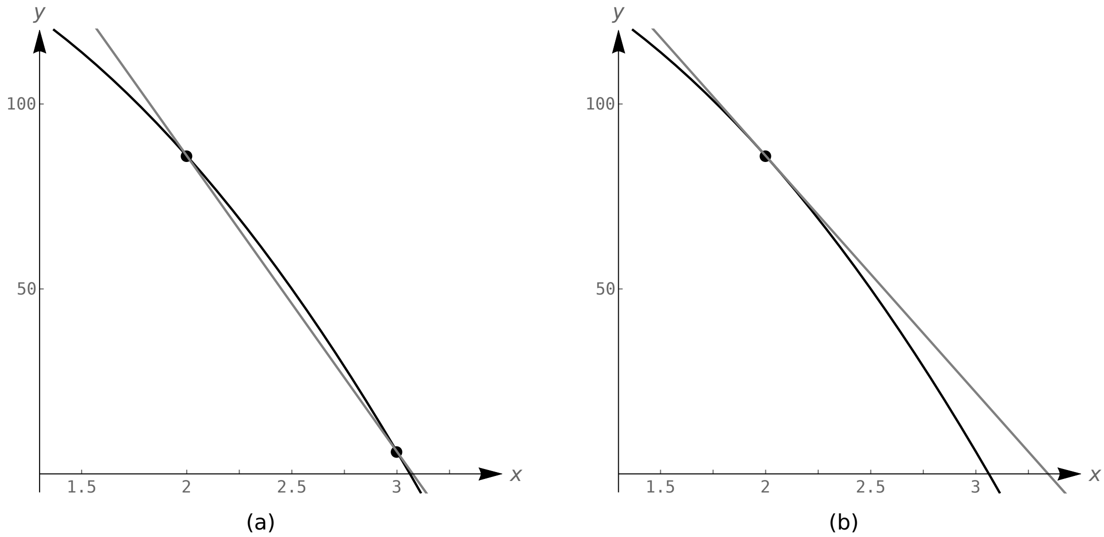


Figure 9.1: The secant line to $f(x)$ with $h = 1$ (a) and the tangent line to f at $x = 2$.

9.1.2 Formalism

Having introduced the derivative in an intuitive way, let us now turn to its formal definition.

Definitie 9.1 (Derivative at a point)
Let f be a continuous function on an open interval I and let c be in I. The **derivative** (*afgeleide*) of f at c, denoted $f'(c)$, is

$$\lim_{h\to 0}\frac{f(c+h)-f(c)}{h},$$

provided the limit exists. If the limit exists, we say that f is differentiable at c; if the limit does not exist, then f is not differentiable at c. If f is **differentiable** (*afleidbaar*) at every point in I, then f is differentiable on I. Furthermore, we call f continuously differentiable over I if f' is continuous over I.

Using this definition, we can also formally define a tangent line to the graph of a function f.

Definitie 9.2 (Tangent line)
Let f be continuous on an open interval I and differentiable at c, for some c in I. The line with equation $y = \ell(x)$

$$y = f'(c)(x-c)+f(c)\,,$$

is the **tangent line** (*raaklijn*) to the graph of f at c; that is, it is the line through $(c,f(c))$ whose slope is the derivative of f at c.

When $f'(c) = 0$, the tangent line is the horizontal line through $\big(c,f(c)\big)$; that is, $y = f(c)$. Moreover, the larger $f'(c)$ the more the tangent lines becomes oriented vertically.

Given the notion of differentiability introduced through Definition 9.1 and our understanding of a tangent line (Definition 9.2), we can give an alternative formulation of a function that is differentiable at a point c. This is accomplished by the next theorem, which is also know as Carathéodory's theorem.

Theorem 9.1 (Differentiability at a point)
Let f be a function on an open interval I. Then, f is differentiable at c if and only if there exists a

function g on I that is continuous at c and satisfies the following :

1. $\forall x \in I : f(x) - f(c) = g(x)(x-c)$
2. $g(c) = f'(c)$

Proof Let us first prove that the conditions above are necessary. For that purpose, suppose f is differentiable at c. Then by Definition 9.1, $f'(c)$ exists. So we can define the function g by:

$$g(x) = \begin{cases} \dfrac{f(x)-f(c)}{x-c}, & x \neq c, x \in I \\ f'(c), & x = c\,. \end{cases}$$

Since we have that $\lim\limits_{x\to c} g(x) = f'(c) = g(c)$, it follows that g is continuous at c. Moreover, for $x \neq c$ we obtain

$$g(x)(x-c) = f(x) - f(c)\,,$$

whereas for $x = c$ both sides of the equation in the first condition are zero.

As a second step, we prove that the conditions in Theorem 9.1 are sufficient. For that purpose, suppose a function g as mentioned in the theorem statement exists. Then for $x \neq c$, we have that

$$g(x) = \frac{f(x)-f(c)}{x-c}.$$

Moreover, since g is continuous at c it follows that

$$g(x) = \lim_{x\to c} g(x) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}\,,$$

which implies that f is differentiable at c with derivative $f'(c) = g(c)$. □

Clearly, from the derivative we can also construct the normal line. It is perpendicular to the tangent line, hence its slope is the opposite–reciprocal of the tangent line's slope.

Definitie 9.3 (Normal line)
Let f be continuous on an open interval I and differentiable at c, for some c in I. The **normal line** (*normaal*) to the graph of f at c is the line with equation

$$n(x) = \frac{-1}{f'(c)}(x-c) + f(c),$$

where $f'(c) \neq 0$.

When $f'(c) = 0$, the normal line is the vertical line through $\big(c, f(c)\big)$; that is, $x = c$.

Some examples will help us understand these definitions.

Example 9.1

Let $f(x) = 3x^2 + 5x - 7$. Find:

1. $f'(1)$.
2. The equation of the tangent line to the graph of f at $x = 1$.
3. The equation of the normal line to the graph of f at $x = 1$.

Solution

1. We compute this directly using Definition 9.1.

$$\begin{aligned} f'(1) &= \lim_{h\to 0} \frac{f(1+h)-f(1)}{h} \\ &= \lim_{h\to 0} \frac{3(1+h)^2+5(1+h)-7-(3(1)^2+5(1)-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+11h}{h} \\ &= \lim_{h\to 0}(3h+11) = 11 \end{aligned}$$

2. The tangent line at $x=1$ has slope $f'(1)$ and goes through the point $(1,f(1)) = (1,1)$. Thus the tangent line has equation, in point-slope form, $y = 11(x-1)+1$. In slope-intercept form, we have $y = 11x-10$.

3. Since $f'(1) = 11$. Hence at $x=1$, the normal line will have slope $-1/11$. An equation for the normal line is
$$n(x) = \frac{-1}{11}(x-1)+1.$$

A graph of f is given in Figure 9.2 along with its tangent and normal lines at $x=1$. Note that in this figure these lines do not seem to perpendicular to one another, but this is a mere consequence of the chose aspect ration of the plot window.

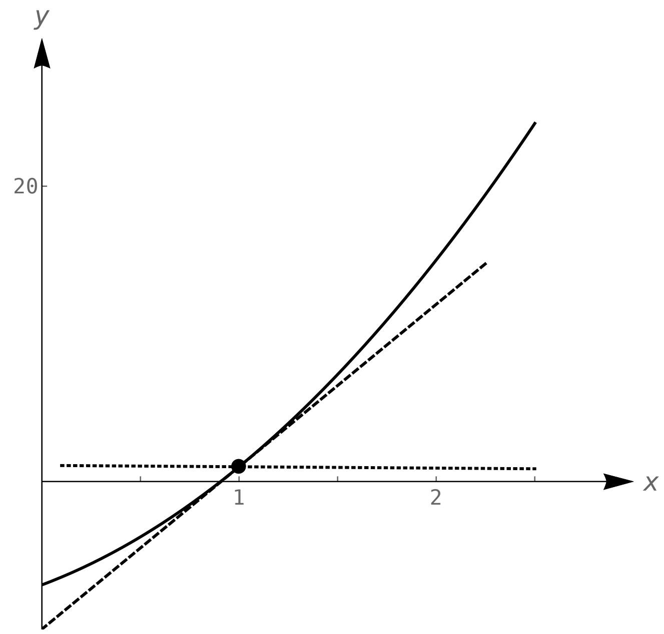


Figure 9.2: A graph of $f(x) = 3x^2+5x-7$ and its tangent (dashed) and normal (dotted) lines at $x=1$.

Linear functions are easy to work with; many functions that arise in the course of solving real problems, however, are not easy to work with. A common practice in mathematical problem solving is to approximate difficult functions with not–so–difficult functions. Lines are a common choice. It turns out that at any given point on the graph of a differentiable function f, the best linear approximation to f is its tangent line. That is one reason we will spend considerable time finding tangent lines to functions.

From Example 9.1, it is clear that we would have to evaluate a limit for every point c at which we want to find the derivative of f. Yet, instead of doing this repeatedly for different values of c, let us do it just once for the variable x. We then take a limit just once. The process now looks like:

$$\text{input variable } x \longrightarrow \boxed{\begin{array}{c}\text{do something}\\ \text{to } f \text{ and } x\end{array}} \longrightarrow \text{return function } f'(x)$$

The output is the derivative function, $f'(x)$. The $f'(x)$ function will take a number c as input and return the derivative of f at c. This gives rise to the following definition.

Definitie 9.4 (Derivative function)
Let f be a differentiable function on an open interval I. The function

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

is the derivative of f.

Note that the following notations all represent the derivative of a function f, if defined as $y = f(x)$:

$$f'(x) \ = \ y' \ = \ \frac{dy}{dx} \ = \ \frac{df}{dx} \ = \ \frac{d}{dx}(f) \ = \ \frac{d}{dx}(y).$$

Example 9.2

Find the derivative of the following functions:

1. $f(x) = 3x^2 + 5x - 7$
2. $f(x) = \sin(x)$

Solution

1. We apply Definition 9.4.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{3(x+h)^2+5(x+h)-7-(3x^2+5x-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+6xh+5h}{h} \\ &= \lim_{h\to 0}(3h+6x+5) \\ &= 6x+5 \end{aligned}$$

So $f'(x) = 6x+5$. Recall earlier we found that $f'(1) = 11$, which is affirmed by our new computation of $f'(x)$. Moreover, we can verify the correctness of our computation using Mathematica. More precisely, we can compute derivatives in Mathematica using the built-in command **D** as follows.

```
In[13]:= D[3*x^2+5*x-7,x]

Out[13]= 5+6x
```

The second argument of the command **D** indicates the variable with respect to which the derivative is computed.

2. We again apply Definition 9.4,

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h} \\
& \quad \left(\begin{array}{c} \text{Use trig identity } \sin(x+h) = \\ \sin(x)\cos(h)+\cos(x)\sin(h) \end{array} \right) \\
&= \lim_{h\to 0} \frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h} && \text{(Sine of sum.)} \\
&= \lim_{h\to 0} \frac{\sin(x)(\cos(h)-1)+\cos(x)\sin(h)}{h} && \text{(Regroup.)} \\
&= \lim_{h\to 0} \left(\frac{\sin(x)(\cos(h)-1)}{h} + \frac{\cos(x)\sin(h)}{h} \right) && \text{(Split into two fractions.)} \\
& \quad \left(\text{use } \lim_{h\to 0} \frac{\cos(h)-1}{h} = 0 \text{ and } \lim_{h\to 0} \frac{\sin(h)}{h} = 1 \right) \\
&= \sin(x)\cdot 0 + \cos(x)\cdot 1 && \text{(Special limits.)} \\
&= \cos(x)
\end{aligned}
$$

We have found that when $f(x) = \sin(x)$, $f'(x) = \cos(x)$. This is not entirely surprising. The sine function is periodic – it repeats itself on regular intervals. Therefore its rate of change also repeats itself on the same regular intervals.

The next example illustrates that the derivative of a function may not always exist.

Example 9.3

Find the derivative of the absolute value function:

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

Its graph is shown in Figures 9.3(a).

Solution

We need to evaluate $\lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h}$. As f is piecewise–defined, we need to consider separately the limits when $x < 0$, $x > 0$ and $x = 0$.

1. When $x < 0$:

$$
\begin{aligned}
\frac{d}{dx}(-x) &= \lim_{h\to 0} \frac{-(x+h)-(-x)}{h} \\
&= \lim_{h\to 0} \frac{-h}{h} \\
&= \lim_{h\to 0}(-1) = -1\,.
\end{aligned}
$$

2. When $x > 0$, a similar computation shows that $\frac{d}{dx}(x) = 1$.
3. We need to also find the derivative at $x = 0$. By the definition of the derivative at a point, we

have

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

Since $x = 0$ is the point where our function's definition switches from one piece to the other, we need to consider left- and right-hand limits. Consider the following, where we compute the left- and right-hand limits side by side.

$$\begin{aligned} &\lim_{\substack{h \to 0 \\ <}} \frac{f(0+h) - f(0)}{h} \\ = &\lim_{\substack{h \to 0 \\ <}} \frac{-h - 0}{h} \\ = &\lim_{\substack{h \to 0 \\ <}} -1 = -1 \end{aligned} \qquad\Bigg|\qquad \begin{aligned} &\lim_{\substack{h \to 0 \\ >}} \frac{f(0+h) - f(0)}{h} \\ = &\lim_{\substack{h \to 0 \\ >}} \frac{h - 0}{h} \\ = &\lim_{\substack{h \to 0 \\ >}} 1 = 1 \end{aligned}$$

Clearly, the left- and right-hand limits are not equal. Therefore the limit does not exist at 0, and f is not differentiable at 0.

Summarising, we have

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

At $x = 0$, $f'(x)$ does not exist; there is a so-called jump discontinuity at 0 (Figure 9.3(b)). So $f(x) = |x|$ is differentiable everywhere except at 0.

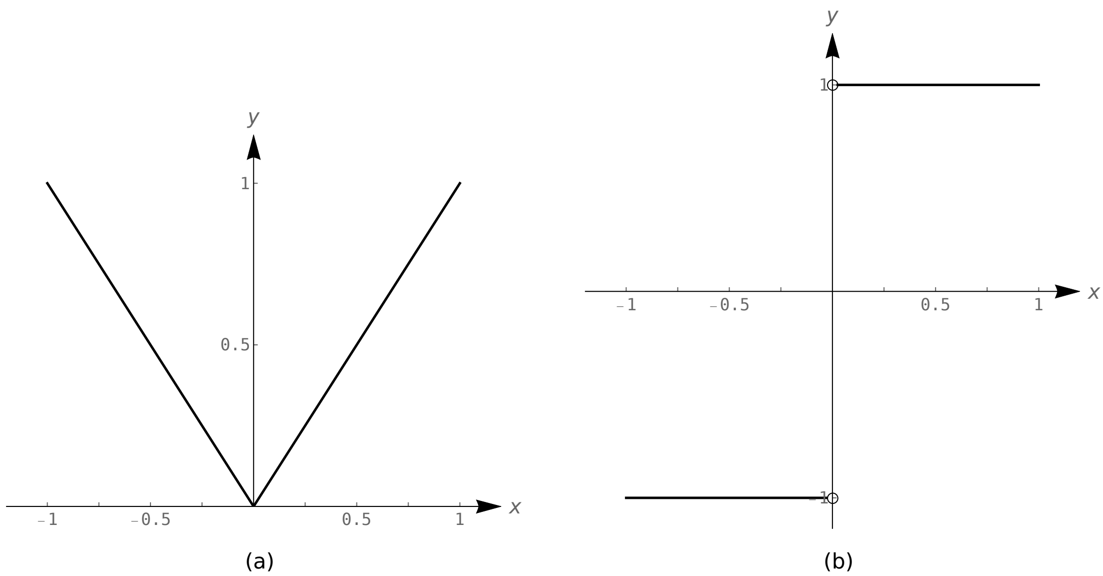


Figure 9.3: The graph of $f(x) = |x|$ (a) and its derivative (b).

The point of non-differentiability came where the piecewise defined function switched from one piece to the other.

Our next example shows that this does, however, not always cause trouble.

Example 9.4

Find the derivative of $f(x)$, given by

$$f(x)=\begin{cases}\sin(x), & x\le\frac{\pi}{2}\\ 1, & x>\frac{\pi}{2}.\end{cases}$$

Its graph is shown in Figure 9.4(a).

Solution

From Example 9.2, we know that when $x<\pi/2$, $f'(x)=\cos(x)$. It is easy to verify that when $x>\pi/2$, $f'(x)=0$; consider:

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

So far, we have

$$f'(x)=\begin{cases}\cos(x) & x<\frac{\pi}{2}\\ 0 & x>\frac{\pi}{2}.\end{cases}$$

We still need to find $f'(\pi/2)$. Notice at $x=\pi/2$ that both pieces of f' are 0, meaning we can state that $f'(\pi/2)=0$.

Being more rigorous, we can again evaluate the difference quotient limit at $x=\pi/2$, utilizing again left- and right–hand limits:

$$\begin{aligned}&\lim_{\substack{h\to 0\\ <}}\frac{f\left(\frac{\pi}{2}+h\right)-f\left(\frac{\pi}{2}\right)}{h}\\ =\ &\lim_{\substack{h\to 0\\ <}}\frac{\sin\left(\frac{\pi}{2}+h\right)-\sin\left(\frac{\pi}{2}\right)}{h}\\ =\ &\lim_{\substack{h\to 0\\ <}}\frac{\sin\left(\frac{\pi}{2}\right)\cos(h)+\sin(h)\cos\left(\frac{\pi}{2}\right)-\sin\left(\frac{\pi}{2}\right)}{h}\\ =\ &\lim_{\substack{h\to 0\\ <}}\frac{1\cdot\cos(h)+\sin(h)\cdot 0-1}{h}\\ =\ &0\,.\end{aligned}$$

$$\begin{aligned}&\lim_{\substack{h\to 0\\ >}}\frac{f\left(\frac{\pi}{2}+h\right)-f\left(\frac{\pi}{2}\right)}{h}\\ =\ &\lim_{\substack{h\to 0\\ >}}\frac{1-1}{h}\\ =\ &\lim_{\substack{h\to 0\\ >}}\frac{0}{h}\\ =\ &0\,.\end{aligned}$$

Since both are 0 at $x=\pi/2$, the limit exists and $f'(\pi/2)$ exists (and is 0). Therefore we can fully write f' as

$$f'(x)=\begin{cases}\cos(x), & x\le\frac{\pi}{2}\\ 0, & x>\frac{\pi}{2}.\end{cases}$$

See Figure 9.4(b) for a graph of this function.

Loosely speaking, we defined a continuous function in Chapter 8 as one in which we could sketch its graph without lifting our pencil. Likewise, it can be understood that a function is differentiable if it is a continuous function that does not have any sharp corners. One such sharp corner is shown in Figure 9.3(a). On the other hand, even though the function f in Example 9.4 is piecewise–defined, the transition is smooth hence it is differentiable.

In general, we are guaranteed that a differentiable function is continuous through the following theo-

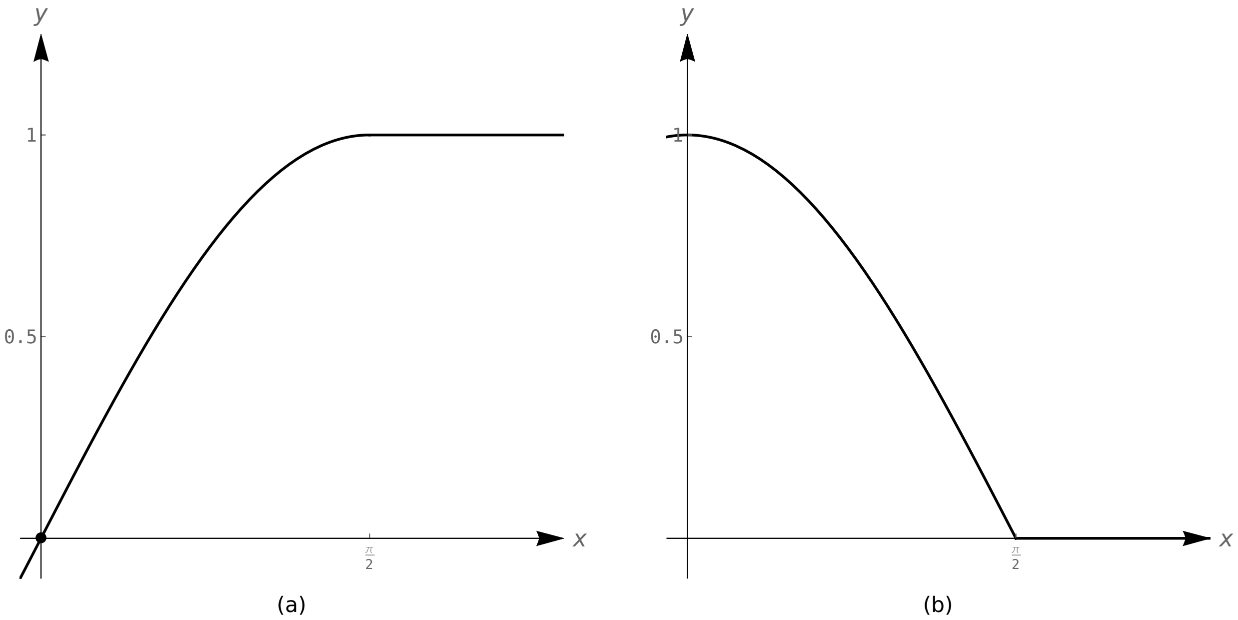


Figure 9.4: The graph of $f(x)$ as defined in Example 9.4 (a) and its derivative (b).

rem. Be aware, however, that continuity does not imply differentiability.

Theorem 9.2 (Differentiable functions are continuous)
Let f be differentiable in $c \in]a,b[$, then is f continuous in c.

Proof From Theorem 9.1, we know that for all $x \in]a,b[$ with $x \neq c$ it holds that:

$$f(x)-f(c) = \frac{f(x)-f(c)}{x-c}(x-c).$$

The limits for $x \longrightarrow c$ of both factors in the right-hand side of this exist, so

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

This allows us to conclude that

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

which on its turn implies that f is continuous in c. □

9.1.3 Differentiability on closed intervals

When we defined the derivative at a point in Definition 9.1, we specified that the interval I over which a function f was defined needed to be an open interval. Open intervals are required so that we can take a limit at any point c in I, meaning we want to approach c from both the left and right.

Recall we also required open intervals in Definition 8.3 when we defined what it meant for a function to be continuous. Later, we used one-sided limits to extend continuity to closed intervals. We now extend differentiability to closed intervals by again considering one-sided limits.

Our motivation for doing this is three-fold. First, we consider common sense. In Example 9.2 we found that when $f(x) = 3x^2+5x-7$, $f'(x) = 6x+5$, and this derivative is defined for all real numbers, hence f is differentiable everywhere. It seems appropriate to also conclude that f is differentiable on closed intervals, like $[0,1]$, as well. After all, $f'(x)$ is defined at both $x = 0$ and $x = 1$. Secondly, consider

$f(x) = \sqrt{x}$. The domain of f is $\mathbb{R}^+$. It is natural to ask ourselves whether f is differentiable on its domain – specifically, is f differentiable at 0? Thirdly, having the derivative defined on closed intervals will prove useful throughout the remainder of this chapter.

Below we give a formal definition of differentiability on a closed interval.

Definitie 9.5 (Differentiability on a closed interval)
Let f be continuous on $[a,b]$ and differentiable on $]a,b[$, and let the one-sided limits

$$\lim_{\substack{h\to 0\\ >}} \frac{f(a+h)-f(a)}{h} \quad \text{and} \quad \lim_{\substack{h\to 0\\ <}} \frac{f(b+h)-f(b)}{h}$$

exist. Then we say f is **differentiable** on $[a,b]$.

Given the notation of Definition 9.5 and in line with the terminology introduced for one-sided limits, we say that the function f is **right differentiable** (*rechts afleidbaar*) in a and **left differentiable** (*links afleidbaar*) in b. Moreover, the one-sided limit

$$\lim_{\substack{h\to 0\\ >}} \frac{f(a+h)-f(a)}{h}$$

is called the **right-derivative** (*rechter afgeleide*) of f in a, while the one-sided limit

$$\lim_{\substack{h\to 0\\ <}} \frac{f(b+h)-f(b)}{h}$$

is referred to as the **left derivative** (*linker afgeleide*) of f in b. Using this terminology, we may say that a function f is differentiable in a if and only if it is left and right differentiable in a and if its left and right derivatives in a are equal.

Example 9.5

Consider the functions $f(x) = \sqrt{x}$ and $g(x) = \sqrt{x^3}$. The domain of these functions is $\mathbb{R}^+$ and it is easy to see that they are differentiable on $\mathbb{R}_0^+$. Determine the differentiability of each at $x = 0$.

Solution

We start by considering f and take the right-hand limit of the difference quotient:

$$\begin{aligned}
\lim_{\substack{h\to 0\\ >}} \frac{f(a+h)-f(a)}{h} &= \lim_{\substack{h\to 0\\ >}} \frac{\sqrt{0+h}-\sqrt{0}}{h} \\
&= \lim_{\substack{h\to 0\\ >}} \frac{\sqrt{h}}{h} \\
&= \lim_{\substack{h\to 0\\ >}} \frac{1}{\sqrt{h}} = +\infty.
\end{aligned}$$

The one-sided limit of the difference quotient does not exist at $x = 0$ for f; therefore f is differentiable on $\mathbb{R}_0^+$ and not differentiable on $\mathbb{R}^+$.

Now consider g:

$$\lim_{\substack{h\to 0\\ >}} \frac{g(a+h)-g(a)}{h} = \lim_{\substack{h\to 0\\ >}} \frac{\sqrt{(0+h)^3}-\sqrt{0}}{h}$$

$$= \lim_{h \underset{>}{\to} 0} \frac{h^{3/2}}{h}$$
$$= \lim_{h \underset{>}{\to} 0} h^{1/2} = 0.$$

As the one-sided limit exists at $x = 0$, we conclude g is differentiable on its domain of $\mathbb{R}^+$.

The two functions are graphed in Figure 9.5. Note how $f(x) = \sqrt{x}$ seems to go vertical as x approaches 0, implying the slopes of its tangent lines are growing toward infinity. Also note how the slopes of the tangent lines to $g(x) = \sqrt{x^3}$ approach 0 as x approaches 0.

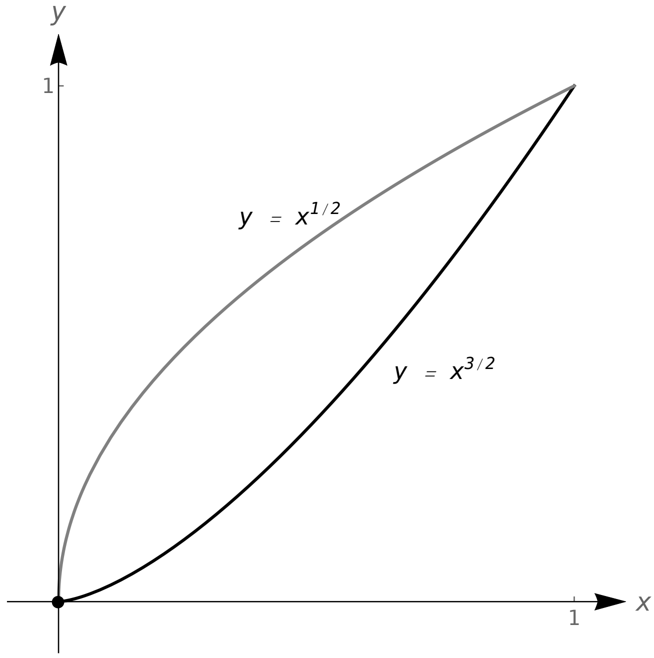


Figure 9.5: A graph of $y = x^{1/2}$ and $y = x^{3/2}$.

9.1.4 Interpretations of the derivative

We offer two interconnected interpretations of the derivative, hopefully explaining why we care about it and why it is worthy of study.

9.1.4.1 Instantaneous rate of change

If f is a function of x, then $f'(x)$ measures the instantaneous rate of change of f with respect to x. It is useful to recognize the units of the derivative function. If y is a function of x, i.e., $y = f(x)$ for some function f, and y is measured in metres and x in seconds, then the units of $y' = f'$ are metres per second. In general, if y is measured in units P and x is measured in units Q, then y' will be measured in units P per Q.

Referring back to the falling amusement–park ride, knowing that at $t = 2$ the velocity was -64 m/s, we could reasonably assume that 1 second later the riders' height would have dropped by about 64 metres. Knowing that the riders were accelerating as they fell would inform us that this is an under–approximation. If all we knew was that $f(2) = 86$ and $f'(2) = -64$, we'd know that we'd have to stop the riders quickly otherwise they would hit the ground.

Example 9.6

Let $P(t)$ represent the world population t minutes after 12:00 a.m., January 1, 2012. It is fairly accurate to say that $P(0) = 7028734178$ (www.prb.org). It is also fairly accurate to state that $P'(0) = 156$; that is, at midnight on January 1, 2012, the population of the world was growing by about 156 people per minute. Twenty days later (or, 28,800 minutes later) we could reasonably assume the population grew by about $28800 \cdot 156 = 4492800$ people.

In this example we made use of the important approximation the rate of change was constant. Notationally, we would say that

$$f(c+h) \approx f(c) + f'(c) \cdot h.$$

This approximation is best when h is small. Small is a relative term; when dealing with the world population, $h = 22$ days = 28,800 minutes is small in comparison to years.

One of the most fundamental applications of the derivative is the study of motion. Let $s(t)$ be a position function, where t is time and $s(t)$ is distance. For instance, s could measure the height of a projectile or the distance an object has travelled. Then $s'(t)$ has units metres per second, it measures the instantaneous rate of distance change – it measures **velocity** (*snelheid*).

Now consider $v(t)$, a velocity function. That is, at time t, $v(t)$ gives the velocity of an object. The derivative of v, $v'(t)$, gives the instantaneous rate of velocity change – **acceleration** (*versnelling*). If velocity is measured in metres per second, and time is measured in seconds, then the units of acceleration are metres per second per second, or (m/s)/s. We often shorten this to metres per second squared, but this tends to obscure the meaning of the units.

9.1.4.2 The slope of the tangent line

Given a function $y = f(x)$, the difference quotient

$$\frac{f(c+h) - f(c)}{h}$$

gives a change in y-values divided by a change in x-values; i.e., it is a measure of the slope of the line that goes through two points on the graph of f: $\big(c, f(c)\big)$ and $\big(c+h, f(c+h)\big)$. As h shrinks to 0, these two points come close together; in the limit we find $f'(c)$, the slope of a special line called the tangent line that intersects f only once near $x = c$. Lines have a constant rate of change, their slope. Nonlinear functions do not have a constant rate of change, but we can measure their instantaneous rate of change at a given x value c by computing $f'(c)$. We can get an idea of how f is behaving by looking at the slopes of its tangent lines.

If we know $f(c)$ and $f'(c)$ for some value $x = c$, then computing the tangent line at $(c, f(c))$ is easy:

$$y = \ell(x) = f'(c)(x-c) + f(c).$$

It can then be used to approximate a value of f. More specifically, Let us use the tangent line at $x = c$ to approximate a value of f near $x = c$; i.e., compute $\ell(c+h)$ to approximate $f(c+h)$, assuming again that h is small. We get

$$y = \ell(c+h) = f'(c)\big((c+h) - c\big) + f(c) = f'(c)h + f(c).$$

9.1.5 Higher-order derivatives

The derivative of a function f is itself a function, therefore we can take its derivative. The following definition gives a name to this concept and introduces its notation.

Definitie 9.6 (Higher-order derivatives)
Let $y = f(x)$ be a differentiable function on I. The following are defined, provided the corresponding limits exist.

1. The **second derivative** (*tweede afgeleide*) of f is:

$$f''(x) = \frac{d}{dx}\Big(f'(x)\Big) = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} = y''.$$

2. The **third derivative** (*derde afgeleide*) of f is:

$$f'''(x) = \frac{d}{dx}\Big(f''(x)\Big) = \frac{d}{dx}\left(\frac{d^2y}{dx^2}\right) = \frac{d^3y}{dx^3} = y'''.$$

3. The **n$^{\text{th}}$ derivative** (*n-de afgeleide*) of f is:

$$f^{(n)}(x) = \frac{d}{dx}\Big(f^{(n-1)}(x)\Big) = \frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) = \frac{d^ny}{dx^n} = y^{(n)}.$$

In general, when finding the fourth derivative and so on, we resort to the $f^{(4)}(x)$ notation, not $f''''(x)$; because after a while, too many ticks is confusing. Moreover, the second derivative notation could be written as

$$\frac{d^2y}{dx^2} = \frac{d^2y}{(dx)^2} = \frac{d^2}{(dx)^2}(y).$$

That is, we take the derivative of y twice (hence d^2), both times with respect to x (hence $(dx)^2 = dx^2$). Also higher-order derivatives can be computed in Mathematica using the built-in command **`D`**. For instance, the second derivative of $y = x^2$ can be computed as follows.

```
In[14]:=  D[x^2,{x,2}]

Out[14]=  2
```

The second argument of the function **`D`** is now a list containing the focal variable and the order of the derivative.

But what do higher order derivatives mean? What is the practical interpretation? Our first answer is

> The second derivative of a function f is the rate of change of the rate of change of f.

One way to grasp this concept is to let f describe a position function. Then, f' describes the rate of position change: velocity. We now consider f'', which describes the rate of velocity change. Its derivative describes the rate of change of the rate of position change, which we know as acceleration. It can, however, be difficult to consider the meaning of the third, and higher order, derivatives. The third derivative is the rate of change of the rate of change of the rate of change of f. That is essentially meaningless to the uninitiated. In the context of our position/velocity/acceleration example, the third derivative is the rate of change of acceleration, commonly referred to as jerk.

Make no mistake: higher-order derivatives have great importance even if their practical interpretations are hard to understand. The mathematical topic of series makes extensive use of higher-order derivatives (Chapter 14).

9.1.6 Smoothness

Some of the graphs we encountered so far have sharp corners, or **cusps**, where the corresponding functions are not differentiable. This leads us to a definition.

Definitie 9.7 (Smoothness)
A **smooth function** (*gladde functie*) is a function that has continuous derivatives up to some desired order over some domain. A function can therefore be said to be smooth over a restricted interval I. Moreover, a function f is **piecewise smooth** (*stuksgewijs gladde functie*) on I if I can be partitioned into subintervals where f is smooth on each subinterval.

The number of continuous derivatives necessary for a function to be considered smooth depends on the problem at hand, and may vary from two to infinity.

The function f is said to be of (differentiability) **class** C^k (*differentieerbaarheidsklasse*) if the derivatives $f', f'', \ldots, f^{(k)}$ exist and are continuous. Note that the continuity is implied by differentiability for all the derivatives except for $f^{(k)}$. The function f is said to be of class C^∞, or **smooth** (*glad*), if it has derivatives of all orders. Such a function is also called a C^∞**-function**. To put it differently, the class C^0 consists of all continuous functions. The class C^1 consists of all differentiable functions whose derivative is continuous; such functions are called continuously differentiable. Thus, a C^1-function is exactly a function whose derivative exists and is of class C^0.

For instance, the function

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

is continuous, but not differentiable at $x = 0$, so it is of class C^0 but not of class C^1.

The function

$$g(x) = \begin{cases} x^2 \sin\left(\dfrac{1}{x}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0 \end{cases}$$

is differentiable, with derivative

$$g'(x) = \begin{cases} -\cos\left(\dfrac{1}{x}\right) + 2x \sin\left(\dfrac{1}{x}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Still, because $\cos\left(x^{-1}\right)$ oscillates as $x \to 0$, $g'(x)$ is not continuous at zero. Therefore, $g(x)$ is differentiable but not of class C^1.

Example 9.7

Determine whether or not the following functions are smooth.

1. $y = x^3$
2. $y = x\,|x|$

Solution

1. We observe that the first derivative is given by $y' = 3x^2$, the second derivative by $y'' = 6x$, the third derivative by 6, while the fourth and higher-derivatives are all 0. Clearly, all these derivatives are continuous, so $y = x^3$ is a smooth function.

2. For the sake of understanding, let us rewrite the equation as

$$y = \begin{cases} x^2, & \text{if } x \geq 0, \\ -x^2, & \text{if } x < 0\,. \end{cases}$$

Its first derivative is given by

$$y' = \begin{cases} 2x, & \text{if } x \geq 0, \\ -2x, & \text{if } x < 0\,, \end{cases}$$

and is continuous everywhere. Yet, its second derivative,

$$y'' = \begin{cases} 2, & \text{if } x \geq 0, \\ -2, & \text{if } x < 0 \end{cases}$$

is not continuous everywhere, as there is a discontinuity at $x = 0$. Consequently, this function is not smooth. More specifically, it is a C^1-function because this functions derivatives exist and are continuous up to order 1. Hence, the first derivative of this function is a C^0-function.

9.2 Basic differentiation rules

The derivative is a powerful tool but is admittedly awkward given its reliance on limits. Fortunately, one thing mathematicians are good at is abstraction.

9.2.1 Derivatives of algebraic and transcendental functions

Let us consider a linear function, $y = mx + b$. What is y'? Without limits, recognize that linear function are characterized by being functions with a constant rate of change (the slope). The derivative, y', gives the instantaneous rate of change; with a linear function, this is constant, m. Thus $y' = m$.

Let us abstract once more. Let us find the derivative of the general quadratic function,

$$f(x) = ax^2 + bx + c\,.$$

Using the definition of the derivative, we have:

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{a(x+h)^2 + b(x+h) + c - (ax^2 + bx + c)}{h} \\ &= \lim_{h \to 0} \frac{ah^2 + 2ahx + bh}{h} \\ &= \lim_{h \to 0} (ah + 2ax + b) \\ &= 2ax + b. \end{aligned}$$

So if $y = 6x^2 + 11x - 13$, we can immediately compute $y' = 12x + 11$.

In a similar way, using Definition 9.4 we can easily find the derivatives of the algebraic and transcendental functions we studied in Chapter 4 and 5, respectively. We find, for the constant function $f(x) = c$

$$\frac{d}{dx}(c) = 0\,,$$

where $c \in \mathbb{R}$. This indicates the logical fact that constant functions have no rate of change. For the other algebraic functions we encountered we have

- $\dfrac{d}{dx}(x^n) = nx^{n-1}$, where $n \in \mathbb{Z}$,
- $\dfrac{d}{dx}(x^a) = ax^{a-1}$, where $a \in \mathbb{R}_0$ and $x > 0$.

Hence, we immediately get

$$\frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}},$$

where $x > 0$.

For what concerns the exponential and logarithmic functions, we get the following derivative functions:

- $\dfrac{d}{dx}(e^x) = e^x$,
- $\dfrac{d}{dx}(a^x) = a^x \ln(a),$ where $a > 0$,
- $\dfrac{d}{dx}\big(\ln(x)\big) = \dfrac{1}{x}$,
- $\dfrac{d}{dx}\big(\log_a(x)\big) = \dfrac{1}{x\ln(a)},$ where $a > 0, a \neq 1$,

while for the trigonometric and hyperbolic functions we get:

- $\dfrac{d}{dx}\big(\sin(x)\big) = \cos(x)$,
- $\dfrac{d}{dx}\big(\cos(x)\big) = -\sin(x)$,
- $\dfrac{d}{dx}\big(\tan(x)\big) = \dfrac{1}{\cos^2(x)} = \sec^2(x)$,
- $\dfrac{d}{dx}\big(\cot(x)\big) = \dfrac{-1}{\sin^2(x)} = -\csc^2(x)$,
- $\dfrac{d}{dx}\big(\sinh(x)\big) = \cosh(x)$,
- $\dfrac{d}{dx}\big(\cosh(x)\big) = \sinh(x)$,
- $\dfrac{d}{dx}\big(\tanh(x)\big) = \dfrac{1}{\cosh^2(x)}$,
- $\dfrac{d}{dx}\big(\coth(x)\big) = \dfrac{-1}{\sinh^2(x)}$.

9.2.2 Properties of the derivative

Using the derivatives of the basic algebraic and transcendental functions, we can easily find the derivative of $y = x^3$, but we cannot compute the derivative of $y = 2x^3$, $y = x^3 + \sin(x)$ nor $y = x^3 \sin(x)$. The following theorem helps with the first two of these examples.

Theorem 9.3 (Properties of the derivative)
Let f and g be differentiable on an open interval I and let c be a real number. Then the following properties hold:

1. ***Sum/Difference rule**:*

$$\frac{d}{dx}\Big(f(x) \pm g(x)\Big) = \frac{d}{dx}\Big(f(x)\Big) \pm \frac{d}{dx}\Big(g(x)\Big) = f'(x) \pm g'(x). \tag{9.1}$$

2. ***Constant multiple rule***:

$$\frac{d}{dx}\Big(c\cdot f(x)\Big)=c\cdot\frac{d}{dx}\Big(f(x)\Big)=c\cdot f'(x). \tag{9.2}$$

Proof The sum rule, for instance, can be proved as follows. Let f and g be two functions that are differentiable in a, and for which $\operatorname{dom} f\cap\operatorname{dom} g\neq\emptyset$. Then, for all $h\neq 0$ and $a+h\in\operatorname{dom} f+g$:

$$\begin{aligned}(f(a)+g(a))' &= \lim_{h\to 0}\frac{(f(a+h)+g(a+h))-(f(a)+g(a))}{h}\\ &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}+\lim_{h\to 0}\frac{g(a+h)-g(a)}{h}\\ &= f'(a)+g'(a).\end{aligned}$$

Consequently, for all $x\in\operatorname{dom} f+g$ Equation (9.1) holds. □

Intuitively, it is clear that the sum rule can be extended to n functions $f_i(x)$; that is:

$$\frac{d}{dx}\left(\sum_{i=1}^{n}f_i(x)\right)=\left(\sum_{i=1}^{n}f_i(x)\right)'=\sum_{i=1}^{n}f_i'(x)\,.$$

Theorem 9.3 allows us to find the derivatives of a wide variety of functions. It can be used in conjunction with the power rule to find the derivatives of any polynomial. Recall in Example 9.2 that we found, using the limit definition, the derivative of $f(x)=3x^2+5x-7$. We can now find its derivative without expressly using limits:

$$\begin{aligned}\frac{d}{dx}\Big(3x^2+5x+7\Big)&=3\frac{d}{dx}\Big(x^2\Big)+5\frac{d}{dx}\Big(x\Big)+\frac{d}{dx}\Big(7\Big)\\ &=3\cdot 2x+5\cdot 1+0\\ &=6x+5.\end{aligned}$$

When having to compute the derivative a product of functions, we may turn to the product rule.

Theorem 9.4 (Product rule)
Let f and g be differentiable functions on an open interval I. Then fg is a differentiable function on I, and

$$\frac{d}{dx}\Big(f(x)g(x)\Big)=f(x)g'(x)+f'(x)g(x).$$

Proof To prove Theorem 9.4, we use the definition of the derivative. More specifically, by the limit definition, we have

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

We now do something a bit unexpected; add 0 to the numerator (so that nothing is changed) in the form of $-f(x+h)g(x)+f(x+h)g(x)$, then do some regrouping as shown.

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

$$
\begin{aligned}
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{\Big(f(x+h)g(x+h) - f(x+h)g(x)\Big) + \Big(f(x+h)g(x) - f(x)g(x)\Big)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} + \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} f(x+h)\frac{g(x+h) - g(x)}{h} + \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} g(x) \\
&= f(x)g'(x) + f'(x)g(x).
\end{aligned}
$$

□

Just as with the sum rule, the product rule can be generalized to n functions $f_i(x)$:

$$
\begin{aligned}
\frac{d}{dx}\left[\prod_{i=1}^{n} f_i(x)\right] &= \sum_{i=1}^{n}\left(\frac{d}{dx}(f_i(x))\prod_{j\neq i} f_j(x)\right) \\
&= \left(\prod_{i=1}^{n} f_i(x)\right)\left(\sum_{i=1}^{n}\frac{f_i'(x)}{f_i(x)}\right) \\
&= f_1'(x)f_2(x)f_3(x)\cdots f_n(x) + f_1(x)f_2'(x)f_3(x)\cdots f_n(x) + \cdots + f_1(x)f_2(x)f_3(x)\cdots f_n'(x)\,.
\end{aligned}
$$

Example 9.8

Find the derivatives of the following:

1. $y = 5x^2 \sin(x)$
2. $y = x^3 \ln(x)\cos(x)$
3. $y = \cosh(x)e^x$

Solution

1. To make our use of the product rule explicit, let us set $f(x) = 5x^2$ and $g(x) = \sin(x)$. We easily compute/recall that $f'(x) = 10x$ and $g'(x) = \cos(x)$. Employing the product rule, we have
$$\frac{d}{dx}\Big(5x^2\sin(x)\Big) = 5x^2\cos(x) + 10x\sin(x).$$
2. We have a product of three functions, so, we get
$$y' = 3x^2\ln(x)\cos(x) + x^3\frac{1}{x}\cos(x) + x^3\ln(x)(-\sin(x))\,.$$
3. Here we again have a product of two functions, so we immediately get
$$y' = \sinh(x)e^x + \cosh(x)e^x\,.$$

We have learned how to compute the derivatives of sums, differences, and products of functions. We now learn how to find the derivative of a quotient of functions.

Theorem 9.5 (Quotient rule)
Let f and g be differentiable functions defined on an open interval I, where $g(x) \neq 0$ on I. Then

f/g is differentiable on I, and

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right)=\frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}.$$

Proof The proof of the quotient rule follows easily using the properties introduced earlier. □

Example 9.9

Find the derivatives of the following:

1. $y=\dfrac{5x^2}{\sin(x)}$
2. $y=\tan(x)$

Solution

1. Directly applying the quotient rule gives:

$$y'=\frac{10x\sin(x)-5x^2\cos(x)}{\sin^2(x)}.$$

2. Though we could resort to the list of derivatives of elementary functions, we can proceed as well by recalling that $\tan(x)=\sin(x)/\cos(x)$, so we can apply the quotient rule.

$$\begin{aligned}\frac{d}{dx}\Big(\tan(x)\Big)&=\frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right)\\&=\frac{\cos(x)\cos(x)-\sin(x)(-\sin(x))}{\cos^2(x)}\\&=\frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)}\\&=\frac{1}{\cos^2(x)}\\&=\sec^2(x).\end{aligned}$$

The derivatives of the cotangent, cosecant and secant functions can all be computed using the known derivatives of the cosine and sine function together with the quotient rule, so there is no need to learn those by heart.

Taking the derivative of many functions is relatively straightforward. It is clear what rules apply and in what order they should be applied. Other functions present multiple paths; different rules may be applied depending on how the function is treated. One of the beautiful things about calculus is that there is not the right way; each path, when applied correctly, leads to the same result, the derivative.

9.3 The chain rule

We have covered almost all of the derivative rules that deal with combinations of two (or more) functions. The operations of addition, subtraction, multiplication (including by a constant) and division led to the sum, difference, constant multiple, power rule, product and quotient rules. To complete the list of differentiation rules, we look at the last way two (or more) functions can be combined: the process of composition.

One example of a composition of functions is $f(x) = \cos(x^2)$. We currently do not know how to compute this derivative. If forced to guess, one would likely guess $f'(x) = -\sin(2x)$, where we recognize $-\sin(x)$ as the derivative of $\cos(x)$ and $2x$ as the derivative of x^2. However, this is not the case; $f'(x) \neq -\sin(2x)$.

Before we define this new rule, recall the notation for composition of functions. We write $(f\circ g)(x)$ or $f(g(x))$, read as f of g of x, to denote composing f with g. In shorthand, we simply write $f\circ g$ or $f(g)$ and read it as f of g. When composing functions, we need to make sure that the new function is actually defined. For instance, consider $f(x) = \sqrt{x}$ and $g(x) = -x^2 - 1$. The domain of f excludes all negative numbers, but the range of g is only negative numbers. Therefore the composition $f\big(g(x)\big) = \sqrt{-x^2-1}$ is not defined for any x, and hence is not differentiable.

The following theorem of the **chain rule** (*kettingregel*) takes care to ensure this problem does not arise. We'll focus more on the derivative result than on the domain/range conditions.

Theorem 9.6 (The chain rule)
Let g be a differentiable function on an interval I, let the range of g be a subset of the interval J, and let f be a differentiable function on J. Then $y = f(g(x))$ is a differentiable function on I, and

$$y' = f'(g(x))\,g'(x).$$

Proof This theorem can be proved by resorting to Theorem 9.1 that provides us with an alternative viewpoint on differentiability.

First of all, putting $z_0 = g(c)$, let us define an auxiliary function q as follows:

$$q : \operatorname{dom} f \to \mathbb{R} : z \mapsto q(z) = \begin{cases} \dfrac{f(z)-f(z_0)}{z-z_0} & , \quad \text{if } z \neq z_0\,, \\ f'(z_0) & , \quad \text{if } z = z_0\,. \end{cases}$$

The function q is continuous at z_0 since it holds that

$$\lim_{z\to z_0} q(z) = f'(z_0) = q(z_0)\,.$$

Now, we may consider the composition $q\circ g$ for which $\operatorname{dom}(q\circ g) = \operatorname{dom}(f\circ g)$ because $\operatorname{dom} q = \operatorname{dom} f$. Given the definition of the function q, we have for this composition that

$$(q\circ g)(x) = \begin{cases} \dfrac{f(g(x))-f(g(c))}{g(x)-g(c)} & , \quad \text{if } x \neq c\,, \\ f'\big(g(c)\big) & , \quad \text{if } x = c\,. \end{cases}$$

Finally, for all $x \neq c$ we may write

$$\begin{aligned} \frac{f(g(x))-f(g(c))}{x-c} &= \frac{f(g(x))-f(g(c))}{g(x)-g(c)}\,\frac{g(x)-g(c)}{x-c} \\ &= (q\circ g)(x)\frac{g(x)-g(c)}{x-c}\,. \end{aligned}$$

As we are looking for an expression for $(f \circ g)'(c)$, we consider the limit of both sides of the last expression for $x \to c$, which yields:

$$(f \circ g)'(c) = f'(g(c))\, g'(c)\,.$$

Acknowledging the chain rule, we can immediately state the so-called generalized power rule.

Theorem 9.7 (Generalized power rule)
Let $g(x)$ be a differentiable function and let $n \neq 0$ be an integer. Then

$$\frac{d}{dx}\Big(g(x)^n\Big) = n\big(g(x)\big)^{n-1} g'(x).$$

It is instructive to understand what the chain rule looks like using $\frac{dy}{dx}$ notation instead of y' notation. Suppose that $y = f(u)$ is a function of u, where $u = g(x)$ is a function of x, as stated in Theorem 9.6. Then, through the composition $f \circ g$, we can think of y as a function of x, as $y = f(g(x))$. Thus the derivative of y with respect to x makes sense; we can talk about $\frac{dy}{dx}$. This leads to an interesting progression of notation:

$$\begin{aligned} y' &= f'(g(x))g'(x) && \\ \frac{dy}{dx} &= y'(u)u'(x) && \text{(Since } y = f(u) \text{ and } u = g(x).) \\ \frac{dy}{dx} &= \frac{dy}{du}\frac{du}{dx} && \text{(Using fractional notation for the derivative.)} \end{aligned}$$

It might seem as though the du terms cancel out, but it is important to realize that we are not cancelling these terms; the derivative notation of $\frac{dy}{du}$ is one symbol. It is equally important to realize that this notation was chosen precisely because of this behaviour. It makes applying the chain rule easy with multiple variables and/or with multiple functions. For instance, if we consider three functions $y = f(u)$, $u = h(v)$ en $v = g(x)$, then we may consider the function composition $y = f(h(g(x)))$. The derivative of y with respect to x is then given by

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx}$$

or equivalently

$$\frac{dy}{dx} = \frac{df(u)}{du}\frac{dh(v)}{dv}\frac{dg(x)}{dx}\,.$$

We now consider some examples that employ the chain rule.

Example 9.10

Find the derivatives of the following:

1. $y = x^5 \sin(2x^3)$
2. $y = \tan^5(6x^3 - 7x)$
3. $y = \dfrac{x\cos(x^{-2}) - \sin^2(e^{4x})}{\ln(x^2 + 5x^4)}$.
4. $y = \tanh\big(\sinh(x)\big)$

Solution

1. We must use the product and chain rules and proceed step–by–step.

$$y' = x^5\big(6x^2\cos(2x^3)\big) + 5x^4\big(\sin(2x^3)\big) = 6x^7\cos(2x^3) + 5x^4\sin(2x^3)\,.$$

2. Recognize that we have the $g(x) = \tan(6x^3 - 7x)$ function inside the $f(x) = x^5$ function. We begin using the generalized power rule; in this first step, we do not fully compute the derivative. Rather, we are approaching this step–by–step.

$$y' = 5\tan^4(6x^3 - 7x)g'(x).$$

We now find $g'(x)$. We again need the chain rule;

$$g'(x) = \sec^2(6x^3 - 7x)(18x^2 - 7).$$

Combine this with what we found above to give

$$\begin{aligned} y' &= 5\tan^4(6x^3 - 7x)\sec^2(6x^3 - 7x)(18x^2 - 7) \\ &= (90x^2 - 35)\sec^2(6x^3 - 7x)\tan^4(6x^3 - 7x). \end{aligned}$$

3. Using the quotient, product and chain rules we get the following answer without simplification:

$$y' = \frac{\ln(x^2+5x^4)\cdot\Big[\big(x(-\sin(x^{-2}))(-2x^{-3}) + 1\cos(x^{-2})\big) - 2\sin(e^{4x})\cos(e^{4x})(4e^{4x})\Big]}{\big(\ln(x^2+5x^4)\big)^2} - \frac{\Big(x\cos(x^{-2}) - \sin^2(e^{4x})\Big)\cdot\frac{2x+20x^3}{x^2+5x^4}}{\big(\ln(x^2+5x^4)\big)^2}$$

4. Direct application of the chain rule with $g(x) = \sinh(x)$ inside the function $f(x) = \tanh(x)$ yields

$$y = \frac{\cosh(x)}{\cosh^2\big(\sinh(x)\big)}\,.$$

This example demonstrates that derivatives can be computed systematically, no matter how arbitrarily complicated the function is. A key to correctly working the considered problems is to break the problem down into smaller, more manageable pieces. For instance, when using the product and chain rules together, consider the first part of the product rule at first: $f(x)g'(x)$. Just rewrite $f(x)$, then find $g'(x)$. Then move on to the $f'(x)g(x)$ part. Do not attempt to figure out both parts at once. Likewise, using the quotient rule, approach the numerator in two steps and handle the denominator after completing that. Only simplify afterwards.

The chain rule also has theoretic value. That is, it can be used to find the derivatives of certain functions.

Example 9.11

Use the chain rule to find the derivative of $y = 2^x$.

Solution

We only know how to find the derivative of one exponential function, $y = e^x$. We can accomplish our goal by rewriting 2 in terms of e. Recalling that e^x and $\ln x$ are inverse functions, so we can write

$$2 = e^{\ln(2)} \quad \text{and so} \quad y = 2^x = \left(e^{\ln(2)}\right)^x = e^{x(\ln(2))}.$$

The function is now the composition $y = f(g(x))$, with $f(u) = e^u$ and $g(x) = x(\ln(2))$. Since $f'(u) = e^u$ and $g'(x) = \ln(2)$, the chain rule gives

$$y' = e^{x(\ln(2))}\ln(2).$$

Recall that the $e^{x(\ln(2))}$ term on the right hand side is just 2^x, our original function. Thus, the derivative contains the original function itself. We have

$$y' = y\ln(2) = 2^x\ln(2).$$

We can extend this process to use any base a, where $a > 0$ and $a \neq 1$. All we need to do is replace each "2" in our work with "a." In this way, the chain rule, coupled with the derivative rule of e^x, allows us to find the derivatives of all exponential functions.

The comment at the end of previous example is important. Let $f(x) = a^x$, for $a > 0, a \neq 1$. Then f is differentiable everywhere and

$$f'(x) = \ln(a)a^x.$$

Likewise, it can be shown that

$$\frac{d}{dx}\big(\log_a(x)\big) = \frac{1}{\ln(a)}\frac{1}{x}.$$

In the next section, we use the chain rule to justify another differentiation technique. There are many curves that we can draw in the plane that fail the vertical line test. See for instance Section 4.4, where we studied amongst other things the equation $x^2 + y^2 = 1$, which describes the unit circle. We may still be interested in finding slopes of tangent lines to the circle at various points. The next section shows how we can find $\frac{dy}{dx}$ without first solving for y. While we can in this instance, in many other instances solving for y is impossible. In these situations, implicit differentiation is indispensable.

9.4 Implicit differentiation

9.4.1 First derivative

In the previous sections we learned to find the derivative when y is given explicitly as a function of x. That is, if we know $y = f(x)$ for some function f, we can find y'. Sometimes the relationship between y and x is not explicit; rather, it is implicit. For instance, we might know that $x^2 - y = 4$. Can we still find y'? In this case, sure; we solve for y to get $y = x^2 - 4$ and then differentiate to get $y' = 2x$. Sometimes, however, the implicit relationship between x and y is complicated. Suppose we are given $\sin(y) + y^3 = 6 - x^3$. In this case there is absolutely no way to solve for y in terms of elementary functions. The surprising thing is, however, that we can still find y' via implicit differentiation.

Implicit differentiation is a technique based on the chain rule that is used to find a derivative when the relationship between the variables is given implicitly rather than explicitly.

Let f and g be functions of x. Then we have according to the chain rule

$$\frac{d}{dx}\Big(f(g(x))\Big) = f'(g(x))\,g'(x).$$

Suppose now that $y = g(x)$. We can rewrite the above as

$$\frac{d}{dx}\Big(f(y)\Big) = f'(y)y', \quad \text{or} \quad \frac{d}{dx}\Big(f(y)\Big) = f'(y)\frac{dy}{dx}. \tag{9.3}$$

These equations look strange; the key concept to learn here is that we can find y' even if we do not exactly know how y and x relate.

Example 9.12

Given that

$$\sin(y) + y^3 = 6 - x^3\,,$$

find y' and the equation of the tangent line at the point $(\sqrt[3]{6}, 0)$.

Solution

We start by taking the derivative of both sides, which maintains the equality. We have :

$$\frac{d}{dx}\Big(\sin(y) + y^3\Big) = \frac{d}{dx}\Big(6 - x^3\Big).$$

The right-hand side is easy; it returns $-3x^2$.

The left-hand side requires more consideration. We take the derivative term–by–term. Using Equation (9.3), we can see that

$$\frac{d}{dx}\Big(\sin(y)\Big) = \cos(y)\,y'.$$

We apply the same process to the y^3 term.

$$\frac{d}{dx}\Big(y^3\Big) = \frac{d}{dx}\Big((y)^3\Big) = 3(y)^2\,y'.$$

Similarly, the derivative of y^3 is $3y^2y'$. Putting all this together with the right hand side, we have

$$\cos(y)y' + 3y^2y' = -3x^2.$$

Now solve for y'.

$$\begin{aligned} & \cos(y)y' + 3y^2y' = -3x^2 \\ \Leftrightarrow\quad & \big(\cos(y) + 3y^2\big)y' = -3x^2 \\ \Leftrightarrow\quad & y' = \frac{-3x^2}{\cos(y) + 3y^2} \end{aligned}$$

We can now find the slope of the tangent line at the point $(\sqrt[3]{6}, 0)$ by substituting $\sqrt[3]{6}$ for x and 0 for y. Thus at the point $(\sqrt[3]{6}, 0)$, we have the slope as

$$y' = \frac{-3(\sqrt[3]{6})^2}{\cos(0) + 3\cdot 0^2} = \frac{-3\sqrt[3]{36}}{1} \approx -9.91.$$

Therefore the equation of the tangent line to the implicitly defined function $\sin y + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$ is

$$y = -3\sqrt[3]{36}(x - \sqrt[3]{6}) + 0 \approx -9.91x + 18.$$

The curve and this tangent line are shown in Figure 9.6.

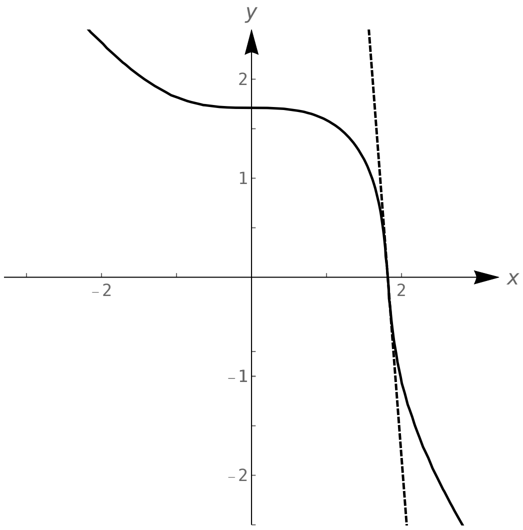


Figure 9.6: The function $\sin y + y^3 = 6 - x^3$ and its tangent line (dashed) at the point $(\sqrt[3]{6}, 0)$.

This example suggests a general method for implicit differentiation. For the steps below assume y is a function of x.

1. Take the derivative of each term in the equation. Treat the x terms like normal. When taking the derivatives of y terms, the usual rules apply except that, because of the chain rule, we need to multiply each term by y'.
2. Get all the y' terms on one side of the equal sign and put the remaining terms on the other side.
3. Factor out y'; solve for y' by dividing.

Example 9.13

Given the implicitly defined function

$$\sin(x^2y^2) + y^3 = x + y,$$

find y'.

Solution

Differentiating term by term, we find the most difficulty in the first term. It requires both the chain and product rules.

$$\begin{aligned}\frac{d}{dx}\Big(\sin(x^2y^2)\Big) &= \cos(x^2y^2)\,\frac{d}{dx}\Big(x^2y^2\Big)\\ &= \cos(x^2y^2)\big(x^2(2yy') + 2xy^2\big)\\ &= 2(x^2yy' + xy^2)\cos(x^2y^2).\end{aligned}$$

We leave the derivatives of the other terms to the reader. After taking the derivatives of both sides, we have

$$2(x^2yy' + xy^2)\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

We now have to be careful to properly solve for y', particularly because of the product on the left. It is best to multiply out the product. Doing this, we get

$$2x^2y\cos(x^2y^2)y' + 2xy^2\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

From here we can safely move around terms to get the following:

$$2x^2y\cos(x^2y^2)y' + 3y^2y' - y' = 1 - 2xy^2\cos(x^2y^2).$$

Then we can solve for y' to get

$$y' = \frac{1 - 2xy^2\cos(x^2y^2)}{2x^2y\cos(x^2y^2) + 3y^2 - 1}.$$

A graph of this implicit function is given in Figure 9.7. It is easy to verify that the points $(0, 1)$ and $(0, -1)$ all lie on the graph. We can find the slopes of the tangent lines at each of these points using our formula for y':

- at $(0, 1)$, the slope is $1/2$;
- at $(0, -1)$, the slope is also $1/2$.

The tangent lines have been added to the graph of the function in Figure 9.7.

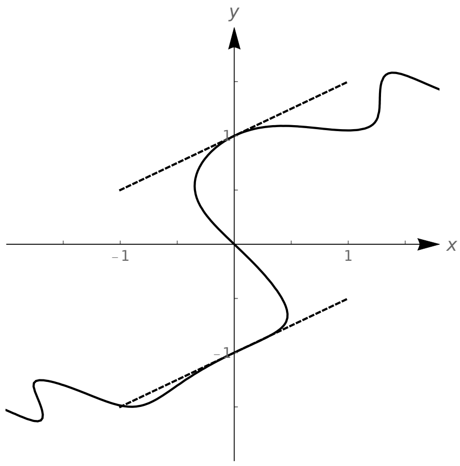


Figure 9.7: A graph of the implicitly defined function $\sin(x^2y^2)+y^3=x+y$ and tangent lines at $(0,1)$ and $(0,-1)$.

We may also use Mathematica to check our answer for what concerns y'. We just have to be careful to explicitly mention the dependence of y on x.

```
In[15]:= D[Sin[x^2 y[x]^2] + y[x]^3 == x + y[x],x]

Out[15]= 3 y[x]^2 y′[x]+Cos[x^2 y[x]^2] (2 x y[x]^2+2 x^2 y[x] y′[x])==1+y′[x]
```

Implicit functions are generally harder to deal with than explicit functions. With an explicit function, given an x value, we have an explicit formula for computing the corresponding y value. With an implicit function, one often has to find x and y values at the same time that satisfy the equation. It is much easier to demonstrate that a given point satisfies the equation than to actually find such a point.

9.4.2 Higher-order derivatives

We can use implicit differentiation to find higher-order derivatives as well. In theory, this is simple: first find $\frac{dy}{dx}$, then take its derivative with respect to x. In practice, it is not hard, but it often requires a bit of algebra. We demonstrate this in an example.

Example 9.14

Given $x^2+y^2=1$, find y''.

Solution

Taking derivatives, we get $2x+2yy'=0$. Solving for y' gives:

$$y' = \frac{-x}{y}.$$

To find y'', we apply implicit differentiation to y'.

$$\begin{aligned} y'' &= \frac{d}{dx}(y') \\ &= \frac{d}{dx}\left(-\frac{x}{y}\right) && \text{(Use the quotient rule.)} \end{aligned}$$

$$= -\frac{y \cdot 1 - x(y')}{y^2}$$

replace y' with $-x/y$:

$$= -\frac{y - x(-x/y)}{y^2}$$
$$= -\frac{y + x^2/y}{y^2}.$$

While this is not a particularly simple expression, it is usable. For instance, we can see that $y'' > 0$ when $y < 0$ and $y'' < 0$ when $y > 0$. In Section 10.4, we will see how this relates to the shape of the graph.

9.4.3 Logarithmic differentiation

Consider the function $y = x^x$; it is graphed in Figure 9.8. It is well defined for $x > 0$ and we might be interested in finding equations of lines tangent and normal to its graph. How do we take its derivative?

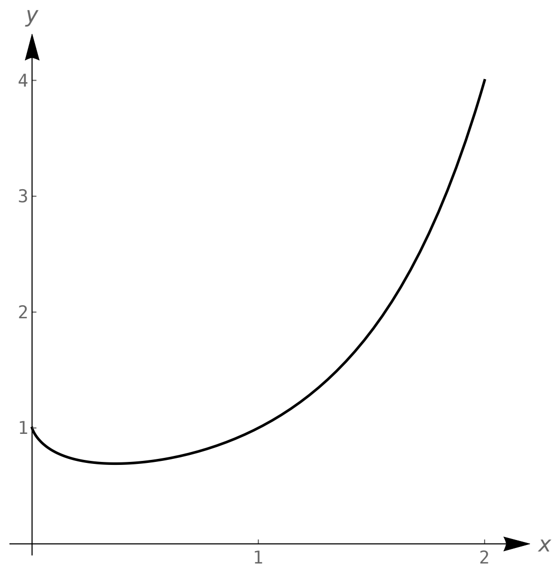


Figure 9.8: A plot of $y = x^x$.

The function is not a power function: it has a power of x, not a constant. It is not an exponential function: it has a base of x, not a constant. A differentiation technique known as **logarithmic differentiation** (*logaritmisch afleiden*) becomes useful here. The basic principle is this: take the natural log of both sides of an equation $y = f(x)$, then use implicit differentiation to find y'. We demonstrate this in the following example.

Example 9.15

Given $y = x^x$, use logarithmic differentiation to find y'.

Solution

We start by taking the natural log of both sides then applying implicit differentiation.

$$\begin{array}{rrcll} & y & = & x^x & \\ \Leftrightarrow & \ln(y) & = & \ln(x^x) & \text{(Apply logarithm rule.)} \\ \Leftrightarrow & \ln(y) & = & x\ln(x) & \text{(Use implicit differentiation.)} \\ \Leftrightarrow & \frac{d}{dx}\Big(\ln(y)\Big) & = & \frac{d}{dx}\Big(x\ln(x)\Big) & \\ \Leftrightarrow & \frac{y'}{y} & = & \ln(x)+x\,\frac{1}{x} & \\ \Leftrightarrow & \frac{y'}{y} & = & \ln(x)+1 & \\ \Leftrightarrow & y' & = & y\big(\ln(x)+1\big) & (y=x^x\,.) \\ \Leftrightarrow & y' & = & x^x\big(\ln(x)+1\big). & \end{array}$$

The attentive reader might wonder at this point, whether logarithmic differentiation still works if $f(x) <=$ because we may not take the logarithm then. It turns out that it does because yo can consider $\ln|f(x)|$, which works for negative and positive values of any function $f(x)$. Where f(x)<0, we have that $\ln|f(x)| = \ln(-f(x))$ and the derivative

$$\frac{d}{dx}\big(\ln(-f(x))\big) = \frac{1}{-f(x)}(-f'(x)) = \frac{f'(x)}{f(x)}\,,$$

which is the same as the derivative of $\ln(f(x))$, but just also works for negative values of $f(x)$.

9.5 Derivatives of inverse functions

Recall that a function $y = f(x)$ is said to be injective if it passes the horizontal line test; that is, for two different x values x_1 and x_2, we do not have $f(x_1) = f(x_2)$. In some cases the domain of f must be restricted so that it is injective. For instance, consider $f(x) = x^2$. Clearly, $f(-1) = f(1)$, so f is not one to one on its regular domain, but by restricting f to $\mathbb{R}_0^+$, f is one to one.

Now recall that injective functions have inverses. That is, if f is one to one, it has an inverse function, denoted by f^{-1}, such that if $f(a) = b$, then $f^{-1}(b) = a$. The domain of f^{-1} is the range of f, and vice-versa. For ease of notation, we set $g = f^{-1}$ and treat g as a function of x.

When the point (a,b) lies on the graph of f, the point (b,a) lies on the graph of g. This made us to discover in Section 3.4 that the graph of g is the reflection of f across the line $y = x$. Because of this relationship, whatever we know about f can quickly be transferred into knowledge about g.

For example, consider Figure 9.9 where the tangent line to f at the point (a,b) is drawn. That line has slope $f'(a)$. Through reflection across $y = x$, we can see that the tangent line to g at the point (b,a) should have slope $\frac{1}{f'(a)}$. This then tells us that $g'(b) = \frac{1}{f'(a)}$.

We have discovered a relationship between f' and g' in a mostly graphical way. We can realize this relationship analytically as well. Let $y = g(x)$, where again $g = f^{-1}$. We want to find y'. Since $y = g(x)$, we know that $f(y) = x$. Using the chain rule and implicit differentiation, take the derivative of both

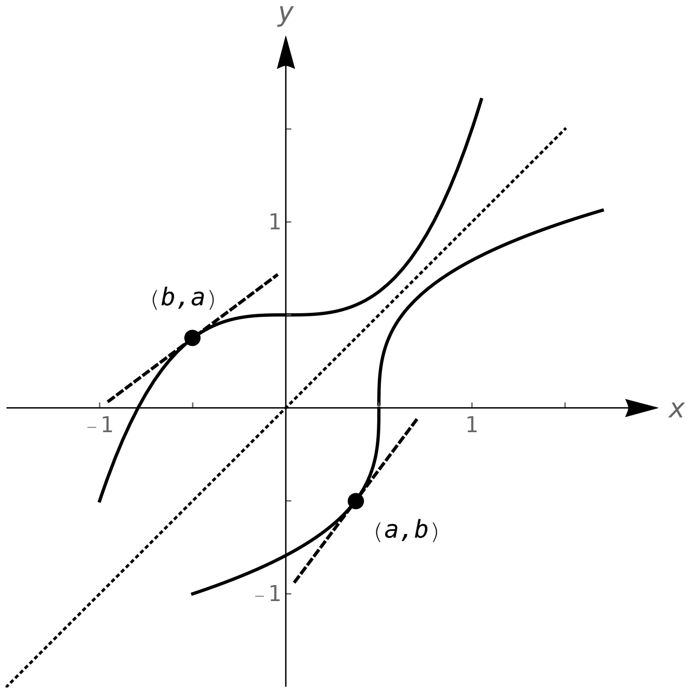


Figure 9.9: Corresponding tangent lines drawn to f and f^{-1}.

sides of this last equality:

$$\begin{aligned} \frac{d}{dx}\Big(f(y)\Big) &= \frac{d}{dx}\Big(x\Big) \\ \Leftrightarrow \quad f'(y)\,y' &= 1 \\ \Leftrightarrow \quad y' &= \frac{1}{f'(y)} \\ \Leftrightarrow \quad y' &= \frac{1}{f'(g(x))}. \end{aligned}$$

This leads us to the following theorem.

Theorem 9.8 (Derivatives of inverse functions)
Let f be differentiable and injective on an open interval I, where $f'(x) \neq 0$ for all x in I, let J be the range of f on I, let g be the inverse function of f, and let $f(a) = b$ for some a in I. Then g is a differentiable function on J, and in particular,

1. $\left(f^{-1}\right)'(b) = g'(b) = \dfrac{1}{f'(a)}$ *and* *2.* $\left(f^{-1}\right)'(x) = g'(x) = \dfrac{1}{f'(g(x))}$

The results of Theorem 9.8 are not trivial; the notation may seem confusing at first. Careful consideration, along with examples, should earn understanding.

In the next example we apply Theorem 9.8 to the arcsine function.

Example 9.16

Let $y = \arcsin(x)$. Find y'.

Solution

Adopting our previously defined notation, let $g(x) = \arcsin(x)$ and $f(x) = \sin(x)$. Consequently,

$f'(x) = \cos(x)$. Applying the theorem, we have

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\cos(\arcsin(x))}.$$

This last expression is not immediately illuminating. Drawing a figure while assuming that $x >$ will help, as shown in Figure 9.10.

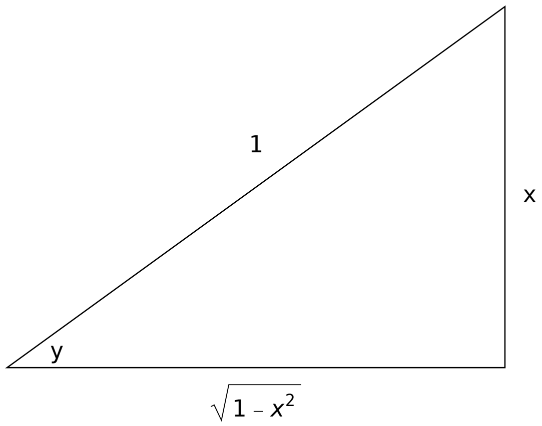


Figure 9.10: A right triangle defined by $y = \arcsin(x/1)$ with the length of the third leg found using the Pythagorean theorem.

Recall that the sine function can be viewed as taking in an angle and returning a ratio of sides of a right triangle, specifically, the ratio opposite over hypotenuse. This means that the arcsine function takes as input a ratio of sides and returns an angle. The equation $y = \arcsin(x)$ can be rewritten as $y = \arcsin(x/1)$; that is, consider a right triangle where the hypotenuse has length 1 and the side opposite of the angle with measure y has length x. This means the final side has length $\sqrt{1-x^2}$, using the Pythagorean theorem.

Therefore

$$\cos(\arcsin(x)) = \cos(y) = \frac{\sqrt{1-x^2}}{1} = \sqrt{1-x^2},$$

resulting in

$$\frac{d}{dx}\big(\arcsin(x)\big) = g'(x) = \frac{1}{\sqrt{1-x^2}}.$$

Remember that the input x of the arcsine function is a ratio of a side of a right triangle to its hypotenuse; the absolute value of this ratio will never be greater than 1. Therefore the inside of the square root will never be negative.

Using similar techniques as in Example 9.16, we can find the derivatives of all the inverse trigonometric and hyperbolic functions.

The inverse trigonometric functions are differentiable on all open sets contained in their domains (as listed in Table 5.5) and their derivatives are as follows:

- $\frac{d}{dx}\big(\arcsin(x)\big) = \frac{1}{\sqrt{1-x^2}}$,
- $\frac{d}{dx}\big(\arccos(x)\big) = -\frac{1}{\sqrt{1-x^2}}$,
- $\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}$.
- $\frac{d}{dx}\big(\operatorname{arccot}(x)\big) = -\frac{1}{1+x^2}$,

The derivatives of the inverse hyperbolic functions are as follows:

- $\frac{d}{dx}\big(\operatorname{arsinh}(x)\big) = \frac{1}{\sqrt{x^2+1}}$,
- $\frac{d}{dx}\big(\operatorname{arcosh}(x)\big) = \frac{1}{\sqrt{x^2-1}}$, for all $x > 1$,
- $\frac{d}{dx}\big(\operatorname{artanh}(x)\big) = \frac{1}{1-x^2}$, for all $|x| < 1$,
- $\frac{d}{dx}\big(\operatorname{arcoth}(x)\big) = \frac{1}{1-x^2}$, for all $|x| > 1$.

In Section 9.2, we stated without proof or explanation that $\frac{d}{dx}\big(\ln(x)\big) = \frac{1}{x}$. We can justify that now using Theorem 9.8, as shown in the following example.

Example 9.17

Use Theorem 9.8 to compute

$$\frac{d}{dx}\big(\ln(x)\big).$$

Solution

View $y = \ln(x)$ as the inverse of $y = e^x$. Therefore, using our standard notation, let $f(x) = e^x$ and $g(x) = \ln(x)$. We wish to find $g'(x)$. Theorem 9.8 gives:

$$\begin{aligned} g'(x) &= \frac{1}{f'(g(x))} \\ &= \frac{1}{e^{\ln(x)}} \\ &= \frac{1}{x}. \end{aligned}$$

9.6 L'Hôpital's rule

Our treatment of limits in Chapter 8 exposed us to the notion of 0/0, an indeterminate form. If $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, we do not conclude that $\lim_{x\to c} f(x)/g(x)$ is $0/0$; rather, we use $0/0$ as notation to describe the fact that both the numerator and denominator approach 0. The expression 0/0 has no numeric value; other work must be done to evaluate the limit.

Other indeterminate forms exist; they are: ∞/∞, $0\cdot\infty$, $\infty-\infty$, 0^0, 1^∞ and ∞^0. Just as 0/0 does not mean divide 0 by 0, the expression ∞/∞ does not mean divide infinity by infinity. Instead, it means a quantity is growing without bound and is being divided by another quantity that is growing without bound. We cannot determine from such a statement what value, if any, results in the limit.

9.6.1 Indeterminate forms 0/0 and ∞/∞

Here, we introduce l'Hôpital's rule, a method of resolving limits that produce the indeterminate forms 0/0 and ∞/∞. We will also show how algebraic manipulation can be used to convert other indeterminate expressions into one of these two forms so that our new rule can be applied.

Theorem 9.9 (L'Hôpital's rule for 0/0)
Let $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, where f and g are differentiable functions on an open interval I containing c, and $g'(x) \neq 0$ on I except possibly at c. Then

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}.$$

Proof In order to prove this theorem, suppose that f and g are continuously differentiable at a real number c, that $f(c) = g(c) = 0$, and that $g'(c) \neq 0$. Then

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f(x)-0}{g(x)-0},$$

and since $f(c) = g(c) = 0$, the latter expression equals

$$\lim_{x\to c}\frac{f(x)-f(c)}{g(x)-g(c)}.$$

Dividing both numerator and denominator of this expression by $x-c$ yields

$$\lim_{x\to c}\frac{\left(\dfrac{f(x)-f(c)}{x-c}\right)}{\left(\dfrac{g(x)-g(c)}{x-c}\right)} = \frac{\lim\limits_{x\to c}\left(\dfrac{f(x)-f(c)}{x-c}\right)}{\lim\limits_{x\to c}\left(\dfrac{g(x)-g(c)}{x-c}\right)},$$

or by relying on the difference quotient definition of the derivative and acknowledging continuity of the derivatives at c, we get

$$\frac{f'(c)}{g'(c)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}.$$

The limit in the conclusion is not indeterminate because $g'(c) \neq 0$. □

We demonstrate the use of l'Hôpital's rule (LHR) in the following examples.

Example 9.18

Evaluate the following limits.

1. $\lim\limits_{x\to 0}\dfrac{\sin(x)}{x}$
2. $\lim\limits_{x\to 1}\dfrac{\sqrt{x+3}-2}{1-x}$
3. $\lim\limits_{x\to 0}\dfrac{x^2}{1-\cos(x)}$

Solution

1. We proved this limit is 1 in Example 8.8 using the squeeze theorem. Here we use l'Hôpital's rule to show its power.
$$\lim_{x\to 0}\frac{\sin(x)}{x} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{\cos(x)}{1} = 1.$$
2. $$\lim_{x\to 1}\frac{\sqrt{x+3}-2}{1-x} \overset{\text{by LHR}}{=} \lim_{x\to 1}\frac{\frac{1}{2}(x+3)^{-1/2}}{-1} = -\frac{1}{4}.$$

3. $$\lim_{x\to 0}\frac{x^2}{1-\cos(x)} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{2x}{\sin(x)}.$$

This latter limit also evaluates to the 0/0 indeterminate form. To evaluate it, we apply l'Hôpital's rule again.

$$\lim_{x\to 0}\frac{2x}{\sin(x)} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{2}{\cos(x)} = 2.$$

Thus $\lim\limits_{x\to 0}\dfrac{x^2}{1-\cos(x)} = 2$.

Note that at each step where l'Hôpital's rule was applied, it was needed: the initial limit returned the indeterminate form of $0/0$. If the initial limit returns, for example, 1/2, then l'Hôpital's rule does not apply.

The following theorem extends our initial version of l'Hôpital's rule in two ways. It allows the technique to be applied to the indeterminate form ∞/∞ and to limits where x approaches $\pm\infty$.

Theorem 9.10 (L'Hôpital's rule for ∞/∞)

1. *Let $\lim\limits_{x\to a} f(x) = \pm\infty$ and $\lim\limits_{x\to a} g(x) = \pm\infty$, where f and g are differentiable on an open interval I containing a. Then*
$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}.$$
2. *Let f and g be differentiable functions on the open interval $]a,+\infty[$ for some value a, where $g'(x)\neq 0$ on $]a,+\infty[$ and $\lim\limits_{x\to+\infty} f(x)/g(x)$ returns either "0/0" or "∞/∞". Then*
$$\lim_{x\to+\infty}\frac{f(x)}{g(x)} = \lim_{x\to+\infty}\frac{f'(x)}{g'(x)}.$$

A similar statement can be made for limits where x approaches $-\infty$.

Example 9.19

Evaluate the following limits.

1. $\lim\limits_{x\to+\infty}\dfrac{3x^2-100x+2}{4x^2+5x-1000}$
2. $\lim\limits_{x\to+\infty}\dfrac{e^x}{x^3}$

Solution

1. We can evaluate this limit already using Theorem 8.13; the answer is 3/4. We apply l'Hôpital's rule to demonstrate its applicability.
$$\lim_{x\to+\infty}\frac{3x^2-100x+2}{4x^2+5x-1000} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{6x-100}{8x+5} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{6}{8} = \frac{3}{4}.$$
2. We directly find that
$$\lim_{x\to+\infty}\frac{e^x}{x^3} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{3x^2} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{6x} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{6} = +\infty.$$

Recall that this means that the limit does not exist; as x approaches $+\infty$, the expression

e^x/x^3 grows without bound. We can infer from this that e^x grows faster than x^3; as x gets large, e^x is far larger than x^3.

9.6.2 Indeterminate forms $0\cdot\infty$ and $\infty-\infty$

L'Hôpital's rule can only be applied to ratios of functions. When faced with an indeterminate form such as $0\cdot\infty$ or $\infty-\infty$, we can sometimes apply algebra to rewrite the limit so that l'Hôpital's rule can be applied. We demonstrate the general idea in the next example.

Example 9.20

Evaluate the following limits.

1. $\lim\limits_{\substack{x\to 0\\ >}}\left(x\,e^{1/x}\right)$
2. $\lim\limits_{\substack{x\to 0\\ <}}\left(x\,e^{1/x}\right)$
3. $\lim\limits_{x\to+\infty}\big(\ln(x+1)-\ln(x)\big)$
4. $\lim\limits_{x\to 0}\left(\dfrac{1}{\tanh(x)}-\dfrac{1}{x}\right)$

Solution

1. As $x \underset{>}{\to} 0$, $x\to 0$ and $e^{1/x}\to e^{+\infty}\to+\infty$. Thus we have the indeterminate form $0\cdot\infty$. We rewrite the expression $x\cdot e^{1/x}$ as
$$\frac{e^{1/x}}{1/x};$$
now, as $x \underset{>}{\to} 0$, we get the indeterminate form ∞/∞ to which l'Hôpital's rule can be applied.
$$\lim_{\substack{x\to 0\\ >}}\left(x\,e^{1/x}\right)=\lim_{\substack{x\to 0\\ >}}\frac{e^{1/x}}{1/x}\overset{\text{by LHR}}{=}\lim_{\substack{x\to 0\\ >}}\frac{(-1/x^2)e^{1/x}}{-1/x^2}=\lim_{\substack{x\to 0\\ >}}e^{1/x}=+\infty.$$
So, we may conclude that $e^{1/x}$ grows faster than x shrinks to zero, meaning their product grows without bound.
2. As $x \underset{<}{\to} 0$, $x\to 0$ and $e^{1/x}\to e^{-\infty}\to 0$. The the limit evaluates to $0\cdot 0$ which is not an indeterminate form. We conclude then that
$$\lim_{\substack{x\to 0\\ <}}\left(x\,e^{1/x}\right)=0.$$
3. This limit initially evaluates to the indeterminate form $\infty-\infty$. By applying a logarithmic rule, we can rewrite the limit as
$$\lim_{x\to+\infty}\big(\ln(x+1)-\ln(x)\big)=\lim_{x\to+\infty}\ln\left(\frac{x+1}{x}\right).$$
As $x\to+\infty$, the argument of the ln term approaches ∞/∞, but
$$\lim_{x\to+\infty}\frac{x+1}{x}=\lim_{x\to+\infty}\frac{x(1+1/x)}{x}=1.$$

Since $x \to +\infty \Rightarrow \frac{x+1}{x} \to 1$, it follows that $x \to +\infty$ implies

$$\ln\left(\frac{x+1}{x}\right) \to \ln 1 = 0.$$

Thus

$$\lim_{x\to+\infty}\big(\ln(x+1) - \ln(x)\big) = \lim_{x\to+\infty}\ln\left(\frac{x+1}{x}\right) = 0.$$

Since this limit evaluates to 0, it means that for large x, there is essentially no difference between $\ln(x+1)$ and $\ln(x)$; their difference is essentially 0.

4. It is straightforward to see that this limit gives rise to the indeterminate form $\infty - \infty$, though by choosing $x\tanh(x)$ as common denominator and reformulate the expression as

$$\lim_{x\to 0}\frac{x - \tanh(x)}{x\tanh(x)},$$

we get the indeterminate form $0/0$. Consequently, we proceed as follows

$$\begin{aligned}
\lim_{x\to 0}\frac{x-\tanh(x)}{x\tanh(x)} &\overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{1-\cosh^{-2}(x)}{\tanh(x) + x\cosh^{-2}(x)} \\
&\overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{2\cosh^{-3}(x)\sinh(x)}{\cosh^{-2}(x)+\cosh^{-2}(x)-2x\cosh^{-3}(x)\sinh(x)} \\
&= 0
\end{aligned}$$

9.6.3 Indeterminate forms 0^0, 1^∞ and ∞^0

When faced with an indeterminate form that involves a power, it often helps to employ the natural logarithmic function. More precisely, we rely on the fact that if $\lim_{x\to c}\ln\big(f(x)\big) = L$, then

$$\lim_{x\to c} f(x) = \lim_{x\to c} e^{\ln(f(x))} = e^L.$$

Example 9.21

Evaluate the following limits.

1. $\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x$

2. $\lim_{\substack{x\to 0\\ >}} x^x$.

Solution

1. This is equivalent to a special limit given in Theorem 8.6. Note that the exponent approaches $+\infty$ while the base approaches 1, leading to the indeterminate form $1^{+\infty}$. Let $f(x) = (1+1/x)^x$; the problem asks to evaluate $\lim_{x\to+\infty} f(x)$. Let us first evaluate $\lim_{x\to+\infty}\ln\big(f(x)\big)$.

$$\lim_{x\to+\infty}\ln\big(f(x)\big) = \lim_{x\to+\infty}\ln\left(1+\frac{1}{x}\right)^x$$

$$= \lim_{x \to +\infty} x \ln\left(1 + \frac{1}{x}\right)$$

$$= \lim_{x \to +\infty} \frac{\ln\left(1 + \frac{1}{x}\right)}{1/x}$$

This produces the indeterminate form 0/0, so we apply l'Hôpital's rule.

$$= \lim_{x \to +\infty} \frac{\frac{1}{1+1/x} \cdot (-1/x^2)}{(-1/x^2)}$$

$$= \lim_{x \to +\infty} \frac{1}{1 + 1/x}$$

$$= 1.$$

Thus $\lim_{x \to +\infty} \ln\big(f(x)\big) = 1$. We finally return to the original limit:

$$\lim_{x \to +\infty} \left(1 + \frac{1}{x}\right)^x = \lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} e^{\ln(f(x))} = e^1 = e.$$

2. This limit leads to the indeterminate form 0^0. Let $f(x) = x^x$ and consider first $\lim_{x \underset{>}{\to} 0} \ln\big(f(x)\big)$.

$$\lim_{x \underset{>}{\to} 0} \ln\big(f(x)\big) = \lim_{x \underset{>}{\to} 0} \ln\left(x^x\right)$$

$$= \lim_{x \underset{>}{\to} 0} x \ln(x)$$

$$= \lim_{x \underset{>}{\to} 0} \frac{\ln(x)}{1/x}.$$

This produces the indeterminate form $\frac{\infty}{\infty}$ so we apply l'Hôpital's rule.

$$= \lim_{x \underset{>}{\to} 0} \frac{1/x}{-1/x^2}$$

$$= \lim_{x \underset{>}{\to} 0} -x$$

$$= 0.$$

Thus $\lim_{x \underset{>}{\to} 0} \ln\big(f(x)\big) = 0$. We finally return to the original limit:

$$\lim_{x \underset{>}{\to} 0} x^x = \lim_{x \underset{>}{\to} 0} f(x) = \lim_{x \underset{>}{\to} 0} e^{\ln(f(x))} = e^0 = 1.$$

This result is supported by the graph of $f(x) = x^x$ given in Figure 9.8.

9.7 Applications of the derivative

9.7.1 Newton's method

Solving equations is one of the most important things we do in mathematics, yet we are surprisingly limited in what we can solve analytically. For instance, equations as simple as $x^5+x+1=0$ or $\cos x = x$ cannot be solved by algebraic methods in terms of familiar functions. Fortunately, there are methods that can give us approximate solutions to equations like these. These methods can usually give an approximation correct to as many decimal places as we like. In Section 8.5 we learned about the bisection method. Here, we focus on another technique (which generally works faster), called **Newton's method** (*methode van Newton*) .

Newton's method is built around tangent lines. The main idea is that if x is sufficiently close to a root of $f(x)$, then the tangent line to the graph at $(x, f(x))$ will cross the x-axis at a point closer to the root than x.

We start Newton's method with an initial guess about roughly where the root is. Call this x_0 (Figure 9.11(a)). Draw the tangent line to the graph at $(x_0, f(x_0))$ and see where it meets the x-axis. Call this point x_1. Then repeat the process – draw the tangent line to the graph at $(x_1, f(x_1))$ and see where it meets the x-axis (Figure 9.11(b)). Call this point x_2. Repeat the process again to get x_3, x_4, etc. This sequence of points will often converge rather quickly to a root of f (Figure 9.11(c)).

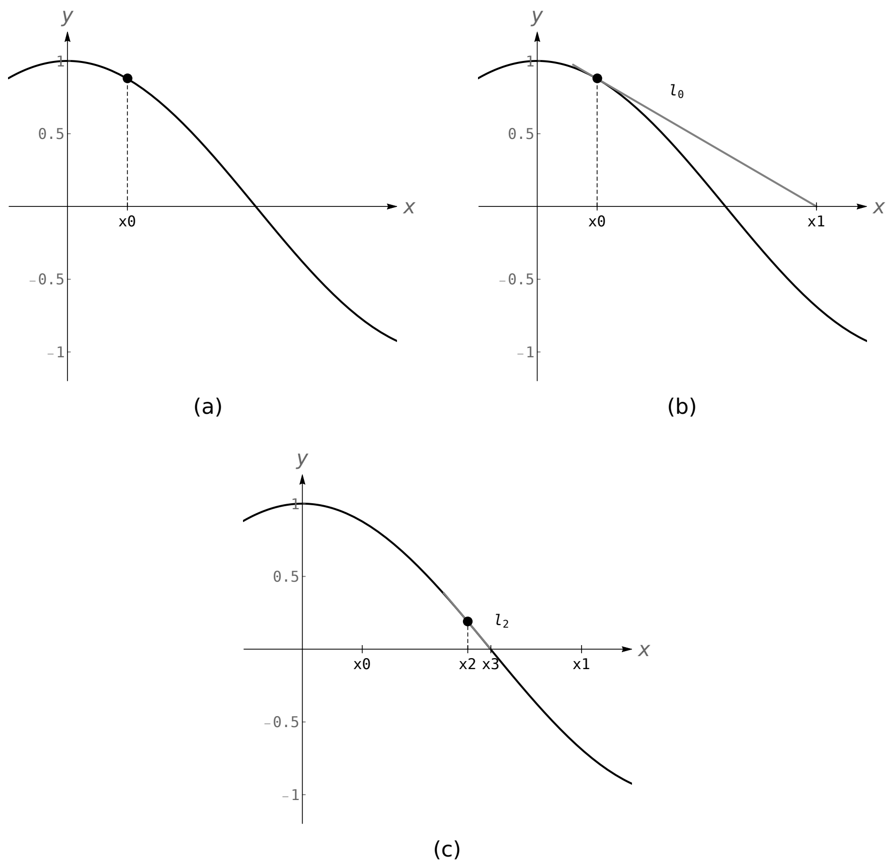


Figure 9.11: The geometric concept behind Newton's method. Note how x_3 is very close to a solution to $f(x)=0$.

We can use this geometric process to create an algebraic process. Let us look at how we found x_1. We started with the tangent line to the graph at $(x_0, f(x_0))$. The slope of this tangent line is $f'(x_0)$ and the equation of the line is

$$y = f'(x_0)(x - x_0) + f(x_0).$$

This line crosses the x-axis when $y = 0$, and the x–value where it crosses is what we called x_1. So let $y = 0$ and replace x with x_1, giving the equation:

$$0 = f'(x_0)(x_1 - x_0) + f(x_0).$$

Now solve for x_1:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Since we repeat the same geometric process to find x_2 from x_1, we have

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

In general, given an approximation x_n, we can find the next approximation, x_{n+1} as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

We summarize this process as follows for a function f that is differentiable on an interval I with a root in I. To approximate the value of the root, accurate to d decimal places:

1. Choose a value x_0 as an initial approximation of the root. This is often done by just looking at a graph of f.
2. Create successive approximations iteratively; given an approximation x_n, compute the next approximation x_{n+1} as
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$
3. Stop the iterations when successive approximations do not differ in the first d places after the decimal point.

Newton's method is not infallible. The sequence of approximate values may not converge, or it may converge so slowly that one is tricked into thinking a certain approximation is better than it actually is. Even though it is not (directly) a method for solving equations like $f(x) = g(x)$, this is not a problem; since we can rewrite the latter equation as $f(x) - g(x) = 0$ and then use Newton's method.

We can of course automate this process on a computer, but for now, let us see how Newton's method works using a concrete example on paper.

Example 9.22

Approximate the real root of $x^3 - x^2 - 1 = 0$, accurate to the first 3 places after the decimal, using Newton's method and an initial approximation of $x_0 = 1$.

Solution

To begin, we compute $f'(x) = 3x^2 - 2x$. Then we apply the Newton's method algorithm.

$$x_1 = 1 - \frac{f(1)}{f'(1)} = 1 - \frac{1^3 - 1^2 - 1}{3 \cdot 1^2 - 2 \cdot 1} = 2,$$

$$x_2 = 2 - \frac{f(2)}{f'(2)} = 2 - \frac{2^3 - 2^2 - 1}{3 \cdot 2^2 - 2 \cdot 2} = 1.625,$$

$$x_3 = 1.625 - \frac{f(1.625)}{f'(1.625)} = 1.625 - \frac{1.625^3 - 1.625^2 - 1}{3 \cdot 1.625^2 - 2 \cdot 1.625} \approx 1.48579.$$

$$x_4 = 1.48579 - \frac{f(1.48579)}{f'(1.48579)} \approx 1.46596$$

$$x_5 = 1.46596 - \frac{f(1.46596)}{f'(1.46596)} \approx 1.46557$$

We performed 5 iterations of Newton's method to find a root accurate to the first 3 places after the decimal; our final approximation is 1.465. The exact value of the root, to six decimal places, is 1.465571; It turns out that our x_5 is accurate to more than just 3 decimal places.

A graph of $f(x)$ is given in Figure 9.12. We can see from the graph that our initial approximation of $x_0 = 1$ was not particularly accurate; a closer guess would have been $x_0 = 1.5$. Our choice was based on ease of initial calculation, and shows that Newton's method can be robust enough that we do not have to make a very accurate initial approximation.

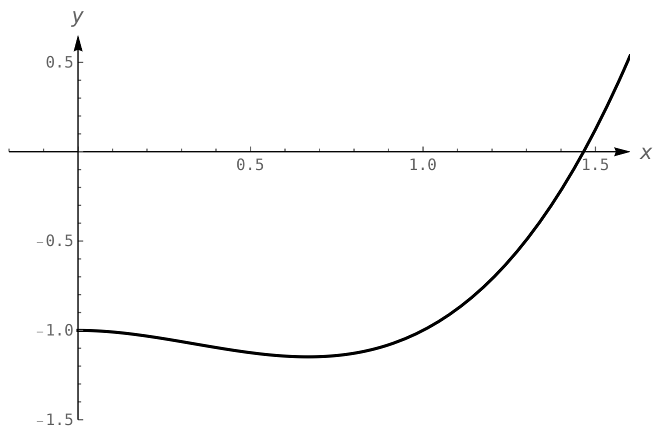


Figure 9.12: A graph of $f(x) = x^3 - x^2 - 1$ in Example 9.22.

What should one use for the initial guess, x_0? Generally, the closer to the actual root the initial guess is, the better. However, some initial guesses should be avoided. For instance, consider Example 9.22 where we sought the root to $f(x) = x^3 - x^2 - 1$. Choosing $x_0 = 0$ would have been a particularly poor choice. Consider Figure 9.13, where $f(x)$ is graphed along with its tangent line at $x = 0$. Since $f'(0) = 0$, the tangent line is horizontal and does not intersect the x–axis. Graphically, we see that Newton's method fails.

We can also see analytically that it fails. Since

$$x_1 = 0 - \frac{f(0)}{f'(0)}$$

and $f'(0) = 0$, we see that x_1 is not well defined. This problem can also occur if, for instance, it turns out that $f'(x_5) = 0$. Adjusting the initial approximation x_0 by a very small amount will likely fix the problem.

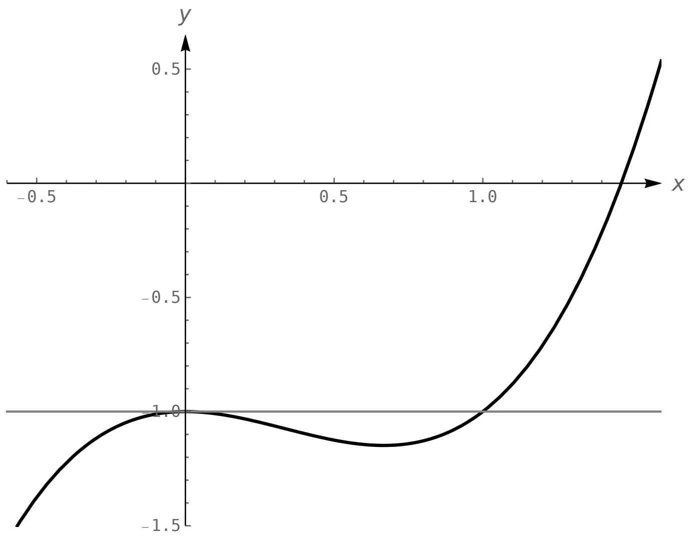


Figure 9.13: A graph of $f(x) = x^3 - x^2 - 1$, showing why an initial approximation of $x_0 = 0$ with Newton's method fails

It is also possible for Newton's method to not converge while each successive approximation is well defined. Consider $f(x) = x^{1/3}$, as shown in Figure 9.14. It is clear that the root is $x = 0$, but let us approximate this with $x_0 = 0.1$. Figure 9.14(a) shows graphically the calculation of x_1; notice how it is farther from the root than x_0. Figures 9.14(b) and 9.14(c) show the calculation of x_2 and x_3, which are even farther away; our successive approximations are getting worse. There is no fix to this problem; Newton's method simply will not work and another method must be used.

While Newton's method does not always work, it does work most of the time, and it is generally very fast. Once the approximations get close to the root, Newton's method can as much as double the number of correct decimal places with each successive approximation.

9.7.2 Related rates

When two quantities are related by an equation, knowing the value of one quantity can determine the value of the other. For instance, the circumference and radius of a circle are related by $C = 2\pi r$; knowing that $C = 6\pi$ centimetres determines the radius must be 3 centimetres.

The topic of related rates takes this one step further: knowing the rate at which one quantity is changing can determine the rate at which another changes. Often, studying related rates will give rise to so-called **differential equations** (*differentiaalvergelijking*).

We demonstrate the concepts of related rates through a few examples.

Example 9.23

The radius of a circle is growing at a rate of 5cm/hr. At what rate is the circumference growing?

Solution

The circumference and radius of a circle are related by $C = 2\pi r$. We are given information about how the length of r changes with respect to time; that is, we are told $\frac{dr}{dt} = 5$cm/hr. We want to know how the length of C changes with respect to time, i.e., we want to know $\frac{dC}{dt}$.

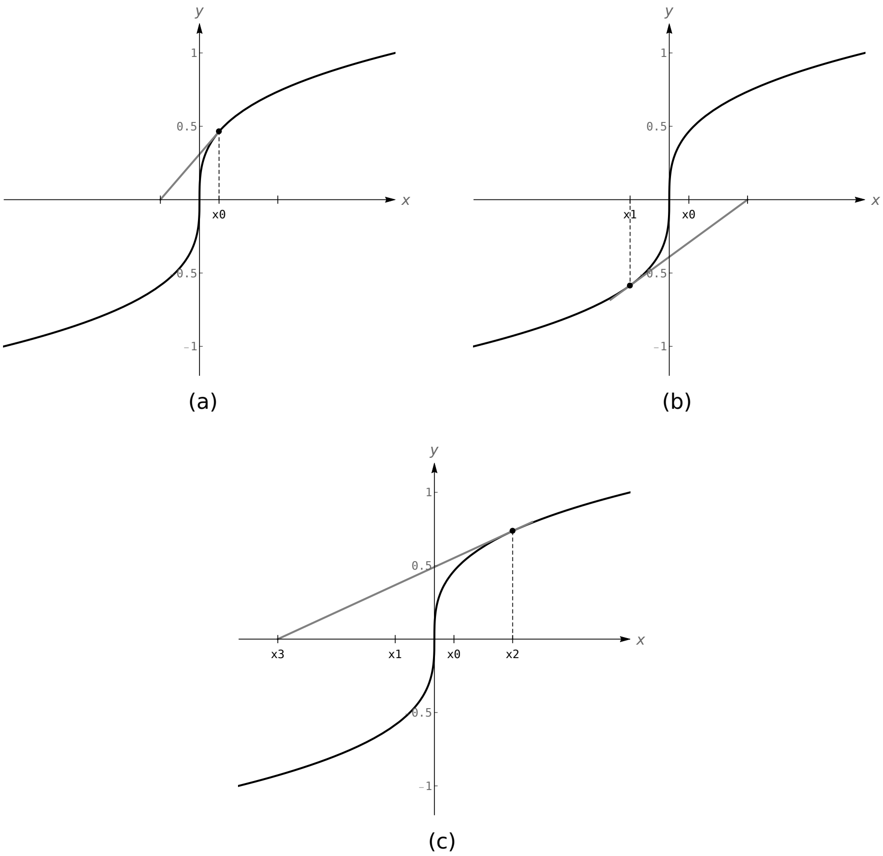


Figure 9.14: Newton's method fails to find a root of $f(x) = x^{1/3}$, regardless of the choice of x_0.

Implicitly differentiate both sides of $C = 2\pi r$ with respect to t:

$$\begin{aligned} C &= 2\pi r \\ \frac{d}{dt}(C) &= \frac{d}{dt}(2\pi r) \\ \frac{dC}{dt} &= 2\pi\frac{dr}{dt}. \end{aligned}$$

As we know $\frac{dr}{dt} = 5$cm/hr, we know

$$\frac{dC}{dt} = 2\pi 5 = 10\pi \approx 31.4\text{cm/hr}.$$

The last equation in Example 9.23 is a typical example of a differential equation, which is an equation involving the derivative of the quantity under study.

Consider another, similar example.

Example 9.24

Radar guns measure the rate of distance change between the gun and the object it is measuring. For instance, a reading of 55 km/hr means the object is moving away from the gun at a rate of 55 kilometres per hour, whereas a measurement of −25 km/hr would mean that the object is approaching the gun at a rate of 25 kilometres per hour.

If the radar gun is moving (say, attached to a police car) then radar readouts are only immediately understandable if the gun and the object are moving along the same line. If a police officer is travelling 60 km/hr and gets a readout of 15 km/hr, he knows that the car ahead of him is moving away at a rate of 15 kilometres an hour, meaning the car is travelling 75 km/hr. This straight–line principle is one reason officers park on the side of the highway and try to shoot straight back down the road. It gives the most accurate reading.

Suppose an officer is driving due north at 30 km/hr and sees a car moving due east, as shown in Figure 9.15. Using his radar gun, he measures a reading of 20 km/hr. By using landmarks, he believes both he and the other car are about 1/2 kilometres from the intersection of their two roads.

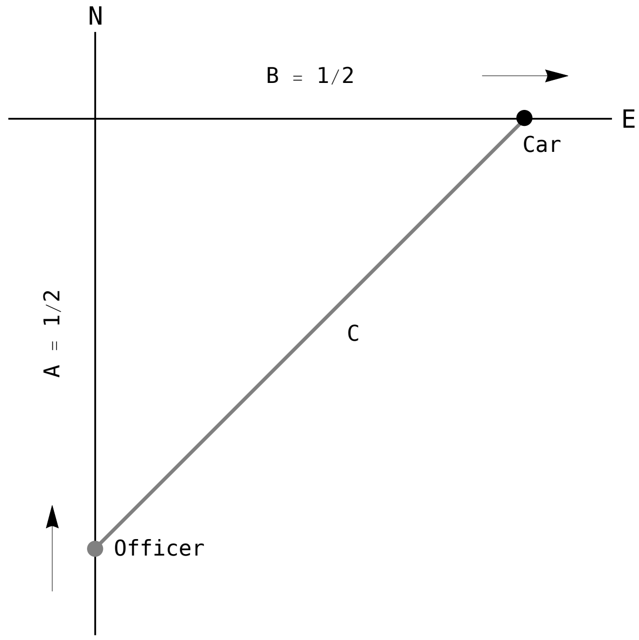


Figure 9.15: A sketch of a police car (at bottom) attempting to measure the speed of a car (at right) in Example 9.24.

If the speed limit on the other road is 55 km/hr, is the other driver speeding?

Solution

Using the diagram in Figure 9.15, let us label what we know about the situation. As both the police officer and other driver are 1/2 kilometres from the intersection, we have $A = 1/2$, $B = 1/2$, and through the Pythagorean theorem, $C = 1/\sqrt{2} \approx 0.707$.

We know the police officer is travelling at 30 km/hr; that is, $\frac{dA}{dt} = -30$. The reason this rate of change is negative is that A is getting smaller; the distance between the officer and the intersection is shrinking. The radar measurement is $\frac{dC}{dt} = 20$. We want to find $\frac{dB}{dt}$.

We need an equation that relates B to A and/or C. The Pythagorean theorem is a good choice: $A^2 + B^2 = C^2$. Differentiate both sides with respect to t:

$$A^2 + B^2 = C^2$$
$$\Rightarrow \frac{d}{dt}\left(A^2 + B^2\right) = \frac{d}{dt}\left(C^2\right)$$
$$\Leftrightarrow 2A\frac{dA}{dt} + 2B\frac{dB}{dt} = 2C\frac{dC}{dt}$$

We have values for everything except $\frac{dB}{dt}$. Solving for this we have

$$\frac{dB}{dt} = \frac{C\dfrac{dC}{dt} - A\dfrac{dA}{dt}}{B} \approx 58.28\text{km/hr}.$$

The other driver appears to be speeding slightly.

The principles presented in the last example are important since many automated vehicles make judgements about other moving objects based on perceived distances, radar–like measurements and the concepts of related rates.

The Verhulst model

One of the most famous models in mathematical biology is the so-called Verhulst model, which was first proposed by Pierre-François Verhulst to describe the change in population size of a certain species.

Letting N [–] represent population size and t [T] represent time, this model is formalized by the following differential equation:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right),$$

where the constant r [T^{-1}] defines the growth rate and K [–] is the carrying capacity, which is the number of individuals that can be supported by the resources available. This model is also known also the logistic growth model.

The solution to the governing equation is

$$P(t) = \frac{KP_0e^{rt}}{K + P_0\left(e^{rt} - 1\right)},$$

where P_0 is the population size at $t = 0$. It can be verified easily that $\lim_{t\to\infty} P(t) = K$, which means that K is the limiting value of P: the highest value that the population can reach given infinite time.

9.7.3 Differentials

Recall that the derivative of a function f can be used to find the slopes of lines tangent to the graph of f. At $x = c$, the tangent line to the graph of f has equation

$$y = f'(c)(x - c) + f(c).$$

The tangent line can be used to find good approximations of $f(x)$ for values of x near c.

We now generalize this concept. Given $f(x)$ and an x value c, the tangent line is

$$y = \ell(x) = f'(c)(x - c) + f(c).$$

Clearly, $f(c) = \ell(c)$. Let Δx be a small number, representing a small change in x value. We assert that:

$$f(c + \Delta x) \approx \ell(c + \Delta x),$$

since the tangent line to a function approximates well the values of that function near $x = c$.

As the x-value changes from c to $c + \Delta x$, the y-value of f changes from $f(c)$ to $f(c + \Delta x)$. We call this change of y-value Δy. That is:

$$\Delta y = f(c + \Delta x) - f(c).$$

Replacing $f(c + \Delta x)$ with its tangent line approximation, we have

$$\Delta y \approx \ell(c + \Delta x) - f(c)$$

$$\begin{aligned} &= f'(c)\big((c+\Delta x)-c\big)+f(c)-f(c) \\ &= f'(c)\Delta x\,. \end{aligned} \tag{9.4}$$

This final equation is important; it becomes the basis of the upcoming definition. In short, it says that when the x-value changes from c to $c+\Delta x$, the y value of a function f changes by about $f'(c)\Delta x$.

We now introduce two new variables, dx and dy in the context of a formal definition.

Definitie 9.8 (Differentials of x and y)
Let $y=f(x)$ be differentiable. The **differential** (*differentiaal*) of x, denoted dx, is any nonzero real number (usually taken to be a small number). The differential of y, denoted dy, is

$$dy=f'(x)dx.$$

We can solve for $f'(x)$ in the above equation: $f'(x)=dy/dx$. This states that the derivative of f with respect to x is the differential of y divided by the differential of x; this is not the alternate notation for the derivative, dy/dx. This latter notation was chosen because of the fraction–like qualities of the derivative, but again, it is one symbol and not a fraction.

In general, if $y=f(x)$ is a differentiable function, we have the following.

1. Let Δx represent a small, nonzero change in x-value.
2. Let dx represent a small, nonzero change in x-value (i.e., $\Delta x = dx$).
3. Let Δy be the change in y value as x changes by Δx; hence
$$\Delta y = f(x+\Delta x)-f(x).$$
4. Let $dy=f'(x)dx$ which, by Equation (9.4), is an approximation of the change in y-value as x changes by Δx; $dy\approx\Delta y$.

Differentials provide both practical and theoretical benefits. We explore both here.

Example 9.25

Consider $f(x)=x^2$. Knowing $f(3)=9$, approximate $f(3.1)$.

Solution

The x-value is changing from $x=3$ to $x=3.1$; therefore, we see that $dx=0.1$. If we know how much the y value changes from $f(3)$ to $f(3.1)$ (i.e., if we know Δy), we will know exactly what $f(3.1)$ is since we already know $f(3)$. We can approximate Δy with dy.

$$\begin{aligned} \Delta y &\approx dy \\ &= f'(3)dx \\ &= 2\cdot 3\cdot 0.1 = 0.6. \end{aligned}$$

We expect the y value to change by about 0.6, so we approximate $f(3.1)\approx 9.6$.

Of course, it is easy to compute the actual answer: $3.1^2=9.61$. So why bother? In most real life situations, we do not know the function that describes a particular behaviour. Instead, we can only take measurements of how things change – measurements of the derivative.

Imagine water flowing down a winding channel. It is easy to measure the speed and direction (i.e., the velocity) of water at any location. It is very hard to create a function that describes the overall flow, hence it is hard to predict where a floating object placed at the beginning of the channel will end up. However, we can approximate the path of an object using differentials. Over small intervals, the path taken by a floating object is essentially linear. Differentials allow us to approximate the true path by piecing together lots of short, linear paths. This technique is called **Euler's method** (*methode van Euler*).

We use differentials once more to approximate the value of a function. Even though calculators are very accessible, it is neat to see how these techniques can sometimes be used to easily compute something that looks rather hard.

Differentials will turn out to be important when we discuss integration (Chapter 12) and proper handling of integrals comes with proper handling of differentials. In light of that, we practice finding differentials in general.

Example 9.26

In each of the following, find the differential dy.

1. $y = \sin(x)$
2. $y = e^x(x^2+2)$
3. $y = \sqrt{x^2+3x-1}$

Solution

1. As $f(x) = \sin(x)$, $f'(x) = \cos(x)$. Thus
$$dy = \cos(x)dx.$$
2. Let $f(x) = e^x(x^2+2)$. We need $f'(x)$, requiring the product rule. We have
$$f'(x) = e^x(x^2+2) + 2xe^x,$$
so
$$dy = \big(e^x(x^2+2) + 2xe^x\big)dx.$$
3. Let $f(x) = \sqrt{x^2+3x-1}$; we need $f'(x)$, requiring the chain rule. We have
$$f'(x) = \frac{1}{2}(x^2+3x-1)^{-\frac{1}{2}}(2x+3) = \frac{2x+3}{2\sqrt{x^2+3x-1}}.$$
Thus
$$dy = \frac{(2x+3)dx}{2\sqrt{x^2+3x-1}}.$$

Finding the differential dy of $y = f(x)$ is really no harder than finding the derivative of f; we just multiply $f'(x)$ by dx. It is important to remember that we are not simply adding the symbol "dx" at the end.

We have seen a practical use of differentials as they offer a good method of making certain approximations. Another use is **error propagation** (*foutenpropagatie*). Suppose a length is measured to be x, although the actual value is $x + \Delta x$ (where Δx is the error, which we hope is small). This measurement of x may be used to compute some other value; we can think of this latter value as $f(x)$ for some function f. As the true length is $x+\Delta x$, one really should have computed $f(x+\Delta x)$. The difference between $f(x)$ and $f(x+\Delta x)$ is the propagated error. How close are $f(x)$ and $f(x+\Delta x)$? This is a difference in y-values:
$$f(x+\Delta x) - f(x) = \Delta y \approx dy.$$

We can approximate the propagated error using differentials.

Example 9.27

A steel ball bearing is to be manufactured with a diameter of 2cm. The manufacturing process has a tolerance of ±0.1mm in the diameter. Given that the density of steel is about 7.85g/cm^3, estimate the propagated error in the mass of the ball bearing.

Solution

The mass of a ball bearing is found using the equation mass = volume × density. In this situation the mass function is a product of the radius of the ball bearing, hence it is $m = 7.85\frac{4}{3}\pi r^3$. The differential of the mass is

$$dm = 31.4\pi r^2 dr.$$

The radius is to be 1cm; the manufacturing tolerance in the radius is ±0.05mm, or ±0.005cm. The propagated error is approximately:

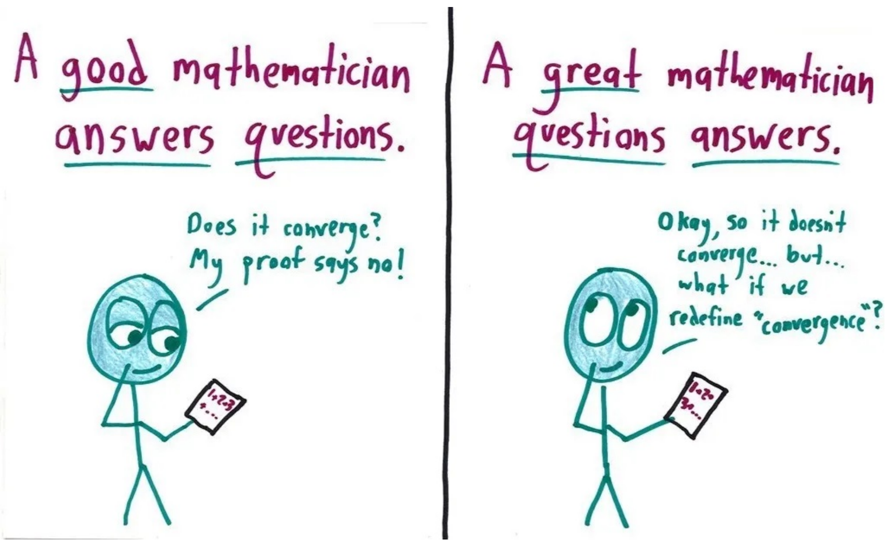
$$\begin{aligned}\Delta m &\approx dm \\ &= 31.4\pi(1)^2(\pm 0.005) \\ &= \pm 0.493\text{g}\end{aligned}$$

Is this error significant? It certainly depends on the application, but we can get an idea by computing the **relative error** (*relatieve fout*). The ratio between amount of error to the total mass is

$$\begin{aligned}\frac{dm}{m} &= \pm\frac{0.493}{7.85\frac{4}{3}\pi} \\ &= \pm\frac{0.493}{32.88} \\ &= \pm 0.015,\end{aligned}$$

or ±1.5%.

If the diameter of the ball was supposed to be 10cm, the same manufacturing tolerance would give a propagated error in mass of ±12.33g, which corresponds to a percent error of ±0.188%. While the amount of error is much greater (12.33 > 0.493), the percent error is much lower.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

9.8 Exercises

9.8.1 Analytical exercises

Definition

Assignment 9.1 — Draw the graph of the function f and determine the continuity and differentiability of f in the given point.

(a) $f(x) = |x|, \quad x = 0$

(b) $f(x) = |x^2 - 1|, \quad x = 1$

(c) $f(x) = |\sin(x)|, \quad x = 0$

(d) $f(x) = \sqrt{x}, \quad x = 0$

(e) $f(x) = \sqrt{1-x^2}, \quad x = -1$

Assignment 9.2 — Consider the function

$$f(x) = \begin{cases} |x|, & \text{als} \quad x \leq 1, \\ 2x^2 - 1, & \text{als} \quad x > 1. \end{cases}$$

Describe the differentiability of f over $[-1,0], [-1,1], [-1,2], [0,1]$ and $[1,2]$.

Assignment 9.3 — Determine the equation of the tangent and normal to graphs of the functions below at the given point.

(a) $f(x) = x^3 - 3x^2 + 2, \quad x = 0$

(b) $f(x) = \dfrac{x}{2+x}, \quad x = -1$

(c) $f(x) = \sin(x), \quad x = \dfrac{\pi}{4}$

(d) $f(x) = \tan(x), \quad x = \dfrac{\pi}{4}$

(e) $f(x) = e^x \left(x^2 + 2\right), \quad x = 0$

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

Assignment 9.4 — Determine the equation(s) of the tangent line(s) at the point with abscissa $x = 0$ to the graph of

(a) $f(x) = |\sin(x)|$

(b) $f(x) = x^{2/3}$

(c) $f(x) = x^{4/3}$

(d) $f(x) = x^3$

Assignment 9.5 — There exist two non-coinciding, intersecting lines that go through $(1,-3)$ and are tangent to $y = x^2$. Determine their equations.

Assignment 9.6 — Suppose that a colony of bacteria reproduces such that at time t the size of the population is $N(t) = N(0)\,2^t$, where $N(0)$ is the size of the population at time $t = 0$. Determine the rate at which the population increases. Show that this rate is directly proportional to the size of the population.

Assignment 9.7 — Give the formula for $f^{(n)}(x)$ given that

(a) $f(x) = \dfrac{1}{x}$

(b) $f(x) = \dfrac{1-x}{1+x}$

(c) $f(x) = \dfrac{1}{\sqrt{x}}$

(d) $f(x) = xe^{ax}$

The chain rule

Assignment 9.8 — Find the first derivative of the functions below.

(a) $f(x) = x^5 - 4x^4 + 3x^2 - 6x + 1$

(b) $f(x) = x\sqrt{x^2-1}$

(c) $f(x) = \dfrac{3x+2}{2x-1}$

(d) $f(x) = \sqrt{\dfrac{x-1}{x+1}}$

(e) $f(x) = \dfrac{1}{(7x^2+3x-6)^2}$

(f) $f(x) = \sqrt[3]{(2x+3)^2}$

(g) $f(x) = x\sqrt[3]{x+1}$

(h) $f(x) = \dfrac{x}{2x + \dfrac{1}{3x+1}}$

(i) $f(x) = \left(1 + \sqrt{\dfrac{x-2}{3}}\right)^4$

Assignment 9.9 — Find the first derivative of the functions below.

(a) $f(x) = x\sin^2(x)$

(b) $f(x) = x^3\cos(x)$

(c) $f(x) = x^5\left(\sec(x) + e^x\right)$

(d) $f(x) = \dfrac{e^{-2x}}{x^2}$

(e) $f(x) = \ln\left(\left(\dfrac{e^x+1}{e^x-1}\right)^{1/2}\right)$

(f) $f(x) = e^{-\arcsin(x)}$

(g) $f(x) = \tan(e^x)$

(h) $f(x) = \log_{10}\left(\sqrt{9-x^2}\right)$

(i) $f(x) = \log_2\left((x^2+x+2)^4\right)$

(j) $f(x) = \log_3\left(\sqrt{\dfrac{2x-1}{2x+1}}\right)$

(k) $f(x) = \arcsin\left(\dfrac{x}{x-1}\right)$

(l) $f(x) = \dfrac{\sin\left(\sqrt{x}\right)}{1+\cos\left(\sqrt{x}\right)}$

(m) $f(x) = \arccos\left(\dfrac{x-b}{a}\right)$

(n) $f(x) = \left(\arcsin(x^2)\right)^{1/2}$

(o) $f(x) = \sqrt{a^2-x^2} + a\arcsin\left(\dfrac{x}{a}\right)$

Assignment 9.10 — Find the first derivative of the functions below, then simplify the result.

(a) $f(x) = \ln\left|\dfrac{1}{\cos(x)} + \tan(x)\right|$

(b) $f(x) = \dfrac{1}{2}\dfrac{\tan(x)}{\cos(x)} + \dfrac{1}{2}\ln\left|\dfrac{1}{\cos(x)} + \tan(x)\right|$

(c) $f(x) = -\dfrac{\sqrt{x^2-a^2}}{x} + \ln\left|x + \sqrt{x^2-a^2}\right|$

(d) $f(x) = \dfrac{x}{8}\left(5-2x^2\right)\sqrt{1-x^2} + \dfrac{3}{8}\arcsin(x)$

(e) $f(x) = \ln\left|\dfrac{\sqrt{x+1}-1}{\sqrt{x+1}+1}\right|$

(f) $f(x) = \dfrac{1}{a^2}\left(\ln|ax+b| + \dfrac{b}{ax+b}\right)$

(g) $f(x) = x\left(2x^2-a^2\right)\sqrt{a^2-x^2} + a^4\arcsin\left(\dfrac{x}{a}\right)$

(h) $f(x) = \dfrac{(x+1)\left(9-2x-x^2\right)}{4}\sqrt{3-2x-x^2} + 6\arcsin\left(\dfrac{x+1}{2}\right)$

Assignment 9.11 — Find the first derivative of the functions below.

(a) $f(x) = x - \tanh(x)$

(b) $f(x) = x\tanh(x)$

(c) $f(x) = 3^{\cosh(x)}$

(d) $f(x) = \ln(\cosh(x)) - \ln(\sinh(x))$

(e) $f(x) = \ln(\cosh(x) + \sinh(x))$

(f) $f(x) = \dfrac{x\sinh(x) - \cosh(x)}{x\cosh(x) - \sinh(x)}$

(g) $f(x) = \dfrac{\sinh(\cosh(x))}{x^2}$

(h) $f(x) = \operatorname{arcosh}\left(\dfrac{1}{x}\right)$

(i) $f(x) = xe^{\operatorname{artanh}(x)}$

(j) $f(x) = \operatorname{arcosh}(\sqrt{x^4+1})$

(k) $f(x) = \operatorname{artanh}\left(\dfrac{x}{\sqrt{x^2+1}}\right)$

(l) $f(x) = \operatorname{arsinh}\left(\dfrac{1-x}{1+x}\right)$

Assignment 9.12 — Express the derivative of the given function in terms of the derivative f' of the differentiable function f.

(a) $\left(f\left(\dfrac{2}{x}\right)\right)^3$

(b) $f\left(2f(3f(x))\right)$

Implicit differentiation

Assignment 9.13 — Consider the implicitly defined functions below. Define y' as a function of x and y.

(a) $x^2e^2 + 2^y = 5$

(b) $\left(3x^2 + 2y^3\right)^4 = 2$

(c) $xy - x + 2y = 1$

(d) $x^3y + xy^5 = 2$

(e) $x^2 + 4(y-1)^2 = 4$

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

(g) $\dfrac{\sin(x)+y}{\cos(y)+x} = 1$

(h) $\ln\left(x^2 + xy + y^2\right) = 1$

Assignment 9.14 — Determine an equation of the tangent to the given curve at the given point.

(a) $\dfrac{x}{y} + \left(\dfrac{y}{x}\right)^3 = 2, \qquad (-1,-1)$

(b) $x\sin(xy - y^2) = x^2 - 1, \qquad (1,1)$

(c) $\left(x^2 + y^2 + x\right)^2 = x^2 + y^2, \qquad (0,1)$

(d) $e^{xy}\ln\left(\dfrac{x}{y}\right) = x + \dfrac{1}{y}, \qquad \left(e, \dfrac{1}{e}\right)$

Assignment 9.15 — We consider $x^2 + 4y^2 = 4$. Determine y'' as a function of x and y.

Assignment 9.16 — Determine y' for the functions below.

(a) $y = \dfrac{x^x}{x+1}$

(b) $y = x^{\sin(x)+2}$

(c) $y = \left(\sin(x)\right)^{\ln(x)}$ with $x \in]0,\pi[$

(d) $y = \left(\dfrac{1}{x}\right)^{\ln(x)}$

(e) $y = (\cos(x))^x - x^{\cos(x)}$

(f) $y = \dfrac{x^{\ln(x)}(\sin(x))^x}{x^x\ln(x)}$

(g) $y = (x^x)^x$

(h) $y = x^{(x^x)}$

Derivatives of inverse functions

Assignment 9.17 — We consider an injective function $f(x)$ for which it holds that $f'(x) = \dfrac{1}{1+x}$. Determine $(f^{-1})'(x)$.

Assignment 9.18 — Consider the functions below and find the required derivative.

(a) $f(x) = 1 + 2x^3, \qquad (f^{-1})'(x)$

(b) $f(x) = x\sqrt{3+x^2}, \qquad (f^{-1})'(-2)$

(c) $f(x) = \dfrac{4x^3}{x^2+1}, \qquad (f^{-1})'(2)$

(d) $f(x) = x^3 + x, \qquad (f^{-1})'(10)$

(e) $f(x) = \sin(2x), \qquad (f^{-1})'(\sqrt{3}/2)$

(f) $f(x) = 6e^{3x}, \qquad (f^{-1})'(6)$

L'Hôpital's rule

Assignment 9.19 — Find the limits below.

(a) $\displaystyle\lim_{x\to 0}\frac{2x-\sin(x)}{x^3}$

(b) $\displaystyle\lim_{x\to 1}\frac{3\ln(x)}{x^5-1}$

(c) $\displaystyle\lim_{x\to+\infty}\frac{x^n}{e^x}$

(d) $\displaystyle\lim_{x\to 0}\frac{\tan(x)-\sin(x)}{x^3}$

(e) $\displaystyle\lim_{x\to+\infty}\left(x\tan\left(\frac{1}{x}\right)\right)$

(f) $\displaystyle\lim_{x\to 0}\left(\frac{1}{x}-\frac{1}{\sin(x)}\right)$

(g) $\displaystyle\lim_{x\to 1}\left(\frac{1}{\ln(x)}-\frac{x}{x-1}\right)$

(h) $\displaystyle\lim_{\substack{x\to\frac{\pi}{2}\\ <}}(\cos(x))^{\frac{\pi}{2}-x}$

(i) $\displaystyle\lim_{\substack{x\to 0\\ >}}x^{\frac{3}{x+\ln(x)}}$

(j) $\displaystyle\lim_{\substack{x\to\frac{\pi}{2}\\ <}}(\tan(x))^{\cos(x)}$

(k) $\displaystyle\lim_{\substack{x\to 0\\ >}}(\csc(x))^{\sin^2(x)}$

(l) $\displaystyle\lim_{x\to+\infty}\left(\cos\left(\sqrt{\frac{3}{x}}\right)\right)^x$

(m) $\displaystyle\lim_{x\to 0}(\cos(x))^{1/x^2}$

(n) $\displaystyle\lim_{x\to 0}x\cot(x)$

(o) $\displaystyle\lim_{x\to 1}\frac{\ln(ex)-1}{\sin(\pi x)}$

(p) $\displaystyle\lim_{\substack{x\to 0\\ >}}x^{\sqrt{x}}$

(q) $\displaystyle\lim_{t\to 0}(\cos(2t))^{1/t^2}$

(r) $\displaystyle\lim_{x\to 0}\frac{\sinh(x)}{x}$

(s) $\displaystyle\lim_{x\to 0}\left(\frac{1}{x}-\frac{1}{\tanh(x)}\right)$

(t) $\displaystyle\lim_{x\to 0}(\cosh(x))^{\frac{1}{\sinh(x)}}$

(u) $\displaystyle\lim_{x\to+\infty}\left(x\,\operatorname{arsinh}\left(\frac{3}{x}\right)\right)$

(v) $\displaystyle\lim_{x\to-\infty}\frac{\cosh(x)-1}{\sinh(x)}$

(w) $\displaystyle\lim_{x\to 0}\frac{\operatorname{arsinh}(5x)}{\ln(1-x)}$

(x) $\displaystyle\lim_{x\to+\infty}\left[x^2\left(1-\cosh\left(\frac{4}{x}\right)\right)\right]$

(y) $\displaystyle\lim_{x\to+\infty}(\operatorname{arcosh}(x)-\operatorname{arsinh}(x))$

(z) $\displaystyle\lim_{x\to 2}[3\sinh(x-2)+1]^{\frac{1}{x-2}}$

Applications of the derivative

Assignment 9.20 — A cube has ribs of length 20 cm. By how much must the length of the ribs decrease so that the volume of the cube decreases with 12 cm^3?

Assignment 9.21 — A spherical balloon is inflated causing the radius to increase in one minute from 20 cm tot 20.2 cm. By how much is the volume going to increase in one minute?

9.8.2 Root finding algorithms

Two methods for finding roots numerically were discussed: the **Bisection method** (*halveringsmethode*) (Section 8.5.2) and **Newton's method** (*methode van Newton*) (Section 9.7.1).

Here, we will implement both methods in Python. For those who have little or no programming experience, a Python Tutorial is provided in Appendix C. Both the code below and the Python Tutorial are available as Jupyter Notebooks.

9.8.2.1 Bisection method

Below is an implementation of the bisection method in Python. First, see how this method translates to executable Python code, then answer the questions below.

```
def bisection(f, interval, eps=10**-6, max_it=100):
        '''
         Bisection method for approximating the root
        of the function f on a given interval [a,b],
        where f(a) and f(b) have a different sign
        Inputs:
        - f: function whose roots should be found
        - interval: interval [a,b]
        - eps: maximum approximation error (default: 10^-6)
        - max_it: maximum number of iterations

        Output:
        - root: approximated root
        '''
        print("Bisection method")
        print("----------------")
        # extract the values of a and b from the interval and calculate the corresponding
    function values
        a = interval[0]
        b = interval[1]
        f_a = f(a)
        f_b = f(b)

        # display error message if the sign of a and b does not differ
        if np.sign(f_a)==np.sign(f_b):
                print("The sign of f(a) and f(b) does not differ!\n")
                return None

        # determine the midpoint of the interval
        m = (a+b)/2
        f_m = f(m)

        # initialize the iterator
        it = 0

        while abs(f_m)>=eps:
                # evaluate sign of m
                if np.sign(f_a)==np.sign(f_m):
                        a=m
                        f_a=f_m
                else:
```

```
                        b=m
                        f_b = f_m
                        m = (a+b)/2
                        f_m = f(m)

                # update the iterator
                it=it+1
                #stop when the maximum number of iterations is reached
                if it==max_it:
                        print("Maximum number of iterations reached!")
                        break
        root = m
        print("Approximated root {}\nwas reached after {} iterations\n".format(root,it))
        print("==================================================")
        return root
```

Question 1.a How does the method in the implementation above differ from that of Example 8.16?

Question 1.b Use the function **bisection** to approximate the root(s) of the following functions with a maximum approximation error of $\epsilon = 10^{-6}$.

- $g_1(x) = \sin(x)$ on the interval $\left[\frac{3\pi}{4}, \frac{3\pi}{2}\right]$
- $g_2(x) = 2x^2 - 2$ on the interval $\left[-\frac{3}{2}, 0\right]$
- $g_3(x) = 2x^2 + 2$ on the interval $[-5, 5]$
- $g_4(x) = \frac{x^5}{2} - 3x^4 + 5x^3 + 6x^2 - 9x - 5$ on the interval $[-1, 1]$

These functions are available in the file `teachingtools`, from which they can be imported as follows.

```
from teachingtools import g_1, g_2, g_3
```

This way, for g_1 we obtain:

```
>>> bisection(g_1, [3*np.pi/4, 3*np.pi/2], eps=10**-6, max_it=100);
        Bisection method
        -----------------
        Approximated root 3.1415919045757357
        was reached after 19 iterations

        ==================================================
```

9.8.2.2 Newton's method

A faster alternative to the bisection method is Newton's method, which was discussed in Section 9.7.1.

Question 2.a Implement Newton's method by completing the code below where you find "...". You can find all functions and techniques necessary for this implementation in the Python Tutorial.

```
def Newton(f, df, x0, eps= 10**-6, max_it=100):
    '''
    Newton's method for the approximation of the root
   of the function f, starting from an initial estimate x0
    Inputs:
    - f: function whose root should be found
    - df: derivative of function whose root should be found
    - x0: initial approximation of the root
    - eps: maximum approximation error (default: 10^-6)
    - maxI: maximum number of iterations (default: 100)

    Output:
    - root: approximated root
    '''
    print("Newton's method")
    print("-------------------------")
    it = 0
    x = x0

    while ...: # to be completed
          ...
          ...

         if ...:  #stop when maximum number of iterations is reached
                  # to be completed

            print("Maximum number of iterations reached before convergence criterion
             was satisfied!")
            print("Verify if there is a root.")
            return None
    root = x
    print("Approximated root {}\nwas reached after {} iterations\n".format(root,it))
    print("==================================================")
    return root
```

Question 2.b Find the derivative of the functions from Question 1 and implement them as the functions dg_i (for i going from 1 to 4).

```
def dg_1(x):
        return ...
def dg_2(x):
        return ...
def dg_3(x):
        return ...
```

Question 2.c Use the function **Newton** to find the root of the functions g_1 through g_4 from Question 1 (with a maximum approximation error of $\epsilon = 10^{-6}$). Choose a value at the border of the specified intervals as starting point $x0$ and compare your result with that of the bisection method. What stands out?

Question 3.a Consider the functions

$$k(x) = -\frac{5(-1+2x-5x^2+2x^5)}{3(4-3x^2+5x^3+x^4)} \quad \text{and} \quad l(x) = 2 - \frac{e^x}{20}.$$

These functions are available in the file `teachingtools`, from where they can be imported.

QUestion 3.b Find the intersection point(s) of the graphs of $k(x)$ and $l(x)$ for $x \in [-0.8, 8]$, if any exist. To that end, first find the derivatives of $k(x)$ and $l(x)$ and implement them as $dk(x)$ and $dl(x)$.

Nothing takes place in the world whose meaning is not that of some maximum or minimum.

— Leonhard Euler —

10

The graphical behaviour of functions

Our study of limits led to continuous functions, a certain class of functions that behave in a particularly nice way. Limits then gave us an even nicer class of functions, functions that are differentiable.

This chapter explores many of the ways we can take advantage of the information that continuous and differentiable functions provide.

10.1 Extreme values

Given any quantity described by a function, we are often interested in the largest and/or smallest values that quantity attains. For instance, if a function describes the speed of an object, it seems reasonable to want to know the fastest/slowest the object traveled. If a function describes the value of a stock, we might want to know the highest/lowest values the stock attained over the past year. We call such values **extreme values** (*extrema*).

Definitie 10.1 (Extreme values)
Let f be defined on an interval I containing c.

1. $f(c)$ is the **minimum** (also, absolute minimum) of f on I if $\forall x \in I : f(c) \leq f(x)$.
2. $f(c)$ is the **maximum** (also, absolute maximum) of f on I if $\forall x \in I : f(c) \geq f(x)$.

The maximum and minimum values are the extreme values, or extrema, of f on I.

We can also define relative minima and maxima, which may be understood as the smallest and largest y-value nearby, respectively. We can make this intuitive understanding more formal as follows.

Definitie 10.2 (Relative minimum and relative maximum)
Let f be defined on an interval I containing c.

1. If there is a $\delta > 0$ such that $f(c) \leq f(x)$ for all x in I where $|x-c| < \delta$, then $f(c)$ is a **relative minimum** (*lokaal minimum*) of f. We also say that f has a relative minimum at $(c, f(c))$.

2. If there is a $\delta > 0$ such that $f(c) \geq f(x)$ for all x in I where $|x-c| < \delta$, then $f(c)$ is a **relative maximum** (*lokaal maximum*) of f. We also say that f has a relative maximum at $(c, f(c))$.

The relative maximum and minimum values comprise the **relative extrema** (*lokaal extremum*) of f.

The function displayed in Figure 10.1(a) has a maximum, but no minimum, as the interval over which the function is defined is open. In Figure 10.1(b), the function has a minimum, but no maximum; there is a discontinuity in the natural place for the maximum to occur. Finally, the function shown in Figure 10.1(c) has both a maximum and a minimum; note that the function is continuous and the interval on which it is defined is closed.

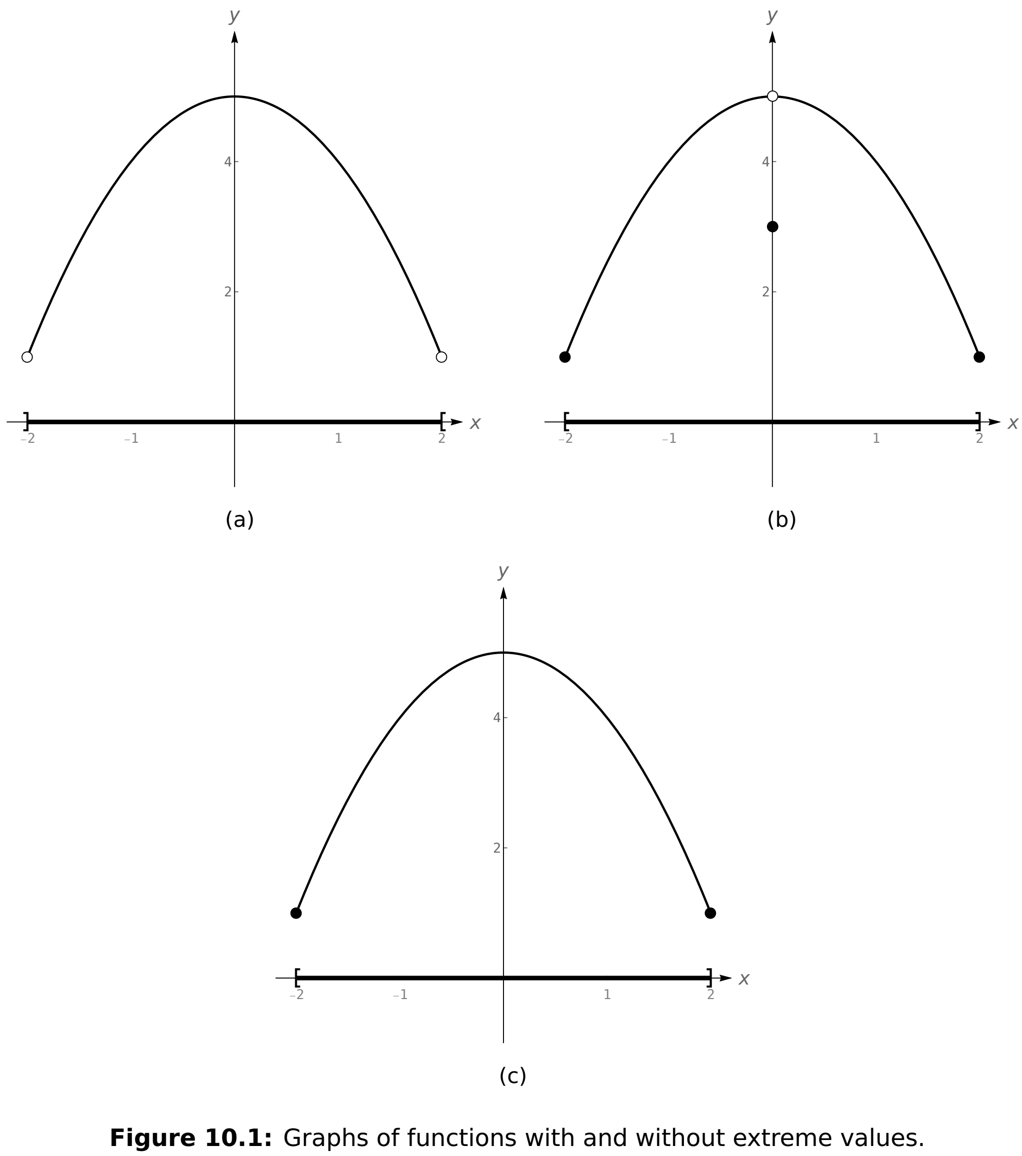


Figure 10.1: Graphs of functions with and without extreme values.

It is possible for discontinuous functions defined on an open interval to have both a maximum and minimum value, but we have just seen examples where they did not. On the other hand, continuous functions on a closed interval always have a maximum and minimum value. This is formalized in the following theorem.

Theorem 10.1 (The extreme value theorem)
Let f be a continuous function defined on a closed interval I. Then f has both a maximum and minimum value on I.

Proof The set $\{y \in \mathbb{R} : y = f(x), x \in [a,b]\}$ is a bounded set. Hence, its least upper bound exists by the least upper bound property of the real numbers. Let M be this supremum, i.e. $M = sup(f(x))$ on $[a,b]$. If there is no point x on $[a,b]$ so that $f(x) = M$, then $f(x) < M$ on $[a,b]$. Therefore,

$$\frac{1}{M - f(x)}$$

is continuous on $[a,b]$.

However, to every positive number ϵ, there is always some x in $[a,b]$ such that $M - f(x) < \epsilon$ because M is the least upper bound. Hence,

$$\frac{1}{M - f(x)} > \frac{1}{\epsilon},$$

which means that $1/(M - f(x))$ is not bounded. Since every continuous function on an interval $[a,b]$ is bounded (Theorem 8.9), this contradicts the conclusion that $1/(M - f(x))$ was continuous on $[a,b]$. Therefore, there must be a point x in $[a,b]$ such that $f(x) = M$. Consequently, the function f attains its maximum value at some point in the interval $[a,b]$.

A similar reasoning is possible for what concerns the minimum. □

This theorem states that f has extreme values, but it does not offer any advice about how/where to find these values. The process can seem to be fairly easy, as the next example illustrates. After the example, we will draw on lessons learned to form a more general and powerful method for finding extreme values.

Example 10.1

Consider the functions

1. $f(x) = \dfrac{3x^4 - 4x^3 - 12x^2 + 5}{5}$,
2. $g(x) = (x-1)^{2/3} + 2$,

as shown in Figure 10.2(a) and 10.2(b), respectively. Approximate the relative extrema of these functions. At each of these points, evaluate the corresponding first derivative.

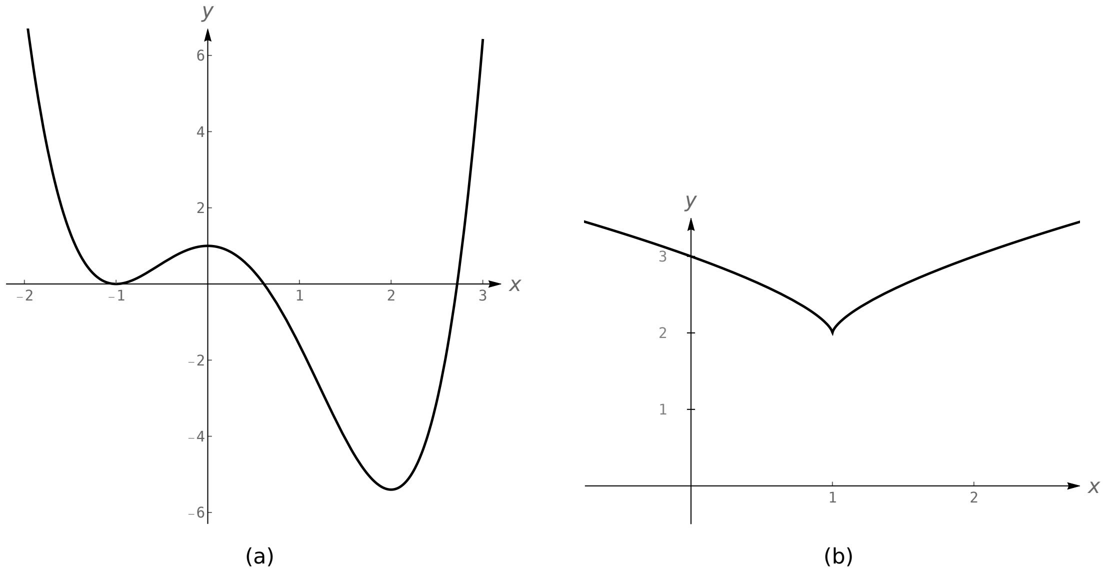


Figure 10.2: A graph of $f(x) = (3x^4 - 4x^3 - 12x^2 + 5)/5$ (a) and $g(x) = (x-1)^{2/3} + 2$ (b).

Solution

We do not yet have the tools to exactly find the relative extrema, but the graphs do allow us to make reasonable approximations.

1. It seems f has relative minima at $x = -1$ and $x = 2$, with values of $f(-1) = 0$ and $f(2) = -5.4$. It also seems that f has a relative maximum at the point $(0, 1)$. We approximate the relative minima to be 0 and -5.4; we approximate the relative maximum to be 1. It is straightforward to evaluate $f'(x) = (12x^3 - 12x^2 - 24x)/5$ at $x = 0, 1$ and 2. In each case, $f'(x) = 0$.
2. Figure 10.2(b) implies that g does not have any relative maxima, but has a relative minimum at $(1, 2)$. The graph suggests that not only is this point a relative minimum, $y = g(1) = 2$ is the absolute minimum value of the function. We compute $g'(x) = 2/3(x-1)^{-1/3}$ note that when $x = 1$, g' is undefined.

What can we learn from the previous two examples? We were able to visually approximate relative extrema, and at each such point, the derivative was either 0 or it was not defined. This observation holds for all functions, leading to a definition and a theorem.

Definitie 10.3 (Critical numbers and critical points)
Let f be defined at c. The value c is a **critical number** (or **critical value** (*kritische waarde*)) of f if $f'(c) = 0$.
If c is a critical number of f, then the point $(c, f(c))$ is a **critical point** (*kritisch punt*) of f.

Definitie 10.4 (Singularities and singular points)
Let f be defined at c. The function has a **singularity** (*singulariteit*) at $x = c$ if $f'(c)$ is not defined.
The point $(c, f(c))$ is called the **singular point** (*singulier punt*).

Theorem 10.2 (Relative extrema and critical points)
Let a function f be defined on an open interval I containing c, and let f have a relative extremum at the point $(c,f(c))$. Then $(c,f(c))$ is a critical or singular point of f.

In case the function f is also differentiable on an open interval I containing c, we restate Theorem 10.2 as follows.

Theorem 10.3 (Fermat's theorem)
Let a function f be defined and differentiable on an open interval I containing c, and let f have a relative extremum at the point $(c,f(c))$. Then $f'(c)=0$.

Proof We can prove Fermat's theorem by assuming that in x_0 there is a local maximum, and then prove that the derivative is 0 (a similar proof applies if there is a local minimum in x_0). Then there exists a $\delta > 0$ such that $]x_0-\delta, x_0+\delta[\subset I$ and such that we have $f(x_0) \geq f(x)$ for all x with $|x-x_0| < \delta$. Hence for any $h \in]0,\delta[$ we notice that the following holds

$$\frac{f(x_0+h)-f(x_0)}{h} \leq 0\,.$$

Since the limit of this ratio as h gets close to 0 from above exists and is equal to $f'(x_0)$ we conclude that $f'(x_0) \leq 0$. On the other hand for $h \in]-\delta, 0[$ we notice that

$$\frac{f(x_0+h)-f(x_0)}{h} \geq 0\,,$$

but again the limit as h gets close to 0 from below exists and is equal to $f'(x_0)$ so we also have $f'(x_0) \geq 0$.

Hence we conclude that $f'(x_0) = 0$. □

Be careful to understand that Theorem 10.2 states that relative extrema on open intervals occur at critical or singular points. It does not say that all such points produce relative extrema. For instance, consider the function $f(x) = x^3$. Since $f'(x) = 3x^2$, it is straightforward to determine that $x = 0$ is a critical number of f. However, f has no relative extrema, as illustrated in Figure 10.3.

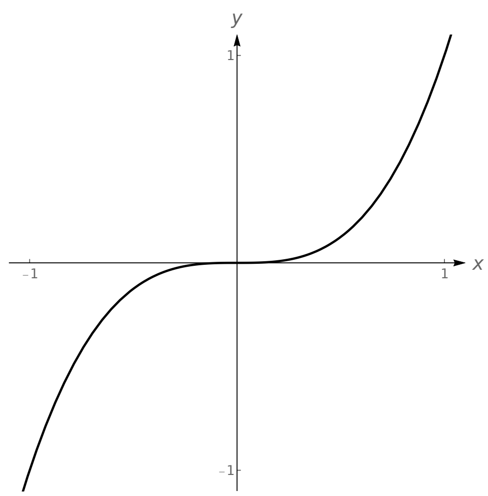


Figure 10.3: A graph of $f(x) = x^3$ which has a critical value of $x = 0$, but no relative extrema.

Theorem 10.1 states that a continuous function on a closed interval will have both an absolute maximum and an absolute minimum. Common sense tells us extrema occur either at the endpoints or somewhere in between. It is easy to check for extrema at endpoints, but there are infinitely many points to check that are in between. Our theory tells us we need only to check at the critical and singular points that are in between the endpoints. We combine these concepts to offer a strategy for finding extrema of a continuous function f defined on a closed interval $[a,b]$.

1. Evaluate f at the endpoints a and b of the interval.
2. Find the critical numbers and singularities of f in $[a,b]$.
3. Evaluate f at each critical number and singularity.
4. The absolute maximum of f is the largest of these values, and the absolute minimum of f is the least of these values.

We practice these ideas in the next examples.

Example 10.2

Find the extrema of the following functions

1. $f(x) = \cos(x^2)$ on $[-2,2]$,
2. $g(x) = \sqrt{1-x^2}$,

which are graphed in Figure 10.4(a) and 10.4(b), respectively.

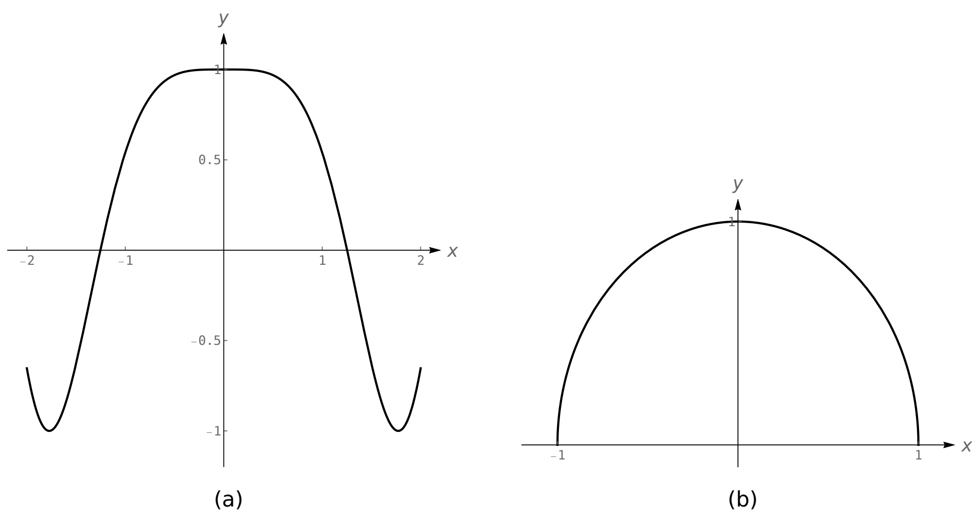


Figure 10.4: A graph of $f(x) = \cos(x^2)$ on $[-2,2]$ (a) and $g(x) = \sqrt{1-x^2}$ (b).

Solution

1. Evaluating f at the endpoints of the interval gives: $f(-2) = f(2) = \cos(4) \approx -0.6536$. We now find the critical values of f. Applying the chain rule, we find $f'(x) = -2x\sin(x^2)$. Set $f'(x) = 0$ and solve for x to find the critical values of f. We do not have to bother about singularities because f' is defined everywhere.

   We have $f'(x) = 0$ when $x = 0$ and when $\sin(x^2) = 0$. In general,

$$\sin(t) = 0 \quad \Leftrightarrow \quad t = \ldots -2\pi, -\pi, 0, \pi, \ldots .$$

Thus $\sin(x^2) = 0$ when $x^2 = 0, \pi, 2\pi, \ldots$ (x^2 is always positive so we ignore $-\pi$, etc.) So $\sin(x^2) = 0$ when $x = 0, \pm\sqrt{\pi}, \pm\sqrt{2\pi}$, etc. The only values to fall in the given interval of $[-2, 2]$ are 0 and $\pm\sqrt{\pi}$, where $\sqrt{\pi} \approx 1.77$. We construct a table for the 5 important values: $x = 0, \pm 2, \pm\sqrt{\pi}$:

x	-2	$-\sqrt{\pi}$	0	$\sqrt{\pi}$	2
$f(x)$	-0.65	-1	1	-1	-0.65

From this table it is clear that the maximum value of f on $[-2, 2]$ is 1 and occurs at $x = 0$; the minimum value is -1 and occurs at $x = \pm\sqrt{\pi}$. The graph of f in Figure 10.4(a) confirms our results.

2. A closed interval is not given, so we find the extreme values of g on its domain. g is defined whenever $1 - x^2 \geq 0$; thus the domain of g is $[-1, 1]$. Evaluating g at either endpoint returns 0. Using the chain rule, we find
$$g'(x) = \frac{-x}{\sqrt{1-x^2}}.$$
The critical points of g are found when $g'(x) = 0$, and its singularities when g' is undefined. It is straightforward to find that $g'(x) = 0$ when $x = 0$, and g' is undefined when $x = \pm 1$, the endpoints of the interval. We get the following table of important values:

x	-1	0	1
$g(x)$	0	1	0

The maximum value is 1 and occurs at $x = 0$. The minimum value is 0 and occurs at $x = \pm 1$.

We can also find extrema of piecewise-defined functions as illustrated in the following example.

Example 10.3

Find the maximum and minimum values of f on $[-4, 2]$, where

$$f(x) = \begin{cases} (x-1)^2, & x \leq 0 \\ x+1, & x > 0. \end{cases}$$

Solution

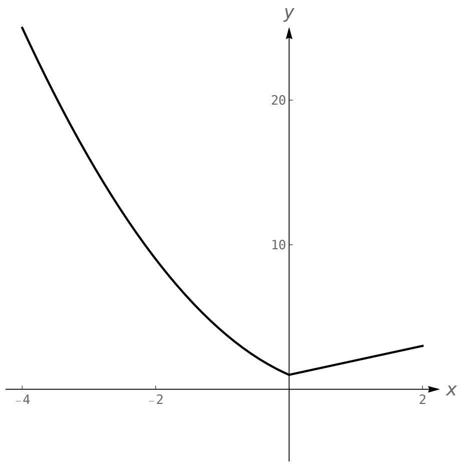
Here f is piecewise–defined, but we can still apply the same approach as before since it is continuous on $[-4, 2]$, i.e. $\lim\limits_{x\to 0} f(x) = f(0)$. Evaluating f at the endpoints gives:

$$f(-4) = 25 \quad \text{and} \quad f(2) = 3.$$

We now find the critical numbers and/or singularities of f. We have to define f' in a piecewise manner; it is
$$f'(x) = \begin{cases} 2(x-1), & x < 0 \\ 1. & x > 0. \end{cases}$$
Note that while f is defined for all of $[-4, 2]$, f' is not, as the derivative of f does not exist when $x = 0$. From the left, the derivative approaches -2; from the right the derivative is 1. Thus f has a singularity at $x = 0$.

We now set $f'(x) = 0$. When $x > 0$, $f'(x)$ is never 0. When $x < 0$, $f'(x)$ is also never 0, so we find no critical values from setting $f'(x) = 0$. So we have three important x values to consider: $x = -4, 2$ and 0. Evaluating f at each gives, respectively, 25, 3 and 1, shown in Table 10.5(b).

Thus the absolute minimum of f is 1, the absolute maximum of f is 25, confirmed by the graph of f in Figure 10.5(a).



x	$f(x)$
-4	25
0	1
2	3

(a) (b)

Figure 10.5: A graph of $f(x)$ on $[-4,2]$ as in Example 10.3 (a) and finding the extrema of f (b).

In the next section, we further our study of the information we can glean from nice functions with the mean value theorem. On a closed interval, we can find the average rate of change of a function. We will see that differentiable functions always have a point at which their instantaneous rate of change is same as the average rate of change. This is surprisingly useful, as we will see.

10.2 The mean value theorem

10.2.1 Introduction

We motivate this section with the following question: Suppose you leave your house and drive to your friend's house in a city 100 kilometres away, completing the trip in two hours. At any point during the trip do you necessarily have to be going 50 kilomteres per hour?

In answering this question, it is clear that the average speed for the entire trip is 50 km/hr, but the question is whether or not your instantaneous speed is ever exactly 50 km/hr. The answer, under some very reasonable assumptions, is yes.

Let us now see why this situation is in a calculus text by translating it into mathematical symbols.

First assume that the function $y = f(t)$ gives the distance (in kilometres) travelled from your home at time t (in hours) where $0 \le t \le 2$. In particular, this gives $f(0) = 0$ and $f(2) = 100$. The slope of the secant line connecting the starting and ending points $(0, f(0))$ and $(2, f(2))$ is therefore

$$\frac{\Delta f}{\Delta t} = \frac{f(2)-f(0)}{2-0} = \frac{100-0}{2} = 50\,\text{km/hr}.$$

The slope at any point on the graph itself is given by the derivative $f'(t)$. So, since the answer to the question above is yes, this means that at some time during the trip, the derivative takes on the value of 50 km/hr. Symbolically,

$$f'(c) = \frac{f(2)-f(0)}{2-0} = 50$$

for some time $0 \leq c \leq 2$.

How about more generally? Given any function $y = f(x)$ and a range $a \leq x \leq b$ does the value of the derivative at some point between a and b have to match the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$? Or equivalently, does the equation

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

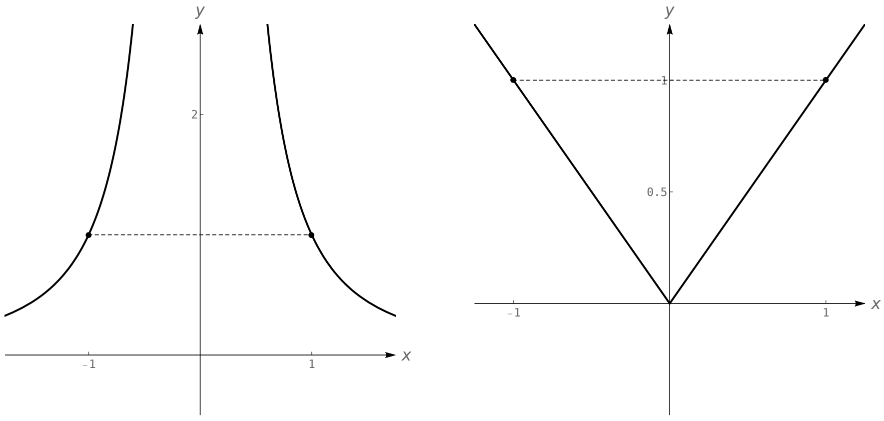
have to hold for some $a < c < b$?

Consider, for instance, the functions

$$f_1(x) = \frac{1}{x^2} \quad \text{and} \quad f_2(x) = |x|$$

with $a = -1$ and $b = 1$ as shown in Figure 10.6 (a) and (b), respectively. Both functions have a value of 1 at a and b. Therefore the slope of the secant line connecting the end points is 0 in each case. But if you look at the plots of each, you can see that there are no points on either graph where the tangent lines have slope zero. Therefore we have found that there is no c in $[-1, 1]$ such that

$$f'(c) = \frac{f(1) - f(-1)}{1 - (-1)} = 0.$$



(a) (b)

Figure 10.6: A graph of $f_1(x) = 1/x^2$ (a) and $f_2(x) = |x|$ (b).

10.2.2 The theorems

So what went wrong? It may not be surprising to find that the discontinuity of f_1 and the corner of f_2 play a role. If our functions had been continuous and differentiable, would we have been able to find that special value c? This is our motivation for the following theorem, which is sometimes also referred to as Lagrange's theorem.

Theorem 10.4 (The mean value theorem of differentiation)
Let $y = f(x)$ be a continuous function on the closed interval $[a, b]$ and differentiable on the open

interval $]a,b[$. *There exists a value* c*, with* $a < c < b$*, such that*

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

That is, there is a value c *in* $]a,b[$ *where the instantaneous rate of change of* f *at* c *is equal to the average rate of change of* f *on* $[a,b]$*.*

Note that the reasons that the functions graphed in Figures 10.6(a) and 10.6(b) fail are indeed that f_1 has a discontinuity on the interval $[-1,1]$ and f_2 is not differentiable at the origin.

We will give a proof of the mean value theorem below. To do so, we use a fact, called Rolle's theorem, stated here.

Theorem 10.5 (Rolle's theorem)
Let f *be continuous on* $[a,b]$ *and differentiable on* $]a,b[$*, where* $f(a) = f(b)$*. There is some* c *in* $]a,b[$ *such that* $f'(c) = 0$.

Consider Figure 10.7 where the graph of a function f is given, where $f(a) = f(b)$. It should make intuitive sense that if f is differentiable (and hence, continuous) that there would be a value c in $]a,b[$ where $f'(c) = 0$; that is, there would be a relative maximum or minimum of f in $]a,b[$. Rolle's theorem guarantees at least one; there may be more.

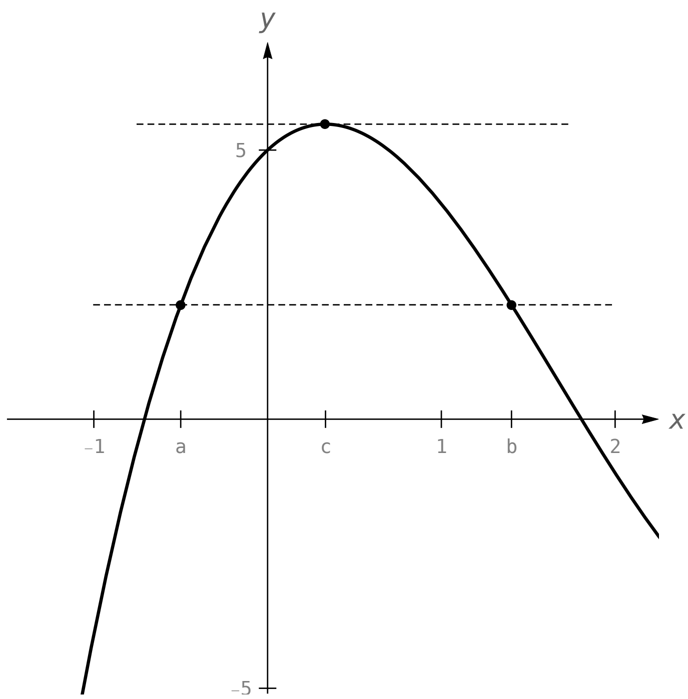


Figure 10.7: A graph of $f(x) = x^3 - 5x^2 + 3x + 5$, where $f(a) = f(b)$. Note the existence of c, where $a < c < b$, where $f'(c) = 0$.

Rolle's theorem is really just a special case of the mean value theorem. If $f(a) = f(b)$, then the average rate of change on $]a,b[$ is 0, and the theorem guarantees some c where $f'(c) = 0$. We will prove Rolle's theorem, then use it to prove the mean value theorem.

Proof of Rolle's theorem

Proof Let f be differentiable on $]a,b[$ where $f(a) = f(b)$. We consider two cases.

Case 1: Consider the case when f is constant on $[a,b]$; that is, $f(x) = f(a) = f(b)$ for all x in $[a,b]$. Then $f'(x) = 0$ for all x in $[a,b]$, showing there is at least one value c in $]a,b[$ where $f'(c) = 0$.

Case 2: Now assume that f is not constant on $[a,b]$. The extreme value theorem guarantees that f has a maximal and minimal value on $[a,b]$, found either at the endpoints or at a critical value in $]a,b[$. Since $f(a) = f(b)$ and f is not constant, it is clear that the maximum and minimum cannot both be found at the endpoints. Assume, without loss of generality, that the maximum of f is not found at the endpoints. Therefore there is a c in $]a,b[$ such that $f(c)$ is the maximum value of f. By Theorem 10.2, c must be a critical number of f; since f is differentiable, we have that $f'(c) = 0$, completing the proof of the theorem.

We can now prove the mean value theorem.

Proof of the mean value theorem

Proof Define the function

$$g(x) = f(x) - \frac{f(b)-f(a)}{b-a}x.$$

We know g is differentiable on $]a,b[$ and continuous on $[a,b]$ since f is. We can show $g(a) = g(b)$, though it is actually easier to show $g(b) - g(a) = 0$, which suffices. We can then apply Rolle's theorem to guarantee the existence of c in $]a,b[$ such that $g'(c) = 0$. But note that

$$0 = g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}$$

hence

$$f'(c) = \frac{f(b)-f(a)}{b-a},$$

which is what we sought to prove. □

Going back to the very beginning of the section, we see that the only assumption we would need about our distance function $f(t)$ is that it be continuous and differentiable for t from 0 to 2 hours (both reasonable assumptions). By the mean value theorem, we are guaranteed a time during the trip where our instantaneous speed is 50 km/hr. This fact is used in practice. Some law enforcement agencies monitor traffic speeds while in aircraft. They do not measure speed with radar, but rather by timing individual cars as they pass over lines painted on the highway whose distances apart are known. The officer is able to measure the average speed of a car between the painted lines; if that average speed is greater than the posted speed limit, the officer is assured that the driver exceeded the speed limit at some time.

Finally, note that the man value theorem is an **existence theorem** (*existentiestelling*). It states that a special value c exists, but it does not give any indication about how to find it. It turns out that when we need the mean value theorem, existence is all we need.

Example 10.4

Consider $f(x) = x^3 + 5x + 5$ on $[-3,3]$. Find c in $]-3,3[$ that satisfies the mean value theorem.

Solution

The average rate of change of f on $[-3,3]$ is:

$$\frac{f(3)-f(-3)}{3-(-3)} = \frac{84}{6} = 14.$$

We want to find c such that $f'(c) = 14$. We find $f'(x) = 3x^2 + 5$. We set this equal to 14 and solve for x.

$$\begin{array}{rrcl} & f'(x) & = & 14 \\ \Rightarrow & 3x^2+5 & = & 14 \\ \Leftrightarrow & x^2 & = & 3 \\ \Leftrightarrow & x & = & \pm\sqrt{3} \approx \pm 1.732 \end{array}$$

We have found 2 values c in $[-3, 3]$ where the instantaneous rate of change is equal to the average rate of change; the mean value theorem guaranteed at least one. In Figure 10.8, f is graphed with a dashed line representing the average rate of change; the lines tangent to f at $x = \pm\sqrt{3}$ are also given. Note how these lines are parallel (i.e., have the same slope) with the dashed line.

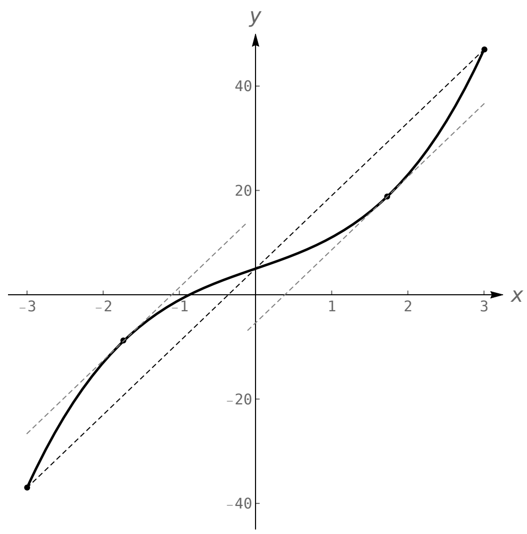


Figure 10.8: Demonstrating the mean value theorem in Example 10.4.

While the mean value theorem has practical use, such as the speed monitoring application mentioned before, it is mostly used to advance other theory. We will use it in the next section to relate the shape of a graph to its derivative.

10.3 Increasing and decreasing functions

Our study of nice functions f in this chapter has so far focused on individual points: points where f is maximal/minimal, points where $f'(x) = 0$ or f' does not exist, and points c where $f'(c)$ is the average rate of change of f on some interval.

In this section we begin to study how functions behave between special points; we begin studying in more detail the shape of their graphs by recalling the following definition from Chapter 3.

Definitie 10.5 (Increasing and decreasing functions)
Let f be a function defined on an interval I.

1. f is **increasing** (*stijgend*) on I if $\big(\forall a, b \in I \mid a < b \Rightarrow f(a) \leq f(b)\big)$.
2. f is **decreasing** (*dalend*) on I if $\big(\forall a, b \in I \mid a < b \Rightarrow f(a) \geq f(b)\big)$.
3. f is **constant** (*constant*) on I if $\big(\forall a, b \in I \mid a < b \Rightarrow f(a) = f(b)\big)$.

Informally, a function is increasing if as x gets larger (i.e., looking left to right) $f(x)$ gets larger, i.e. it does not decrease. Also recall that if the order $\leq$ in the definition of an increasing function is replaced by $<$, we say that f is **strictly increasing** (*strikt stijgend*) on the interval I, and likewise for a **strictly decreasing** (*strikt dalend*) function.

Our interest lies in finding intervals in the domain of f on which f is either increasing or decreasing. Such information should seem useful. For instance, if f describes the speed of an object, we might want to know when the speed was increasing or decreasing (i.e., when the object was accelerating vs. decelerating). If f describes the population of a city, we should be interested in when the population is growing or declining.

To find such intervals, we again consider secant lines. Let f be a strictly increasing, differentiable function on an open interval I, such as the one shown in Figure 10.9, and let $a < b$ be given in I. The secant line on the graph of f from $x = a$ to $x = b$ is drawn; it has a slope of $(f(b)-f(a))/(b-a)$. But note:

$$\frac{f(b)-f(a)}{b-a} \quad \Rightarrow \quad \frac{\text{numerator } > 0}{\text{denominator } > 0} \quad \Rightarrow \quad \begin{array}{c}\text{slope of the}\\ \text{secant line}\\ > 0\end{array} \quad \Rightarrow \quad \begin{array}{c}\text{Average rate}\\ \text{of change of}\\ f \text{ on } [a,b] \text{ is}\\ > 0.\end{array}$$

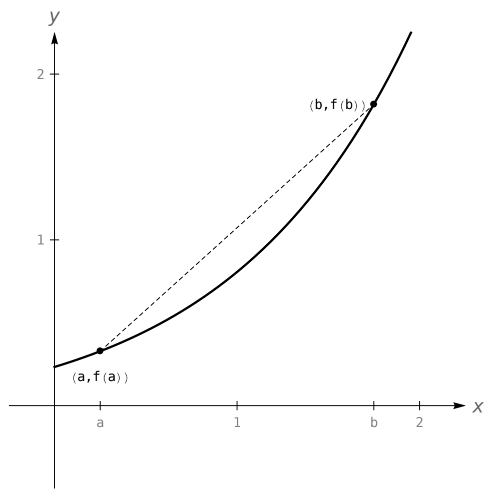


Figure 10.9: Examining the secant line of an increasing function.

We have shown mathematically what may have already been obvious: when f is strictly increasing, its secant lines will have a positive slope. Now recall the mean value theorem guarantees that there is a number c, where $a < c < b$, such that

$$f'(c) = \frac{f(b)-f(a)}{b-a} > 0.$$

By considering all such secant lines in I, we strongly imply that $f'(x) > 0$ on I. A similar statement can be made for strictly decreasing functions.

Our above logic can be summarized as If f is strictly increasing, then f' is probably positive. Theorem 10.6 turns this around by stating If f' is positive, then f is strictly increasing. This leads us to a method for finding when functions are strictly increasing and decreasing.

Theorem 10.6 (Test for increasing/decreasing functions)
Let f be a continuous function on $[a,b]$ and differentiable on $]a,b[$.

1. *If $f'(c) > 0$ for all c in $]a,b[$, then f is strictly increasing on $[a,b]$.*

2. *If $f'(c) < 0$ for all c in $]a,b[$, then f is strictly decreasing on $[a,b]$.*

3. *If $f'(c) = 0$ for all c in $]a,b[$, then f is constant on $[a,b]$.*

Proof The proof of this fact uses the mean value theorem we introduced in the preceding section.

Let us start with the first statement in the theorem, namely that if $f'(c) > 0$ for all c in $]a,b[$, then f is strictly increasing on $[a,b]$.

Let x_1 and x_2 be in $[a,b]$ and suppose that $x_1 < x_2$. Now, using the mean value theorem on $[x_1,x_2]$ means there is a number c such that $x_1 < c < x_2$ and,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) .$$

Because $x_1 < c < x_2$ we know that c must also be in $[a,b]$ and so we know that $f'(c) > 0$. We also know that $x_2 - x_1 > 0$. So, this means that we have,

$$f(x_2) - f(x_1) > 0 .$$

Rewriting this gives

$$f(x_2) > f(x_1) ,$$

and so, by definition, since x_1 and x_2 were two arbitrary numbers in $[a,b]$, $f(x)$ must be strictly increasing on this interval.

This proof for the two other statements in Theorem 10.6 is nearly identical. □

Let f be differentiable on an interval I and let a and b be in I where $f'(a) > 0$ and $f'(b) < 0$. If f' is continuous on $[a,b]$, it follows from the intermediate value theorem (Theorem 8.11) that there must be some value c between a and b where $f'(c) = 0$. It turns out that this is still true even if f' is not continuous on $[a,b]$. This leads us to the following method for finding intervals on which a function is strictly increasing or decreasing.

Let f be a differentiable function on an interval I. To find intervals on which f is increasing and decreasing:

1. Find the critical values and singular points of f. That is, find all c in I where $f'(c) = 0$ or f' is not defined.
2. Use the critical values and singular points to divide I into subintervals.
3. Pick any point p in each subinterval, and find the sign of $f'(p)$.
   (a) If $f'(p) > 0$, then f is strictly increasing on that subinterval.
   (b) If $f'(p) < 0$, then f is strictly decreasing on that subinterval.

Note that parts 1 & 2 of Theorem 10.6 also hold if $f'(c) = 0$ for a finite number of values of c in I. Hence, acknowledging the difference between increasing and strictly increasing functions, we may say that

1. if $f'(p) \geq 0$, then f is increasing.
2. if $f'(p) \leq 0$, then f is decreasing.

We demonstrate using this process in the following example.

Example 10.5

Let $f(x) = x^3 + x^2 - x + 1$. Find intervals on which f is strictly increasing or decreasing.

Solution

Following the method outlined above, we first find the critical values of f. Hence, we have $f'(x) = 3x^2 + 2x - 1 = (3x - 1)(x + 1)$, so $f'(x) = 0$ when $x = -1$ and when $x = 1/3$. f' is never undefined.

Since an interval was not specified for us to consider, we consider the entire domain of f which is $\mathbb{R}$. We thus break the whole real line into three intervals based on the two critical values we just found: $]-\infty, -1[$, $]-1, 1/3[$ and $]1/3, +\infty[$.

We now pick a value p in each interval and find the sign of $f'(p)$. All we care about is the sign, so we do not actually have to fully compute $f'(p)$; pick nice values that make this simple.

Interval 1: $]-\infty, -1[$

We (arbitrarily) pick $p = -2$. We can compute $f'(-2)$ directly:

$$f'(-2) = 3(-2)^2 + 2(-2) - 1 = 7 > 0.$$

We conclude that f is increasing on $]-\infty, -1[$.

Note we can arrive at the same conclusion without computation. For instance, we could choose $p = -100$. The first term in $f'(-100)$, i.e., $3(-100)^2$ is clearly positive and very large. The other terms are small in comparison, so we know $f'(-100) > 0$. All we need is the sign.

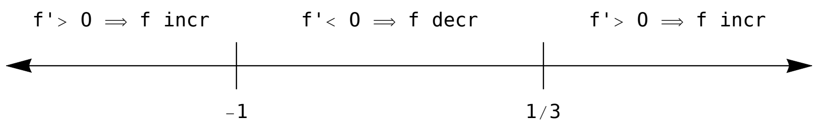
Interval 2: $]-1, 1/3[$

We pick $p = 0$ since that value seems easy to deal with. $f'(0) = -1 < 0$. We conclude f is decreasing on $]-1, 1/3[$.

Interval 3: $]1/3, +\infty[$

Pick an arbitrarily large value for $p > 1/3$ and note that $f'(p) = 3p^2 + 2p - 1 > 0$. We conclude that f is increasing on $]1/3, \infty[$.

In summary, we find:



We can verify our calculations by considering Figure 10.10, where f is graphed. The graph also presents f'; note how $f' > 0$ when f is strictly increasing and $f' < 0$ when f is strictly decreasing.

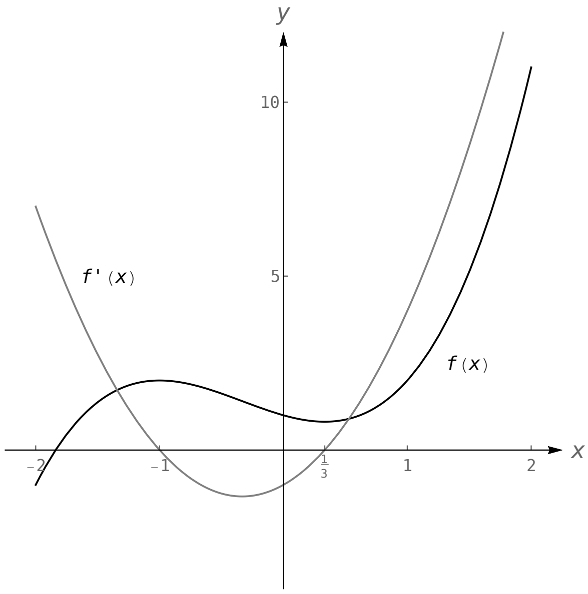


Figure 10.10: A graph of $f(x)$ (black) and $f'(x)$ (gray) in Example 10.5, showing where f is increasing and decreasing.

In Section 10.1 we learned the definition of relative maxima and minima and found that they occur at critical points. We are now learning from Example 10.5 that functions can switch from increasing to decreasing (and vice–versa) at critical points. This new understanding of increasing and decreasing creates a great method of determining whether a critical point corresponds to a maximum, minimum, or neither. Imagine a function increasing until a critical point at $x = c$, after which it decreases. A quick sketch helps confirm that $f(c)$ must be a relative maximum. A similar statement can be made for relative minimums. We formalize this concept in a theorem.

Theorem 10.7 (First derivative test)
Let f be differentiable on an interval I and let c be a critical number in I.

1. *If the sign of f' switches from positive to negative at c, then $f(c)$ is a relative maximum of f.*
2. *If the sign of f' switches from negative to positive at c, then $f(c)$ is a relative minimum of f.*
3. *If f' is positive (or, negative) before and after c, then $f(c)$ is not a relative extremum of f.*

Proof We only prove the first statement in this theorem because the proofs of the other statements are similar.

In this case, we have that f' is positive on $]a,c[$ and negative on $]c,b[$. Let x be an arbitrary point in $]a,c[$. Since f is differentiable on $]a,c[$ and continuous at c, it is continuous on $[x,c]$ and differentiable on $]x,c[$. So, by the mean value theorem, there is a x_0 in $]x,c[$ such that

$$\frac{f(c)-f(x)}{c-x} = f'(x_0)\,.$$

Because x_0 is in $]x,c[$ too, it holds that $f'(x_0) > 0$. And since we also have that $c-x > 0$, it immediately follows that $f(c)-f(x) > 0$, or $f(x) < f(c)$. A similar argument shows that for all x in $]c,b[$, we have that $f(x) < f(c)$. Let $h = \min(c-a, b-c)$. Consequently, for all x in $]c-h,c[\cup]c,c+h[$ it holds that $f(x) < f(c)$. Hence, for all x in $]c-h,c+h[$, we have that $f(x) \leq f(c)$, which implies that f has a local maximum at c. □

Example 10.6

Find the intervals on which f is increasing and decreasing, and determine the relative extrema of f, where

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

Solution

We start by noting the domain of f: $]-\infty,1[\cup]1,+\infty[$. Since the domain of f in this example is the union of two intervals, we apply Theorem 10.7 to both intervals of the domain of f.

Since f is not defined at $x=1$, the increasing/decreasing nature of f could switch at this value. At this point f manifests a singularity, so we should keep track of it.

Using the quotient rule, we find

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

We can now find the critical values and possible further singular points of f; we want to know when $f'(x)=0$ and when f' is not defined. That latter is straightforward: when the denominator of $f'(x)$ is 0, f' is undefined. That occurs when $x=1$, which we have already recognized as an important value.

$f'(x)=0$ when the numerator of $f'(x)$ is 0. That occurs when $x^2-2x-3=(x-3)(x+1)=0$; i.e., when $x=-1,3$.

We have found that f has two critical numbers, $x=-1,3$, and at $x=1$ something important might also happen. These three numbers divide the real number line into 4 subintervals:

$$]-\infty,-1[,\quad]-1,1[,\quad]1,3[\quad \text{and} \quad]3,+\infty[.$$

Pick a number p from each subinterval and test the sign of f' at p to determine whether f is increasing or decreasing on that interval. Again, we do well to avoid complicated computations; notice that the denominator of f' is always positive so we can ignore it during our work.

Interval 1: $]-\infty,-1[$

Choosing a very small number (i.e., a negative number with a large magnitude) p returns p^2-2p-3 in the numerator of f'; that will be positive. Hence f is increasing on $]-\infty,-1[$.

Interval 2: $]-1,1[$

Choosing 0 seems simple: $f'(0)=-3<0$. We conclude f is decreasing on $]-1,1[$.

Interval 3: $]1,3[$

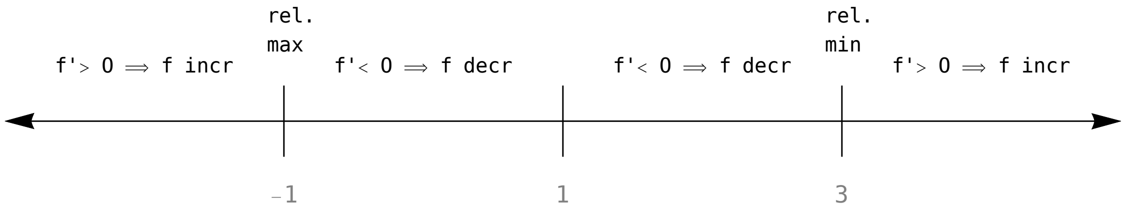
Choosing 2 seems simple: $f'(2)=-3<0$. Again, f is decreasing.

Interval 4: $]3,+\infty[$

Choosing an very large number p from this subinterval will give a positive numerator and (of course) a positive denominator. So f is increasing on $]3,+\infty[$.

In summary, f is increasing on the intervals $]-\infty,-1[$ and $]3,+\infty[$ and is decreasing on the intervals $]-1,1[$ and $]1,3[$. Since at $x=-1$, the sign of f' switched from positive to negative, Theorem 10.7 states that $f(-1)$ is a relative maximum of f. At $x=3$, the sign of f' switched from negative to positive, meaning $f(3)$ is a relative minimum. At $x=1$, f is not defined, so there is no relative extremum at $x=1$.

This is summarized in the number line below. Also, Figure 10.11 shows a graph of f, confirming our calculations. This figure also shows f', again demonstrating that f is increasing when $f' > 0$ and decreasing when $f' < 0$.



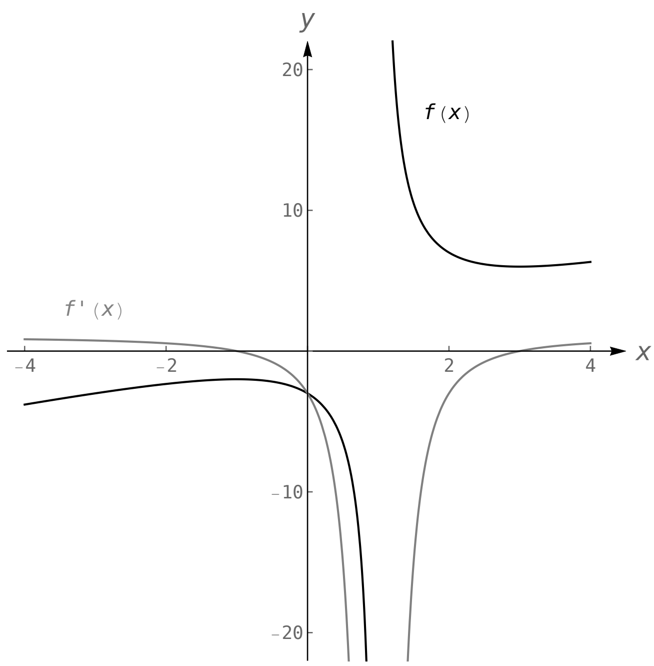


Figure 10.11: A graph of $f(x)$ and $f'(x)$ in Example 10.6, showing where f is increasing and decreasing.

We examine one example.

Example 10.7

Find the intervals on which $f(x) = x^{8/3} - 4x^{2/3}$ is increasing and decreasing and identify the relative extrema.

Solution

We start with taking a derivative. Since we know we want to solve $f'(x) = 0$, we will do some algebra after taking the derivative.

$$\begin{aligned} f(x) &= x^{\frac{8}{3}} - 4x^{\frac{2}{3}} \\ \Rightarrow \quad f'(x) &= \frac{8}{3}x^{\frac{5}{3}} - \frac{8}{3}x^{-\frac{1}{3}} \\ &= \frac{8}{3}x^{-\frac{1}{3}}\left(x^{\frac{6}{3}} - 1\right) \end{aligned}$$

$$= \frac{8}{3}x^{-\frac{1}{3}}(x^2-1)$$

$$= \frac{8}{3}x^{-\frac{1}{3}}(x-1)(x+1).$$

This derivation of f' shows that $f'(x) = 0$ when $x = \pm 1$ and f' is not defined when $x = 0$. Thus we have 2 critical values and one singular point, breaking the number line into 4 subintervals.

Interval 1: $]-\infty, -1[$

We choose $p = -2$; we can easily verify that $f'(-2) < 0$. So f is decreasing on $]-\infty, -1[$.

Interval 2: $]-1, 0[$

Choose $p = -1/2$. We can once more find the sign of $f'(p)$ without computing an actual value. We have $f'(p) = (8/3)p^{-1/3}(p-1)(p+1)$; find the sign of each of the three terms.

$$f'(p) = \frac{8}{3} \cdot \underbrace{p^{-\frac{1}{3}}}_{<0} \cdot \underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}.$$

Consequently, f is increasing on $]-1, 0[$.

Interval 3: $]0, 1[$

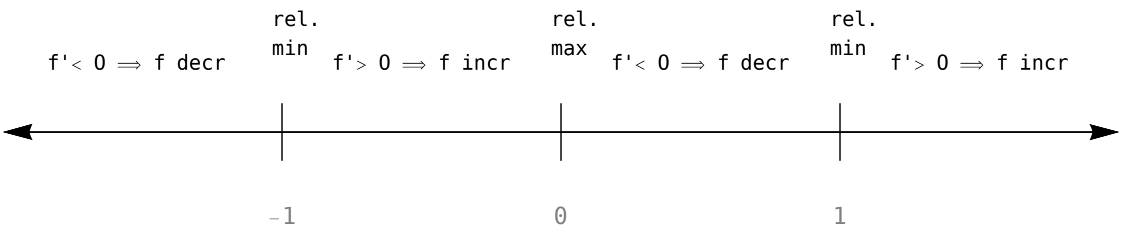
We do a similar sign analysis as before, using p in $]0, 1[$.

$$f'(p) = \frac{8}{3} \cdot \underbrace{p^{-\frac{1}{3}}}_{>0} \cdot \underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}.$$

We have 2 positive factors and one negative factor; $f'(p) < 0$ and so f is decreasing on $]0, 1[$.

Interval 4: $]1, +\infty[$ Similar work to that done for the other three intervals shows that $f'(x) > 0$ on $]1, +\infty[$, so f is increasing on this interval.

Finally, we have:



Consequently, we conclude by stating that f is increasing on the intervals $]-1, 0[$ and $]1, +\infty[$ and decreasing on the intervals $]-\infty, -1[$ and $]0, 1[$. The sign of f' changes from negative to positive around $x = -1$ and $x = 1$, meaning by Theorem 10.7 that $f(-1)$ and $f(1)$ are relative minima of f. As the sign of f' changes from positive to negative at $x = 0$, we have a relative maximum at $f(0)$. Figure 10.12 shows a graph of f, confirming our result. We also graph f', highlighting once more that f is increasing when $f' > 0$ and is decreasing when $f' < 0$.

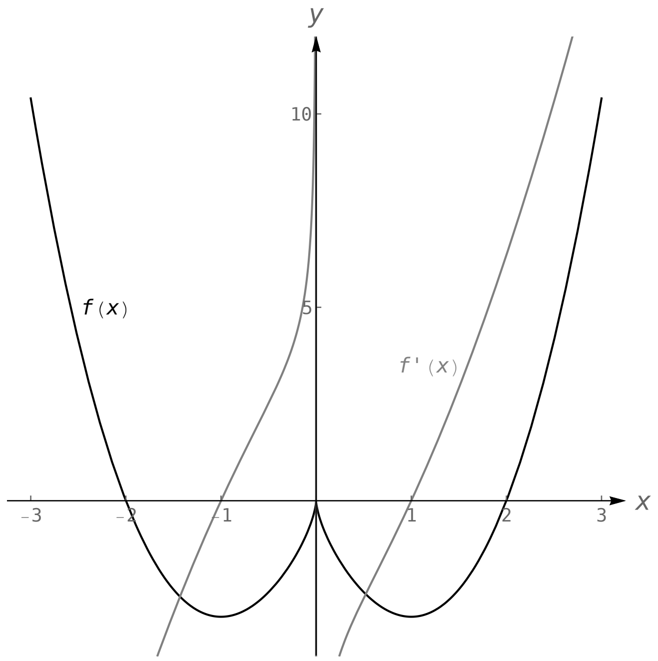


Figure 10.12: A graph of $f(x)$ (black) and $f'(x)$ (gray) in Example 10.7, showing where f is increasing and decreasing.

We have seen how the first derivative of a function helps determine when the function is going up or down. In the next section, we will see how the second derivative helps determine how the graph of a function curves.

10.4 Concavity and the second derivative

Our study of nice functions continues. The previous section showed how the first derivative of a function, f', can relay important information about f. We now apply the same technique to f' itself, and learn what this tells us about f.

The key to studying f' is to consider its derivative, namely f'', which is the second derivative of f. When $f'' \geq 0$, f' is increasing. When $f'' \leq 0$, f' is decreasing. f' has relative maxima and minima where $f'' = 0$ or is undefined.

This section explores how knowing information about f'' gives information about f.

10.4.1 Concavity

We begin with a definition, then explore its meaning.

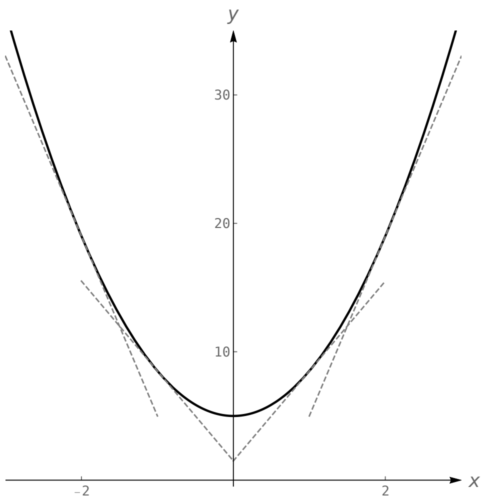
Definitie 10.6 (Concave up and concave down)
Let f be differentiable on an interval I.

1. The graph of f is **concave up** (*convex*) on I if f' is increasing.
2. The graph of f is **concave down** (*concaaf*) on I if f' is decreasing.
3. If f' is constant then the graph of f is said to have no **concavity** (*concaviteit*).

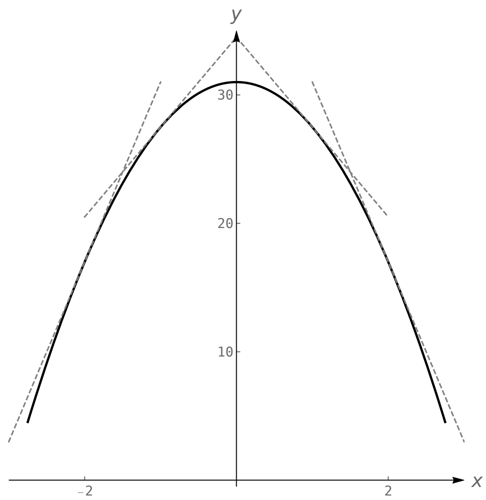
Note that we often state that f is concave up instead of the graph of f is concave up for simplicity. Besides, in agreement with the terminology used for increasing and decreasing functions (Definition 10.5), we call a function f strictly concave up or down if f' is strictly increasing or decreasing, respectively.

The graph of a function f is concave up when f' is increasing. That means as one looks at a concave up graph from left to right, the slopes of the tangent lines will be increasing. Consider Figure 10.13(a), where a concave up graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, downward, corresponding to a small value of f'. On the right, the tangent line is steep, upward, corresponding to a large value of f'. If a function is decreasing and concave up, then its rate of decrease is slowing; it is levelling off. If the function is increasing and concave up, then the rate of increase is increasing.

Now consider a function which is concave down. We essentially repeat the above paragraphs with slight variation. The graph of a function f is concave down when f' is decreasing. That means as one looks at a concave down graph from left to right, the slopes of the tangent lines will be decreasing. Consider Figure 10.13(b), where a concave down graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, upward, corresponding to a large value of f'. On the right, the tangent line is steep, downward, corresponding to a small value of f'. If a function is increasing and concave down, then its rate of increase is slowing; it is levelling off. If the function is decreasing and concave down, then the rate of decrease is decreasing. The function is decreasing at a faster and faster rate. Geometrically speaking it is clear that a function is concave up if its graph lies above its tangent lines. A function is concave down if its graph lies below its tangent lines.



(a)



(b)

Figure 10.13: A function f with a concave up (a) and concave down (b) graph together with some tangent lines (dashed).

Our definition of concave up and concave down is given in terms of when the first derivative is increasing or decreasing. We can apply the results of Section 10.3 to find intervals on which a graph is concave up or down. That is, we recognize that f' is increasing when $f'' \geq 0$, etc.

Theorem 10.8 (Test for concavity)
Let f be twice differentiable on an interval I. The graph of f is concave up if $f'' \geq 0$ on I, and is concave down if $f'' \leq 0$ on I.

Proof We will only prove the concave up part of the theorem as the proof of the concave down part is nearly identical.

Let a be any number in the interval I. The tangent line to $f(x)$ at $x = a$ is,

$$y = \ell(x) = f(a) + f'(a)(x-a) .$$

To show that $f(x)$ is concave up on I then we need to show that for any x, $x \neq a$, in I that,

$$f(x) > f(a) + f'(a)(x-a) ,$$

or in other words, the tangent line is always below the graph of $f(x)$ on I. Note that we require $x \neq a$ because at that point we know that $f(x) = f(a)$ since we are talking about the tangent line.

Let us start the proof off by first assuming that $x > a$. Using the mean value theorem on $[a, x]$ means there is a number c such that $a < c < x$ and,

$$f(x) - f(a) = f'(c)(x-a) ,$$

or equivalently

$$f(x) = f(a) + f'(c)(x-a) . \tag{10.1}$$

Next, let us use the fact that $f''(x) > 0$ for every x on I. This means that the first derivative, $f'(x)$, must be increasing. Now, we know from the mean value theorem that $a < c$ and so because $f'(x)$ is increasing we must have,

$$f'(a) < f'(c) . \tag{10.2}$$

Recall as well that we are assuming x>a and so $x - a > 0$. If we now multiply Inequality (10.2) by $x - a$ (which is positive and so the inequality stays the same) we get,

$$f'(a)(x-a) < f'(c)(x-a) .$$

However, by Equation (10.1), the right side of this is nothing more than $f(x) - f(a)$ and so we have,

$$f(a) + f'(a)(x-a) < f(x) ,$$

but this is exactly what we wanted to show. So, provided $x > a$ the tangent line is in fact below the graph of $f(x)$.

We now need to assume $x < a$. Using the mean value theorem on $[x, a]$ means there is a number c such that $x < c < a$ and

$$f(a) - f(x) = f'(c)(a-x) .$$

If we multiply both sides of this by -1 and then adding $f(a)$ to both sides and we again arrive at Equation (10.1).

Now, from the mean value theorem we know that $c < a$ and because $f''(x) > 0$ for every x on I we know that the derivative is still increasing and so we have, $f'(c) < f'(a)$. Let us now multiply this by $x - a$, which is now a negative number since $x < a$. This gives

$$f'(c)(x-a) > f'(a)(x-a) .$$

Notice that we had to switch the direction of the inequality since we were multiplying by a negative number. If we now add $f(a)$ to both sides of this and then substitute Equation (10.1) into the results we arrive at,

$$\begin{aligned} f(a)+f'(c)(x-a) &> f(a)+f'(a)(x-a) \\ \Leftrightarrow \qquad f(x) &> f(a)+f'(a)(x-a) \end{aligned}$$

So, again we have shown that the tangent line is always below the graph of $f(x)$. We have now shown that if x is any number in I, with $x \neq a$ the tangent lines are always below the graph of $f(x)$ on I and so $f(x)$ is concave up on I.

□

Figure 10.14 demonstrates the four ways that concavity interacts with increasing/decreasing, along with the relationships with the first and second derivatives.

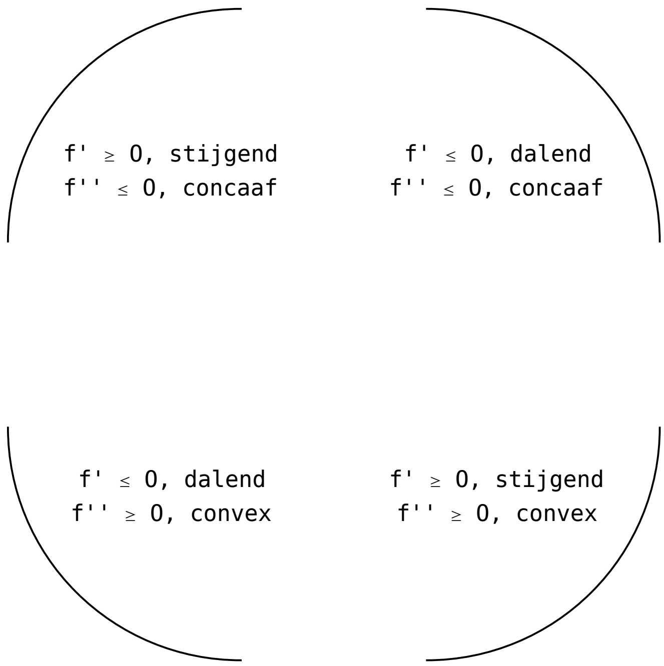


Figure 10.14: Demonstrating the 4 ways that concavity interacts with increasing/decreasing, along with the relationships with the first and second derivatives.

If knowing where a graph is concave up/down is important, it makes sense that the places where the graph changes from one to the other is also important. This leads us to a definition.

Definitie 10.7 (Point of inflection)
A **point of inflection** (*buigpunt*) is a point on the graph of f at which the concavity of f changes.

If the concavity of f changes at a point $(c, f(c))$, then f' is changing from increasing to decreasing (or, decreasing to increasing) at $x = c$. That means that the sign of f'' is changing from positive to negative (or, negative to positive) at $x = c$. This leads to the following theorem.

Theorem 10.9 (Points of inflection)
If $(c, f(c))$ is a point of inflection on the graph of f, then either $f''(c) = 0$ or f'' is not defined at c.

We have identified the concepts of concavity and points of inflection. It is now time to practice using these concepts; given a function, we should be able to find its points of inflection and identify intervals on which it is concave up or down. We do so in the following example.

Example 10.8

Let

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

Find the inflection points of f and the intervals on which it is concave up/down.

Solution

We need to find f' and f''. Using the quotient rule and simplifying, we find

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

To find the possible points of inflection, we seek to find where $f''(x) = 0$ and where f'' is not defined. Solving $f''(x) = 0$ reduces to solving $2x(x^2+3) = 0$; we find $x = 0$. We find that f'' is not defined when $x = \pm 1$, for then the denominator of f'' is 0. We also note that f itself is not defined at $x = \pm 1$, having a domain of $]-\infty,-1[\,\cup\,]-1,1[\,\cup\,]1,+\infty[$. Since the domain of f is the union of three intervals, it makes sense that the concavity of f could switch across intervals. We technically cannot say that f has a point of inflection at $x = \pm 1$ as they are not part of the domain, but we must still consider these x-values to be important and will include them in our number line.

The important x-values at which concavity might switch are $x = -1$, $x = 0$ and $x = 1$, which split the number line into four intervals.

We determine the concavity on each. Keep in mind that all we are concerned with is the sign of f'' on the interval.

Interval 1: $]-\infty,-1[$

Select a number c in this interval with a large magnitude (for instance, $c = -100$). The denominator of $f''(x)$ will be positive. In the numerator, the (c^2+3) will be positive and the $2c$ term will be negative. Thus the numerator is negative and $f''(c)$ is negative. We conclude f is concave down on $]-\infty,-1[$.

Interval 2: $]-1,0[$

For any number c in this interval, the term $2c$ in the numerator will be negative, the term (c^2+3) in the numerator will be positive, and the term $(c^2-1)^3$ in the denominator will be negative. Thus $f''(c) > 0$ and f is concave up on this interval.

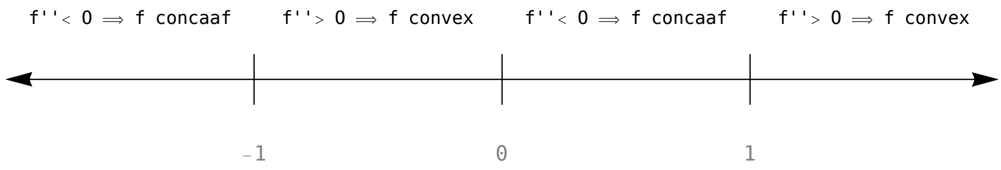
Interval 3: $]0,1[$

Any number c in this interval will be positive and small. Thus the numerator is positive while the denominator is negative. Thus $f''(c) < 0$ and f is concave down on this interval.

Interval 4: $]1,+\infty[$

Choose a large value for c. It is evident that $f''(c) > 0$, so we conclude that f is concave up on $]1,+\infty[$.

Since, we get



we conclude that f is concave up on $]-1,0[$ and $]1,+\infty[$ and concave down on $]-\infty,-1[$ and $]0,1[$. There is only one point of inflection, $]0,0[$, as f is not defined at $x=\pm1$. Our work is confirmed by the graph of f in Figure 10.15.

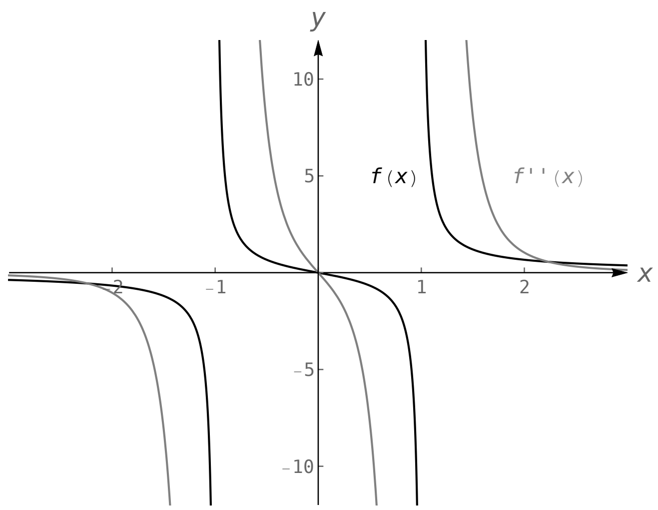


Figure 10.15: A graph of $f(x)$ (black) and $f''(x)$ (gray) in Example 10.8.

Recall that relative maxima and minima of f are found at critical points of f; that is, they are found when $f'(x)=0$ or when f' is undefined. Likewise, the relative maxima and minima of f' are found when $f''(x)=0$ or when f'' is undefined; note that these are the inflection points of f.

What does a relative maximum of f' mean? The derivative measures the rate of change of f; maximizing f' means finding where f is increasing the most – where f has the steepest tangent line. A similar statement can be made for minimizing f'; it corresponds to where f has the steepest negatively–sloped tangent line.

We utilize this concept in the next example.

Example 10.9

The sales of a certain product over a three-year span are modelled by $S(t)=t^4-8t^2+20$, where t is the time in years. Over the first two years, sales are decreasing. Find the point at which sales are decreasing at their greatest rate.

Solution

We want to maximize the rate of decrease, which is to say, we want to find where S' has a minimum. To do this, we find where S'' is 0. We find $S'(t)=4t^3-16t$ and $S''(t)=12t^2-16$. Setting $S''(t)=0$ and solving, we get $t=\sqrt{4/3}\approx1.16$. Note that we ignore the negative value of t since it does not lie in the domain of our function S.

This is both the inflection point and the point of maximum decrease. This is the point at which things first start looking up for the company. After the inflection point, it will still take some time before sales start to increase, but at least sales are not decreasing quite as quickly as they had been.

A graph of $S(t)$ and $S'(t)$ is given in Figure 10.16. When $S'(t) < 0$, sales are decreasing; note how at $t \approx 1.16$, $S'(t)$ is minimized. That is, sales are decreasing at the fastest rate at $t \approx 1.16$. On the interval of $(1.16, 2)$, S is decreasing but concave up, so the decline in sales is levelling off.

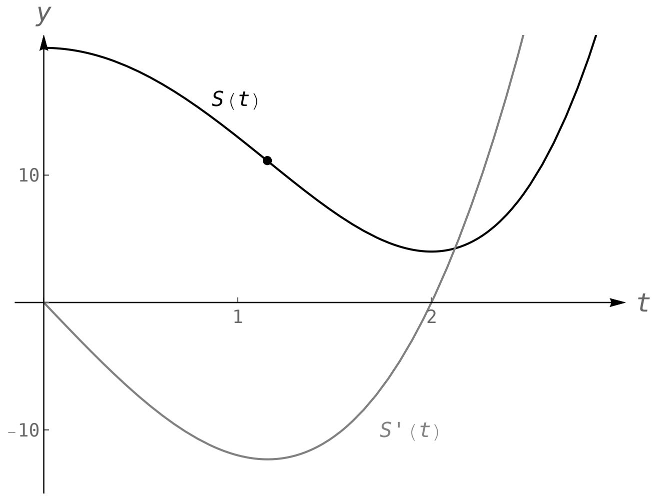


Figure 10.16: A graph of $S(t)$ (black) in Example 10.9 along with $S'(t)$ (gray).

Not every critical point corresponds to a relative extrema; $f(x) = x^3$ has a critical point at $(0, 0)$ but no relative maximum or minimum (Figure 10.3). Likewise, just because $f''(x) = 0$ we cannot conclude concavity changes at that point. We were careful before to use terminology possible point of inflection since we needed to check to see if the concavity changed. The canonical example of $f''(x) = 0$ without concavity changing is $f(x) = x^4$. At $x = 0$, $f''(x) = 0$ but f is always concave up, as shown in Figure 10.17.

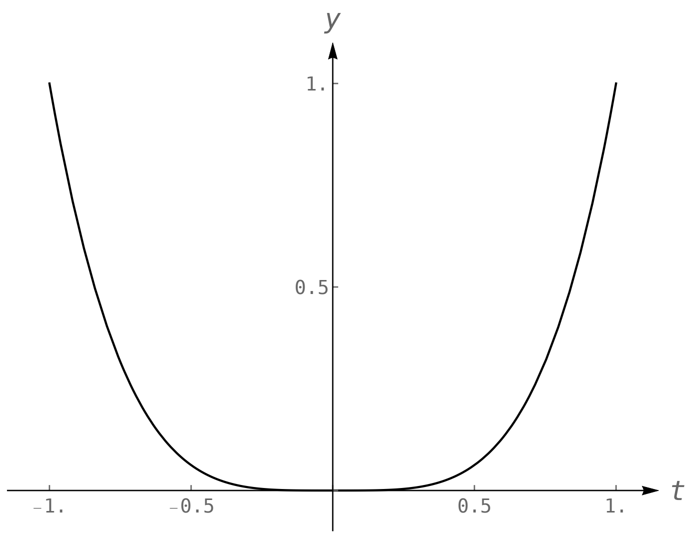


Figure 10.17: A graph of $f(x) = x^4$.

10.4.2 The second derivative test

The first derivative of a function gave us a test to find if a critical value corresponded to a relative maximum, minimum, or neither. The second derivative gives us another way to test if a critical point is a local maximum or minimum. The following theorem states something that is intuitive: if a critical value occurs in a region where a function f is concave up, then that critical value must correspond to a relative minimum of f, etc (Figure 10.18).

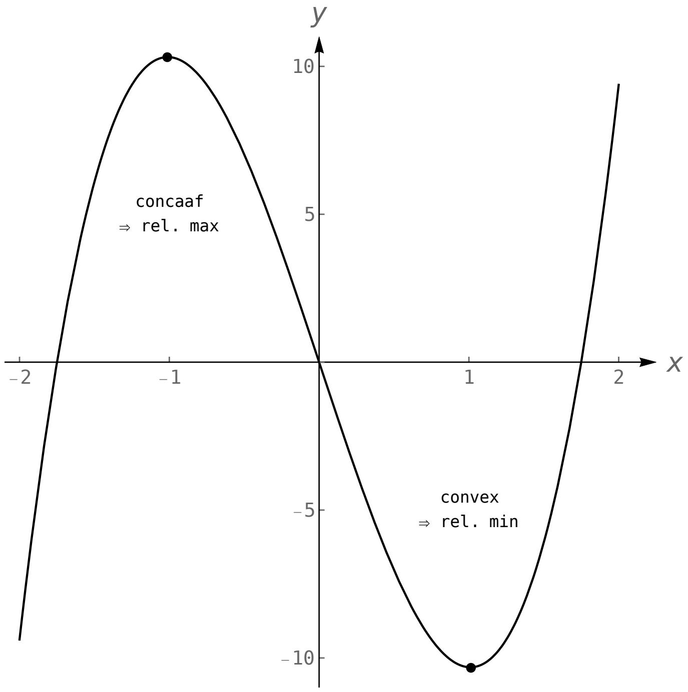


Figure 10.18: Demonstrating the second derivative test.

Theorem 10.10 (The second derivative test)
Let c be a critical value of f where $f''(c)$ is defined.

1. *If $f''(c) < 0$, then f has a local maximum at $(c, f(c))$.*
2. *If $f''(c) > 0$, then f has a local minimum at $(c, f(c))$.*
3. *If $f''(c) = 0$ then $x = c$ can be a local maximum, relative minimum or neither.*

Proof First let us assume that $f''(x)$ is continuous in a region around $x = c$, so that we can assume that in fact $f''(c) < 0$ is also true in some open region, say $]a, b[$ around $x = c$, i.e. $a < c < b$.

Now let x be any number such that $a < x < c$, we are going to use the mean value theorem on $[x, c]$. However, instead of using it on the function itself we are going to use it on the first derivative. So, the mean value theorem tells us that there is a number d for which $x < d < c$ such that,

$$f'(c) - f'(x) = f''(d)(c - x) .$$

Now, because $a < x < d < c$ we know that $f''(d) < 0$ and we also know that $c - x > 0$, so we then get that $f'(c) - f'(x) < 0$. However, we also assumed that $f'(c) = 0$ and so we have that $f'(x) > 0$. Or, in other words to the left of $x = c$ the function is increasing.

Let us now turn things around and let x be any number such that $c < x < b$ and use the mean value theorem on $[c, x]$ and the first derivative. The mean value theorem tells us that there is a number $c < d < x$ such that,

$$f'(x) - f'(c) = f''(d)(x - c) .$$

Now, because $c < d < x < b$ we know that $f''(d) < 0$ and we also know that $x - c > 0$ so we then get that $f'(x) - f'(c) < 0$. Again, use the fact that we also assumed that $f'(c) = 0$ to get $f'(x) < 0$.

Consequently, we now know that to the right of $x = c$ the function is decreasing.

So, to the left of $x = c$ the function is increasing and to the right of $x = c$ the function is decreasing so by the first derivative test this means that $x = c$ must be a relative maximum. □

The second derivative test relates to the first derivative test in the following way. If $f''(c) > 0$, then the graph is concave up at a critical point c and f' itself is growing. Since $f'(c) = 0$ and f' is growing at c, then it must go from negative to positive at c. This means the function goes from decreasing to increasing, indicating a local minimum at c.

Example 10.10

Let

$$f(x) = \frac{100}{x} + x.$$

Find the critical points of f and label them as relative maxima or minima.

Solution

We find

$$f'(x) = -\frac{100}{x^2} + 1$$

and

$$f''(x) = \frac{200}{x^3}.$$

We set $f'(x) = 0$ and solve for x to find the critical values. Note that f' is not defined at $x = 0$, but neither is f so this is not a critical value. We find the critical values are $x = \pm 10$. Evaluating f'' at $x = 10$ gives $0.2 > 0$, so there is a local minimum at $x = 10$. Evaluating $f''(-10) = -0.2 < 0$, determining a relative maximum at $x = -10$. These results are confirmed in Figure 10.19.

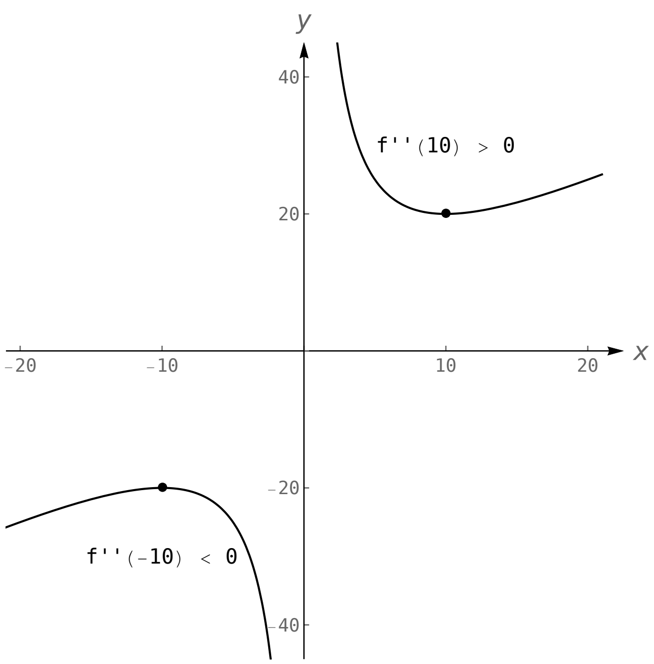


Figure 10.19: A graph of $f(x) = 100/x + x$ in Example 10.10.

We have been learning how the first and second derivatives of a function relate information about the graph of that function. We have found intervals of increasing and decreasing, intervals where the graph is concave up and down, along with the locations of relative extrema and inflection points. In

Chapter 8 we saw how limits explained asymptotic behaviour. In the next section we combine all of this information to produce accurate sketches of functions.

10.5 Curve sketching

We have been learning how we can understand the behaviour of a function based on its first and second derivatives. While we have been treating the properties of a function separately, we combine them here to produce an accurate graph of the function without plotting lots of extraneous points. Why bother? Graphing utilities are very accessible, whether on a computer, a hand–held calculator, or a smartphone. These resources are usually very fast and accurate. For instance, remember that we can plot an explicitly defined function in Mathematica using the built-in command **`Plot`**, while **`ContourPlot`** may be used to plot implicitly defined functions (see Chapter 3). We will see that our method is not particularly fast – it will require time. We are attempting to understand the behavior of a function f based on the information given by its derivatives. While all of a function's derivatives relay information about it, it turns out that most of the behaviour we care about is explained by f' and f''. Understanding the interactions between the graph of f and f' and f'' is important. To gain this understanding, one might argue that all that is needed is to look at lots of graphs. This is true to a point, but is somewhat similar to stating that one understands how an engine works after looking only at pictures. It is true that the basic ideas will be conveyed, but hands–on access increases understanding.

To produce an accurate sketch of a given function f, take the following steps.

1. Find the domain of f. Generally, we assume that the domain is the entire real line then find restrictions, such as where a denominator is 0 or where negatives appear under the radical.
2. Find symmetries and intercepts.
3. Find the location of any asymptotes of f:
   (a) vertical
   (b) horizontal
   (c) slant asymptotes
4. Find the critical and singular points of f.
5. Find the possible points of inflection of f.
6. Create a number line that includes all critical points, possible points of inflection, and locations of vertical asymptotes. For each interval created, determine whether f is increasing or decreasing, concave up or down.
7. Evaluate f at each critical point and possible point of inflection. Plot these points on a set of axes. Connect these points with curves exhibiting the proper concavity. Sketch asymptotes and x- and y-intercepts where applicable.

Example 10.11

Sketch $f(x) = 3x^3 - 10x^2 + 7x + 5$.

Solution

We follow the steps outlined above.

1. The domain of f is the entire real line; there are no values x for which $f(x)$ is not defined.

2. It can be verified easily that the function is neither even nor odd. Besides, the x-intercept is about $x = -0.424$ while the y-intercept is $y = 5$.

3. (a) There are no vertical asymptotes.

   (b) We determine the end behaviour using limits as x approaches $\pm\infty$.

$$\lim_{x\to-\infty} f(x) = -\infty \qquad \text{and} \qquad \lim_{x\to+\infty} f(x) = +\infty$$

   So, we do not have any horizontal asymptotes.

   (c) There is no slant asymptote since

$$a = \lim_{x\to\pm\infty} \frac{f(x)}{x} = \lim_{x\to\pm\infty} \left(3x^2 - 10x + 7 + \frac{5}{x}\right) = +\infty.$$

4. Find the critical values of f. We compute $f'(x) = 9x^2 - 20x + 7$. Use the quadratic formula to find the roots of f':

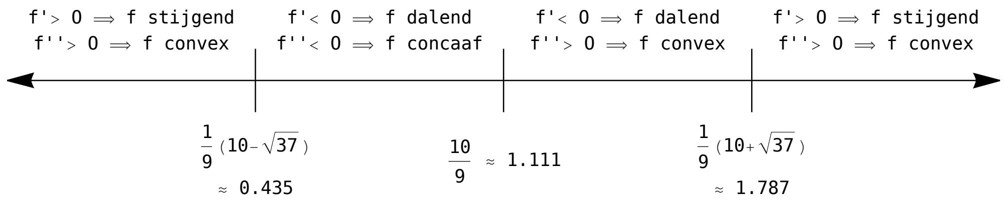
$$x = \frac{20 \pm \sqrt{(-20)^2 - 4(9)(7)}}{2(9)} = \frac{1}{9}\left(10 \pm \sqrt{37}\right),$$

   so we have $x \approx 0.435$ or $x \approx 1.787$.

5. Find the possible points of inflection of f. Compute $f''(x) = 18x - 20$. We have

$$f''(x) = 0 \quad \Rightarrow \quad x = \frac{10}{9} \approx 1.111.$$

6. We place the values $x = (10 \pm \sqrt{37})/9$ and $x = 10/9$ on a number line and we mark each interval as increasing or decreasing, concave up or down:



7. We now plot the appropriate points and connect the points in such a way that the proper concavity is demonstrated. Our curve crosses the y-axis at $y = 5$ and crosses the x-axis near $x = -0.424$ (Figure 10.20).

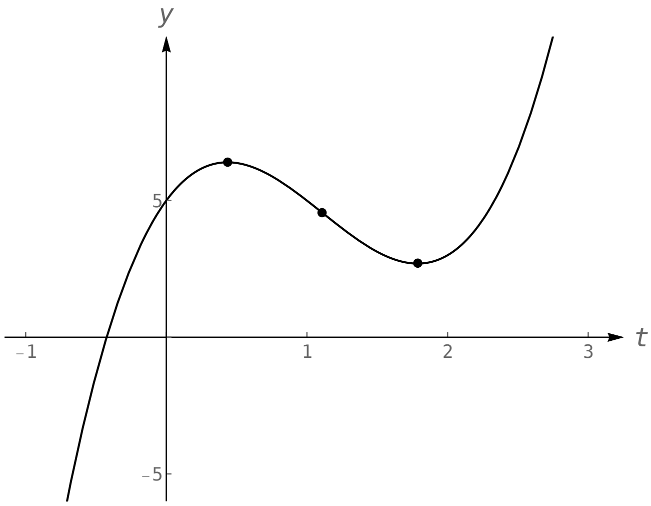


Figure 10.20: A sketch of $f(x) = 3x^3 - 10x^2 + 7x + 5$ in Example 10.11.

Example 10.12

Sketch

$$f(x) = \frac{x^2 - x - 2}{x^2 - x - 6}.$$

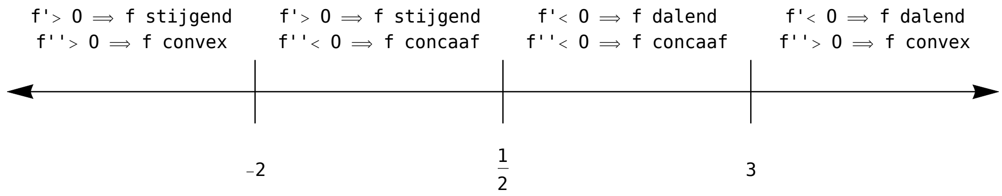
Solution

We again follow the steps outlined above.

1. In determining the domain, we assume it is all real numbers and look for restrictions. We find that at $x = -2$ and $x = 3$, $f(x)$ is not defined because the denominator of $f(x)$ is 0 at those points. So,
$$\operatorname{dom} f = \{\text{real numbers } x \mid x \neq -2, 3\}.$$
2. It can be verified easily that the function is neither even nor odd. Besides, the x-intercepts are $x = -1$ and $x = 2$ while the y-intercept is $y = 1/3$.
3. (a) The vertical asymptotes of f are at $x = -2$ and $x = 3$, the places where f is undefined and the numerator of $f(x)$ is not zero.
   (b) There is a horizontal asymptote of $y = 1$, as $\lim_{x\to-\infty} f(x) = 1$ and $\lim_{x\to+\infty} f(x) = 1$.
   (c) There are no slant asymptotes because there are already horizontal ones.
4. To find the critical values of f, we first find $f'(x)$. Using the quotient rule, we find
$$f'(x) = \frac{-8x+4}{(x^2+x-6)^2} = \frac{-8x+4}{(x-3)^2(x+2)^2}.$$
$f'(x) = 0$ when $x = 1/2$, and f' is undefined when $x = -2$ or $x = 3$. Since f' is undefined only when f is, these are not singular values. The only critical value is $x = 1/2$.
5. To find the possible points of inflection, we find $f''(x)$, again employing the Quotient Rule:
$$f''(x) = \frac{24x^2 - 24x + 56}{(x-3)^3(x+2)^3}.$$

We find that $f''(x)$ is never 0 (setting the numerator equal to 0 and solving for x, we find the only roots to this quadratic are complex) and f'' is undefined when $x = -2$ or $x = 3$. Thus concavity will possibly only change at $x = -2$ and $x = 3$.

6. We place the values $x = 1/2$, $x = -2$ and $x = 3$ on a number line and we mark in each interval whether f is increasing or decreasing, concave up or down:



We see that f has a relative maximum at $x = 1/2$; concavity changes only at the vertical asymptotes.

7. In Figure 10.21, we plot the points from the number line on a set of axes and connect them in such a way that we get the appropriate concavity. We also show f crossing the x-axis at $x = -1$ and $x = 2$.

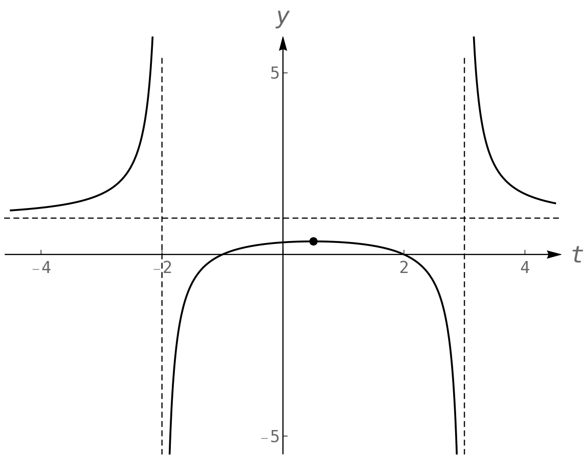


Figure 10.21: A sketch of $f(x) = \frac{x^2-x-2}{x^2-x-6}$ in Example 10.12.

Now why are computer graphics so good at curve sketching? It is not because computers are smarter than we are. Rather, it is largely because computers are much faster at computing than we are. In general, computers graph functions plot equally spaced points, then connect the dots using lines. By using lots of points, the connecting lines are short and the graph looks smooth. This does a fine job of graphing in most cases. However, in regions where the graph is very curvy, this can generate noticeable sharp edges on the graph unless a large number of points are used. High quality computer algebra systems, such as Mathematica, use special algorithms to plot lots of points only where the graph is curvy.

In Figure 10.22, a graph of $y = \sin(x)$ is given, generated by Mathematica using the Mathematica-function **`Plot`**. The small points represent each of the places Mathematica sampled the function. Notice how at the bends of $\sin(x)$, lots of points are used; where $\sin(x)$ is relatively straight, fewer points are used. Moreover, many points are also used at the endpoints to ensure the end behavior is accurate.

How does Mathematica know where the graph is curvy? Calculus. When we study curvature in a later chapter, we will see how the first and second derivatives of a function work together to provide a measurement of curviness. Mathematica employs algorithms to determine regions of high curvature

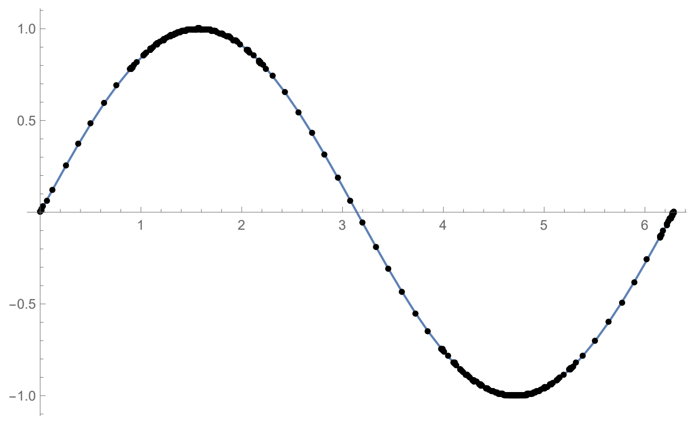


Figure 10.22: A graph of $y = \sin(x)$ generated by Mathematica.

and plots extra points there. Again, the goal of this section is to understand that the shape of the graph of a function is largely determined by understanding the behaviour of the function at a few key places. For instance, in Example 10.12, we were able to accurately sketch a complicated graph using only a few points and knowledge of asymptotes!

Computer algebra systems

A computer algebra system is any mathematical software with the ability to manipulate mathematical expressions in a way similar to the traditional manual computations of mathematicians and scientists. The development of such systems started in the second half of the previous century.

The first popular computer algebra systems were muMATH, Reduce, Derive, and Macsyma. Today, the most popular commercial systems are Mathematica and Maple, which are commonly used by research mathematicians, scientists, and engineers. Freely available alternatives include SageMath[a] and SymPy.

[a] http://www.sagemath.org/

In the next chapters, we will consider the reverse problem to computing the derivative: given a function f, can we find a function whose derivative is f? Being able to do so opens up an incredible world of mathematics and applications.

10.6 Exercises

Extreme values

Assignment 10.1 — Find the extremum of the function $y = x^{(x^2)}$. Prove that it is a minimum.

Assignment 10.2 — The Maxwell-Bolzmann distribution describes the distribution of velocities of gas molecules in an ideal gas. The probability that a molecule with mass m in a gas at temperature T, has velocity v is

$$f(v) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} v^2\, e^{-\frac{mv^2}{2kT}},$$

with k a constant. Determine the velocity v for which $f(v)$ is maximal.

The mean value theorem

Assignment 10.3 — For the functions listed below, verify whether the mean value theorem (theorem 10.4) can be applied to the given interval. Determine, if possible, a $c \in]a,b[$ that is guaranteed by the theorem.

(a) $f(x) = x^2 + 3x - 1, \qquad [-2,2]$

(b) $f(x) = \sqrt{9 - x^2}, \qquad [0,3]$

(c) $f(x) = \dfrac{x^2-9}{x^2-1}, \qquad [0,2]$

Assignment 10.4 — For the functions listed below, verify that Rolle's theorem (stelling 10.5) can be applied to the given interval. Determine, if possible, a $c \in]a,b[$ such that $f'(c) = 0$.

(a) $f(x) = x^2 + x - 6, \qquad [-3,2]$

(b) $f(x) = x^2 + x, \qquad [-2,2]$

(c) $f(x) = \cos(x), \qquad [0,\pi]$

Curve sketching

Assignment 10.5 —

(a) Examine the graph of the function $y = xe^{-kx^2}$, with $k \in \mathbb{R}$. To do this, determine its domain, zeros, symmetries, asymptotes, extrema and inflection points. When determining the asymptotes, discuss the different values of k ($k < 0$, $k = 0$, $k > 0$).

(b) Determine the value of the parameter k such that a maximum occurs at $x = 1$.

(c) Make a sign table of the function and a sketch of the graph for the value of k.

Assignment 10.6 — Sketch the graph of the following functions.

(a) $f(x) = \sqrt{x^2 - 4x + 3}$

(b) $f(x) = \dfrac{x+7}{\sqrt{x^2-3}}$

(c) $f(x) = e^{-\frac{x^2}{2}}$

(d) $f(x) = \dfrac{e^{-x}}{x^3}$

(e) $f(x) = \ln(2^x - 1)$

(f) $f(x) = \dfrac{\ln(x)}{x^2}$

(g) $f(x) = \ln\left(\sqrt{e^x + e^{-x}}\right)$

(h) $f(x) = \dfrac{2\ln(x)}{1-\ln(x)}$

(i) $f(x) = \ln\big(\cos(x)\big)$

(j) $f(x) = \arctan(\ln(x))$

(k) $f(x) = \arctan\left(\dfrac{1}{x}\right)$

(l) $f(x) = x^x$

(m) $f(x) = (x^2)^x$

(n) $f(x) = x - 2\sin(x)$

(o) $f(x) = e^{-x}\sin(x), \quad (x \geq 0)$

(p) $f(x) = x + \sin(x)$

(q) $f(x) = \dfrac{|1+x|-1}{x}$

(r) $f(x) = \left|2 - \sqrt{2x+4}\right|$

(s) $f(x) = \dfrac{x^2}{x|x|+1}$

(t) $f(x) = \left|(x-2)^2 - 4\right|$

(u) $f(x) = \sinh(x) - x$

(v) $f(x) = e^x \sinh(x)$

(w) $f(x) = \coth(x) + x$

(x) $f(x) = \operatorname{arcosh}(\sqrt{x-2})$

(y) $f(x) = \operatorname{artanh}\left(\dfrac{4}{x}\right)$

Review exercises

Assignment 10.7 — Figure 10.23 shows the graph of a function f, its derivative functions f' and f'' and a function g. Which graph corresponds to which function?

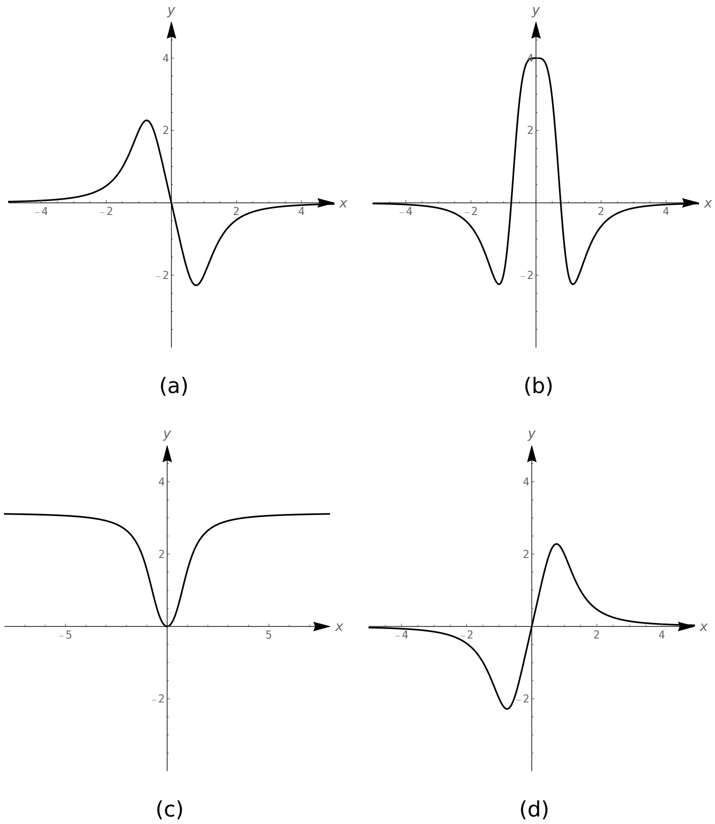
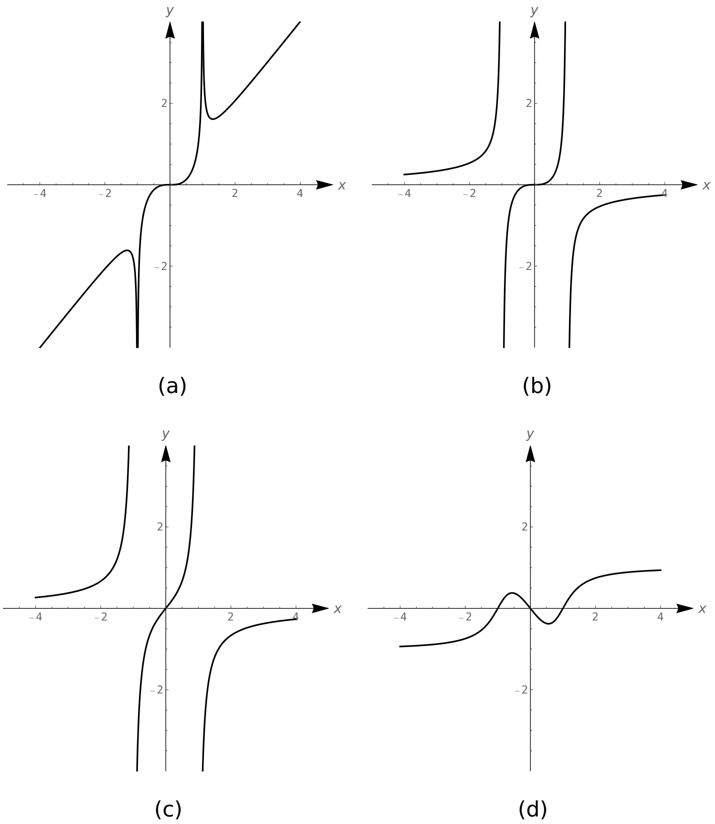


Figure 10.23: Figure belonging to Exercise 10.7

Assignment 10.8 — Figure 10.24 shows the graphs of four functions:

$$f(x) = \frac{x}{1-x^2}, \qquad g(x) = \frac{x^3}{1-x^4}, \qquad h(x) = \frac{x^3-x}{\sqrt{x^6+1}} \qquad \text{and} \qquad k(x) = \frac{x^3}{\sqrt{|x^4-1|}}.$$

Which graph corresponds to which function?



(a) (b) (c) (d)

Figure 10.24: Figure corresponding to exercise 10.8.

Assignment 10.9 — For the functions below, if possible, determine the local maxima and/or minima and any inflection points. Also determine the intervals where the function is convex or concave.

(a) $f(x) = 2x^3 - 3x^2 + 9x + 5$

(b) $f(x) = \dfrac{1}{x^2+1}$

(c) $f(x) = \sin(x) + \cos(x), \qquad x \in]-\pi, \pi[$

(d) $f(x) = x^2 \ln(x)$

Assignment 10.10 — Describe the critical points and possible inflection points of the function f as a function of a and b.

(a) $f(x) = \dfrac{a}{x^2+b^2}$

(b) $f(x) = \sin(ax+b)$

At its heart, engineering is about using science to find creative, practical solutions. It is a noble profession.

— Queen Elizabeth II —

11

Polar coordinates and parametric equations

11.1 Polar coordinates

11.1.1 Definition

In Section 3.1, we introduced the Cartesian coordinates of a point in the plane as a means of assigning ordered pairs of numbers to points in the plane. We defined the Cartesian coordinate plane using two number lines – one horizontal and one vertical – which intersect at right angles at a point we called the origin. For this reason, the Cartesian coordinates of a point are often called **rectangular coordinates** (*rechthoekige coördinaten*) . In this section, we introduce a new system for assigning coordinates to points in the plane – **polar coordinates** (*poolcoördinaten*). We start with an origin point, called the **pole** (*pool*), and a ray called the **polar axis** (*poolas*). We then locate a point P using two coordinates, (r,θ), where r represents a directed distance from the pole and θ is a measure of rotation from the polar axis (Figure 11.1). Roughly speaking, the polar coordinates (r,θ) of a point measure how far out the point is from the pole (that is r), and how far to rotate from the polar axis, (that is θ).

For example, if we wish to plot the point P with polar coordinates $\left(4,\frac{5\pi}{6}\right)$, we would start at the pole, move out along the polar axis 4 units, then rotate $\frac{5\pi}{6}$ radians counter-clockwise. We may also consider this process by thinking of the rotation first. To plot $P\left(4,\frac{5\pi}{6}\right)$ this way, we rotate $\frac{5\pi}{6}$ counter-clockwise from the polar axis, then move outwards from the pole 4 units.

If $r<0$, we begin by moving in the opposite direction on the polar axis from the pole and as you may have guessed, $\theta<0$ means the rotation away from the polar axis is clockwise instead of counter-clockwise. Furthermore, it should come as no surprise that any given point expressed in polar coordinates has infinitely many other representations in polar coordinates. More formally, suppose (r,θ) and $\left(\tilde{r},\tilde{\theta}\right)$ are polar coordinates where $r\neq 0$, $\tilde{r}\neq 0$ and the angles are measured in radians. Then (r,θ) and $\left(\tilde{r},\tilde{\theta}\right)$ determine the same point P if and only if one of the following is true:

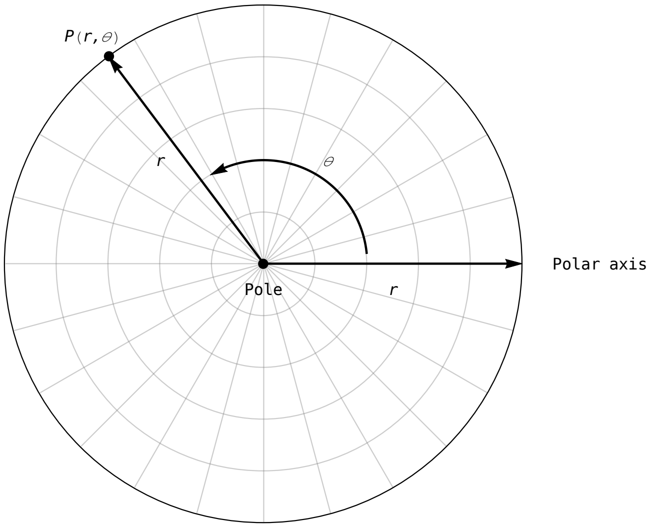


Figure 11.1: Polar coordinate system.

- $\tilde{r} = r$ and $\tilde{\theta} = \theta + 2\pi k$ for some integer k,
- $\tilde{r} = -r$ and $\tilde{\theta} = \theta + (2k+1)\pi$ for some integer k.

Moreover, all polar coordinates of the form $(0, \theta)$ represent the pole regardless of the value of θ.

Polar coordinates in aviation

Aircraft use a slightly modified version of the polar coordinates for navigation. In this system, the one generally used for any sort of navigation, the zero-degree ray is generally called heading 360, and the angles continue in a clockwise direction, rather than counterclockwise, as in the mathematical system. Heading 360 corresponds to magnetic north, while headings 90, 180, and 270 correspond to magnetic east, south, and west, respectively. Thus, an aircraft traveling 5 nautical miles due east will be traveling 5 units at heading 90.

11.1.2 Linking polar and rectangular coordinates

To marry the polar coordinate system with the Cartesian (rectangular) coordinate system we identify the pole and polar axis in the polar system with the origin and positive x-axis, respectively, in the rectangular system. We get the following result.

Theorem 11.1 (Conversion between rectangular and polar coordinates)
Suppose a point P is represented in rectangular coordinates as (x, y) and in polar coordinates as (r, θ). Then

- $x = r\cos(\theta)$ *and* $y = r\sin(\theta)$;
- $x^2 + y^2 = r^2$ *and* $\tan(\theta) = \dfrac{y}{x}$ *(provided* $x \neq 0$*).*

In the case $r > 0$, Theorem 11.1 is an immediate consequence of Theorem 5.5 along with the definition of the tangent. If $r < 0$, then we know an alternate representation for (r, θ) is $(-r, \theta + \pi)$. Since we have that $\cos(\theta+\pi) = -\cos(\theta)$ and $\sin(\theta+\pi) = -\sin(\theta)$, applying the theorem to $(-r, \theta+\pi)$ gives

$$\begin{cases} x &= (-r)\cos(\theta+\pi) = (-r)(-\cos(\theta)) = r\cos(\theta) \\ y &= (-r)\sin(\theta+\pi) = (-r)(-\sin(\theta)) = r\sin(\theta). \end{cases}$$

Moreover, $x^2+y^2 = (-r)^2 = r^2$, and $\frac{y}{x} = \tan(\theta+\pi) = \tan(\theta)$, so the theorem is true in this case, too. The remaining case is $r = 0$, in which case $(r,\theta) = (0,\theta)$ is the pole. Since the pole is identified with the origin $(0,0)$ in rectangular coordinates, the theorem in this case amounts to checking $0 = 0$. The following example puts Theorem 11.1 to good use.

Example 11.1

Convert each point in rectangular coordinates given below into polar coordinates with $r \geq 0$ and $0 \leq \theta < 2\pi$.

1. $P\left(2, -2\sqrt{3}\right)$
2. $R(0,-3)$

Solution

1. The point $P\left(2,-2\sqrt{3}\right)$ lies in Quadrant IV. With $x = 2$ and $y = -2\sqrt{3}$, we get

$$r^2 = x^2 + y^2 = (2)^2 + \left(-2\sqrt{3}\right)^2 = 4 + 12 = 16\,,$$

so $r = \pm 4$. Since we are asked for $r \geq 0$, we choose $r = 4$. To find θ, we have that

$$\tan(\theta) = \frac{y}{x} = \frac{-2\sqrt{3}}{2} = -\sqrt{3}\,.$$

This tells us θ has a reference angle of $-\frac{\pi}{3}$, and since P lies in Quadrant IV, we know θ is a Quadrant IV angle. We are asked to have $0 \leq \theta < 2\pi$, so we choose $\theta = \frac{5\pi}{3}$. Hence, our answer is $\left(4, \frac{5\pi}{3}\right)$ (Figure 11.2(a)).

2. The point $Q(0,-3)$ lies along the negative y-axis. While we could go through the usual computations to find the polar form of R, in this case we can find the polar coordinates of Q using the definition. Since the pole is identified with the origin, we can easily tell the point Q is 3 units from the pole, which means in the polar representation (r,θ) of Q we know $r = \pm 3$. Since we require $r \geq 0$, we choose $r = 3$. Concerning θ, the angle $\theta = \frac{3\pi}{2}$ satisfies $0 \leq \theta < 2\pi$ with its terminal side along the negative y-axis, so our answer is $\left(3, \frac{3\pi}{2}\right)$ (Figure 11.2(b)).

From the previous example, it is clear that it is important to know in which quadrant the point under investigation lies in order to infer the corresponding θ. Instead of using the arctan-function for that purpose and then figure out the correct angle, it is often more useful to use the atan2-function (2-argument tangent). The atan2 is defined as the angle in the Euclidean plane, given in radians, between the positive x-axis and the ray to the point $(x,y) \neq (0,0)$. The angles are signed, with counter-clockwise angles being positive, and clockwise ones being negative. In other words, $\operatorname{atan2}(y,x)$ is in the interval

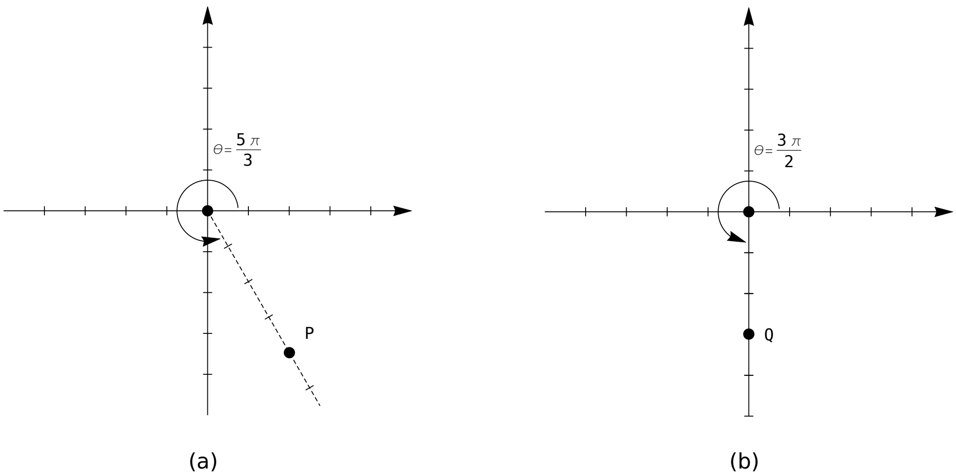


Figure 11.2: The location of the point P with rectangular coordinates $\left(2,-2\sqrt{3}\right)$ and polar coordinates $\left(4,\frac{5\pi}{3}\right)$ (a) and the point Q with rectangular coordinates $(0,-3)$ and polar coordinates $\left(3\sqrt{2},\frac{5\pi}{4}\right)$ (b).

$[0,\pi]$ when $y>0$ and in $]-\pi,0[$ when $y<0$. The function is defined as:

$$\operatorname{atan2}(y,x)=\begin{cases}\arctan\left(\frac{y}{x}\right) & \text{, if } x>0\\ \arctan\left(\frac{y}{x}\right)+\pi & \text{, if } x<0\wedge y\geq 0\\ \arctan\left(\frac{y}{x}\right)-\pi & \text{, if } x<0\wedge y<0\\ \frac{\pi}{2} & \text{, if } x=0\wedge y>0\\ -\frac{\pi}{2} & \text{, if } x=0\wedge y<0\\ \text{undefined} & \text{, if } x=0\wedge y=0\end{cases} \tag{11.1}$$

Of course, we do not have to restrict to points when converting from the rectangular to the polar coordinate system. We can do the same with equations using Theorem 11.1. In polar coordinates, we will end up with equations in the variables r and θ. The obvious strategy to convert an equation from rectangular to polar coordinates is to replace every occurrence of x with $r\cos(\theta)$ and every occurrence of y with $r\sin(\theta)$ and use identities to simplify. On the other hand, converting equations from polar to rectangular coordinates is not as straightforward. We could solve $r^2=x^2+y^2$ for r to get $r=\pm\sqrt{x^2+y^2}$ and solving $\tan(\theta)=\frac{y}{x}$ requires the arctangent function to get $\theta=\arctan\left(\frac{y}{x}\right)+\pi k$ for integers k. Still, since neither of these expressions for r and θ are especially user-friendly, we might resort to a second strategy involving the rearrangement of the given polar equation so that the expressions $r^2=x^2+y^2$, $r\cos(\theta)=x$, $r\sin(\theta)=y$ and/or $\tan(\theta)=\frac{y}{x}$ present themselves.

Example 11.2

1. Convert each equation in rectangular coordinates into an equation in polar coordinates.

   (a) $(x-3)^2+y^2=9$

   (b) $y=-x$

2. Convert each equation in polar coordinates into an equation in rectangular coordinates.

   (a) $r=-3$

(b) $r = 1 - \cos(\theta)$

Solution

1. (a) We start by substituting $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $(x-3)^2 + y^2 = 9$ and then simplify. With no real direction in which to proceed, we follow our mathematical instincts and see where they take us.

$$\begin{array}{rrcll} & (r\cos(\theta)-3)^2 + (r\sin(\theta))^2 & = & 9 & \\ \Leftrightarrow & r^2\cos^2(\theta) - 6r\cos(\theta) + 9 + r^2\sin^2(\theta) & = & 9 & \\ \Leftrightarrow & r^2\left(\cos^2(\theta) + \sin^2(\theta)\right) - 6r\cos(\theta) & = & 0 & \text{(Subtract 9 from both sides.)} \\ \Leftrightarrow & r^2 - 6r\cos(\theta) & = & 0 & \text{(Since } \cos^2(\theta)+\sin^2(\theta) = 1.\text{)} \\ \Leftrightarrow & r(r - 6\cos(\theta)) & = & 0 & \text{(Factor.)} \end{array}$$

We get $r = 0$ or $r = 6\cos(\theta)$. From Section 4.4 we know that the equation $(x-3)^2 + y^2 = 9$ describes a circle, and since $r = 0$ describes just a point (namely the pole/origin), we choose $r = 6\cos(\theta)$ for our final answer.

(b) Substituting $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $y = -x$ gives $r\sin(\theta) = -r\cos(\theta)$. Rearranging, we get $r\cos(\theta) + r\sin(\theta) = 0$ or $r(\cos(\theta) + \sin(\theta)) = 0$. This gives $r = 0$ or $\cos(\theta) + \sin(\theta) = 0$. Solving the latter equation for θ, we get $\theta = -\frac{\pi}{4} + \pi k$ for integers k. We know $y = -x$ describes a line through the origin. As before, $r = 0$ describes the origin, but nothing else. Consider the equation $\theta = -\frac{\pi}{4}$. In this equation, the variable r is free, meaning it can assume any and all values including $r = 0$. If we imagine plotting points $(r, -\frac{\pi}{4})$ for all conceivable values of r (positive, negative and zero), we are essentially drawing the line containing the terminal side of $\theta = -\frac{\pi}{4}$ which is none other than $y = -x$. Hence, we can take as our final answer $\theta = -\frac{\pi}{4}$ here.

2. (a) Starting with $r = -3$, we can square both sides to get $r^2 = (-3)^2$ or $r^2 = 9$. We may now substitute $r^2 = x^2 + y^2$ to get the equation $x^2 + y^2 = 9$. As we have seen, squaring an equation does not, in general, produce an equivalent equation. The concern here is that the equation $r^2 = 9$ might be satisfied by more points than $r = -3$. On the surface, this appears to be the case since $r^2 = 9$ is equivalent to $r = \pm 3$, and not just $r = -3$. However, any point with polar coordinates $(3, \theta)$ can be represented as $(-3, \theta + \pi)$, which means any point (r, θ) whose polar coordinates satisfy the relation $r = \pm 3$ has an equivalent representation – meaning that they represent the same point in the plane – which satisfies $r = -3$.

(b) Once again, we need to manipulate $r = 1 - \cos(\theta)$ a bit before using the conversion formulas given in Theorem 11.1. We could square both sides of this equation to obtain an r^2 on the left hand side, but that does not result in something helpful for the right hand side. Instead, we multiply both sides by r to obtain $r^2 = r - r\cos(\theta)$. We now have an r^2 and an $r\cos(\theta)$ in the equation, which we can easily handle, but we also have another r to deal with. Rewriting the equation as $r = r^2 + r\cos(\theta)$ and squaring both sides yields $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Substituting $r^2 = x^2 + y^2$ and $r\cos(\theta) = x$ gives $x^2 + y^2 = \left(x^2 + y^2 + x\right)^2$. Once again, we have performed some algebraic manoeuvres which may have altered the set of points described by the original equation. First, we multiplied both sides by r. This means that now $r = 0$ is a viable solution to the equation. In the original equation, $r = 1 - \cos(\theta)$, we see that $\theta = 0$ gives $r = 0$, so the multiplication by r does not introduce any new points.

The squaring of both sides of this equation is also a reason to pause. Are there points with coordinates (r,θ) which satisfy $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ but do not satisfy $r = r^2 + r\cos(\theta)$? Suppose $\left(r',\theta'\right)$ satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Then $r' = \pm\left((r')^2 + r'\cos(\theta')\right)$. If we have that $r' = (r')^2 + r'\cos(\theta')$, we are done. What if $r' = -\left((r')^2 + r'\cos(\theta')\right) = -(r')^2 - r'\cos(\theta')$? We claim that the coordinates $(-r',\theta'+\pi)$, which determine the same point as (r',θ'), satisfy $r = r^2 + r\cos(\theta)$. We substitute $r = -r'$ and $\theta = \theta' + \pi$ into $r = r^2 + r\cos(\theta)$ to see if we get a true statement.

$$\begin{array}{rrcll}
 & -r' & \stackrel{?}{=} & \left(-r'\right)^2 + \left(-r'\cos(\theta'+\pi)\right) & \\
\Leftrightarrow & -\left(-(r')^2 - r'\cos(\theta')\right) & \stackrel{?}{=} & (r')^2 - r'\cos(\theta'+\pi) & \text{(Since } r' = -(r')^2 - r'\cos(\theta')\text{.)} \\
\Leftrightarrow & (r')^2 + r'\cos(\theta') & \stackrel{?}{=} & (r')^2 - r'(-\cos(\theta')) & \text{(Since } \cos(\theta'+\pi) = -\cos(\theta')\text{.)} \\
\Leftrightarrow & (r')^2 + r'\cos(\theta') & \stackrel{\checkmark}{=} & (r')^2 + r'\cos(\theta') &
\end{array}$$

Since both sides worked out to be equal, $(-r',\theta'+\pi)$ satisfies $r = r^2 + r\cos(\theta)$, which means that any point (r,θ) that satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ has a representation which satisfies $r = r^2 + r\cos(\theta)$, and we are done.

In practice, much of the pedantic verification of the equivalence of equations in Example 11.2 is left unsaid. Indeed, in most textbooks, squaring equations like $r = -3$ to arrive at $r^2 = 9$ happens without a second thought. If you take anything away from Example 11.2, it should be that relatively nice things in rectangular coordinates, such as $y = x^2$, can turn ugly in polar coordinates, and vice-versa.

11.1.3 Graphs of polar equations

Having introduced polar coordinates and equations expressed therein, we now discuss how to graph equations in polar coordinates on the rectangular coordinate plane. Since any given point in the plane has infinitely many different representations in polar coordinates, we have the following fundamental graphing principle.

Definitie 11.1 (The fundamental graphing principle for polar equations)
The graph of an equation in polar coordinates is the set of points that satisfy the equation. That is, a point $P(r,\theta)$ is on the graph of an equation if and only if there is a representation of P, say $\left(\tilde{r},\tilde{\theta}\right)$, such that $\tilde{r}$ and $\tilde{\theta}$ satisfy the equation.

Our first example focuses on some of the more structurally simple polar equations.

Example 11.3

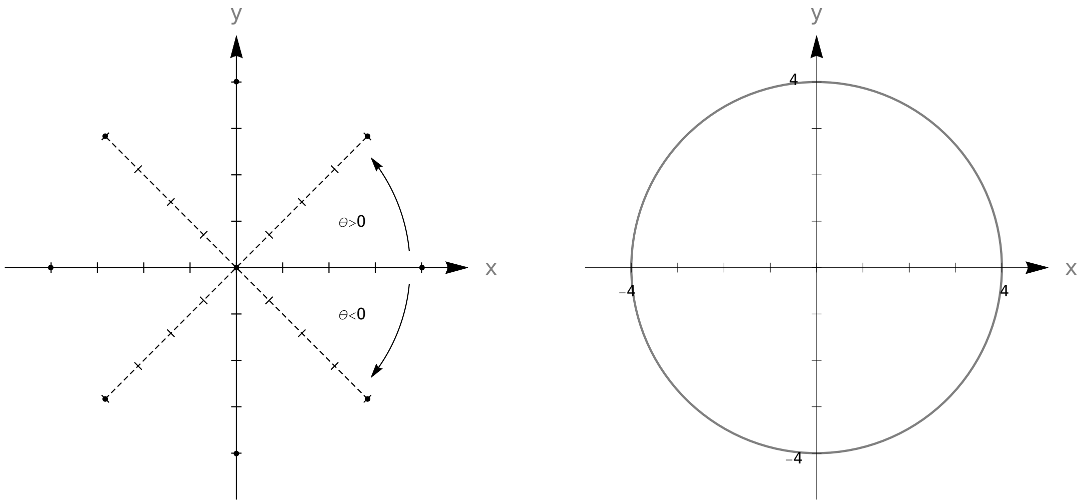
Graph the following polar equations.

1. $r = 4$
2. $\theta = -\dfrac{3\pi}{2}$

Solution

1. In the r equation, θ is free. Its graph is, therefore, all points which have a polar coordinate representation $(4,\theta)$, for any choice of θ (Figure 11.3(a)). Graphically this translates into

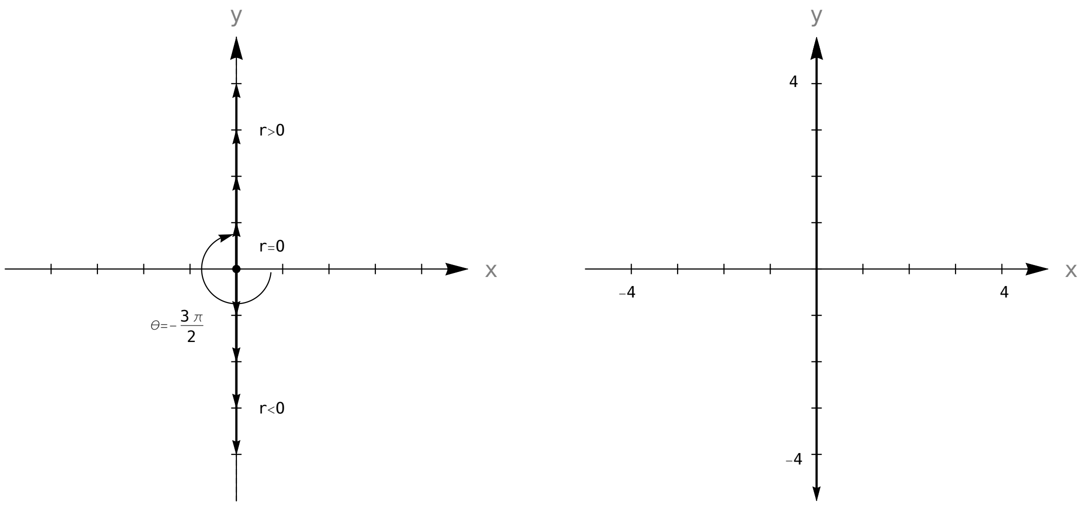
tracing out all of the points 4 units away from the origin. This is exactly the definition of circle, centred at the origin, with a radius of 4 (Figure 11.3(b)).



(a) In $r = 4$, θ is free. (b) The graph of $r = 4$.

Figure 11.3: Constructing the graph of $r = 4$.

2. Here, the variable r is free (Figure 11.4(a)). Plotting $\left(r, -\frac{3\pi}{2}\right)$ for various values of r shows us that we are tracing out the y-axis (Figure 11.4(b)).



(a) In $\theta = -\frac{3\pi}{2}$, r is free. (b) The graph of $\theta = -\frac{3\pi}{2}$.

Figure 11.4: Constructing the graph of $\theta = -\frac{3\pi}{2}$.

Our experience in Example 11.3 makes the following clear.

- The graph of the polar equation $r = a$ on the Cartesian plane is a circle centred at the origin of radius $|a|$.
- The graph of the polar equation $\theta = \alpha$ on the Cartesian plane is the line containing the terminal side of α when plotted in standard position.

Since it gets way more involved to construct the graphs of generic polar equations, we will resort to Mathematica for that purpose. This programme also allows us to check analytically, for instance, intersection of the graphs of multiple polar equations. More specifically, we can rely on the built-in function `PolarPlot` to construct the graph of polar equations.

Example 11.4

Graph the following polar equations.

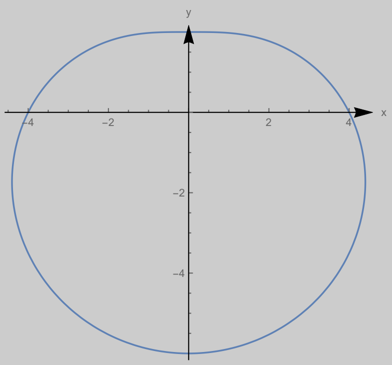
1. $r = 4 - 2\sin(\theta)$
2. $r^2 = 16\cos(2\theta)$

Solution

1. We proceed in Mathematica using the following instruction, where we chose to indicate the axis labels and directions.

```
In[16]:= PolarPlot[4-2 Sin[theta],{theta,0,2*Pi}, AxesLabel→{"x","y"},
         AxesStyle→Arrowheads[{0,0.05}]]
```

Out[16]=



Note that we chose θ to vary between 0 and 2π because the period of the concerned polar equation is 2π.

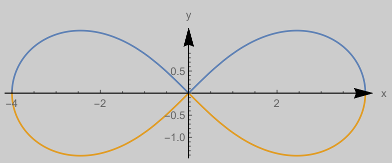
2. Graphing $r^2 = 16\cos(2\theta)$ is complicated by the r^2, so we solve to get

$$r = \pm\sqrt{16\cos(2\theta)} = \pm 4\sqrt{\cos(2\theta)}.$$

Since the period of the involved cosine is π, we plot both functions together using `PolarPlot` for $\theta \in [0,\pi]$.

```
In[17]:= PolarPlot[{Sqrt[16 Cos[2* theta]],- Sqrt[16 Cos[2*theta]]},{theta,0,Pi},
         AxesLabel→{"x","y"},AxesStyle→Arrowheads[{0,0.05}]]
```

Out[17]=



Also note that we may plot values of θ outside of the interval $[0,\pi]$, but then we will find ourselves retracing parts of the curve we already had obtained.

The previous example makes us appreciate the symmetry that is a common occurrence when graphing polar equations. Indeed, it can be verified easily that $r = f(\theta)$ is symmetric about the x-axis if f is even because then we have that $f(\theta) = f(-\theta)$ (e.g. Example 11.4.2), symmetric about the y-axis if $f(\pi - \theta) = f(\theta)$ (e.g. Example 11.4.1 and 2) and symmetric about the origin if f is odd because then we have that $f(-\theta) = -f(\theta)$ (e.g. Example 11.4.2). In addition these usual kinds of symmetry, it is possible to talk about rotational symmetry. More specifically, if $f(\theta - \alpha) = f(\theta)$ it will be rotationally symmetric by α clockwise and counter-clockwise about the pole.

In our next example, we are given the task of finding the intersection points of polar curves.

Example 11.5

Find the points of intersection of the graphs of the following polar equations.

1. $r = 2\sin(\theta)$ and $r = 2 - 2\sin(\theta)$
2. $r = 3$ and $r = 6\cos(2\theta)$

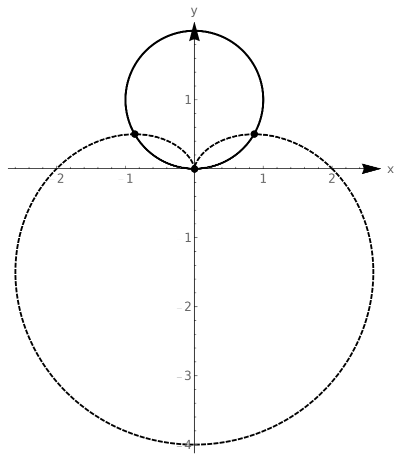
Solution

1. We first try to see if we can find any points which have a single representation $P(r,\theta)$ that satisfies both equations. Assuming such a pair (r,θ) exists, then equating the expressions for r gives

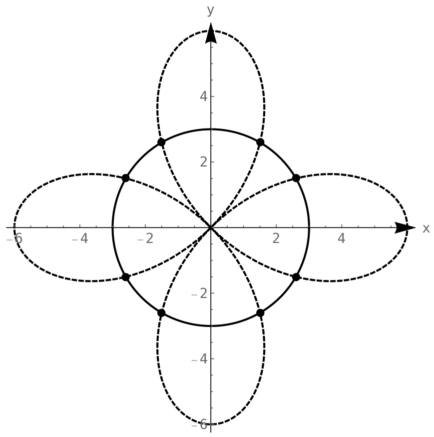
$$\begin{aligned} 2\sin(\theta) = 2 - 2\sin(\theta) \quad &\Leftrightarrow \quad \sin(\theta) = \frac{1}{2} \\ &\Leftrightarrow \quad \theta = \frac{\pi}{6} + 2\pi k \text{ or } \theta = \frac{5\pi}{6} + 2\pi k \end{aligned}$$

for integers k. Plugging $\theta = \frac{\pi}{6}$ into $r = 2\sin(\theta)$, we get $r = 2\sin\left(\frac{\pi}{6}\right) = 2\left(\frac{1}{2}\right) = 1$, which is also the value we obtain when we substitute it into $r = 2 - 2\sin(\theta)$. Hence, $\left(1, \frac{\pi}{6}\right)$ is one representation for the point of intersection in quadrant I. For the point of intersection in Quadrant II, we try $\theta = \frac{5\pi}{6}$. Both equations give us the point $\left(1, \frac{5\pi}{6}\right)$, so this is our answer here. Now, let us check graphically whether we indeed found all intersection points of the graphs of the involved polar equations. (Figure 11.5(a)).

From this graph it appears that there are three intersection points: one in Quadrant I, one in Quadrant II, and the origin. So, how to find the latter algebraically? We know that the pole may be represented as $(0,\theta)$ for any angle θ. On the graph of $r = 2\sin(\theta)$, we start at the origin when $\theta = 0$ and return to it at $\theta = \pi$. Actually, we are at the origin exactly when $\theta = \pi k$ for integers k. On the curve $r = 2 - 2\sin(\theta)$, however, we reach the origin when $\theta = \frac{\pi}{2}$, and more generally, when $\theta = \frac{\pi}{2} + 2\pi k$ for integers k. There is no integer value of k for which $\pi k = \frac{\pi}{2} + 2\pi k$, which means while the origin is on both graphs, the point is never reached simultaneously. In any case, we have determined the three points of intersection to be $\left(1, \frac{\pi}{6}\right)$, $\left(1, \frac{5\pi}{6}\right)$ and the origin.



(a) $r = 2\sin(\theta)$ (solid) and $r = 2 - 2\sin(\theta)$ (dashed).



(b) $r = 3$ (solid) and $r = 6\cos(2\theta)$ (dashed).

Figure 11.5: Points of intersection of the graphs of two polar equations.

2. Let us graph the equations to get an idea of how many intersection points to expect and where they lie. The graph of $r = 3$ is a circle centred at the origin with a radius of 3 and the graph of $r = 6\cos(2\theta)$ is a four-leafed rose (Figure 11.5(b)).

   It appears as if there are eight points of intersection - two in each quadrant. We first look to see if there any points $P(r,\theta)$ with a representation that satisfies both equations. For these points,

$$6\cos(2\theta) = 3 \iff \cos(2\theta) = \frac{1}{2}$$
$$\iff \theta = \frac{\pi}{6} + \pi k \text{ or } \theta = \frac{5\pi}{6} + \pi k$$

   for integers k. Out of all of these solutions, we obtain just four distinct points represented by $\left(3, \frac{\pi}{6}\right)$, $\left(3, \frac{5\pi}{6}\right)$, $\left(3, \frac{7\pi}{6}\right)$ and $\left(3, \frac{11\pi}{6}\right)$. To determine the coordinates of the remaining four points, we have to consider how the representations of the points of intersection can differ. We know from the beginning of this section that if (r,θ) and (r',θ') represent the same point and $r \neq 0$, then either $r = r'$ or $r = -r'$. If $r = r'$, then $\theta' = \theta + 2\pi k$, so one possibility is that an intersection point P has a representation (r,θ) which satisfies $r = 3$ and another representation $(r, \theta + 2\pi k)$ for some integer k, which satisfies $r = 6\cos(2\theta)$. At this point, if we replace every occurrence of θ in the equation $r = 6\cos(2\theta)$ with $(\theta + 2\pi k)$ and then see if, by equating the resulting expressions for r, we get any more solutions for θ. Since $\cos(2(\theta + 2\pi k)) = \cos(2\theta + 4\pi k) = \cos(2\theta)$ for every integer k, however, the equation $r = 6\cos(2(\theta + 2\pi k))$ reduces to the same equation we had before, $r = 6\cos(2\theta)$, which means we get no additional solutions.
   Moving on to the case where $r = -r'$, we have that $\theta' = \theta + (2k+1)\pi$ for integers k. We look to see if we can find points P which have a representation (r,θ) that satisfies $r = 3$ and another, $(-r, \theta + (2k+1)\pi)$, that satisfies $r = 6\cos(2\theta)$. To do this, we substitute $(-r)$ for r and $(\theta + (2k+1)\pi)$ for θ in the equation $r = 6\cos(2\theta)$ and get $-r = 6\cos(2(\theta + (2k+1)\pi))$. Since $\cos(2(\theta + (2k+1)\pi)) = \cos(2\theta + (2k+1)(2\pi)) = \cos(2\theta)$ for all integers k, the equation $-r = 6\cos(2(\theta + (2k+1)\pi))$ reduces to $-r = 6\cos(2\theta)$, or $r = -6\cos(2\theta)$. Coupling this equation with $r = 3$ gives

$$-6\cos(2\theta) = 3 \iff \cos(2\theta) = -\frac{1}{2}$$

$$\iff \quad \theta = \frac{\pi}{3} + \pi k \text{ or } \theta = \frac{2\pi}{3} + \pi k$$

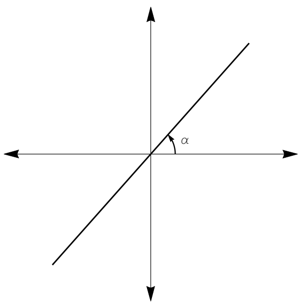
From these solutions, we obtain the remaining four intersection points with representations $\left(-3, \frac{\pi}{3}\right)$, $\left(-3, \frac{2\pi}{3}\right)$, $\left(-3, \frac{4\pi}{3}\right)$ and $\left(-3, \frac{5\pi}{3}\right)$

There are a number of basic and classic polar curves, famous for their beauty and/or applicability in science. For that reason, this section ends with a small gallery of some of these graphs.

Lines

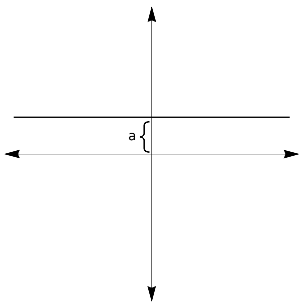
Through the origin:

$\theta = \alpha$



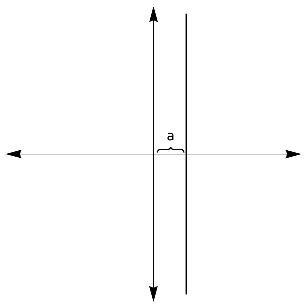
Horizontal line:

$r = a\csc(\theta)$



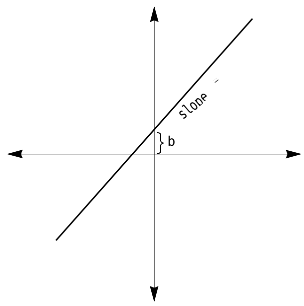
Vertical line:

$r = a\sec(\theta)$



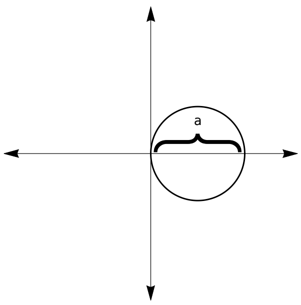
Not through origin:

$r = \dfrac{b}{\sin(\theta) - m\cos(\theta)}$

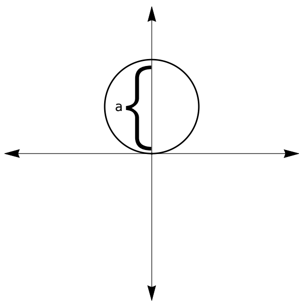


Circles

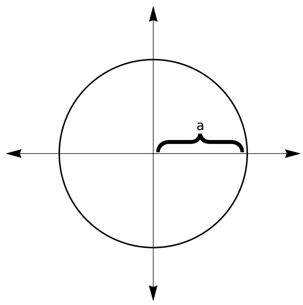
Centred on x-axis:
$r = a\cos(\theta)$



Centred on y-axis:
$r = a\sin(\theta)$

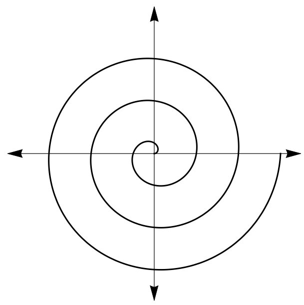


Centred on origin:
$r = a$



Spiral

Archimedean spiral
$r = \theta$

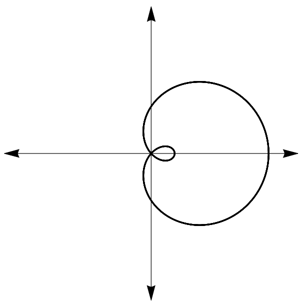


Limaçons

Symmetric about x-axis: $r = a \pm b\cos(\theta)$; Symmetric about y-axis: $r = a \pm b\sin(\theta)$; $a, b > 0$

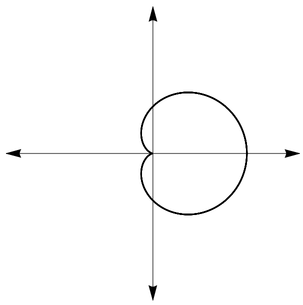
With inner loop:

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



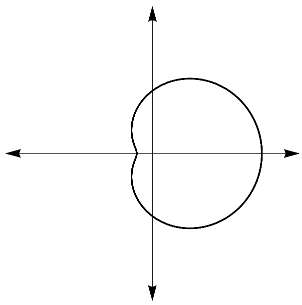
Cardioid:

$\frac{a}{b} = 1$



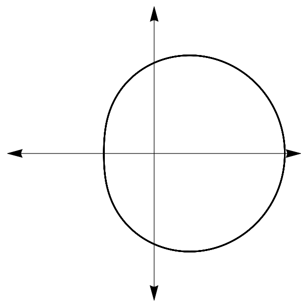
Dimpled:

$1 < \frac{a}{b} < 2$



Convex:

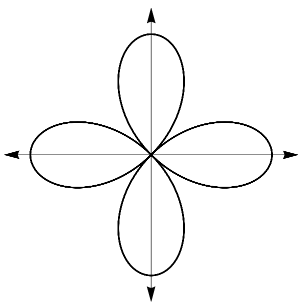
$\frac{a}{b} > 2$



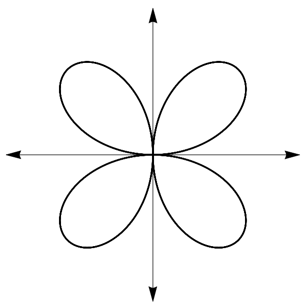
Rose Curves

Symmetric about x-axis: $r = a\cos(n\theta)$; Symmetric about y-axis: $r = a\sin(n\theta)$
Curve contains $2n$ petals when n is even and n petals when n is odd.

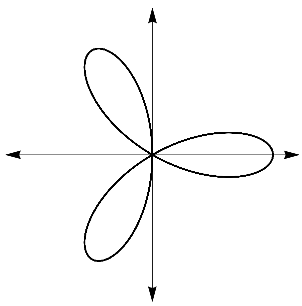
$r = a\cos(2\theta)$



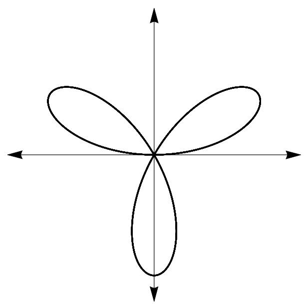
$r = a\sin(2\theta)$



$r = a\cos(3\theta)$



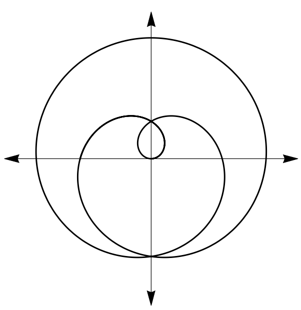
$r = a\sin(3\theta)$



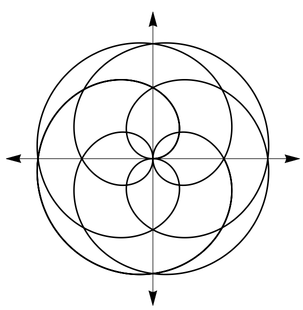
Special Curves

Rose curves

$r = a\sin(\theta/5)$

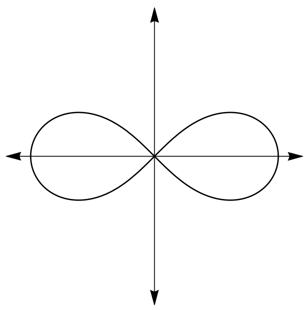


$r = a\sin(2\theta/5)$



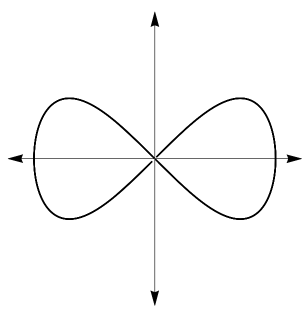
Lemniscate:

$r^2 = a^2\cos(2\theta)$



Eight Curve:

$r^2 = a^2\sec^4(\theta)\cos(2\theta)$



11.2 Parametric equations

11.2.1 Definition

As we have seen in Section 11.1 there are interesting curves which, when plotted in the xy-plane, neither represent y as a function of x nor x as a function of y. Here, we present a new concept which allows us to use functions to study these kinds of curves. To motivate the idea, we imagine a bug crawling across a table top starting at the point O and tracing out a curve C in the plane, as shown in Figure 11.6.

The curve C does not represent y as a function of x because it fails the vertical line test and it does not represent x as a function of y because it fails the horizontal line test. However, since the bug can be in only one place $P(x, y)$ at any given time t, we can define the x-coordinate of P as a function of t and the y-coordinate of P as a different function of t. The independent variable t in this case is called a **parameter** (*parameter*) and the system of equations

$$\begin{cases} x &= f(t) \\ y &= g(t) \end{cases}$$

is called a **system of parametric equations** (*stelsel parametervergelijkingen*) or a **parametrization** (*parametervoorstelling*) of the curve C.

The parametrization of C endows it with an **orientation** (*zin*) and the arrows on C indicate motion in the direction of increasing values of t. In this case, our bug starts at the point O, travels upwards to the left, then loops back around to cross its path at the point Q and finally heads off into the first quadrant. It is important to note that the curve itself is a set of points and as such is devoid of any orientation. It is the parametrization that determines the orientation and as we shall see, different parametrizations can determine different orientations. Actually, the system of equations

$$\begin{cases} x &= \cos(t) \\ y &= \sin(t) \end{cases}$$

parametrizes the unit circle, giving it a counter-clockwise orientation. More generally, the equations of circular motion (Theorem 5.5)

$$\begin{cases} x &= r\cos(\omega t) \\ y &= r\sin(\omega t) \end{cases}$$

are parametric equations that trace out a circle of radius r centred at the origin. If $\omega > 0$, the orientation is counter-clockwise; if $\omega < 0$, the orientation is clockwise. The angular frequency ω determines how fast the object moves around the circle.

11.2.2 Graphing parametric equations

Graphing parametric equations is pretty straightforward in the sense that we just choose some friendly values of t, plot the corresponding points and connect the results in a pleasing fashion. For instance, consider the following system of parametric equations:

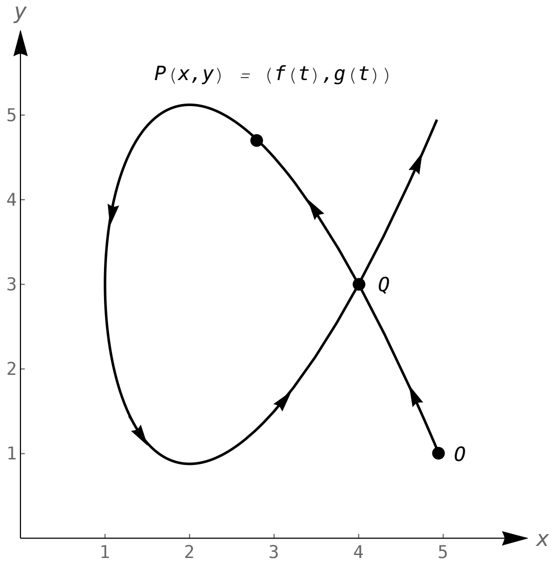


Figure 11.6: A bug crawling tracing out a curve C in the plane.

$$\begin{cases} x = t^2 - 3 \\ y = 2t - 1 \end{cases} \tag{11.2}$$

for $t \geq -2$.

Since we are told $t \geq -2$, we start there and evaluate $x(t)$ and $y(t)$, yielding the following values:

t	$x(t)$	$y(t)$
-2	1	-5
-1	-2	-3
0	-3	-1
1	-2	1
2	1	3
3	6	5

Then we plot the successive points in Figure 11.7 and we draw an arrow to indicate the direction of the path for increasing values of t. The curve looks like a parabola. To verify this we may eliminate the parameter t from Equation (11.2). To do so, we choose to solve the equation $y = 2t - 1$ for t to get $t = \frac{y+1}{2}$. Substituting this into the equation $x = t^2 - 3$ yields

$$x = \left(\frac{y+1}{2}\right)^2 - 3$$

or, after some rearrangement,

$$(y+1)^2 = 4(x+3). \tag{11.3}$$

The graph of this equation is a parabola with vertex $(-3,-1)$ which opens to the right. Technically speaking, Equation (11.3) describes the entire parabola, while the parametric equations (Equation (11.2)) for $t \geq -2$ describe only a portion of the parabola. In this case, we can remedy this situation by restricting the bounds on y. Since the portion of the parabola we want is exactly the part where $y \geq -5$, Equation (11.3) coupled with the restriction $y \geq -5$ describes the same curve as the given parametric equations. The one piece of information we can never recover after eliminating the parameter is the orientation of the curve.

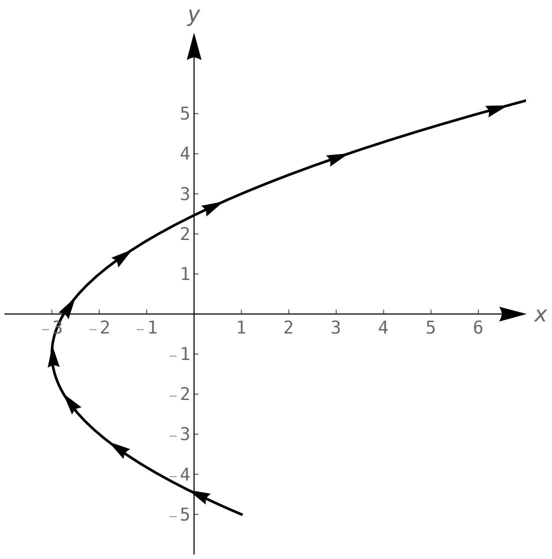


Figure 11.7: The curve described by Equation (11.2).

In Mathematica, we can use the built-in function **ParametricPlot** to construct the graph of a parametric equation. For instance, for Equation (11.2) this can be achieved as follows.

```
In[18]:= ParametricPlot[{t^2-3,2* t-1},{t,-2,3.2},AxesLabel→{"x","y"},
         AxesStyle→Arrowheads[{0,0.05}]]
```

Example 11.6

Sketch the curves described by the following parametric equations.

1. For $-1 \leq t \leq 1$
$$\begin{cases} x &= t^3 \\ y &= 2t^2, \end{cases}$$

2. For $0 \leq t \leq \dfrac{3\pi}{2}$
$$\begin{cases} x &= 1 + 3\cos(t) \\ y &= 2\sin(t), \end{cases}$$

Solution

1. We first sketch the graphs of $x = t^3$ and $y = 2t^2$ over the interval $[-1, 1]$ (Figures 11.8(a) and 11.8(b)). We note that as t takes on values in $[-1, 1]$, $x = t^3$ ranges between -1 and 1, and $y = 2t^2$ ranges between 0 and 2. This means that all of the action is happening on a portion of the plane, namely $\{(x, y) \mid -1 \leq x \leq 1,\ 0 \leq y \leq 2\}$. Next, we plot a few points. Certainly, $t = -1$ and $t = 1$ are good values to pick since these are the extreme values of t. We also choose $t = 0$ as this is a relative minimum on the graph of $y = 2t^2$. Plugging in $t = -1$ gives the point $(-1, 2)$, $t = 0$ gives $(0, 0)$ and $t = 1$ gives $(1, 2)$. More generally, we see that $x = t^3$ is increasing over the entire interval $[-1, 1]$ whereas $y = 2t^2$ is decreasing over the interval $[-1, 0]$ and then increasing over $[0, 1]$. Geometrically, we start at $(-1, 2)$ (where $t = -1$), then move to the right (since x is increasing) and down (since y is decreasing) to

$(0,0)$ (where $t=0$). We continue to move to the right (since x is still increasing) but now move upwards (since y is now increasing) until we reach $(1,2)$ (where $t=1$). Finally, we eliminate the parameter. Solving $x=t^3$ for t, we get $t=\sqrt[3]{x}$. Substituting this into $y=2t^2$ gives $y=2(\sqrt[3]{x})^2=2x^{2/3}$. The final graph is shown in Figure 11.8(c).

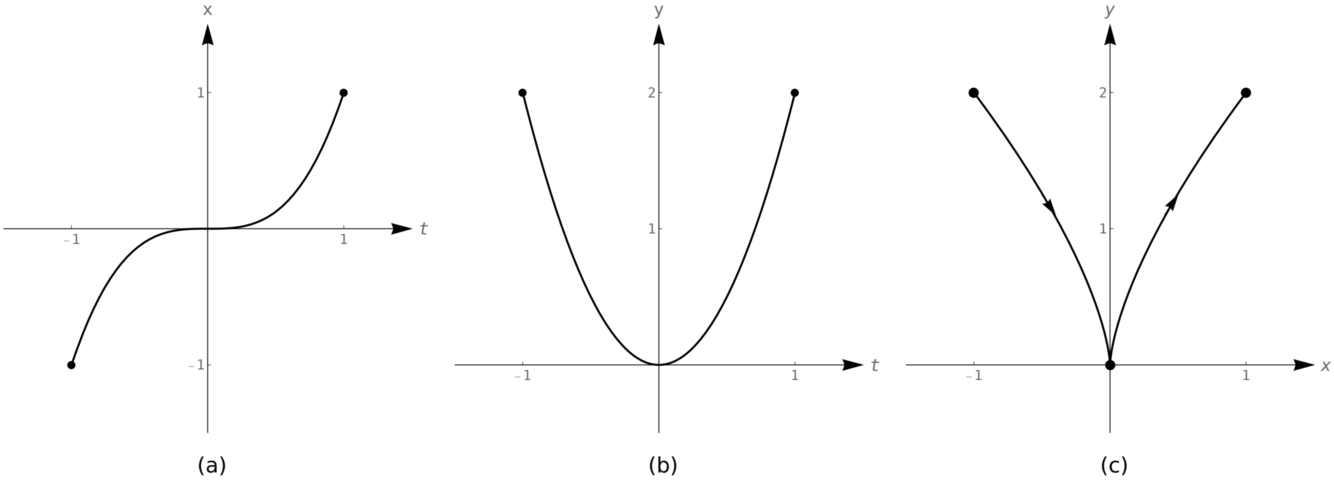


Figure 11.8: The graph of $x=t^3$ (a), $y=2t^2$ (b) and $x=t^3$, $y=2t^2$ (c) for $-1\le t\le 1$.

2. Proceeding as above, we set about graphing this system by first graphing $x=1+3\cos(t)$ and $y=2\sin(t)$ on the interval $\left[0,\frac{3\pi}{2}\right]$ (Figures 11.9(a) and 11.9(b)). We see that x ranges from -2 to 4 and y ranges from -2 to 2. Plugging in $t=0$, $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$ gives the points $(4,0)$, $(1,2)$, $(-2,0)$ and $(1,-2)$, respectively. As t ranges from 0 to $\frac{\pi}{2}$, $x=1+3\cos(t)$ is decreasing, while $y=2\sin(t)$ is increasing. This means that we start tracing out our answer at $(4,0)$ and continue moving to the left and upwards towards $(1,2)$. For $\frac{\pi}{2}\le t\le\pi$, x is decreasing, as is y, so the motion is still right to left, but now is downwards from $(1,2)$ to $(-2,0)$. On the interval $\left[\pi,\frac{3\pi}{2}\right]$, x begins to increase, while y continues to decrease. Hence, the motion becomes left to right but continues downwards, connecting $(-2,0)$ to $(1,-2)$. To eliminate the parameter here, we use the Pythagorean identity. Hence, we solve $x=1+3\cos(t)$ for $\cos(t)$ to get $\cos(t)=\frac{x-1}{3}$, and we solve $y=2\sin(t)$ for $\sin(t)$ to get $\sin(t)=\frac{y}{2}$. Substituting these expressions into $\cos^2(t)+\sin^2(t)=1$ gives

$$\left(\frac{x-1}{3}\right)^2+\left(\frac{y}{2}\right)^2=1,$$

or

$$\frac{(x-1)^2}{9}+\frac{y^2}{4}=1.$$

The graph of this equation is an ellipse centred at $(1,0)$ with vertices at $(-2,0)$ and $(4,0)$. Our parametric equations here are tracing out three-quarters of this ellipse, in a counter-clockwise direction (Figure 11.9(c)).

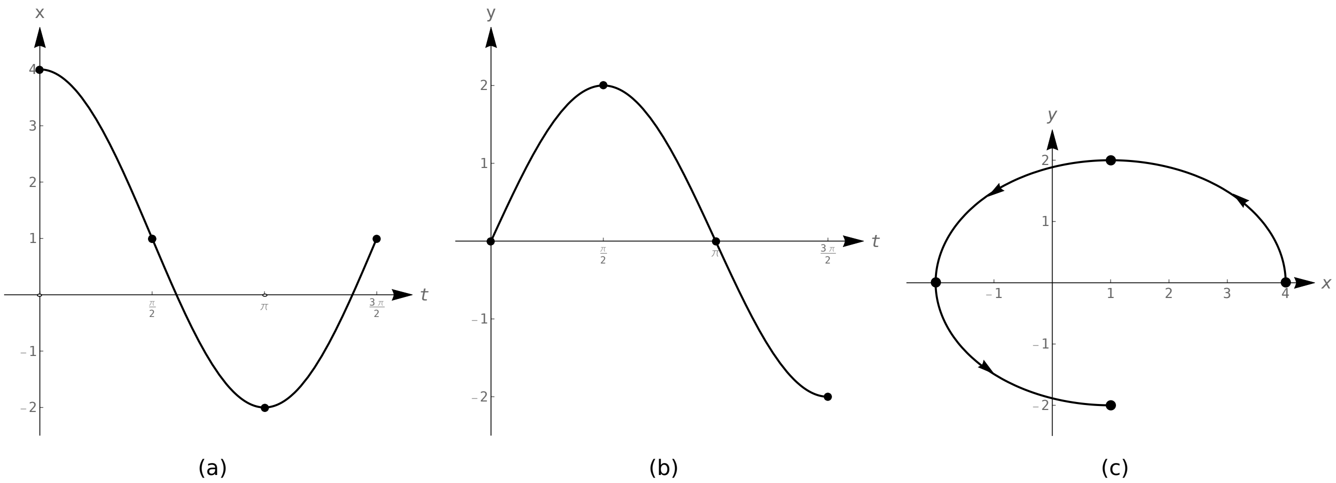


Figure 11.9: The graph of $x = 1 + 3\cos(t)$ (a), $y = 2\sin(t)$ (b) and $x = 1 + 3\cos(t)$, $y = 2\sin(t)$ (c) for $0 \le t \le \frac{3\pi}{2}$.

11.2.3 Parametrising curves

Now that we have had some good practice sketching the graphs of parametric equations, we turn to the problem of finding parametric representations of curves. For that purpose, we have the following guidelines.

- To parametrize $y = f(x)$ as x runs through some interval I, let $x = t$ and $y = f(t)$ and let t run through I.
- To parametrize $x = g(y)$ as y runs through some interval I, let $y = t$ and $x = g(t)$ and let t run through I.
- To parametrize a directed line segment with initial point (x_0, y_0) and terminal point (x_1, y_1), let $x = x_0 + (x_1 - x_0)t$ and $y = y_0 + (y_1 - y_0)t$ for $0 \le t \le 1$.
- To parametrize $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1$ where $a, b > 0$, let $x = x_0 + a\cos(t)$ and $y = y_0 + b\sin(t)$ for $0 \le t < 2\pi$. This will impart a counter-clockwise orientation.

Example 11.7

Find a parametrization for each of the following curves.

1. $y = x^2$ from $x = -3$ to $x = 2$.
2. The line segment which starts at $(2, -3)$ and ends at $(1, 5)$.
3. The circle $x^2 + 2x + y^2 - 4y = 4$.
4. The left half of the ellipse
$$\frac{x^2}{4} + \frac{y^2}{9} = 1.$$

Solution

1. Since $y = x^2$ is written in the form $y = f(x)$, we let $x = t$ and $y = f(t) = t^2$. Since $x = t$, the

bounds on t match precisely the bounds on x so we get

$$\begin{cases} x = t \\ y = t^2, \end{cases}$$

for $-3 \le t \le 2$.

2. To find the equation for x, we have that the line segment starts at $x = 2$ and ends at $x = 1$. This means the displacement in the x-direction is $(1-2) = -1$. Hence, the equation for x is $x = 2 + (-1)t = 2 - t$. For y, we note that the line segment starts at $y = -3$ and ends at $y = 5$. Hence, the displacement in the y-direction is $(5-(-3)) = 8$, so we get $y = -3 + 8t$. Our final answer is

$$\begin{cases} x = 2 - t \\ y = -3 + 8t, \end{cases}$$

for $0 \le t \le 1$.

3. In order to use the formulas above to parametrize the circle $x^2 + 2x + y^2 - 4y = 4$, we first need to put it into the correct form. We complete the square and get $(x+1)^2 + (y-2)^2 = 9$, or

$$\frac{(x+1)^2}{9} + \frac{(y-2)^2}{9} = 1.$$

In this equation, we identify $\cos(t) = \frac{x+1}{3}$ and $\sin(t) = \frac{y-2}{3}$. Rearranging these last two equations, we get $x = -1 + 3\cos(t)$ and $y = 2 + 3\sin(t)$. In order to complete one revolution around the circle, we let t range through the interval $[0, 2\pi[$. We get as our final answer

$$\begin{cases} x = -1 + 3\cos(t) \\ y = 2 + 3\sin(t), \end{cases}$$

for $0 \le t < 2\pi$.

4. We immediately get $x = 2\cos(t)$ and $y = 3\sin(t)$. The normal range on the parameter in this case is $0 \le t < 2\pi$, but since we are interested in only the left half of the ellipse, we restrict t to the values which correspond to Quadrant II and Quadrant III angles, namely $\frac{\pi}{2} \le t \le \frac{3\pi}{2}$. Our final answer is

$$\begin{cases} x = 2\cos(t) \\ y = 3\sin(t), \end{cases}$$

for $\frac{\pi}{2} \le t \le \frac{3\pi}{2}$.

We note that the parametrisation approach offers only one of literally infinitely many ways to parametrize the concerning curves. Essentially, there are two easy ways to alter parametrizations.

- **Reversing Orientation:** Replacing every occurrence of t with $-t$ in a parametric description for a curve reverses the orientation of the curve.

- **Shift of Parameter:** Replacing every occurrence of t with $(t-c)$ in a parametric description for a curve shifts the start of the parameter t ahead by c units.

These techniques are illustrated in the following example.

Example 11.8

Find a parametrization for the unit circle, oriented clockwise, with $t = 0$ corresponding to $(0, -1)$.

Solution

We know that

$$\begin{cases} x &= \cos(t) \\ y &= \sin(t), \end{cases}$$

for $0 \leq t < 2\pi$ gives a counter-clockwise parametrization of the unit circle with $t = 0$ corresponding to $(1, 0)$, so the first order of business is to reverse the orientation. Replacing t with $-t$ gives

$$\begin{cases} x &= \cos(-t) \\ y &= \sin(-t), \end{cases}$$

for $0 \leq t < 2\pi$ which simplifies to

$$\begin{cases} x &= \cos(t) \\ y &= -\sin(t), \end{cases}$$

for $0 \leq t < 2\pi$.

This parametrization gives a clockwise orientation, but $t = 0$ still corresponds to the point $(1, 0)$; the point $(0, -1)$ is reached when $t = -\frac{3\pi}{2}$. Our strategy is to first get the parametrization to start at the point $(0, -1)$ and then shift the parameter accordingly so the start coincides with $t = 0$. We know that any interval of length 2π will parametrize the entire circle, so we start the parameter t at $-\frac{3\pi}{2}$, and find the upper bound by adding 2π so $-\frac{3\pi}{2} \leq t < \frac{\pi}{2}$. We now shift the parameter by introducing a time delay of $\frac{3\pi}{2}$ units by replacing every occurrence of t with $\left(t - \frac{3\pi}{2}\right)$, i.e.

$$\begin{cases} x &= \cos\left(t - \frac{3\pi}{2}\right) \\ y &= -\sin\left(t - \frac{3\pi}{2}\right), \end{cases}$$

for $-\frac{3\pi}{2} \leq t - \frac{3\pi}{2} < \frac{\pi}{2}$. This simplifies to

$$\begin{cases} x &= -\sin(t) \\ y &= -\cos(t), \end{cases} \tag{11.4}$$

for $0 \leq t < 2\pi$, as required.

We can now use our answer to Example 11.8 to derive the equation of a **cycloid** (*cycloïde*). Suppose a circle of radius r rolls along the positive x-axis at a constant velocity v as pictured in Figure 11.10. Let θ be the angle in radians which measures the amount of clockwise rotation experienced by the radius highlighted in the figure.

Our goal is to find parametric equations for the coordinates of the point $P(x, y)$ in terms of θ. From Example 11.8, we know that clockwise motion along the unit circle starting at the point $(0, -1)$ can be modelled by Equation (11.4) . To model this motion on a circle of radius r, all we need to do is multiply both x and y by the factor r which yields

$$\begin{cases} x &= -r\sin(\theta) \\ y &= -r\cos(\theta). \end{cases}$$

We now need to adjust for the fact that the circle is not stationary with centre $(0, 0)$, but rather, is rolling along the positive x-axis. Since the velocity v is constant, we know that at time t, the centre of the circle has travelled a distance vt down the positive x-axis. Furthermore, since the radius of the circle is r and the circle is not moving vertically, we know that the centre of the circle is always r units above the x-axis. Putting these two facts together, we have that at time t, the centre of the circle is at the point (vt, r). We know from Chapter 5 that $v = \frac{r\theta}{t}$, or $vt = r\theta$. Hence, the centre of the circle, in terms of the parameter θ, is $(r\theta, r)$. As a result, we need to modify the governing equations by shifting the x-coordinate to the right $r\theta$ units and the y-coordinate up r units. In this way, we get

$$\begin{cases} x &= -r\sin(\theta)+r\theta \\ y &= -r\cos(\theta)+r, \end{cases}$$

which can be written as

$$\begin{cases} x &= r(\theta-\sin(\theta)) \\ y &= r(1-\cos(\theta)). \end{cases}$$

Since the motion starts at $\theta = 0$ and proceeds indefinitely, we set $\theta \geq 0$.

Figure 11.11 gives a small gallery of interesting and famous curves along with the parametric equations that produce them.

11.2.4 Conic sections continued even further

For completeness, we conclude this chapter by listing the parametric representations of the conic sections we introduced in Section 4.4.

The parametric representation of a circle with centre at the origin and radius r is given by

$$\begin{cases} x = r\cos(t) \\ y = r\sin(t), \end{cases} \tag{11.5}$$

for $0 \leq t \leq 2\pi$. Likewise, the parametric representation of an ellipse centred at the origin and with semi-major and conjugate axis of a and b respectively ($a > b$), is given by

$$\begin{cases} x = a\cos(t) \\ y = b\sin(t), \end{cases} \tag{11.6}$$

for $0 \leq t \leq 2\pi$.

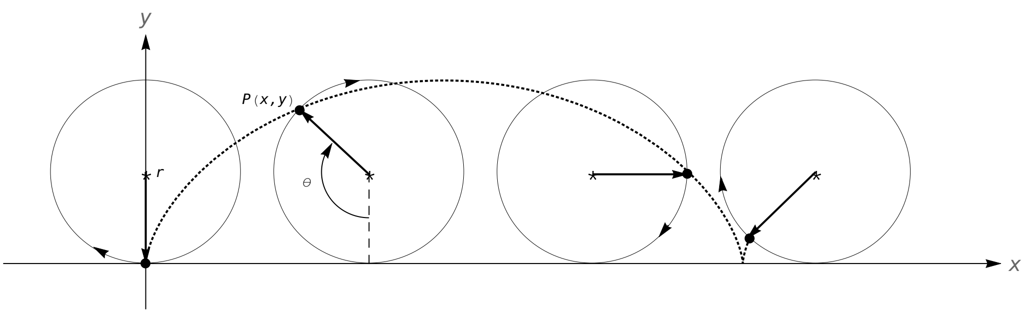


Figure 11.10: Constructing a cycloid from a rolling circle with radius r.

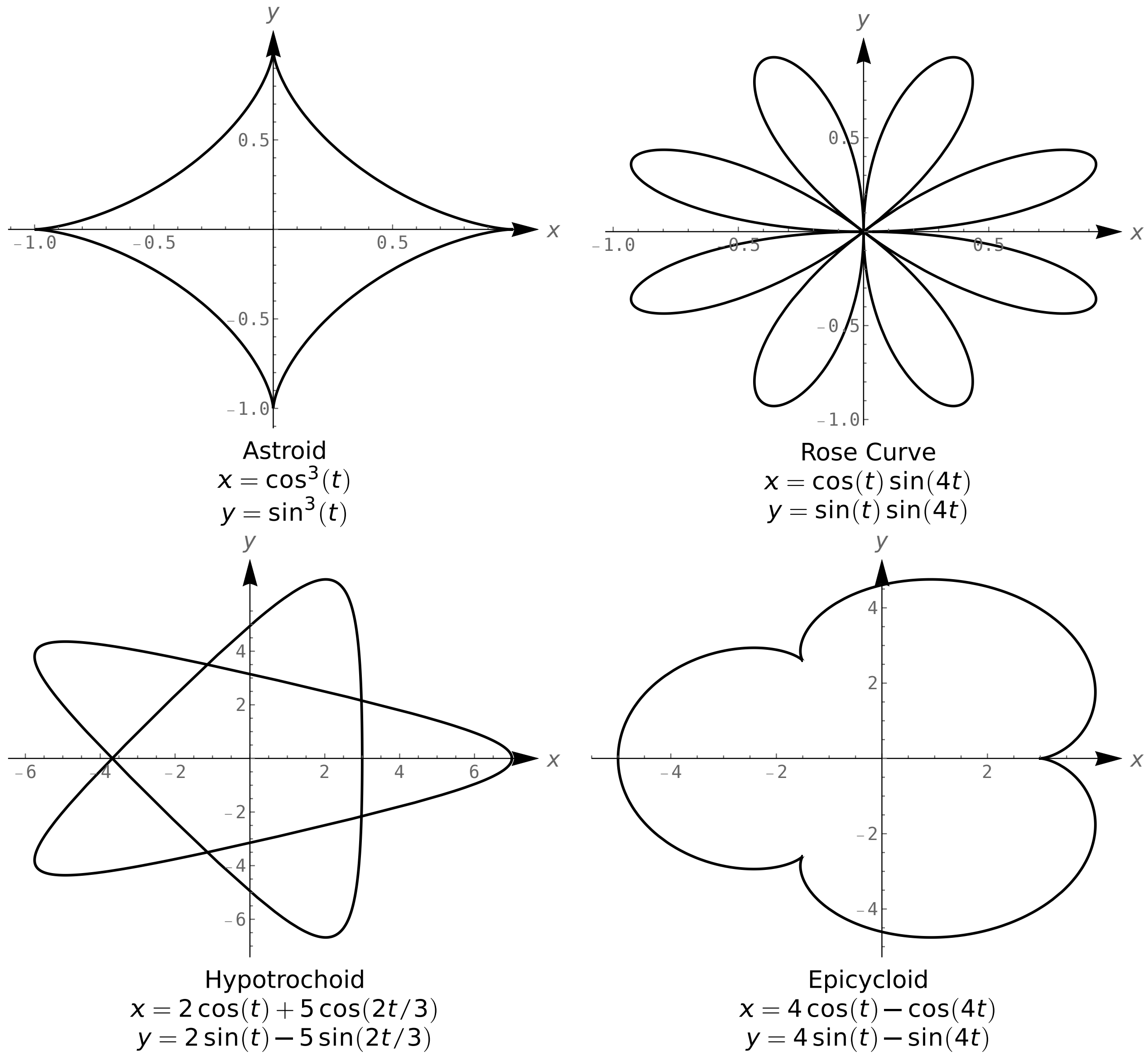


Figure 11.11: A gallery of interesting planar curves.

11.3 Derivatives and parametric and polar equations

11.3.1 Parametric equations

Here we will exemplify the techniques of calculus to study curves given by a set of parametric equations. Amongst other things, we are interested in lines tangent to points on such a curve. They describe how the y-values are changing with respect to the x-values, they are useful in making approximations, and they indicate instantaneous direction of travel.

The slope of the tangent line is still $\frac{dy}{dx}$, and the chain rule allows us to calculate this in the context of parametric equations. If $x = f(t)$ and $y = g(t)$, the chain rule states that

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}.$$

Solving for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{g'(t)}{f'(t)}, \tag{11.7}$$

provided that $f'(t) \neq 0$, and we also assume that f and g are differentiable on some open interval I.

These pieces of information allow us to define the tangent and normal lines to a curve C.

Definitie 11.2 (Tangent and normal lines)
Let a curve C be parametrized by $x = f(t)$ and $y = g(t)$, where f and g are differentiable functions on some interval I containing $t = t_0$. The **tangent line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope

$$m = \frac{g'(t_0)}{f'(t_0)},$$

provided $f'(t_0) \neq 0$.

The **normal line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope $m = -f'(t_0)/g'(t_0)$, provided $g'(t_0) \neq 0$.

This definition leaves two special cases to consider. When the tangent line is horizontal, the normal line is undefined by the above definition as $g'(t_0) = 0$. Likewise, when the normal line is horizontal, the tangent line is undefined. It seems reasonable that these lines be defined, so we add the following to the above definition.

1. If the tangent line at $t = t_0$ has a slope of 0, the normal line to C at $t = t_0$ is the line $x = f(t_0)$.
2. If the normal line at $t = t_0$ has a slope of 0, the tangent line to C at $t = t_0$ is the line $x = f(t_0)$.

Example 11.9

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$, and let C be the curve defined by these equations.

1. Find the equations of the tangent and normal lines to C at $t = 3$.
2. Find where C has vertical and horizontal tangent lines.

Solution

1. We start by computing $f'(t) = 10t - 6$ and $g'(t) = 2t + 6$. Thus

$$\frac{dy}{dx} = \frac{2t+6}{10t-6}.$$

Make note of something that might seem unusual: $\frac{dy}{dx}$ is a function of t, not x. Just as points on the curve are found in terms of t, so are the slopes of the tangent lines.

The point on C at $t = 3$ is $(31, 26)$. The slope of the tangent line is $m = 1/2$ and the slope of the normal line is $m = -2$. Thus,

- the equation of the tangent line is $y = \frac{1}{2}(x - 31) + 26$, and
- the equation of the normal line is $y = -2(x - 31) + 26$.

This is illustrated in Figure 11.12.

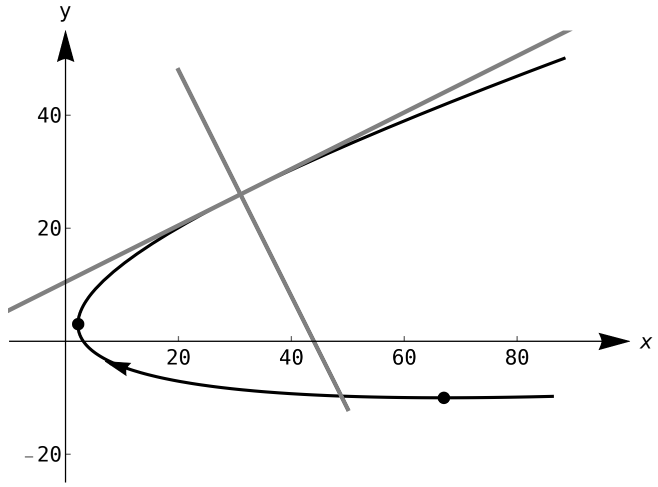


Figure 11.12: Graphing tangent and normal lines in Example 11.9.

2. To find where C has a horizontal tangent line, we set $\frac{dy}{dx} = 0$ and solve for t. In this case, this amounts to setting $g'(t) = 0$ and solving for t (and making sure that $f'(t) \neq 0$):

$$g'(t) = 0 \quad \Rightarrow \quad 2t + 6 = 0 \quad \Leftrightarrow \quad t = -3.$$

The point on C corresponding to $t = -3$ is $(67, -10)$; the tangent line at that point is horizontal (hence with equation $y = -10$).

To find where C has a vertical tangent line, we find where it has a horizontal normal line, and set $-\frac{f'(t)}{g'(t)} = 0$. This amounts to setting $f'(t) = 0$ and solving for t and making sure that $g'(t) \neq 0$.

$$f'(t) = 0 \quad \Rightarrow \quad 10t - 6 = 0 \quad \Leftrightarrow \quad t = 0.6.$$

The point on C corresponding to $t = 0.6$ is $(2.2, 2.96)$. The tangent line at that point is $x = 2.2$.

Example 11.10

Find the equation of the tangent line to the astroid $x = \cos^3(t)$, $y = \sin^3(t)$ at $t = 0$ shown in Figure 11.11.

Solution

We start by finding $x'(t)$ and $y'(t)$:

$$x'(t) = -3\sin(t)\cos^2(t), \qquad \text{and} \qquad y'(t) = 3\cos(t)\sin^2(t).$$

Note that both of these are 0 at $t = 0$; the curve is not smooth at $t = 0$ forming a cusp on the graph. Evaluating $\frac{dy}{dx}$ at this point returns the indeterminate form of 0/0. $\frac{dy}{dx}$:

$$\frac{dy}{dx} = \frac{3\cos(t)\sin^2(t)}{-3\sin(t)\cos^2(t)} = -\frac{\sin(t)}{\cos(t)},$$

as long as $\cos(t) \neq 0$ and $\sin(t) \neq 0$. When $t = 0$, it is tempting to declare that

$$\frac{dy}{dx} = -\frac{\sin(0)}{\cos(0)} = 0,$$

but this overlooks the fact that we cancelled earlier with the stipulation that $\sin(t) \neq 0$. In fact, the graph of the curve has a cusp at $t = 0$, as both $x' = 0$ and $y' = 0$.

We can, however, examine the slopes of tangent lines near $t = 0$, and take the limit as $t \to 0$.

$$\begin{aligned}
\lim_{t\to 0} \frac{y'(t)}{x'(t)} &= \lim_{t\to 0} \frac{3\cos(t)\sin^2(t)}{-3\sin(t)\cos^2(t)} && \text{(We can cancel as } t \neq 0.\text{)}\\
&= \lim_{t\to 0} \left(-\frac{\sin(t)}{\cos(t)}\right)\\
&= 0\,.
\end{aligned}$$

We have accomplished something significant. When the derivative $\frac{dy}{dx}$ returns an indeterminate form at $t = t_0$, we can define its value by setting it to be $\lim\limits_{t\to t_0} \frac{dy}{dx}$, if that limit exists. This allows us to find slopes of tangent lines at cusps, which can be very beneficial.

We found the slope of the tangent line at $t = 0$ to be 0; therefore the tangent line is $y = 0$, the x-axis.

11.3.2 Polar equations

A basis for much of what is done in this section is the ability to turn a polar function $r = f(\theta)$ into a set of parametric equations. Using the identities $x = r\cos(\theta)$ and $y = r\sin(\theta)$, we can create the parametric equations $x = f(\theta)\cos(\theta)$, $y = f(\theta)\sin(\theta)$ and continue our work with those.

For instance, if we are asked to construct the tangent line to a curve described by $r = f(\theta)$, we will use $x = f(\theta)\cos(\theta)$, $y = f(\theta)\sin(\theta)$ to compute $\frac{dy}{dx}$. Using Equation (11.7) we have

$$\frac{dy}{dx} = \frac{\dfrac{dy}{d\theta}}{\dfrac{dx}{d\theta}}\,.$$

Each of the two derivatives on the right hand side of the equality requires the use of the product rule to arrive at

$$\frac{dy}{dx} = \frac{f'(\theta)\sin(\theta) + f(\theta)\cos(\theta)}{f'(\theta)\cos(\theta) - f(\theta)\sin(\theta)}. \tag{11.8}$$

Example 11.11

Consider the limaçon $r = 1 + 2\sin(\theta)$ on $[0, 2\pi]$. Find the equations of the tangent and normal lines to the graph at $\theta = \pi/4$.

Solution

We start by computing $\frac{dy}{dx}$. With $f'(\theta) = 2\cos(\theta)$, we have

$$\begin{aligned}\frac{dy}{dx} &= \frac{2\cos(\theta)\sin(\theta) + \cos(\theta)(1+2\sin(\theta))}{2\cos^2(\theta) - \sin(\theta)(1+2\sin(\theta))} \\ &= \frac{\cos(\theta)(4\sin(\theta)+1)}{2(\cos^2(\theta)-\sin^2(\theta)) - \sin(\theta)}.\end{aligned}$$

When $\theta = \pi/4$, $\frac{dy}{dx} = -2\sqrt{2} - 1$. In rectangular coordinates, the point on the graph at $\theta = \pi/4$ is $(1+\sqrt{2}/2, 1+\sqrt{2}/2)$. Thus the rectangular equation of the line tangent to the limaçon at $\theta = \pi/4$ is

$$y = (-2\sqrt{2}-1)\left(x - \left(1 + \frac{\sqrt{2}}{2}\right)\right) + 1 + \frac{\sqrt{2}}{2} \approx -3.83x + 8.24\,.$$

The limaçon and the tangent line are graphed in Figure 11.13.

The normal line has the opposite–reciprocal slope as the tangent line, so its equation is

$$y \approx \frac{1}{3.83}x + 1.26.$$

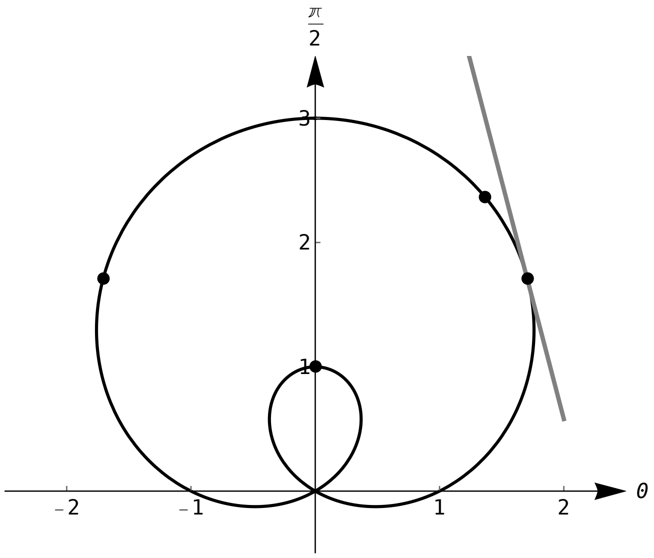


Figure 11.13: The limaçon in Example 11.11 with its tangent line at $\theta = \pi/4$ and points of vertical and horizontal tangency.

11.3.3 Smoothness

For what concerns the smoothness of parametric curves, we have the following – stricter – definition in order to arrive at curves without corners.

Definitie 11.3 (Smoothness of a parametric curve)
A curve C defined by $x = f(t)$, $y = g(t)$ is **smooth** (*glad*) on an interval I if f' and g' are continuous on I and not simultaneously 0 (except possibly at the endpoints of I). A curve is **piecewise smooth** (*stuksgewijs glad*) on I if I can be partitioned into subintervals where C is smooth on each subinterval.

The continuity condition is in agreement with Definition 9.7 and relates to parameterizations that could fail to be differentiable at a point. The second condition, however, relates to parameterizations that could slow to a stop, and then start up again in a completely different direction. Indeed, if a curve is not smooth at $t = t_0$, it means that $x'(t_0) = y'(t_0) = 0$ as defined. This, in turn, means that rate of change of x (and y) is 0; that is, at that instant, neither x nor y is changing. If the parametric equations describe the path of some object, this means the object is at rest at t_0. An object at rest can make a sharp change in direction, whereas moving objects tend to change direction in a smooth fashion.

Consider the astroid, given by $x = \cos^3(t)$, $y = \sin^3(t)$ (Figure 11.11). Taking derivatives, we have:

$$x' = -3\cos^2(t)\sin(t) \quad \text{and} \quad y' = 3\sin^2(t)\cos(t).$$

It is clear that each is 0 when $t = 0,\ \pi/2,\ \pi, \ldots$. Thus the astroid is not smooth at these points, corresponding to the cusps seen in Figure 11.11.

Example 11.12

Let a curve C be defined by the parametric equations $x = t^3 - 12t + 17$ and $y = t^2 - 4t + 8$. Determine the points, if any, where it is not smooth.

Solution

We begin by taking derivatives.

$$x' = 3t^2 - 12, \text{ and } y' = 2t - 4.$$

We set each equal to 0. It follows that

$$x' = 0 \quad \Leftrightarrow \quad 3t^2 - 12 = 0 \quad \Leftrightarrow \quad t = \pm 2,$$

and

$$y' = 0 \quad \Leftrightarrow \quad 2t - 4 = 0 \quad \Leftrightarrow \quad t = 2.$$

We see that at $t = 2$ both x' and y' are 0; thus C is not smooth at $t = 2$, corresponding to the point $(1, 4)$. The curve is graphed in Figure 11.14, illustrating the cusp at $(1, 4)$.

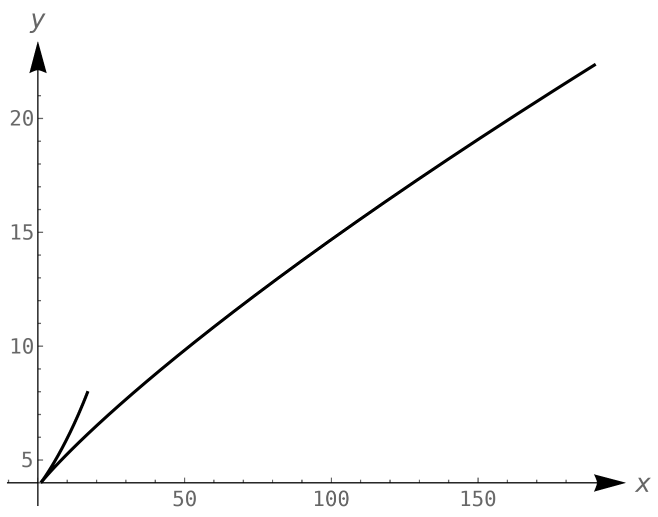


Figure 11.14: Graphing the curve in Example 11.12; note it is not smooth at $(1, 4)$.

One should be careful to note that a sharp corner does not have to occur when a curve is not smooth. For instance, one can verify that $x = t^3$ and $y = t^6$ produce the familiar $y = x^2$ parabola. However, in this parametrization, the curve is not smooth. A particle travelling along the parabola according to the given parametric equations comes to rest at $t = 0$, though no sharp point is created.

11.3.4 Concavity

For what concerns curves in the plane described by means of parametric equations, we may also consider their concavity; that is, we are interested in $\frac{d^2y}{dx^2}$. To find this, we need to find the derivative of $\frac{dy}{dx}$ with respect to x; that is,

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right],$$

but recall that $\frac{dy}{dx}$ is a function of t, not x, making this computation not straightforward.

Let now $h(t) = \frac{dy}{dx}$. We want $\frac{d}{dx}[h(t)]$, which follows from the chain rule. Indeed, we have

$$\frac{dh}{dt} = \frac{dh}{dx}\cdot\frac{dx}{dt} \quad\Rightarrow\quad \frac{dh}{dx} = \frac{\frac{dh}{dt}}{\frac{dx}{dt}}\,.$$

Hence, this leads to

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left[\frac{dy}{dx}\right]}{\frac{dx}{dt}} = \frac{\frac{d}{dt}\left[\frac{dy}{dx}\right]}{f'(t)}\,. \tag{11.9}$$

An example will help us understand this.

Example 11.13

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$ as in Example 11.9. Determine the t-intervals on which the graph is concave up/down.

Solution

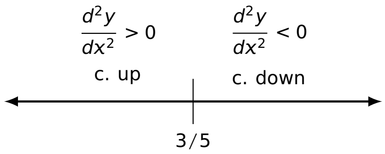
Concavity is determined by the second derivative of y with respect to x, $\frac{d^2y}{dx^2}$, so we compute that here following Equation (11.9).

In Example 11.9, we found $\frac{dy}{dx} = \frac{2t+6}{10t-6}$ and $f'(t) = 10t - 6$. So:

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{\frac{d}{dt}\left[\frac{2t+6}{10t-6}\right]}{10t-6}\\ &= -\frac{\frac{72}{(10t-6)^2}}{10t-6}\\ &= -\frac{72}{(10t-6)^3}\\ &= -\frac{9}{(5t-3)^3}.\end{aligned}$$

The graph of the parametric functions is concave up when $\frac{d^2y}{dx^2} > 0$ and concave down when $\frac{d^2y}{dx^2} < 0$. We determine the intervals when the second derivative is greater/less than 0 by first finding when it is 0 or undefined.

As the numerator of $-\frac{9}{(5t-3)^3}$ is never 0, $\frac{d^2y}{dx^2} \neq 0$ for all t. It is undefined when $5t-3=0$; that is, when $t = 3/5$. Following the work established in Section 10.4, we look at values of t greater/less than $3/5$ on a number line:



Reviewing Example 11.9, we see that when $t = 3/5 = 0.6$, the graph of the parametric equations has a vertical tangent line. This point is also a point of inflection for the graph, illustrated in Figure 11.15.

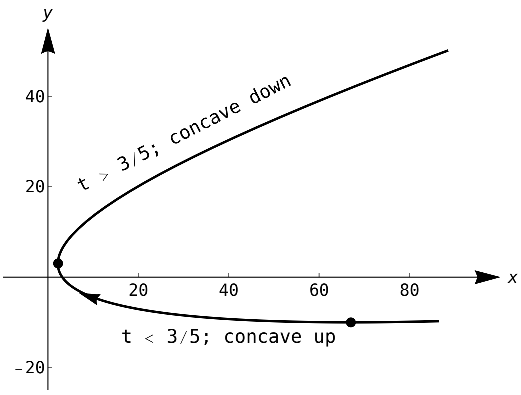


Figure 11.15: Graphing the parametric equations in Example 11.13 to demonstrate concavity.

11.4 Exercises

Polar coordinates

Assignment 11.1 — Which of the following pairs of polar coordinates represents the same point?

(a) $(3,0)=(-3,\pi)$

(b) $(-3,0)=(-3,2\pi)$

(c) $\left(2,\dfrac{2\pi}{3}\right)=\left(-2,-\dfrac{\pi}{3}\right)$

(d) $\left(2,\dfrac{7\pi}{3}\right)=\left(2,\dfrac{\pi}{3}\right)=\left(2,\dfrac{13\pi}{3}\right)$

Assignment 11.2 — Determine the Cartesian coordinates for the following points given in polar coordinates.

(a) $\left(\sqrt{2},\dfrac{\pi}{4}\right)$

(b) $(1,0)$

(c) $(0,\pi)$

(d) $\left(-\sqrt{2},\dfrac{\pi}{4}\right)$

(e) $\left(2\sqrt{3},\dfrac{2\pi}{3}\right)$

Assignment 11.3 — Convert the given polar equation into a Cartesian equation and name the curve.

(a) $\theta=\dfrac{\pi}{4}$

(b) $r=\dfrac{7}{5\sin(\theta)-2\cos(\theta)}$

(c) $r=2\cos(\theta)$

(d) $r=-4\sin(\theta)$

(e) $r=\sin(\theta)+\cos(\theta)$

(f) $r=\dfrac{2}{\sqrt{\cos^2(\theta)+4\sin^2(\theta)}}$

(g) $r=\dfrac{1}{1-\cos(\theta)}$

(h) $r=\dfrac{1}{1-2\sin(\theta)}$

Assignment 11.4 — Sketch the graph of the curves below.

(a) $r = 2 \qquad \left(0 \le \theta \le \dfrac{\pi}{2}\right)$

(b) $r = 2 - \sin(\theta)$

(c) $r = \dfrac{3}{2\cos(\theta) - \sin(\theta)}$

(d) $r = 2 + 4\cos(\theta)$

(e) $r = 5\sin(2\theta)$

(f) $r = 2\sin(2\theta)$

(g) $r = \cos\left(\dfrac{2\theta}{3}\right) \qquad (0 \le \theta \le 6\pi)$

(h) $r = 3\sin(\theta) \qquad (0 \le \theta \le \pi)$

(i) $r = 3\csc(\theta) \qquad (0 < \theta < \pi)$

(j) $r^2 = 4\sin(2\theta)$

(k) $r^2 = 4\cos(3\theta)$

(l) $r^2 = \sin(3\theta)$

(m) $r = a\sqrt{\cos(2\theta)} \qquad (a > 0)$

(n) $r = a\cos(n\theta) \qquad (a > 0, n \in \mathbb{Z})$

(o) $r = \dfrac{a}{\theta} \qquad (a > 0)$

Assignment 11.5 — Determine the intersection(s) of the graphs represented by the polar equations below.

(a) $r = 3\cos(\theta), \quad r = 1 + \cos(\theta)$

(b) $r = \sec^2\left(\dfrac{\theta}{2}\right), \quad r = 3\csc^2\left(\dfrac{\theta}{2}\right)$

(c) $r = \sin(\theta), \quad r = 1 - \sin(\theta)$

(d) $r = \sqrt{3}\cos(\theta), \quad r = \sin(\theta)$

(e) $r^2 = 2\cos(2\theta), \quad r = 1$

(f) $r = \sin(3\theta), \quad r = \cos(3\theta), \quad [0, \pi]$

(g) $r = 1 - \cos(\theta), \quad r = 1 + \sin(\theta), \ [0, 2\pi]$

Parametric equations

Assignment 11.6 — Determine the Cartesian equation of the given parameter representation and draw the corresponding curve.

(a) $\begin{cases} x = 2 - t \\ y = t + 1 \end{cases} \qquad (0 < t < +\infty)$

(b) $\begin{cases} x = \dfrac{1}{t} \\ y = t - 1 \end{cases} \qquad (0 < t < 4)$

(c) $\begin{cases} x = \dfrac{1}{1 + t^2} \\ y = \dfrac{t}{1 + t^2} \end{cases} \qquad (t \in \mathbb{R})$

(d) $\begin{cases} x = 3\sin(2t) \\ y = 3\cos(2t) \end{cases} \qquad \left(0 \le t \le \dfrac{\pi}{3}\right)$

(e) $\begin{cases} x = 1 - \sqrt{4 - t^2} \\ y = 2 + t \end{cases} \qquad (-2 \le t \le 2)$

(f) $\begin{cases} x = \sec(t) \\ y = \tan(t) \end{cases} \qquad (t \in \mathbb{R})$

(g) $\begin{cases} x = e^t \\ y = e^{3t} - 3 \end{cases} \qquad (t \in \mathbb{R})$

(h) $\begin{cases} x = \cos(\sin(s)) \\ y = \sin(\sin(s)) \end{cases} \qquad (s \in \mathbb{R})$

(i) $\begin{cases} x = \cosh(t) \\ y = \sinh(t) \end{cases} \qquad (t \in \mathbb{R})$

Assignment 11.7 — Determine a parametrization of the curves below.

(a) The lower half of the parabola $y^2 = x - 1$.

(b) $x^{2/3} + y^{2/3} = 6^{2/3}$

Assignment 11.8 — Use $t = y$ to paramterize the intersection of the planes $y = 2x - 4$ and $z = 3x + 1$ between $(2,0,7)$ and $(3,2,10)$.

Assignment 11.9 — The curve of intersection of the plane $x + y = 1$ with the paraboloid $z = x^2 + y^2$ is a parabola. Parameterize this parabola using $t = x$ as a parameter. Can you use $t = y$ as well? What about $t = z$?

Assignment 11.10 — Parameterize the curve that defines the intersection between the given curves.

(a) $x^2 + y^2 = 9 \quad \text{and} \quad z = x + y$

(b) $z = \sqrt{1 - x^2 - y^2} \quad \text{and} \quad x + y = 1$

(c) $z = x^2 + y^2 \quad \text{and} \quad 2x - 4y - z - 1 = 0$

Assignment 11.11 — Sketch the graph of the curves given by the parameter representations below.

(a) $\begin{cases} x = t^2 \\ y = 2 \end{cases} \quad (-2 \leq t \leq 2)$

(b) $\begin{cases} x = t - 1 \\ y = 2t + 3 \end{cases} \quad (-\infty < t < +\infty)$

(c) $\begin{cases} x = t^2 \\ y = t - 3 \end{cases} \quad (-\infty < t < +\infty)$

(d) $\begin{cases} x = \sin(t) \\ y = \cos^2(t) \end{cases} \quad \left(0 \leq t \leq \dfrac{3\pi}{2}\right)$

(e) $\begin{cases} x = -2\sin(t) \\ y = 3\cos(t) \end{cases} \quad (0 \leq t \leq 3\pi)$

(f) $\begin{cases} x = 2t \\ y = t^3 + 4 \end{cases} \quad (-2 \leq t \leq 2)$

Assignment 11.12 — Describe the similarities and differences between the graphs belonging to the parametric equations below.

(a) $x = \cos(t), \quad y = \sin(t) \qquad (0 \leq t \leq 2\pi)$

(b) $x = \cos\left(t^2\right), \quad y = \sin\left(t^2\right) \qquad (0 \leq t \leq 2\pi)$

(c) $x = \cos(1/t), \quad y = \sin(1/t) \qquad (0 < t < 1)$

(d) $x = \cos\big(\cos(t)\big), \quad y = \sin\big(\cos(t)\big) \qquad (0 \leq t \leq 2\pi)$

Assignment 11.13 — Determine which graphs of $x = f(t)$ and $y = f(t)$ in Figure 11.20 belong to the graphs of the parametric curves in Figure 11.21.

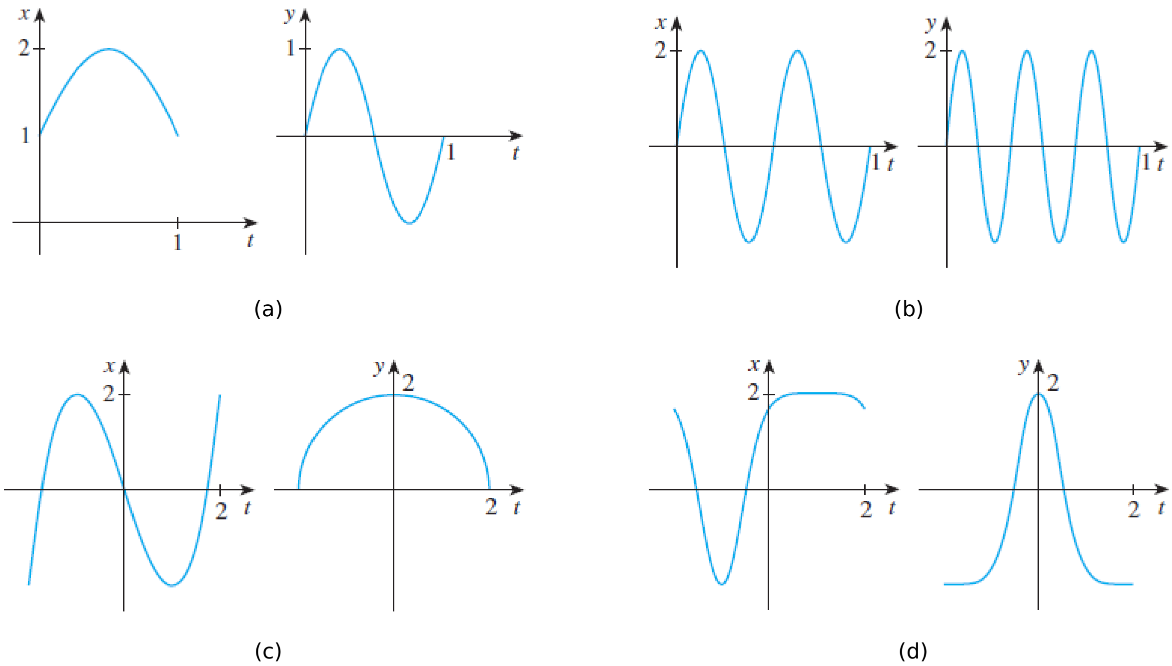


Figure 11.20: Graphs of the parametric equations from Exercise 13.

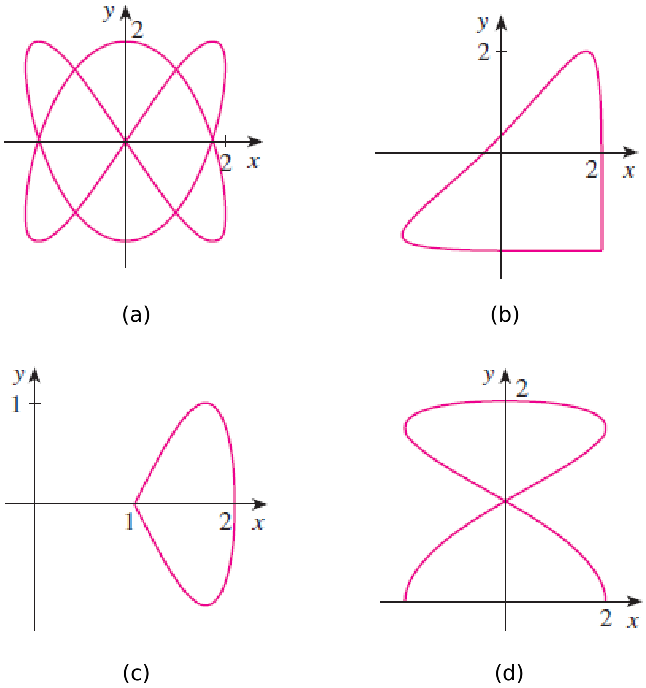


Figure 11.21: Graphs of the parameter curves from Exercise 13.

Assignment 11.14 — A hypocycloid is the curve that describes how a points moves on a circle that rolls without slipping in a larger circle. Suppose that the smallest circle has radius b and the largest circle has radius $a > b$, while the center of the latter is at the origin. The curve starts in $(a, 0)$.

Show that a hypocycloid is given by

$$\begin{cases} x = (a-b)\cos(t) + b\cos\left(\dfrac{a-b}{b}t\right) \\ y = (a-b)\sin(t) - b\sin\left(\dfrac{a-b}{b}t\right) \end{cases}$$

where t is the angle between the positive x-axis and the line through the origin and the point where the rolling circle touches the largest circle (see Figure 11.23). These parametric equations are expressions for the coordinates of the point P_t. Make use of this figure and determine the following to arrive at the parametric equations.

(a) Determine the coordinates of the point C_t as a function of a, b and t.

(b) Derive an expression for the x-coordinate of the point P_t as a function of a, b, t and θ_t. To this end, write x as the difference between two distances, being: $d(x_{C_t}, x_0)$ en $d(x_{P_t}, x_{C_t})$, where x_{C_t} represents the x-coordinate of the point C_t.

(c) Derive in the same manner an expression for the y-coordinate of the point P_t as a function of a, b, t and θ_t.

(d) Write θ_t as a function of a, b and t by using the arc length of $\widehat{AT_t}$ en $\widehat{T_tP_t}$. From this, the requested parametric equations follow.

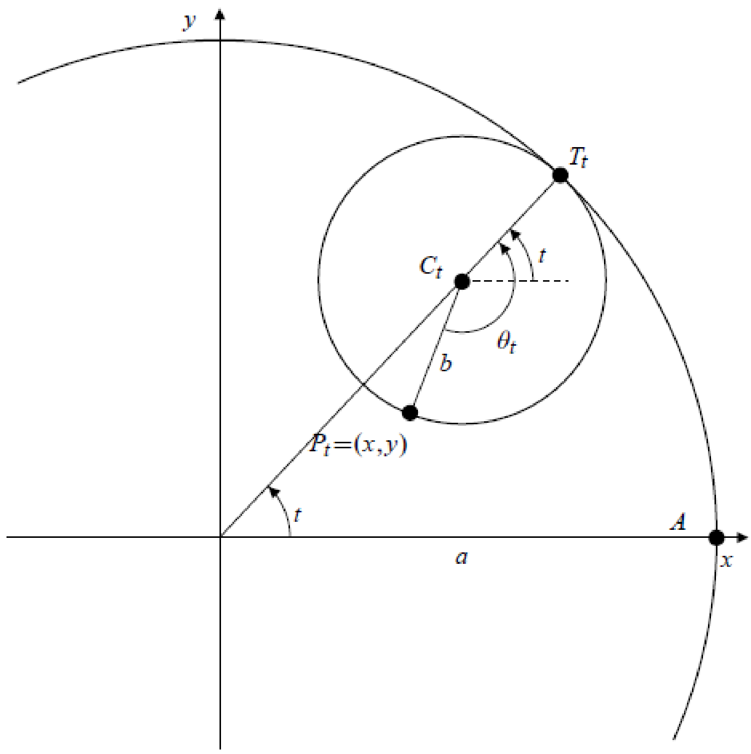


Figure 11.23: Figure from Exercise 11.14.

Prove that if $a = 2$ and $b = 1$ the hypocycloid becomes a line and if $a = 4$ and $b = 1$ the hypocycloid becomes an asteroid.

Derivatives and parametric and polar equations

Assignment 11.15 — Determine the slope of the tangent to the given curve at the given point.

(a) $r = 1 - 3\cos(\theta) \quad \text{in } \theta = \dfrac{3\pi}{4}$

(b) $r = \sin(4\theta) \quad \text{in } \theta = \dfrac{\pi}{3}$

(c) $x = t^3 + t, \quad y = 1 - t^3 \quad \text{in } t = 1$

(d) $x = e^{2t}, \quad y = te^{2t} \quad \text{in } t = -2$

Assignment 11.16 — Determine an equation of the tangent and normal at the given point to the given curve.

(a) $r = 1 + \sin(\theta) \quad \text{in } \theta = \dfrac{\pi}{6}$

(b) $r = \dfrac{1}{\sin(\theta) - \cos(\theta)} \quad \text{in } \theta = \pi$

(c) $x = t^2 - t, \quad y = t^2 + t \quad \text{in } t = 1$

(d) $x = \cos(t), \quad y = \sin(2t) \quad \big(t \in [0, 2\pi]\big) \quad \text{in } t = \pi/4$

(e) $x = e^{t/10}\cos(t), \quad y = e^{t/10}\sin(t) \quad \text{in } t = \pi/2$

Assignment 11.17 — Determine the coordinates of the points where the given curve has (a) a horizontal and (b) a vertical tangent.

(a) $x = t^3 - 3t, \quad y = 2t^3 + 3t^2$

(b) $x = \sin(t), \quad y = \sin(t) - t\cos(t)$

(c) $x = \dfrac{3t}{1+t^3}, \quad y = \dfrac{3t^2}{1+t^3}$

(d) $x = \sec(t), \quad y = \tan(t) \quad \left(-\dfrac{\pi}{2} < t < \dfrac{\pi}{2}\right)$

(e) $x = \cos(t)\sin(2t), \quad y = \sin(t)\sin(2t)$

(f) $r = 1 + \cos(\theta)$

(g) $r^2 = \cos(2\theta)$

(h) $r = 2(1 - \sin(\theta))$

Assignment 11.18 — Determine the values of t for which the given curve is not smooth.

(a) $x = t^2 - 4t, \quad y = t^3 - 2t^2 - 4t$

(b) $x = t\sin(t), \quad y = t^3$

(c) $x = 2\cos(t) - \cos(2t), \quad y = 2\sin(t) - \sin(2t)$

(d) $x = \dfrac{1}{t^2+1}, \quad y = t^3$

(e) $x = t^3 - 3t^2 + 3t - 1, \quad y = t^2 - 2t + 1$

(f) $x = \cos^2(t), \quad y = 1 - \sin^2(t)$

Assignment 11.19 — Sketch the graph of the given curve based on the first and second derivatives.

(a) $x = t^2 - 2t, \quad y = t^2 - 4t$

(b) $x = t^3 - 3t, \quad y = \dfrac{2}{1+t^2}$

(c) $x = \cos(t) + t\sin(t), \quad y = \sin(t) - t\cos(t) \qquad (t \geq 0)$

(d) $x = t^2 + t, \quad y = 1 - t^2 \qquad (-3 \leq t \leq 3)$

Nature laughs at the difficulties of integration.

— Pierre-Simon Laplace —

12

Integration

We have spent considerable time considering the derivatives of a function and their applications. In the following chapters, we are going to starting thinking in the other direction. That is, given a function $f(x)$, we are going to consider functions $F(x)$ such that $F'(x) = f(x)$. These functions will help us compute area, volume, mass, force, pressure, work, and much more.

12.1 Antiderivatives and (in)definite integration

12.1.1 Antiderivatives and indefinite integration

Given a function $y = f(x)$, a **differential equation** (*differentiaalvergelijking*) is one that incorporates y, x, and the derivatives of y. For instance, a simple differential equation is:

$$y' = 2x.$$

Solving a differential equation amounts to finding a function y that satisfies the given equation. Take a moment and consider that equation; can you find a function y such that $y' = 2x$?

Hopefully one was able to come up with at least one solution: $y = x^2$. Finding another may have seemed impossible until one realizes that a function like $y = x^2 + 1$ also has a derivative of $2x$. Once that discovery is made, finding yet another is not difficult; the function $y = x^2 + 123\,456\,789$ also has a derivative of $2x$. The differential equation $y' = 2x$ has many solutions. This leads us to some definitions.

Definitie 12.1 (Antiderivatives and indefinite integrals)
Let a function $f(x)$ be given. An **antiderivative** (*primitieve functie*) of $f(x)$ is a function $F(x)$ such that $F'(x) = f(x)$.

The set of all antiderivatives of $f(x)$ is the **indefinite integral** (*onbepaalde integraal*) of f, denoted by

$$\int f(x)\ dx.$$

Note that we refer to an antiderivative of f, as opposed to the antiderivative of f, since there is always an infinite number of them. We often use upper-case letters to denote antiderivatives. Besides, knowing one antiderivative of f allows us to find infinitely more, simply by adding a constant. Not only does this give us more antiderivatives, it gives us all of them.

Theorem 12.1 (Antiderivative forms)
Let $F(x)$ and $G(x)$ be antiderivatives of $f(x)$ on an interval I. Then there exists a constant C such that, on I,

$$G(x) = F(x) + C.$$

Given a function f defined on an interval I and one of its antiderivatives F, we know all antiderivatives of f on I have the form $F(x)+C$ for some constant C. Using Definition 12.1, we can say that

$$\int f(x)\ dx = F(x) + C.$$

The integration symbol, $\int$, is in reality an elongated S, representing summing. We will later see how sums and antiderivatives are related. The function we want to find an antiderivative of is called the **integrand** (*integrand*). It contains the differential of the variable we are integrating with respect to.

Let us now use our notice to evaluate

$$\int \sin(x)\ dx.$$

Essentially, this means that we should find all functions $F(x)$ such that $F'(x) = \sin(x)$. Of course, some thought leads us to one solution: $F(x) = -\cos(x)$, because $\frac{d}{dx}(-\cos(x)) = \sin(x)$. The indefinite integral of $\sin(x)$ is thus $-\cos(x)$, plus a constant of integration C. So:

$$\int \sin(x)\ dx = -\cos(x) + C.$$

To fully understand what is happening, it is important to realise that the process of antidifferentiation is really solving a differential question. The integral

$$\int \sin(x)\ dx$$

presents us with a differential, $dy = \sin(x)\ dx$. It is asking: What is y? We found lots of solutions, all of the form $y = -\cos(x) + C$.

Letting $dy = \sin(x)\ dx$, rewrite

$$\int \sin(x)\ dx \quad \text{as} \quad \int dy.$$

This is asking: “What functions have a differential of the form dy?” The answer is “Functions of the form $y + C$, where C is a constant.” What is y? We have lots of choices, all differing by a constant; the simplest choice is $y = -\cos(x)$.

In Mathematica, we can use the command **Integrate** to evaluate an indefinite integral. For instance,

$$\int (3x^2 + 4x + 5)\ dx.$$

can be evaluated as follows.

```
In[19]:= Integrate[3*x^2+4*x+5, x]

Out[19]= 5x +2x^2 +x^3
```

We can also see that taking the derivative of our answer returns the function in the integrand. Thus we can say that:

$$\frac{d}{dx}\left(\int f(x)\ dx\right) = f(x).$$

Differentiation undoes the work done by antidifferentiation.

Taking into account the lists of derivatives of algebraic and transcendental functions presented in Chapter 9, we may now state some important antiderivatives. We easily see that

$$\int 0\ dx = C\,,$$

and

$$\int 1\ dx = \int dx = x + C\,,$$

from which we can infer the following more general integral rule:

$$\int x^n\ dx = \frac{1}{n+1}x^{n+1} + C,$$

for $n \neq -1$.

For what concerns the exponential and logarithmic functions, we get the following derivative functions:

- $\int e^x\ dx = e^x + C,$
- $\int a^x\ dx = \frac{1}{\ln(a)}a^x + C,$
- $\int \frac{1}{x}\ dx = \ln|x| + C,$

while for the trigonometric and hyperbolic functions we get:

- $\int \sin(x)\ dx = -\cos(x) + C$
- $\int \cos(x)\ dx = \sin(x) + C$
- $\int \sec^2(x)\ dx = \tan(x) + C$
- $\int \csc^2(x)\ dx = -\cot(x) + C$
- $\int \sinh(x)\ dx = \cosh(x) + C$
- $\int \cosh(x)\ dx = \sinh(x) + C$
- $\int \frac{1}{\cosh^2(x)}\ dx = \tanh(x) + C$
- $\int \frac{1}{\sinh^2(x)}\ dx = -\coth(x) + C$

Besides, we have the following properties, which are completely in line with those for derivatives (Theorem 9.3)

Theorem 12.2 (Properties of the antiderivative)
Let f and g be differentiable on an open interval I and let k be a real number. Then:

1. *Sum/Difference rule:*
$$\int \big(f(x) \pm g(x)\big)\,dx = \int f(x)\,dx \pm \int g(x)\,dx\,. \tag{12.1}$$

2. *Constant multiple rule:*
$$\int k f(x)\,dx = k \int f(x)\,dx\,. \tag{12.2}$$

Proof For the sake of illustration, We will prove the sum rule. The proofs of the other properties proceed in a similar way.

Suppose that $F(x)$ is an anti-derivative of $f(x)$ and that $G(x)$ is an anti-derivative of $g(x)$. So we have that $F'(x) = f(x)$ and $G'(x) = g(x)$. Basic properties of derivatives also tell us that

$$\big(F(x) + G(x)\big)' = F'(x) + G'(x) = f(x) + g(x)\,,$$

and so $F(x) + G(x)$ is an anti-derivative of $f(x) + g(x)$. In other words,

$$\int f(x) + g(x)\,dx = F(x) + G(x) + C = \int f(x)\,dx + \int g(x)\,dx\,.$$

□

In Section 9.1.4 we saw that the derivative of a position function gave a velocity function, and the derivative of a velocity function describes acceleration. We can now go the other way: the antiderivative of an acceleration function gives a velocity function, etc. While there is just one derivative of a given function, there are infinitely many antiderivatives. Therefore we cannot ask “What is the velocity of an object whose acceleration is -32m/s^2?”, since there is more than one answer.

We can find the answer if we provide more information with the question, as done in the following example. Often the additional information comes in the form of an initial value, a value of the function that one knows beforehand.

Example 12.1

The acceleration due to gravity of a falling object is -9 m/s^2. At time $t = 3$, a falling object had a velocity of -10 m/s. Find the equation of the object’s velocity.

Solution

We want to know a velocity function, $v(t)$. We know two things:

- The acceleration, i.e., $v'(t) = -9$, and
- the velocity at a specific time, i.e., $v(3) = -10$.

Using the first piece of information, we know that $v(t)$ is an antiderivative of $v'(t) = -9$. So we begin by finding the indefinite integral of −9:

$$\int v'(t)\ dt = \int (-9)\ dt = -9t + C = v(t).$$

Now we use the fact that $v(3) = -10$ by plugging in this point in the equation we just got for $v(t)$:

$$-9 \cdot (3) + C = -10\,,$$

for which it directly follows that $C = 17$.

Thus $v(t) = -9t + 17$. We can use this equation to understand the motion of the object: when $t = 0$, the object had a velocity of $v(0) = 17$ m/s. Since the velocity is positive, the object was moving upward.

In the remainder of this section, we will see how position and velocity are unexpectedly related by the areas of certain regions on a graph of the velocity function.

12.1.2 The definite integral

We start with an easy problem. An object travels in a straight line at a constant velocity of 5m/s for 10 seconds. How far away from its starting point is the object?

Since, we have that Distance = Rate × Time, it follows that this distance is 50 metres. This solution can be represented graphically. Consider Figure 12.1(a), where the constant velocity of 5m/s is graphed on the axes. Shading the area under the line from $t = 0$ to $t = 10$ gives a rectangle with an area of 50 square units; when one considers the units of the axes, we can say this area represents 50 m.

Now consider a slightly harder situation (and not particularly realistic): an object travels in a straight line with a constant velocity of 5m/s for 10 seconds, then instantly reverses course at a rate of 2m/s for 4 seconds. How far away from the starting point is the object – what is its displacement?

Here, we get:

$$\text{Distance} \;\; = 5 \cdot 10 + (-2) \cdot 4 = 42 \text{ m}.$$

Hence the object is 42 metres from its starting location.

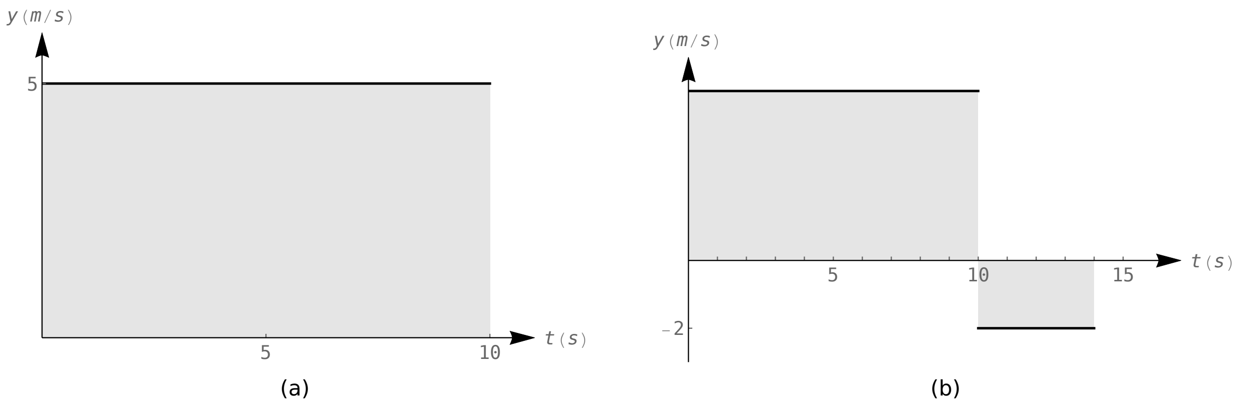


Figure 12.1: The total displacement of an object travelling in a straight line at a constant velocity of 5m/s for 10 seconds (a) and an object travelling a straight line with a constant velocity of 5m/s for 10 seconds, and then instantly reversing course at a rate of 2m/s for 4 seconds (b).

We can again depict this situation graphically. In Figure 12.1(b) we have the velocities graphed as straight lines on $[0, 10]$ and $[10, 14]$, respectively. The displacement of the object is given by

$$\text{Area above the } t\text{–axis} \quad - \quad \text{Area below the } t\text{–axis},$$

which is easy to calculate as $50 - 8 = 42$ metres.

These examples do not prove a relationship between area under a velocity function and displacement, but it does imply a relationship exists. Section 12.3 will fully establish fact that the area under a velocity function is displacement.

Anyhow, given a graph of a function $y = f(x)$, we will find that there is great use in computing the area between the curve $y = f(x)$ and the x-axis. Because of this, we need to define some terms.

Definitie 12.2 (The definite integral, total signed area)
Let $y = f(x)$ be defined on a closed interval $[a, b]$. The total signed area from $x = a$ to $x = b$ between f and the x-axis is:

$$(\text{area under } f \text{ and above the } x\text{–axis on } [a, b]) - (\text{area above } f \text{ and under the } x\text{–axis on } [a, b]).$$

The **definite integral** (*bepaalde integraal*) of f on $[a, b]$ is the total signed area of f on $[a, b]$, denoted

$$\int_a^b f(x)\, dx,$$

where a and b are the bounds of integration.

By our definition, the definite integral gives the signed area under f. We usually drop the word signed when talking about the definite integral, and simply say the definite integral gives the area under f or, more commonly, the area under the curve. The indefinite integral and definite integral are very much related, as we will see in Section 12.3.

Let us now practice this definition.

Example 12.2

Consider the function f given in Figure 12.2.

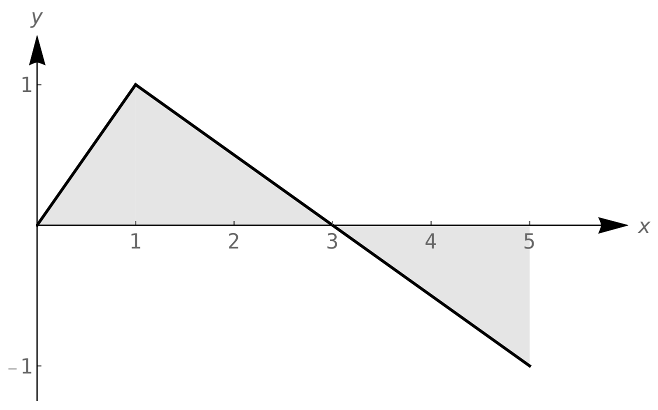


Figure 12.2: A graph of $f(x)$ in Example 12.2.

Find:

1. $\displaystyle\int_0^3 f(x)\ dx$
2. $\displaystyle\int_3^5 f(x)\ dx$
3. $\displaystyle\int_0^5 f(x)\ dx$
4. $\displaystyle\int_0^3 5f(x)\ dx$
5. $\displaystyle\int_1^1 f(x)\ dx$

Solution

1. This definite integral is the area under f on the interval $[0,3]$. This region is a triangle, so the area is
$$\int_0^3 f(x)\ dx = \frac{1}{2}(3)(1) = 1.5.$$
2. This definite integral represents the area of the triangle found under the x–axis on $[3,5]$. The area is $1/2(2)(1) = 1$; since it is found under the x–axis, this is negative area. So,
$$\int_3^5 f(x)\ dx = -1.$$
3. This definite integral is the total signed area under f on $[0,5]$. This is $1.5 + (-1) = 0.5$.
4. This definite integral is the area under $5f$ on $[0,3]$. Again, the region is a triangle, with height 5 times that of the height of the original triangle. Thus the area is
$$\int_0^3 5f(x)\ dx = \frac{1}{2}(15)(1) = 7.5.$$
5. This definite integral is the area under f on the interval $[1,1]$. This describes a line segment, not a region; it has no width. Therefore the area is 0.

This example illustrates some of the properties of the definite integral, listed in the following theorem.

Theorem 12.3 (Properties of the definite integral)
Let f and g be defined on a closed interval I that contains the values a, b and c, and let k be a constant. The following hold:

1. $\displaystyle\int_a^a f(x)\ dx = 0,$
2. $\displaystyle\int_a^b f(x)\ dx + \int_b^c f(x)\ dx = \int_a^c f(x)\ dx,$

3. $\displaystyle\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$,

4. $\displaystyle\int_a^b \big(f(x) \pm g(x)\big)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$,

5. $\displaystyle\int_a^b kf(x)\,dx = k \cdot \int_a^b f(x)\,dx$.

The proofs of these properties will be provided at the end of the next section once we have a better understanding of definite integrals through the conceptualisation of Riemann sums.

The area definition of the definite integral allows us to use geometry to compute the definite integral of some simple functions.

Example 12.3

Evaluate the following definite integrals:

1. $\displaystyle\int_{-2}^{5} (2x-4)\,dx$

2. $\displaystyle\int_{-3}^{3} \sqrt{9-x^2}\,dx$.

Solution

1. It is useful to sketch the function in the integrand, as shown in Figure 12.3(a). We see we need to compute the areas of two regions, which we have labelled R_1 and R_2. Both are triangles, so the area computation is straightforward:

$$R_1:\ \frac{1}{2}(4)(8) = 16 \qquad\qquad R_2:\ \frac{1}{2}(3)6 = 9.$$

Region R_1 lies under the x–axis, hence it is counted as negative area, so

$$\int_{-2}^{5} (2x-4)\,dx = -16+9 = -7.$$

We may check this answer in Mathematica as follows

```
In[20]:= Integrate[2*x-4, x,-2,5]

Out[20]= -7
```

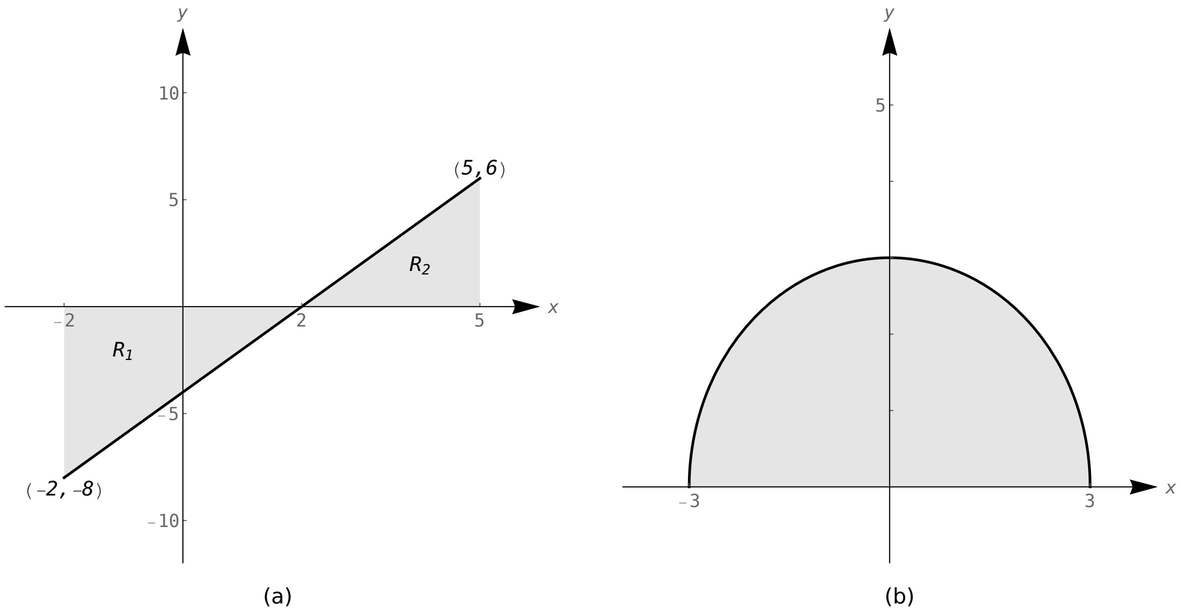


Figure 12.3: A graph of $f(x) = 2x - 4$ in (a) and $f(x) = \sqrt{9 - x^2}$ in (b), from Example 12.3.

2. Recognize that the integrand of this definite integral describes a half circle, as sketched in Figure 12.3(b), with radius 3. Thus the area is:

$$\int_{-3}^{3} \sqrt{9 - x^2}\, dx = \frac{1}{2}\pi r^2 = \frac{9}{2}\pi.$$

12.2 Riemann sums

In our previous examples, we have either found the areas of regions that have nice geometric shapes or the areas were given to us. But what is, for instance, the area of a region below $y = x^2$? The function $y = x^2$ is relatively simple, yet the shape it defines has an area that is not simple to find geometrically. In this section we will explore how to find the areas of such regions.

12.2.1 Approximating areas

Consider the region given in Figure 12.4, which is the area under $y = 4x - x^2$ on $[0, 4]$. What is the signed area of this region – i.e., what is $\int_0^4 (4x - x^2)\, dx$? We start by approximating. We can surround the region with a rectangle with height and width of 4 and find the area is approximately 16 square units. This is obviously an over–approximation; we are including area in the rectangle that is not under the parabola.

We have an approximation of the area, using one rectangle. How can we refine our approximation to make it better? The key to this section is this answer: use more rectangles. Let us use 4 rectangles with an equal width of 1. This partitions the interval $[0, 4]$ into 4 subintervals, $[0, 1]$, $[1, 2]$, $[2, 3]$ and $[3, 4]$. On each subinterval we will draw a rectangle.

There are three common ways to determine the height of these rectangles: the **left hand rule** (*linkerhand regel*), the **right hand rule** (*rechterhand regel*), and the **midpoint rule** (*midpoint regel*). The

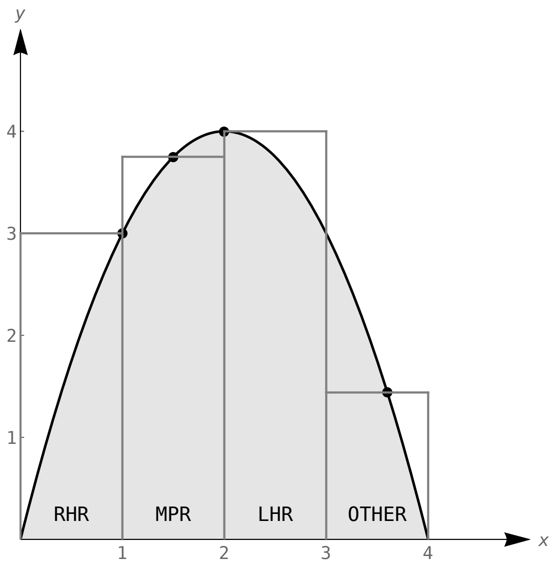


Figure 12.4: A graph of $f(x) = 4x - x^2$ and approximating $\int_0^4 (4x - x^2)\, dx$ using rectangles.

left hand rule says to evaluate the function at the left–hand endpoint of the subinterval and make the rectangle that height. In Figure 12.4, the rectangle drawn on the interval $[2,3]$ has height determined by the left hand rule (LHR); it has a height of $f(2)$.

The right hand rule (RHR) says the opposite: on each subinterval, evaluate the function at the right endpoint and make the rectangle that height. In Figure 12.4, the rectangle drawn on $[0,1]$ is drawn using $f(1)$ as its height. The midpoint rule (MPR) says to evaluate the function at the midpoint of each subinterval, and to make the rectangle that height. The rectangle drawn on $[1,2]$ was made using the midpoint rule, with a height of $f(1.5)$.

These are the three most common rules for determining the heights of approximating rectangles, but one is not forced to use one of these three methods. The rectangle on $[3,4]$ has a height of approximately $f(3.53)$, very close to the midpoint rule. It was chosen so that the area of the rectangle is exactly the area of the region under f on $[3,4]$.

It is hard to tell at this moment which is a better approximation. We can continue to refine our approximation by using more rectangles.

12.2.2 Riemann sums

Consider again $\int_0^4 (4x - x^2)\, dx$. We divide or partition the number line of $[0,4]$ into 16 equally spaced subintervals. We denote 0 as x_1 , so in general, we have

$$x_i = x_1 + (i-1)\Delta x\,,$$

where $i = 1, 2, \ldots, 16$ For the sake of simplicity, we will often write $\Delta x = \Delta x_i$, where Δx_i is the width of the i^{th} subinterval, whenever the width of the subintervals is the same.

Given any subdivision of $[0,4]$, the first subinterval is $[x_1, x_2]$; the second is $[x_2, x_3]$; the i^{th} subinterval is $[x_i, x_{i+1}]$. Hence, when using the left hand rule, the height of the i^{th} rectangle will be $f(x_i)$. When using the right hand rule, the height of the i^{th} rectangle will be $f(x_{i+1})$, and finally, when using the midpoint rule, the height of the i^{th} rectangle will be

$$f\left(\frac{x_i + x_{i+1}}{2}\right).$$

We illustrate this in the next example.

Example 12.4

Approximate

$$\int_0^4 (4x-x^2)\ dx$$

using the right hand rule with 16 and 1000 equally spaced intervals.

Solution

Using 16 equally spaced intervals and the right hand rule, we can approximate the definite integral as

$$\sum_{i=1}^{16} f(x_{i+1})\Delta x,$$

where we have $\Delta x = 4/16 = 0.25$. Moreover, since $x_1 = 0$, we have

$$\begin{aligned} x_{i+1} &= 0 + \big((i+1)-1\big)\Delta x \\ &= i\Delta x\,. \end{aligned}$$

Using summation formulas, we may now consider:

$$\begin{aligned} \int_0^4 (4x-x^2)\ dx \approx \sum_{i=1}^{16} f(x_{i+1})\Delta x &= \sum_{i=1}^{16} f(i\Delta x)\Delta x \\ &= \sum_{i=1}^{16}\big(4i\Delta x-(i\Delta x)^2\big)\Delta x = \sum_{i=1}^{16}(4i\Delta x^2 - i^2\Delta x^3) \\ &= (4\Delta x^2)\sum_{i=1}^{16} i - \Delta x^3\sum_{i=1}^{16} i^2 \qquad (12.3) \\ &= (4\Delta x^2)\frac{16\cdot 17}{2} - \Delta x^3\frac{16(17)(33)}{6} = 10.625 \qquad (\Delta x = 0.25) \end{aligned}$$

(12.4)

We were able to sum up the areas of 16 rectangles with very little computation. In Figure 12.5 the function and the 16 rectangles are graphed. While some rectangles over–approximate the area, other under–approximate the area by about the same amount. Thus our approximate area of 10.625 is likely a fairly good approximation.

For what concerns the approximation based on 1000 equally spaced, we can just use Equation (12.3); after replacing the 16's to 1000's and appropriately changing the value of Δx.

We do so here, skipping from the original summand to the equivalent of Equation (12.3) to save space. Note that $\Delta x = 4/1000 = 0.004$.

$$\int_0^4 (4x-x^2)\ dx \approx \sum_{i=1}^{1000} f(x_{i+1})\Delta x$$

$$
\begin{aligned}
&= (4\Delta x^2) \sum_{i=1}^{1000} i - \Delta x^3 \sum_{i=1}^{1000} i^2 \\
&= (4\Delta x^2) \frac{1000 \cdot 1001}{2} - \Delta x^3 \frac{1000(1001)(2001)}{6} \\
&= 10.666656
\end{aligned}
$$

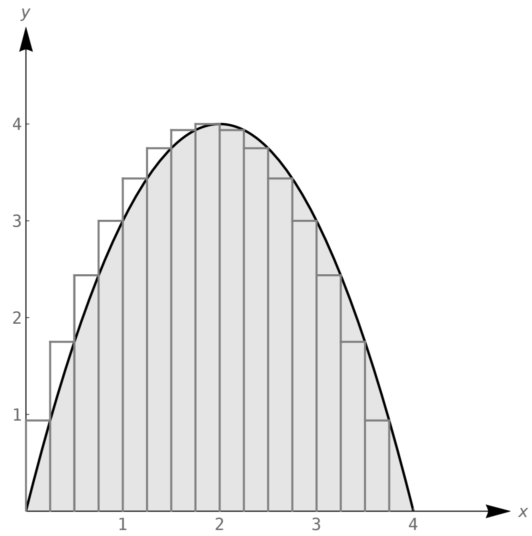


Figure 12.5: Approximating $\int_0^4 (4x - x^2)\ dx$ with the right hand rule and 16 evenly spaced subintervals.

Using many, many rectangles, we have a likely good approximation of $\int_0^4 (4x - x^2)\Delta x$. That is,

$$\int_0^4 (4x - x^2)\ dx \approx 10.666656.$$

Instead of approximating a definite integral using rectangles of the same width and height determined by evaluating f at a particular point in each consecutive subinterval, we could partition an interval $[a,b]$ with subintervals that do not have the same size. We refer to the length of the i^{th} subinterval as Δx_i. Also, one could determine each rectangle's height by evaluating f at any point c_i in the i^{th} subinterval. Thus the height of the i^{th} subinterval would be $f(c_i)$, and the area of the i^{th} rectangle would then be $f(c_i)\Delta x_i$.

These ideas are formally defined below.

Definitie 12.3 (Partition)
A **partition** (*partitie*) of a closed interval $[a,b]$ is a set of numbers $x_1, x_2, \ldots x_{n+1}$ where

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b.$$

The length of the i^{th} subinterval, $[x_i, x_{i+1}]$, is $\Delta x_i = x_{i+1} - x_i$. If $[a,b]$ is partitioned into subintervals of equal length, we let Δx_i represent the length of each subinterval.

The size of the partition, denoted $\mathcal{L}$, is the length of the largest subinterval of the partition, i.e. $\mathcal{L} = \max_i (\Delta x_i)$.

Summations of rectangles with area $f(c_i)\Delta x_i$ are named after mathematician Georg Friedrich Bernhard Riemann, as given in the following definition.

Definitie 12.4 (Riemann sum)
Let f be defined on a closed interval $[a,b]$, let $\{x_1, x_2, \ldots, x_{n+1}\}$ be a partition of $[a,b]$ and let c_i denote any value in the i^{th} subinterval.

The sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i$$

is a **Riemann sum** (*Riemann som*) of f on $[a,b]$.

Usually Riemann sums are calculated using one of the three methods we have introduced. The uniformity of construction makes computations easier. So

$$\int_a^b f(x)\, dx$$

is typically approximated by means of the following Riemann sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i\,,$$

for which we take the following steps.

1. Divide the interval $[a,b]$ in n subintervals have equal length, such that

   $$\Delta x_i = \Delta x = \frac{b-a}{n}$$

   and the i^{th} term of the equally spaced partition is

   $$x_i = a + (i-1)\Delta x\,.$$

   Thus $x_1 = a$ and $x_{n+1} = b$.

2. Evaluate one of the following summations:

   (a) using the left hand rule we get the so-called **left Riemann sum** (*linker Riemann som*):

   $$\sum_{i=1}^{n} f(x_i)\Delta x,$$

   (b) using the right hand rule we get the so-called **right Riemann sum** (*rechter Riemann som*):

   $$\sum_{i=1}^{n} f(x_{i+1})\Delta x,$$

   (c) and using the midpoint rule we get the **middle Riemann sum** (*midden Riemann som*):

   $$\sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x\,.$$

Example 12.5

Revisit

$$\int_0^4 (4x - x^2)\ dx$$

yet again. Approximate this definite integral using the right hand rule with n equally spaced subintervals.

Solution

We know $\Delta x = (4-0)/n = 4/n$. We also find $x_i = 0 + \Delta x(i-1) = 4(i-1)/n$. The right hand rule uses x_{i+1}, which is $x_{i+1} = 4i/n$.

We construct the right Riemann sum as follows.

$$\begin{aligned}
\int_0^4 (4x - x^2)\ dx &\approx \sum_{i=1}^{n} f(x_{i+1})\Delta x \\
&= \sum_{i=1}^{n} f\left(\frac{4i}{n}\right)\Delta x \\
&= \sum_{i=1}^{n} \left[4\frac{4i}{n} - \left(\frac{4i}{n}\right)^2\right]\Delta x \\
&= \sum_{i=1}^{n}\left(\frac{16\Delta x}{n}\right)i - \sum_{i=1}^{n}\left(\frac{16\Delta x}{n^2}\right)i^2 \\
&= \left(\frac{16\Delta x}{n}\right)\sum_{i=1}^{n} i - \left(\frac{16\Delta x}{n^2}\right)\sum_{i=1}^{n} i^2 \\
&= \left(\frac{16\Delta x}{n}\right)\cdot\frac{n(n+1)}{2} - \left(\frac{16\Delta x}{n^2}\right)\frac{n(n+1)(2n+1)}{6} \\
&= \frac{32(n+1)}{n} - \frac{32(n+1)(2n+1)}{3n^2} \\
&= \frac{32}{3}\left(1 - \frac{1}{n^2}\right)
\end{aligned}$$

The result is an amazing, easy to use formula. To approximate the definite integral with 10 equally spaced subintervals and the right hand rule, set $n = 10$ and compute

$$\int_0^4 (4x - x^2)\ dx \approx \frac{32}{3}\left(1 - \frac{1}{10^2}\right) = 10.56.$$

Recall how earlier we approximated the definite integral with 4 subintervals; with $n = 4$, the formula gives 10, our answer as before.

We now take an important leap. More precisely, for any finite n, we know that

$$\int_0^4 (4x - x^2)\, dx \approx \frac{32}{3}\left(1 - \frac{1}{n^2}\right).$$

Both common sense and high–level mathematics tell us that as n gets large, the approximation gets better. In fact, if we take the limit as $n \to +\infty$, we get the exact are we are looking for, that is:

$$\begin{aligned}\int_0^4 (4x - x^2)\, dx &= \lim_{n\to+\infty} \frac{32}{3}\left(1 - \frac{1}{n^2}\right)\\ &= \frac{32}{3}(1-0)\\ &= \frac{32}{3}.\end{aligned}$$

This is a fantastic result. By considering n equally–spaced subintervals, we obtained a formula for an approximation of the definite integral that involved our variable n. As n grows large – without bound – the error shrinks to zero and we obtain the exact area.

In addition to the left, right and middle Riemann sums, also **upper and lower Riemann sums** (*boven en onder Riemann som*) can be defined. For that purpose, we consider a partition as before, and note that f has both a minimum and maximum on $[x_i, x_{i+1}]$, so there are numbers l_i and u_i in $[x_i, x_{i+1}]$ such that

$$f(l_i) \le f(x) \le f(u_i)$$

for all x in $[x_i, x_{i+1}]$. If $f(x) \ge 0$, $f(l_i)\Delta x_i$ and $f(u_i)\Delta x_i$ represent the areas of rectangles having the interval $[x_i, x_{i+1}]$ as basis and having tops passing through the lowest and highest points on the graph of f on that interval (Figure 12.6). Clearly, if A_i is the area under the graph of f and above the horizontal axis, enclosed between the straight lines $x = x_i$ and $x + x_{i+1}$, then it holds that

$$f(l_i)\Delta x_i \le A_i \le f(u_i)\Delta x_i.$$

If f is not restricted to the positive half plan, then either one or both $f(l_i)\Delta x_i$ and $f(u_i)\Delta x_i$ can be negative and will then represent the area of a rectangle lying below the x-axis. Anyhow, it will always hold that $f(l_i)\Delta x_i \le f(u_i)\Delta x_i$.

With this notation in place we can define the lower Riemann sum as

$$S_l(n) = \sum_{i=1}^{n} f(l_i)\Delta x\,,$$

and the upper Riemann sum as

$$S_u(n) = \sum_{i=1}^{n} f(u_i)\Delta x\,.$$

To illustrate the subtle difference between the left and lower Riemann sums, on the one hand, and the lower Riemann sum, for instance, on the other hand, consider Figure 12.7, where the area under the sine curve between $x = 0$ and $x = \pi$ is approximated using the latter. From this figure, it should be clear that lower Riemann sum agrees with the left Riemann sum where the sine is increasing, whereas it corresponds with the right Riemann sum on the interval where the sine curve is decreasing.

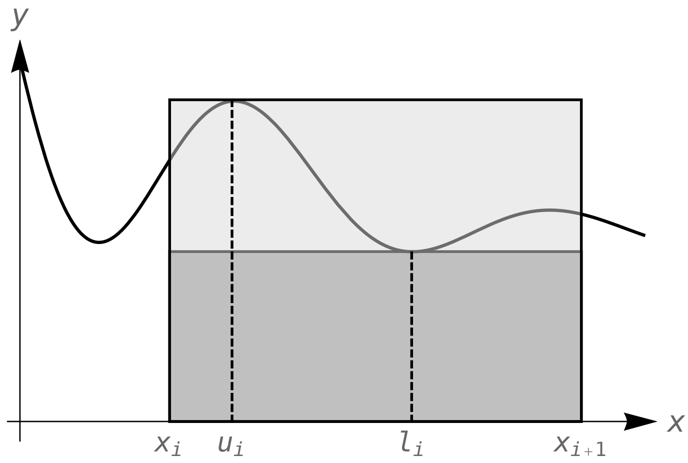


Figure 12.6: f has both a minimum and maximum on $[x_i, x_{i+1}]$.

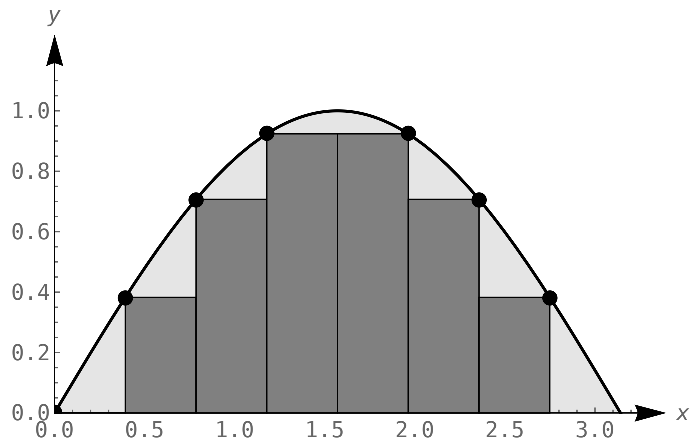


Figure 12.7: Distinction between the left and right Riemann sums and the lower Riemann sum.

12.2.3 Limits of Riemann sums

We have used limits to evaluate given definite integrals. Will this always work? We will show, given not–very–restrictive conditions, that yes, it will always work.

The previous example has shown us how we can think of a summation as a function of n. More precisely, given a definite integral $\int_a^b f(x)\,dx$, we let:

- $S_L(n) = \sum_{i=1}^{n} f(x_i)\Delta x$, be the left Riemann sum,
- $S_R(n) = \sum_{i=1}^{n} f(x_{i+1})\Delta x$, be the right Riemann sum,
- $S_M(n) = \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x$, be the sum of equally spaced rectangles formed using the midpoint rule,

and likewise for the lower and upper Riemann sums. Now, recall that the definition of the limit $\lim_{n\to+\infty} S_L(n) = K$ implies that given any $\epsilon > 0$, there exists $N > 0$ such that

$$\left|S_L(n) - K\right| < \epsilon,$$

when $n \geq N$.

The following theorem states that we can use any of our three rules to find the exact value of a definite integral.

Theorem 12.4 (Definite integrals and the limit of Riemann sums)
Let f be continuous on the closed interval $[a,b]$ and let $S_L(n)$, $S_R(n)$, $S_M(n)$, Δx, Δx_i and c_i be defined as before. Then:

1. $\lim\limits_{n\to+\infty} S_L(n) = \lim\limits_{n\to+\infty} S_R(n) = \lim\limits_{n\to+\infty} S_M(n) = \lim\limits_{n\to+\infty} \sum\limits_{i=1}^{n} f(c_i)\Delta x,$

2. $\lim\limits_{n\to+\infty} \sum\limits_{i=1}^{n} f(c_i)\Delta x = \int\limits_a^b f(x)\ dx$, *and*

3. $\lim\limits_{\mathcal{L}\to 0} \sum\limits_{i=1}^{n} f(c_i)\Delta x_i = \int\limits_a^b f(x)\ dx.$

This theorem also goes two steps further. It states that the height of each rectangle does not have to be determined following a specific rule, but could be $f(c_i)$, where c_i is any point in the i^{th} subinterval. Furthermore, it goes on to state that the rectangles do not need to be of the same width.

Let $\mathcal{L}$ represent the length of the largest subinterval in the partition: that is, $\mathcal{L}$ is the largest of all the Δx_i's. If $\mathcal{L}$ is small, then $[a,b]$ must be partitioned into many subintervals, since all subintervals must have small lengths. Taking the limit as $\mathcal{L}$ goes to zero implies that the number n of subintervals in the partition is growing to infinity, as the largest subinterval length is becoming arbitrarily small. We then interpret the expression

$$\lim_{\mathcal{L}\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i$$

as the limit of the sum of the areas of rectangles, where the width of each rectangle can be different but getting small, and the height of each rectangle is not necessarily determined by a particular rule. The theorem states that this Riemann sum also gives the value of the definite integral of f over $[a,b]$.

Having a better understanding of the definite integral in terms of Riemann sums, we are now ready to prove the properties listed in Theorem 12.3.

Proof (of Theorem 12.3) We prove the fourth statement in this theorem, namely that

$$\int\limits_a^b \big(f(x) \pm g(x)\big)\ dx = \int\limits_a^b f(x)\ dx \pm \int\limits_a^b g(x)\ dx\,.$$

The proofs of the other properties proceed in a similar way.

First we will prove the sum rule. From the definition of the definite integral we have,

$$\begin{aligned}
\int\limits_a^b \big(f(x)+g(x)\big)\ dx &= \lim_{n\to+\infty} \sum_{i=1}^{n} \Big(f\left(x_i^*\right) + g\left(x_i^*\right)\Big)\Delta x \\
&= \lim_{n\to+\infty} \left(\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x + \sum_{i=1}^{n} g\left(x_i^*\right)\Delta x\right) \\
&= \lim_{n\to+\infty} \sum_{i=1}^{n} f\left(x_i^*\right)\Delta x + \lim_{n\to+\infty} \sum_{i=1}^{n} g\left(x_i^*\right)\Delta x
\end{aligned}$$

$$= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$

To prove the difference formula we can either redo the above work with a minus sign instead of a plus sign or we can use the fact that we now know this is true with a plus and using the properties proved above as follows.

$$\begin{aligned}\int_a^b \big(f(x) - g(x)\big)\,dx &= \int_a^b f(x) + \big(-g(x)\big)\,dx \\ &= \int_a^b f(x)\,dx + \int_a^b \big(-g(x)\big)\,dx \\ &= \int_a^b f(x)\,dx - \int_a^b g(x)\,dx\end{aligned}$$ □

By resorting to Riemann sums we can also prove some properties related to the magnitude of a definite integral. These are listed in the following theorem.

Theorem 12.5 (Properties of the magnitude of a definite integral)
Let f and g be defined on a closed interval I that contains the values a and b, and let m and M be constants. The following hold:

1. *If $f(x) \geq 0$ for $a \leq x \leq b$, then*
$$\int_a^b f(x)\,dx \geq 0.$$
2. *If $f(x) \geq g(x)$ for $a \leq x \leq b$ then*
$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$
3. *If $m \leq f(x) \leq M$ for $a \leq x \leq b$ then*
$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$
4. $$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.$$

Proof We will prove the first property in this theorem. From the definition of the definite integral we have

$$\int_a^b f(x)\,dx = \lim_{n\to+\infty} \sum_{i=1}^{n} f\left(x_i^*\right)\Delta x,$$

where $\Delta x = \frac{b-a}{n}$. Now, by assumption $f(x) \geq 0$ and we also have $\Delta x > 0$ and so we know that

$$\sum_{i=1}^{n} f\left(x_i^*\right) \Delta x \geq 0\,.$$

So, from the basic properties of limits we then have

$$\lim_{n \to +\infty} \sum_{i=1}^{n} f\left(x_i^*\right) \Delta x \geq \lim_{n \to +\infty} 0 = 0\,.$$

But the left side is exactly the definition of the integral and so we have

$$\int_a^b f(x)\, dx = \lim_{n \to +\infty} \sum_{i=1}^{n} f\left(x_i^*\right) \Delta x \geq 0\,.$$

We now know of a way to evaluate a definite integral using limits; in the next section we will see how the fundamental theorem of calculus makes the process simpler. The key feature of this theorem is its connection between the indefinite integral and the definite integral.

Lebesgue integration

The integral we study within the framework of this course, the so-called Riemann integral, is just one kind of integral that has been proposed. While the Riemann integral considers the area under a curve as made out of vertical rectangles, the Lebesgue definition considers horizontal slabs that are not necessarily just rectangles, and so it is more flexible. For this reason, the Lebesgue definition makes it possible to calculate integrals for a broader class of functions. How the Lebesgue integral differs from the Riemann integral is illustrated in Figure 12.8 for a function f. Essentially, to compute the latter, one partitions the domain of f into subintervals, while for the latter one partitions the range of f.

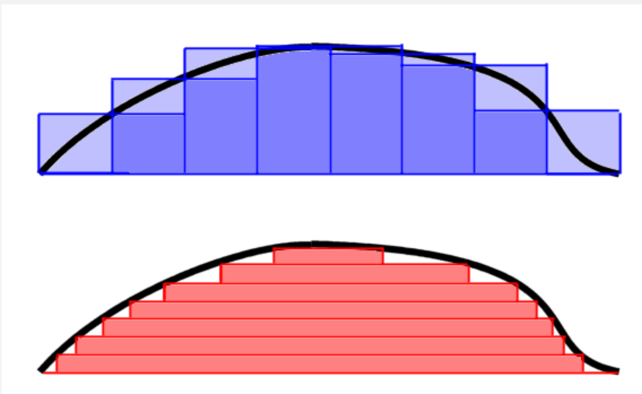


Figure 12.8: Riemann (top) versus Lebesgue integration (bottom).

12.3 The fundamental theorem of calculus

12.3.1 Mean value theorem for definite integrals

Consider the graph of a function f in Figure 12.9(a) and the area defined by $\int_1^4 f(x)\, dx$. In Figure 12.9(b), the height of the rectangle is greater than f on $[1, 4]$, hence the area of this rectangle is greater than $\int_1^4 f(x)\, dx$. In Figure 12.9(c), the height of the rectangle is smaller than f on $[1, 4]$, hence the area of this rectangle is less than $\int_1^4 f(x)\ dx$. Finally, in Figure 12.9(d) the height of the rectangle is such that the area of the rectangle is exactly that of $\int_1^4 f(x)\ dx$. Since rectangles that are too big, as in Figure 12.9(b), and rectangles that are too little, as in Figure 12.9(c), give areas greater/lesser than $\int_1^4 f(x)\ dx$, it makes sense that there is a rectangle, whose top intersects $f(x)$ somewhere on $[1,4]$, whose area is exactly that of the definite integral.

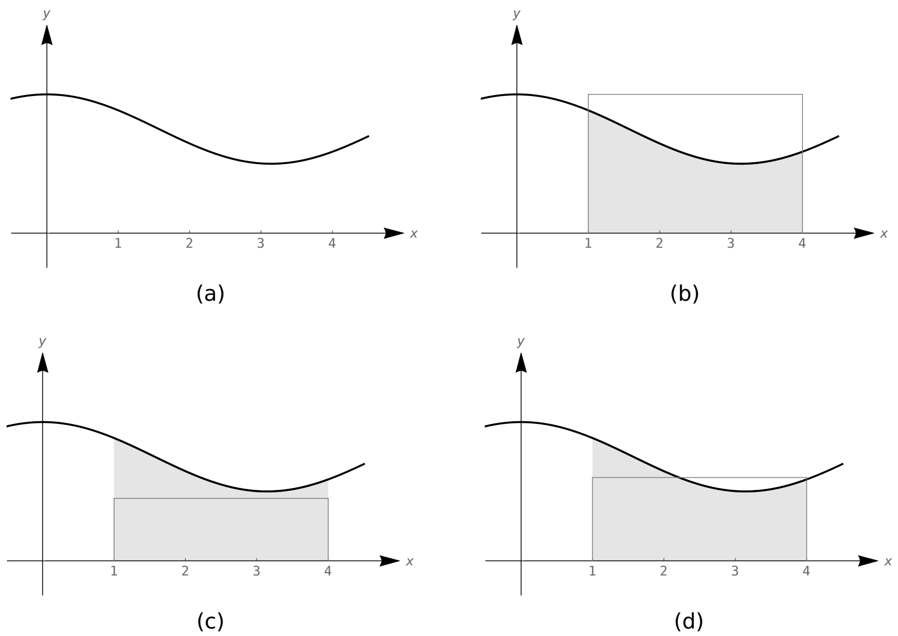


Figure 12.9: The graph of a function f (a) and differently sized rectangles give upper and lower bounds on $\int_1^4 f(x)\ dx$ (b-c).

We state this idea formally in a theorem.

Theorem 12.6 (The mean value theorem of integration)
Let f be continuous on $[a,b]$. There exists a value c in $]a,b[$ such that

$$\int_a^b f(x)\ dx = f(c)(b-a).$$

This is an existential statement; c exists, but we do not provide a method of finding it. Theorem 12.6 is directly connected to the mean value theorem of differentiation (Theorem 10.4).

Proof Let us prove this theorem by considering a more general formulation. Namely, if $f:[a,b]\to\mathbb{R}$ is continuous and g is an integrable function that does not change sign on $[a,b]$, then there exists c in $]a,b[$ such that

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx. \tag{12.5}$$

Clearly, we obtain the expression used in Theorem 12.6 by letting $g(x)=1$.

Now to prove Equation (12.5). Let us assume that $f:[a,b]\to\mathbb{R}$ is continuous and g is a nonnegative integrable function on $[a,b]$. By the extreme value theorem (Theorem 10.1), there exists m and M such that for each x in $[a,b]$, it holds that $m\le f(x)\le M$ and $f[a,b]=[m,M]$. Since g is nonnegative, we may write that

$$m\int_a^b g(x)\,dx \leqslant \int_a^b f(x)g(x)\,dx \leqslant M\int_a^b g(x)\,dx.$$

Now let

$$I = \int_a^b g(x)\,dx.$$

Obviously, if $I = 0$, we are done since

$$0 \leqslant \int_a^b f(x)g(x)\,dx \leqslant 0$$

means

$$\int_a^b f(x)g(x)\,dx = 0,$$

so for any c in $]\,a,b\,[$,

$$\int_a^b f(x)g(x)\,dx = f(c)I = 0.$$

If $I \neq 0$, then

$$m \leqslant \frac{1}{I}\int_a^b f(x)g(x)\,dx \leqslant M.$$

By the intermediate value theorem, f attains every value of the interval $[m,M]$, so for some c in $]\,a,b\,[$ we have

$$f(c) = \frac{1}{I}\int_a^b f(x)g(x)\,dx,$$

that is,

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx.$$

Finally, if g is negative on $[a,b]$, then

$$M\int_a^b g(x)\,dx \leqslant \int_a^b f(x)g(x)\,dx \leqslant m\int_a^b g(x)\,dx,$$

and we still get the same result as above. □

12.3.2 Main theorems

Let $f(t)$ be a continuous function defined on $[a,b]$. The definite integral $\int_a^b f(x)\,dx$ is the area under f on $[a,b]$. We can turn this concept into a function by letting the upper (or lower) bound vary.

Let $F(x) = \int_a^x f(t)\,dt$. It computes the area under f on $[a,x]$ as illustrated in Figure 12.10. We can study this function using our knowledge of the definite integral.

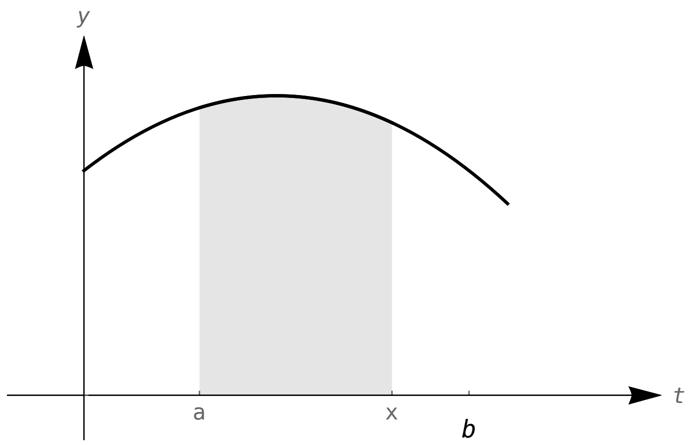


Figure 12.10: The area of the shaded region is $F(x) = \int_a^x f(t)\, dt$.

We can also apply calculus ideas to $F(x)$; in particular, we can compute its derivative. While this may seem like an innocuous thing to do, it has far–reaching implications, as demonstrated by the fact that the result is given as an important theorem.

Theorem 12.7 (The fundamental theorem of calculus, Part 1)
Let f be continuous on $[a,b]$ and let $F(x) = \int_a^x f(t)\,dt$. Then F is a differentiable function on $]a,b[$, and

$$F'(x) = f(x).$$

Proof For a given $f(t)$, let us define the function $F(x)$ as

$$F(x) = \int_a^x f(t)\,dt.$$

For any two numbers x_1 and $x_1 + \Delta x$ in $[a,b]$, we have

$$F(x_1) = \int_a^{x_1} f(t)\,dt$$

and likewise

$$F(x_1 + \Delta x) = \int_a^{x_1+\Delta x} f(t)\,dt.$$

Subtracting these two equalities yields

$$F(x_1 + \Delta x) - F(x_1) = \int_a^{x_1+\Delta x} f(t)\,dt - \int_a^{x_1} f(t)\,dt. \tag{12.6}$$

Using Theorem 12.3 we can rewrite the right hand side of Equation (12.6) as

$$\int_a^{x_1+\Delta x} f(t)\,dt - \int_a^{x_1} f(t)\,dt = \int_a^{x_1+\Delta x} f(t)\,dt + \int_{x_1}^{a} f(t)\,dt = \int_{x_1}^{x_1+\Delta x} f(t)\,dt.$$

Hence, Equation (12.6) becomes

$$F(x_1 + \Delta x) - F(x_1) = \int_{x_1}^{x_1+\Delta x} f(t)\,dt. \tag{12.7}$$

According to the mean value theorem for integration (Theorem 12.6), there exists a real number $c \in [x_1, x_1 + \Delta x]$ such that

$$\int_{x_1}^{x_1+\Delta x} f(t)\,dt = f(c) \cdot \Delta x.$$

This expression allows us to rewrite Equation (12.7) as

$$F(x_1 + \Delta x) - F(x_1) = f(c) \cdot \Delta x. \tag{12.8}$$

Note that we just write c in order not to overload the notation, but one should keep in mind that, for a given function f, the value of c depends on x_1 and on Δx, though it is always confined to the interval $[x_1, x_1 + \Delta x]$.

Now, dividing both sides of Equation (12.8) by Δx gives

$$\frac{F(x_1+\Delta x)-F(x_1)}{\Delta x}=f(c),$$

whose left side is the difference quotient for F at x_1. So, let us take the limit as $\Delta x \to 0$ on both sides of the equation. This yields:

$$\lim_{\Delta x\to 0}\frac{F(x_1+\Delta x)-F(x_1)}{\Delta x}=\lim_{\Delta x\to 0}f(c). \tag{12.9}$$

Clearly, the expression on the left side of the resulting equation is the definition of the derivative of F at x_1, so we may rewrite Equation (12.9) as

$$F'(x_1)=\lim_{\Delta x\to 0}f(c). \tag{12.10}$$

To find the limit on the right side of Equation (12.10), we resort to the squeeze theorem (Theorem 8.5). The number c is in the interval $[x_1, x_1+\Delta x]$, so $x_1 \le c \le x_1+\Delta x$. Besides, it holds that

$$\lim_{\Delta x\to 0}x_1=x_1$$

and

$$\lim_{\Delta x\to 0}(x_1+\Delta x)=x_1.$$

Therefore, according to the squeeze theorem, it must hold that

$$\lim_{\Delta x\to 0}c=x_1.$$

Consequently, we may rewrite Equation (12.10) as

$$F'(x_1)=\lim_{c\to x_1}f(c).$$

The function f is continuous at c, so the limit can be taken inside the function. In this way, we get

$$F'(x_1)=f(x_1),$$

which completes the proof. □

To illustrate this theorem, let us consider

$$F(x)=\int_{-5}^{x}(t^2+\sin(t))\,dt$$

and try to find $F'(x)$.

Using Theorem 12.7, we immediately have $F'(x)=x^2+\sin(x)$. This simple example reveals that $F(x)$ is an antiderivative of $x^2+\sin(x)$! Therefore, $F(x)=x^3/3-\cos(x)+C$ for some value of C. We have done more, however, than found a complicated way of computing an antiderivative. Consider a function f defined on an open interval containing a, b and c. Suppose we want to compute $\int_a^b f(t)\,dt$. First, let

$$F(x)=\int_c^x f(t)\,dt.$$

Using the properties of the definite integral (Theorem 12.3), we know

$$\begin{aligned}\int_a^b f(t)\,dt &= \int_a^c f(t)\,dt + \int_c^b f(t)\,dt\\ &= -\int_c^a f(t)\,dt + \int_c^b f(t)\,dt\\ &= -F(a) + F(b)\\ &= F(b) - F(a).\end{aligned}$$

We now see how indefinite integrals and definite integrals are related: we can evaluate a definite integral using antiderivatives. This is the second part of the fundamental theorem of calculus.

Theorem 12.8 (The fundamental theorem of calculus, Part 2)
Let f be continuous on $[a,b]$ and let F be any antiderivative of f. Then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Proof We will rely on Riemann sums to prove this theorem in a more rigorous way. For that purpose, let f be integrable on the interval $[a,b]$, and let f admit an antiderivative F on $[a,b]$. Consider the quantity $F(b)-F(a)$ and let there be a partition of size $\mathcal{L}$ with numbers $x_1, x_2, \ldots, x_{n+1}$ such that

$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b.$$

Clearly,

$$F(b) - F(a) = F(x_{n+1}) - F(x_1).$$

Now, for $i = 2, \ldots, n$ we add each $F(x_i)$ along with its additive inverse, so that the resulting quantity is equal:

$$\begin{aligned}F(b) - F(a) &= F(x_{n+1}) + [-F(x_n) + F(x_n)] + \cdots + [-F(x_2) + F(x_2)] - F(x_1)\\ &= [F(x_{n+1}) - F(x_n)] + [F(x_n) - F(x_{n-1})] + \cdots + [F(x_3) - F(x_2)] + [F(x_2) - F(x_1)].\end{aligned}$$

Or shorter:

$$F(b) - F(a) = \sum_{i=1}^{n} [F(x_{i+1}) - F(x_i)]. \tag{12.11}$$

Inspecting the right hand side of this equation reminds us to the mean value theorem of differentiation (Theorem 10.4). Indeed, it tells us that

$$F'(c) = \frac{F(b) - F(a)}{b-a},$$

where $c \in [a,b]$, or equivalently

$$F(b) - F(a) = F'(c)(b-a).$$

Hence, since the function F is differentiable on the interval $[a,b]$ and hence differentiable and continuous on each interval $[x_{i-1}, x_i]$, we can rewrite the terms appearing in right hand side of Equation (12.11) as

$$F(x_{i+1}) - F(x_i) = F'(c_i)(x_{i+1} - x_i),$$

where $c_i \in [x_{i+1}, x_i]$. This allows us to rewrite Equation (12.11) as

$$F(b)-F(a)=\sum_{i=1}^{n}[F'(c_i)(x_{i+1}-x_i)].$$

Moreover, our assumption that F is an antiderivative of f implies that $F'(c_i)=f(c_i)$. Hence, letting $x_{i+1}-x_i=\Delta x_i$, we get

$$F(b)-F(a)=\sum_{i=1}^{n}[f(c_i)(\Delta x_i)]. \tag{12.12}$$

Essentially, we are describing with this expression the area of a rectangle, with the width times the height, and we are adding the areas together. By taking the limit of the expression as the norm of the partitions approaches zero, we arrive at the Riemann integral. We know that this limit exists because f was assumed to be integrable. So, we take the limit on both sides of Equation (12.12), to obtain

$$\lim_{\mathcal{L}\to 0}\big(F(b)-F(a)\big)=\lim_{\mathcal{L}\to 0}\sum_{i=1}^{n}[f(c_i)(\Delta x_i)].$$

Neither $F(b)$ nor $F(a)$ is dependent on $\mathcal{L}$, so the limit on the left side remains $F(b)-F(a)$. Furthermore, the expression on the right side of the equation defines the integral over f from a to b (Theorem 12.4). Therefore, we obtain

$$F(b)-F(a)=\int_a^b f(x)\,dx,$$

which completes the proof. □

Example 12.6

We spent a great deal of time in the previous section studying $\int_0^4 (4x-x^2)\,dx$. Using the fundamental theorem of calculus, evaluate this definite integral.

Solution

We need an antiderivative of $f(x)=4x-x^2$. All antiderivatives of f have the form

$$F(x)=2x^2-\frac{1}{3}x^3+C;$$

for simplicity, choose $C=0$.

The fundamental theorem of calculus states

$$\int_0^4 (4x-x^2)\,dx=F(4)-F(0)=\left(2(4)^2-\frac{1}{3}4^3-\big(0-0\big)\right)=32-\frac{64}{3}=\frac{32}{3}.$$

This is the same answer we obtained using limits in the previous section, just with much less work.

A special notation is often used in the process of evaluating definite integrals using the fundamental theorem of calculus. Instead of explicitly writing $F(b)-F(a)$, the notation $F(x)\Big|_a^b$ is used. Also note that any antiderivative $F(x)$ can be chosen when using the fundamental theorem of calculus to evaluate a definite integral, meaning any value of C can be picked. The constant always cancels out of the expression when evaluating $F(b)-F(a)$, so it does not matter what value is picked. This being the case, we might as well let $C=0$.

Example 12.7

Evaluate the following definite integrals.

1. $\displaystyle\int_{-2}^{2} x^3\,dx$ 2. $\displaystyle\int_{0}^{\pi} \sin(x)\,dx$ 3. $\displaystyle\int_{0}^{5} e^t\,dt$ 4. $\displaystyle\int_{4}^{9} \sqrt{u}\,du$ 5. $\displaystyle\int_{1}^{5} 2\,dx$

Solution

1. $\displaystyle\int_{-2}^{2} x^3\,dx = \frac{1}{4}x^4\Big|_{-2}^{2} = \left(\frac{1}{4}2^4\right) - \left(\frac{1}{4}(-2)^4\right) = 0.$
2. $\displaystyle\int_{0}^{\pi} \sin(x)\,dx = -\cos(x)\Big|_{0}^{\pi} = -\cos(\pi) - \big(-\cos 0\big) = 1+1 = 2$. So, the area under one hump of a sine curve is 2.
3. $\displaystyle\int_{0}^{5} e^t\,dt = e^t\Big|_{0}^{5} = e^5 - e^0 = e^5 - 1 \approx 147.41.$
4. $\displaystyle\int_{4}^{9} \sqrt{u}\,du = \int_4^9 u^{\frac{1}{2}}\,du = \frac{2}{3}u^{\frac{3}{2}}\Big|_{4}^{9} = \frac{2}{3}\left(9^{\frac{3}{2}} - 4^{\frac{3}{2}}\right) = \frac{2}{3}\big(27-8\big) = \frac{38}{3}.$
5. $\displaystyle\int_{1}^{5} 2\,dx = 2x\Big|_{1}^{5} = 2(5) - 2 = 2(5-1) = 8.$

This last integral in Example 12.7 is interesting; the integrand is a constant function, hence we are finding the area of a rectangle with width $(5-1) = 4$ and height 2. Notice how the evaluation of the definite integral led to $2(4) = 8$.

In general, if c is a constant, then

$$\int_a^b c\,dx = c(b-a).$$

12.3.3 Motion and the fundamental theorem of calculus

We established, starting in Section 9.1.4, that the derivative of a position function is a velocity function, and the derivative of a velocity function is an acceleration function. Now consider definite integrals of velocity and acceleration functions. Specifically, if $v(t)$ is a velocity function, what does $\int_a^b v(t)\,dt$ mean?

The fundamental theorem of calculus states that

$$\int_a^b v(t)\,dt = V(b) - V(a),$$

where $V(t)$ is any antiderivative of $v(t)$. Since $v(t)$ is a velocity function, $V(t)$ must be a position function, and $V(b) - V(a)$ measures a change in position, or **displacement** (*verplaatsing*).

Example 12.8

A ball is thrown straight up with velocity given by $v(t) = -32t + 20$m/s, where t is measured in seconds. Find, and interpret, $\int_0^1 v(t)\,dt$.

Solution

Using the fundamental theorem of calculus, we have

$$\begin{aligned}\int_0^1 v(t)\,dt &= \int_0^1 (-32t + 20)\,dt \\ &= -16t^2 + 20t\Big|_0^1 \\ &= 4.\end{aligned}$$

Thus if a ball is thrown straight up into the air with velocity $v(t) = -32t + 20$, the height of the ball, 1 second later, will be 4 metres above the initial height.

Integrating a rate of change function gives total change. Velocity is the rate of position change; integrating velocity gives the total change of position, i.e., displacement.

Integrating a speed function gives a similar, though different, result. Speed is also the rate of position change, but does not account for direction. So integrating a speed function gives total change of position, without the possibility of negative position change. Hence the integral of a speed function gives **distance travelled** (*afgelegde afstand*).

12.3.4 The fundamental theorem of calculus and the chain rule

Using other notation, we may write Part 1 of the fundamental theorem of calculus as

$$\frac{d}{dx}\big(F(x)\big) = f(x).$$

While we have just practised evaluating definite integrals, sometimes finding antiderivatives is impossible and we need to rely on other techniques to approximate the value of a definite integral. Functions written as

$$F(x) = \int_a^x f(t)\,dt$$

are useful in such situations.

It may be of further use to compose such a function with another. As an example, we may compose $F(x)$ with $g(x)$ to get

$$F\big(g(x)\big) = \int_a^{g(x)} f(t)\,dt.$$

What is the derivative of such a function? The chain rule can be employed to find

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'\big(g(x)\big)g'(x) = f\big(g(x)\big)g'(x).$$

An example will help us understand this.

Example 12.9

Find the derivative of

1. $F(x) = \int_2^{x^2} \ln(t)\, dt$
2. $F(x) = \int_{\cos(x)}^{5} t^3\, dt.$

Solution

1. We can view $F(x)$ as being the function $G(x) = \int_2^x \ln(t)\, dt$ composed with $h(x) = x^2$; that is, $F(x) = G\big(h(x)\big)$. The fundamental theorem of calculus states that $G'(x) = \ln(x)$. The chain rule gives us

$$\begin{aligned} F'(x) &= G'\big(h(x)\big)h'(x) \\ &= \ln(h(x))h'(x) \\ &= \ln(x^2)2x \\ &= 2x\ln(x^2). \end{aligned}$$

Normally, of course, the steps defining $G(x)$ and $h(x)$ are skipped.

2. Note that $F(x) = -\int_5^{\cos(x)} t^3\, dt$. Viewed this way, the derivative of F is straightforward:

$$F'(x) = \sin(x)\cos^3(x).$$

12.3.5 Average value

Recognize that the mean value theorem can be rewritten as

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\, dx,$$

for some value of c in $[a,b]$. Next, partition the interval $[a,b]$ into n equally spaced subintervals, $a = x_1 < x_2 < \ldots < x_{n+1} = b$ and choose any c_i in $[x_i, x_{i+1}]$. The average of the numbers $f(c_1)$, $f(c_2)$, $\ldots$, $f(c_n)$ is:

$$\frac{1}{n}\Big(f(c_1) + f(c_2) + \ldots + f(c_n)\Big) = \frac{1}{n}\sum_{i=1}^{n} f(c_i).$$

Multiply this last expression by 1 in the form of $\frac{(b-a)}{(b-a)}$:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} f(c_i) &= \sum_{i=1}^{n} f(c_i)\frac{1}{n}\frac{(b-a)}{(b-a)} \\ &= \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\frac{b-a}{n} \end{aligned}$$

$$= \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x$$

where $\Delta x = (b-a)/n$. Now take the limit as $n \to +\infty$:

$$\lim_{n\to+\infty} \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x = \frac{1}{b-a}\int_a^b f(x)\ dx = f(c).$$

This tells us this: when we evaluate f at n (somewhat) equally spaced points in $[a,b]$, the average value of these samples is $f(c)$ as $n \to +\infty$.

This leads us to a definition.

Definitie 12.5 (The average value of *f* on *[a,b]*)
Let f be continuous on $[a,b]$. The **average value of** f (*gemiddelde functiewaarde*) on $[a,b]$ is $f(c)$, where c is a value in $[a,b]$ guaranteed by the mean value theorem. I.e.,

$$\text{Average Value of } f \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\ dx.$$

An application of this definition is given in the following example.

Example 12.10

An object moves back and forth along a straight line with a velocity given by $v(t) = (t-1)^2$ on $[0,3]$, where t is measured in seconds and $v(t)$ is measured in m/s.

1. What is the average velocity of the object?
2. What was the displacement of the object?

Solution

1. By Definition 12.5, the average velocity is:

$$\frac{1}{3-0}\int_0^3 (t-1)^2\ dt = \frac{1}{3}\int_0^3 \left(t^2-2t+1\right) dt = \frac{1}{3}\left(\frac{1}{3}t^3 - t^2 + t\right)\Bigg|_0^3 = 1 \text{ m/s}.$$

2. We calculate this by integrating its velocity function: $\int_0^3 (t-1)^2\ dt = 3$ m. Its final position was 3 meter from its initial position after 3 seconds: its average velocity was 1 m/s.

This section has laid the groundwork for a lot of great mathematics to follow. The most important lesson is this: definite integrals can be evaluated using antiderivatives. Since the previous section established that definite integrals are the limit of Riemann sums, we can later create Riemann sums to approximate values other than area under the curve, convert the sums to definite integrals, then evaluate these using the fundamental theorem of calculus. This will allow us to compute the work done by a variable force, the volume of certain solids, the arc length of curves, and more.

The downside is this: generally speaking, computing antiderivatives is much more difficult than computing derivatives. For that reason, we will see in Section 12.6.2 how to approximate the value of definite integrals beyond using the left hand, right hand and midpoint rules. These techniques are invaluable when antiderivatives cannot be computed, or when the actual function f is unknown and all we know is the value of f at certain x-values. But first, we will study techniques of finding antiderivatives analytically so that a wide variety of definite integrals can be evaluated.

12.4 Techniques of antidifferentiation

This chapter is devoted to exploring techniques of antidifferentiation. While not every function has an antiderivative in terms of elementary functions like polynomial, exponential or trigonometric functions, we can still find antiderivatives of a wide variety of functions. Nowadays there are also several websites that allow you to calculate integrals with steps[1].

12.4.1 Substitution

12.4.1.1 Rationale

Essentially, integration by **substitution** (*substitutie*) allows us to undo the chain rule. Its underlying principle is to rewrite a complicated integral of the form $\int f(x)\ dx$ as a not–so–complicated integral $\int h(u)\ du$.

For instance, consider

$$\int (20x+30)(x^2+3x-5)^9\ dx.$$

Arguably the most complicated part of the integrand is $(x^2+3x-5)^9$. We wish to make this simpler; we do so through a substitution. Let $u = x^2+3x-5$. Thus

$$(x^2+3x-5)^9 = u^9.$$

We have established u as a function of x, so now consider the differential of u:

$$du = (2x+3)dx.$$

Let us return now to the original integral and do some substitutions through algebra:

$$\begin{aligned}
\int (20x+30)(x^2+3x-5)^9\ dx &= \int 10(2x+3)(x^2+3x-5)^9\ dx \\
&= \int 10(\underbrace{x^2+3x-5}_{u})^9\ \underbrace{(2x+3)\ dx}_{du} \\
&= \int 10u^9\ du \\
&= u^{10}+C \qquad \text{(Replace } u \text{ with } x^2+3x-5.) \\
&= (x^2+3x-5)^{10}+C
\end{aligned}$$

In general, let $F(x)$ and $g(x)$ be differentiable functions and consider the derivative of their composition:

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'(g(x))g'(x).$$

Thus

$$\int F'(g(x))g'(x)\ dx = F(g(x))+C.$$

[1] https://www.integral-calculator.com/

Integration by substitution works by recognizing the inside function $g(x)$ and replacing it with a variable. By setting $u = g(x)$, we can rewrite the derivative as

$$\frac{d}{dx}\Big(F(u)\Big) = F'(u)u'.$$

Since $du = g'(x)dx$, we can rewrite the above integral as

$$\int F'(g(x))g'(x)\ dx = \int F'(u)du = F(u) + C = F(g(x)) + C.$$

The point of substitution is to make the integration step easy. Indeed, the step $\int F'(u)\ du = F(u) + C$ looks easy, as the antiderivative of the derivative of F is just F, plus a constant. The work involved is making the proper substitution. There is not a step–by–step process that one can memorize; rather, experience will be one's guide. To gain experience, we now embark on some examples.

Example 12.11

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{7}{-3x+1}\ dx$,
2. $\displaystyle\int x\sin(x^2+5)\ dx$,
3. $\displaystyle\int x\sqrt{x+3}\ dx$.

Solution

1. View the integrand as the composition of functions $f(g(x))$, where $f(x) = 7/x$ and $g(x) = -3x+1$. Then, we let $u = -3x+1$, the inside function. Thus $du = -3dx$. The integrand lacks a -3; hence divide the previous equation by -3 to obtain $-du/3 = dx$. We can now evaluate the integral through substitution.

$$\begin{aligned}\int \frac{7}{-3x+1}\ dx &= \int \frac{7}{u}\frac{du}{(-3)} \\ &= \frac{-7}{3}\int \frac{du}{u} \\ &= \frac{-7}{3}\ln|u| + C \\ &= -\frac{7}{3}\ln|-3x+1| + C.\end{aligned}$$

2. We choose to let u be the inside function of $\sin(x^2+5)$. So, let $u = x^2+5$, hence $du = 2x\,dx$. The integrand has an $x\,dx$ term, but not a $2x\,dx$ term. We can divide both sides of the du expression by 2:

$$du = 2x\,dx \quad \Rightarrow \quad \frac{1}{2}du = x\,dx.$$

We can now substitute.

$$\int x\sin(x^2+5)\ dx = \int \sin(\underbrace{x^2+5}_{u})\underbrace{x\ dx}_{\frac{1}{2}du}$$

$$= \int \frac{1}{2}\sin(u)\,du$$
$$= -\frac{1}{2}\cos(u) + C \qquad \text{(Now replace } u \text{ with } x^2+5.)$$
$$= -\frac{1}{2}\cos(x^2+5) + C.$$

Thus
$$\int x\sin(x^2+5)\,dx = -\frac{1}{2}\cos(x^2+5) + C.$$

3. Recognizing the composition of functions, set $u = x+3$. Then $du = dx$, giving what seems initially to be a simple substitution. But at this stage, we have:

$$\int x\sqrt{x+3}\,dx = \int x\sqrt{u}\,du.$$

We cannot evaluate an integral that has both an x and an u in it. We need to convert the x to an expression involving just u.

Since we set $u = x+3$, we can also state that $u-3 = x$. Thus we can replace x in the integrand with $u-3$. It will also be helpful to rewrite $\sqrt{u}$ as $u^{\frac{1}{2}}$.

$$\begin{aligned}\int x\sqrt{x+3}\,dx &= \int (u-3)u^{\frac{1}{2}}\,du\\ &= \int \left(u^{\frac{3}{2}} - 3u^{\frac{1}{2}}\right)du\\ &= \frac{2}{5}u^{\frac{5}{2}} - 2u^{\frac{3}{2}} + C\\ &= \frac{2}{5}(x+3)^{\frac{5}{2}} - 2(x+3)^{\frac{3}{2}} + C\end{aligned}$$

12.4.1.2 Integrals involving trigonometric functions

Integration by substitution can also be used to unveil the antiderivatives of the tangent, cotangent, secant and cosecant. For instance, consider the following example concerning the former function.

Example 12.12

Evaluate
$$\int \tan(x)\,dx.$$

Solution

Rewrite $\tan(x)$ as $\sin(x)/\cos(x)$. While the presence of a composition of functions may not be immediately obvious, recognize that $\cos(x)$ is inside the $1/x$ function. Therefore, we see if setting $u = \cos(x)$ returns usable results. We have that $du = -\sin(x)\,dx$, hence $-du = \sin(x)\,dx$. We can integrate:

$$\begin{aligned}\int \tan(x)\, dx &= \int \frac{\sin(x)}{\cos(x)}\, dx\\ &= \int \underbrace{\frac{1}{\cos(x)}}_{1/u} \underbrace{\sin(x)\, dx}_{-du}\\ &= \int \frac{-1}{u}\, du\\ &= -\ln|u| + C\\ &= -\ln|\cos(x)| + C.\end{aligned}$$

This can be simplified even further by bringing the -1 inside the logarithm as a power of $\cos(x)$, as in:

$$\begin{aligned}-\ln|\cos(x)| + C &= \ln|(\cos(x))^{-1}| + C\\ &= \ln\left|\frac{1}{\cos(x)}\right| + C\\ &= \ln|\sec(x)| + C.\end{aligned}$$

Thus the result they give is $\int \tan(x)\, dx = \ln|\sec(x)| + C$.

We can use similar techniques in Example 12.12 to find antiderivatives of the other trigonometric functions. In this way, one finds:

1. $\displaystyle\int \tan(x)\, dx = -\ln|\cos(x)| + C$
2. $\displaystyle\int \cot(x)\, dx = \ln|\sin(x)| + C$
3. $\displaystyle\int \sec(x)\, dx = \ln|\sec(x) + \tan(x)| + C$
4. $\displaystyle\int \csc(x)\, dx = -\ln|\csc(x) + \cot(x)| + C$

Likewise, we can find antiderivatives of hyperbolic functions:

1. $\displaystyle\int \tanh(x)\, dx = \ln(\cosh(x)) + C$
2. $\displaystyle\int \coth(x)\, dx = \ln|\sinh(x)| + C$

Using the power-reducing formulas we have seen in Chapter 5 (Theorem 5.12), we can also tackle integrals involving powers of trigonometric and hyperbolic functions.

Example 12.13

Evaluate

$$\int \cos^2(x)\, dx.$$

Solution

We have a composition of functions as $\cos^2(x) = \big(\cos(x)\big)^2$. However, setting $u = \cos(x)$ means $du = -\sin(x)\ dx$, which we do not have in the integral. So, let us use Theorem 5.12, which states

$$\cos^2(x) = \frac{1+\cos(2x)}{2}.$$

The right hand side of this equation is not difficult to integrate. We have:

$$\begin{aligned}\int \cos^2(x)\ dx &= \int \frac{1+\cos(2x)}{2}\ dx \\ &= \int \left(\frac{1}{2} + \frac{1}{2}\cos(2x)\right)\ dx.\end{aligned}$$

So, we easily find:

$$\begin{aligned}&= \frac{1}{2}x + \frac{1}{2}\frac{\sin(2x)}{2} + C \\ &= \frac{1}{2}x + \frac{\sin(2x)}{4} + C.\end{aligned}$$

We will make significant use of this power–reducing technique in future sections.

12.4.1.3 Integrals leading to inverse trigonometric and hyperbolic functions

When studying derivatives of inverse functions, we learned that

$$\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}.$$

Applying the chain rule to this is not difficult. For instance, in general, we have

$$\frac{d}{dx}\big(\arctan(ax)\big) = \frac{a}{1+a^2x^2}.$$

This result can be used to evaluate

$$\int \frac{1}{a^2+x^2}\ dx$$

.

For that purpose, we rewrite this integral as

$$\frac{1}{a^2}\int \frac{1}{1+\left(\frac{x}{a}\right)^2}\ dx.$$

This can now be integrated using substitution. Set $u = x/a$, hence $du = dx/a$ or $dx = adu$. Thus

$$\begin{aligned}\int \frac{1}{a^2+x^2}\ dx &= \frac{1}{a}\int \frac{1}{1+u^2}\ du \\ &= \frac{1}{a}\arctan(u) + C \\ &= \frac{1}{a}\arctan\left(\frac{x}{a}\right) + C\end{aligned}$$

This demonstrates a general technique that can be applied to other integrands that result in inverse trigonometric functions. More specifically, for $a > 0$, we have

$$\int \frac{1}{a^2+x^2}\,dx = \frac{1}{a}\arctan\left(\frac{x}{a}\right)+C \tag{12.13}$$

$$\int \frac{1}{\sqrt{a^2-x^2}}\,dx = \arcsin\left(\frac{x}{a}\right)+C \tag{12.14}$$

Of course, given the link between trigonometric and hyperbolic functions, similar integrands result in inverse hyperbolic functions:

$$\int \frac{1}{\sqrt{x^2-a^2}}\,dx = \operatorname{arcosh}\left(\frac{x}{a}\right)+C = \ln\left|x+\sqrt{x^2-a^2}\right|+C, \qquad \text{for } 0 < a < x, \tag{12.15}$$

$$\int \frac{1}{\sqrt{x^2+a^2}}\,dx = \operatorname{arsinh}\left(\frac{x}{a}\right)+C = \ln\left|x+\sqrt{x^2+a^2}\right|+C, \qquad \text{for } a > 0, \tag{12.16}$$

$$\int \frac{1}{a^2-x^2}\,dx = \begin{cases} \dfrac{1}{a}\operatorname{artanh}\left(\dfrac{x}{a}\right)+C\,, & x^2 < a^2, \\ \dfrac{1}{a}\operatorname{arcoth}\left(\dfrac{x}{a}\right)+C\,, & a^2 < x^2 \end{cases} \tag{12.17}$$

$$= \frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right|+C\,, \tag{12.18}$$

Example 12.14

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{x^2-4x+13}\,dx,$
2. $\displaystyle\int \frac{4-x}{\sqrt{16-x^2}}\,dx,$
3. $\displaystyle\int \frac{1}{x^2-1}\,dx,$
4. $\displaystyle\int \frac{1}{\sqrt{9x^2+10}}\,dx.$

Solution

1. We start by completing the square in the denominator, i.e.

$$\frac{1}{x^2-4x+13} = \frac{1}{(x-2)^2+9}$$

We can now integrate, to arrive at

$$\int \frac{1}{x^2-4x+13}\,dx = \int \frac{1}{(x-2)^2+9}\,dx = \frac{1}{3}\arctan\left(\frac{x-2}{3}\right)+C.$$

2. This integral requires two different methods to evaluate it. We get to those methods by splitting up the integral:

$$\int \frac{4-x}{\sqrt{16-x^2}}\,dx = \int \frac{4}{\sqrt{16-x^2}}\,dx - \int \frac{x}{\sqrt{16-x^2}}\,dx.$$

The first integral is easy to handle; the second integral is handled by substitution, with $u=16-x^2$. We handle each separately.

$\displaystyle\int \frac{4}{\sqrt{16-x^2}}\,dx = 4\arcsin\left(\frac{x}{4}\right)+C.$

$\displaystyle\int \frac{x}{\sqrt{16-x^2}}\,dx$: Set $u=16-x^2$, so $du=-2xdx$ and $xdx=-du/2$. We have

$$\begin{aligned}\int \frac{x}{\sqrt{16-x^2}}\,dx &= \int \frac{-du/2}{\sqrt{u}} \\ &= -\frac{1}{2}\int \frac{1}{\sqrt{u}}\,du \\ &= -\sqrt{u}+C \\ &= -\sqrt{16-x^2}+C.\end{aligned}$$

Combining these together, we have

$$\int \frac{4-x}{\sqrt{16-x^2}}\,dx = 4\arcsin\left(\frac{x}{4}\right)+\sqrt{16-x^2}+C.$$

3. Multiplying the numerator and denominator by (-1) gives:

$$\int \frac{1}{x^2-1}\,dx = \int \frac{-1}{1-x^2}\,dx.$$

The second integral can be solved directly using Equation (12.18), with $a = 1$. Thus

$$\begin{aligned}
\int \frac{1}{x^2-1}\,dx &= -\int \frac{1}{1-x^2}\,dx \\
&= \begin{cases} -\operatorname{artanh}(x) + C & x^2 < 1 \\ -\operatorname{arcoth}(x) + C & 1 < x^2 \end{cases} \\
&= -\frac{1}{2}\ln\left|\frac{x+1}{x-1}\right| + C \\
&= \frac{1}{2}\ln\left|\frac{x-1}{x+1}\right| + C. \qquad (12.19)
\end{aligned}$$

4. This requires a substitution, then Equation (12.16) can be used.

Let $u = 3x$, hence $du = 3dx$. We have

$$\int \frac{1}{\sqrt{9x^2+10}}\,dx = \frac{1}{3}\int \frac{1}{\sqrt{u^2+10}}\,du.$$

Note $a^2 = 10$, hence $a = \sqrt{10}$. We immediately obtain

$$\int \frac{1}{\sqrt{9x^2+10}}\,dx = \frac{1}{3}\operatorname{arsinh}\left(\frac{3x}{\sqrt{10}}\right) + C = \frac{1}{3}\ln\left|3x + \sqrt{9x^2+10}\right| + C$$

12.4.1.4 Substitution and definite integration

Definite integrals that require substitution can be calculated using the following workflow:

1. Start with a definite integral $\int_a^b f(x)\,dx$ that requires substitution.
2. Ignore the bounds; use substitution to evaluate $\int f(x)\,dx$ and find an antiderivative $F(x)$.
3. Evaluate $F(x)$ at the bounds; that is, evaluate $F(x)\Big|_a^b = F(b) - F(a)$.

This workflow works fine, but substitution offers an alternative that is powerful and time saving. Since substitution converts integrals of the form $\int F'(g(x))g'(x)\,dx$ into an integral of the form $\int F'(u)\,du$ with the substitution of $u = g(x)$, we just have to appropriately change the bounds of a definite integral, i.e.

$$\int\limits_a^b F'\big(g(x)\big)g'(x)\,dx = \int\limits_{g(a)}^{g(b)} F'(u)\,du.$$

This indicates that once you convert to integrating with respect to u, you do not need to switch back to evaluating with respect to x.

Example 12.15

Evaluate

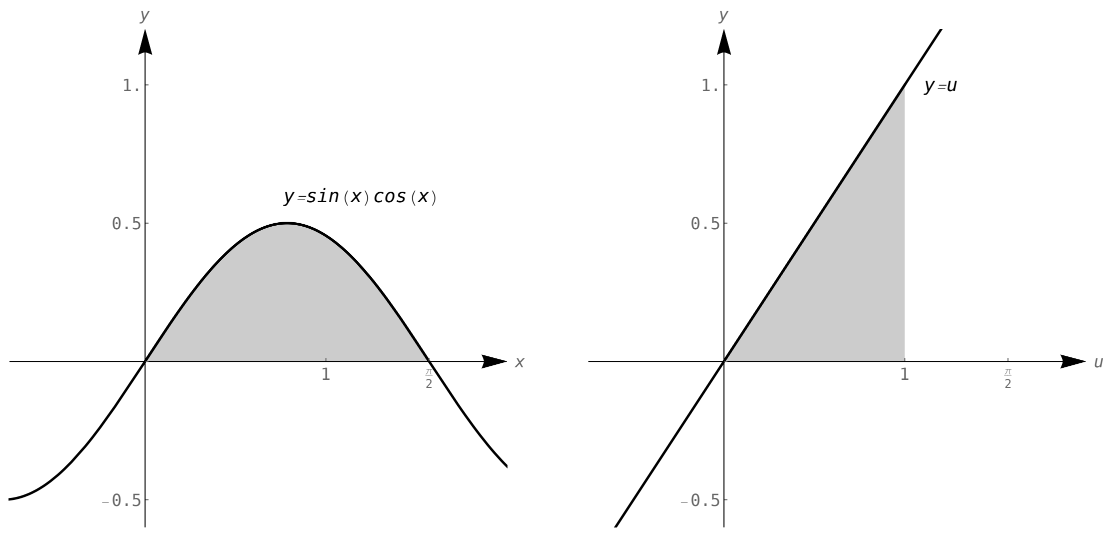
$$\int_0^{\pi/2} \sin(x)\cos(x)\ dx.$$

Solution

Let $u = g(x) = \cos(x)$, giving $du = -\sin(x)\ dx$ and hence $\sin(x)\ dx = -du$. The new upper bound is $g(\pi/2) = 0$; the new lower bound is $g(0) = 1$. Note how the lower bound is actually larger than the upper bound now. We have

$$\begin{aligned}\int_0^{\pi/2} \sin(x)\cos(x)\ dx &= \int_1^0 -u\ du \\ &= \int_0^1 u\ du \\ &= \frac{1}{2}u^2\Big|_0^1 = \frac{1}{2}.\end{aligned}$$

In Figure 12.11 we have graphed the two regions defined by our definite integrals. They bear no resemblance to each other, but they have the same area.



(a) (b)

Figure 12.11: Graphing the areas defined by the definite integrals of Example 12.15.

12.4.1.5 Tangent half-angle substitution

The tangent half-angle substitution, also known as the Weierstrass substitution after Karl Weierstrass, is a substitution used for finding antiderivatives of rational functions of trigonometric functions.

For this substitution we let $t = \tan\left(\frac{x}{2}\right)$. By the double-angle formula for the sine function, we get

$$\begin{aligned} \sin(x) &= 2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right) \\ &= 2t\cos^2\left(\frac{x}{2}\right) \\ &= \frac{2t}{\sec^2\left(\frac{x}{2}\right)} \\ &= \frac{2t}{1+t^2}. \end{aligned}$$

Similarly, by the double-angle formula for the cosine function, we easily find

$$\begin{aligned} \cos(x) &= 1 - 2\sin^2\left(\frac{x}{2}\right) \\ &= 1 - 2t^2\cos^2\left(\frac{x}{2}\right) \\ &= 1 - \frac{2t^2}{\sec^2\left(\frac{x}{2}\right)} \\ &= 1 - \frac{2t^2}{1+t^2} \\ &= \frac{1-t^2}{1+t^2}. \end{aligned}$$

Moreover, since

$$\begin{aligned} \frac{dt}{dx} &= \frac{1}{2}\sec^2\left(\frac{x}{2}\right) \\ &= \frac{1+t^2}{2}, \end{aligned}$$

we get the following expression for dx:

$$dx = \frac{2}{1+t^2}\,dt\,.$$

Example 12.16

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{1+\sin(x)}\,dx$
2. $\displaystyle\int \frac{\sin(x)}{2+\cos^2(x)}\,dx$

Solution

1. Using the Weierstrass substitution, we easily find

$$\int \frac{dx}{1+\sin(x)} = \int \frac{1}{1+\left(\frac{2t}{1+t^2}\right)} \frac{2}{1+t^2}\, dt = \int \frac{2}{(t+1)^2}\, dt\,,$$

where the last integral can be evaluated by a change a variables. Indeed, letting $u = t+1$, we find

$$\int \frac{2}{(t+1)^2}\, dt = \frac{-2}{1+t} + C\,,$$

or in terms of the original variable x where we started from:

$$\int \frac{dx}{1+\sin(x)} = \frac{-2}{1+\tan\left(\frac{x}{2}\right)} + C\,.$$

2. Again, it is clear that the Weierstrass substitution will help us out:

$$\begin{aligned} \int \frac{\sin(x)}{2+\cos^2(x)}\, dx &= \int \frac{\left(\frac{2t}{1+t^2}\right)}{2+\left(\frac{1-t^2}{1+t^2}\right)^2} \left(\frac{2}{1+t^2}\, dt\right) \\ &= \int \frac{4t}{3t^4+2t^2+3}\, dt\,. \end{aligned}$$

It should be clear that we can recast the last integral in order to arrive an arctangent function. This can be done, by letting $v = t^2$, to get

$$\int \frac{2}{3v^2+2v+3}\, dv\,,$$

which can be rewritten, after some algebra, as

$$\frac{3}{4} \int \frac{1}{\left(\frac{3v+1}{2\sqrt{2}}\right)^2+1}\, dv\,.$$

Using the substitution $w = \frac{3v+1}{2\sqrt{2}}$, the latter integral on its turn becomes

$$\frac{\sqrt{2}}{2} \int \frac{1}{w^2+1}\, dw\,,$$

which in terms of v evaluates to

$$\frac{\sqrt{2}}{2} \arctan\left(\frac{3v+1}{2\sqrt{2}}\right)\,.$$

Consequently, in terms of the original variable x, we arrive at

$$\int \frac{\sin(x)}{2+\cos^2(x)}\, dx = \frac{\sqrt{2}}{2} \arctan\left(\frac{3\left(\tan(\frac{x}{2})\right)^2+1}{2\sqrt{2}}\right)\,.$$

12.4.2 Integration by parts

Here is a simple integral that we can not yet evaluate:

$$\int x\cos(x)\,dx.$$

It's a simple matter to take the derivative of the integrand using the product rule, but there is no such rule for integrals. However, this section introduces **integration by parts** (*partiële integratie*), a method of integration that is based on the product rule for derivatives. It will enable us to evaluate this integral.

The product rule says that if u and v are functions of x, then $(uv)' = u'v + uv'$. For simplicity, we have written u for $u(x)$ and v for $v(x)$. Suppose we integrate both sides with respect to x. This gives

$$\int (uv)'\,dx = \int (u'v + uv')\,dx.$$

By the fundamental theorem of calculus, the left side integrates to uv. The right side can be broken up into two integrals, and we have

$$uv = \int u'v\,dx + \int uv'\,dx.$$

Solving for the second integral we have

$$\int uv'\,dx = uv - \int u'v\,dx.$$

Using differential notation, we can write $du = u'(x)dx$ and $dv = v'(x)dx$ and the expression above can be written as follows:

$$\int u\,dv = uv - \int v\,du. \tag{12.20}$$

If our problem concerns a definite integral, we likewise arrive at

$$\int_{x=a}^{x=b} u\ dv = uv\Big|_a^b - \int_{x=a}^{x=b} v\ du.$$

Typically, we try to identify u and dv in the integral we are given, and the key is that we usually want to choose u and dv so that du is simpler than u and v is hopefully not too much more complicated than dv. This will mean that the integral on the right side of the integration by parts formula, $\int v\,du$ will be simpler to integrate than the original integral $\int u\,dv$.

Example 12.17

Evaluate the following indefinite integrals:

1. $\int x^2\cos(x)\,dx$
2. $\int e^x\cos(x)\,dx$
3. $\int \arctan(x)\,dx$
4. $\int \cos(\ln(x))\ dx$

Solution

1. Let $u = x^2$ so that $dv = \cos(x)\,dx$. Then, it follows that $du = 2x\,dx$ and $v = \sin(x)$. Equation (12.20) leads to

$$\int x^2 \cos(x)\,dx = x^2 \sin(x) - \int 2x \sin(x)\,dx.$$

At this point, the integral on the right is indeed simpler than the one we started with, but to evaluate it, we need to do integration by parts again. Here we choose $u = 2x$ and $dv = \sin x$, so that $du = 2dx$ and $v = -\cos(x)$. Through Equation (12.20) this yields:

$$\int x^2 \cos(x)\,dx = x^2 \sin(x) - \left(-2x\cos(x) - \int -2\cos(x)\,dx\right).$$

The integral all the way on the right is now something we can evaluate. It evaluates to $-2\sin x$. Then going through and simplifying, being careful to keep all the signs straight, our answer is

$$\int x^2 \cos(x)\ dx = x^2 \sin(x) + 2x\cos(x) - 2\sin(x) + C.$$

2. This is a classic problem. In this particular example, one can let u be either $\cos(x)$ or e^x; we choose $u = e^x$ and hence $dv = \cos(x)\,dx$. Then $du = e^x\,dx$ and $v = \sin(x)$ as shown below. Using Equation (12.20) yields

$$\int e^x \cos(x)\ dx = e^x \sin(x) - \int e^x \sin(x)\,dx.$$

The integral on the right is not much different than the one we started with, so it seems like we have gotten nowhere. Let us nonetheless keep working and apply integration by parts to the new integral, using $u = e^x$ and $dv = \sin(x)\,dx$. Then we get $du = e^x\,dx$ and $v = -\cos(x)$ and this leads us to the following:

$$\begin{aligned}\int e^x \cos(x)\,dx &= e^x \sin(x) - \left(-e^x\cos(x) - \int -e^x \cos(x)\,dx\right)\\ &= e^x\sin(x) + e^x\cos(x) - \int e^x \cos(x)\ dx.\end{aligned}$$

It seems we are back right where we started, as the right hand side contains $\int e^x \cos(x)\,dx$. But this is actually a good thing.

Add $\int e^x \cos(x)\ dx$ to both sides. This gives

$$2\int e^x \cos(x)\ dx = e^x \sin(x) + e^x\cos(x)$$

Now divide both sides by 2:

$$\int e^x \cos(x)\ dx = \frac{1}{2}\big(e^x\sin(x) + e^x\cos(x)\big).$$

Simplifying a little and adding the constant of integration, our answer is thus

$$\int e^x \cos(x)\ dx = \frac{1}{2}e^x\big(\sin(x) + \cos(x)\big) + C.$$

3. Let $u = \arctan(x)$ and $dv = dx$. Then $du = 1/(1+x^2)\,dx$ and $v = x$. Using Equation (12.20) yields

$$\int \arctan(x)\,dx = x\arctan(x) - \int \frac{x}{1+x^2}\,dx.$$

The integral on the right can be solved by substitution. Taking $u = 1+x^2$, we get $du = 2x\,dx$. The integral then becomes

$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\int \frac{1}{u}\,du.$$

The integral on the right evaluates to $\ln|u| + C$, which becomes $\ln(1+x^2)+C$, as we may drop the absolute values as $1+x^2$ is always positive. Therefore, the answer is

$$\int \arctan(x)\;dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

4. The integrand contains a composition of functions, leading us to think integration by parts would be beneficial. Letting $u = \cos\big(\ln(x)\big)$, we have $du = -\sin\big(\ln(x)\big)/x\;dx$, and consequently $dv = dx$ and $v = x$. We then have

$$\begin{aligned}\int \cos(\ln(x))\;dx &= x\cos(\ln(x)) + \int \sin(\ln(x))\;dx\\ &= x\cos(\ln(x)) + x\sin(\ln(x)) - \int \cos(\ln(x))\;dx\,.\end{aligned}$$

So, we see that

$$\int \cos(\ln(x))\;dx = \frac{1}{2}x\big(\sin(\ln(x)) + \cos(\ln(x))\big) + C.$$

In general, integration by parts is useful for integrating certain products of functions, like $\int xe^x\,dx$ or $\int x^3\sin(x)\,dx$. It is also useful for integrals involving logarithms and inverse trigonometric functions.

As stated before, integration is generally more difficult than differentiation. We are developing tools for handling a large array of integrals, and experience will tell us when one tool is preferable/necessary over another. For instance, consider the three similar–looking integrals

$$\int xe^x\,dx, \qquad \int xe^{x^2}\,dx \qquad \text{and} \qquad \int xe^{x^3}\,dx.$$

While the first is calculated easily with integration by parts, the second is best approached with substitution. Taking things one step further, the third integral has no answer in terms of elementary functions, so none of the methods we learn in calculus will get us the exact answer.

12.4.3 Trigonometric integrals

Functions involving trigonometric functions are useful as they are good at describing periodic behavior. Here, we describe several techniques for finding antiderivatives of certain combinations of trigonometric functions.

12.4.3.1 Integrals of the form $\int \sin^m(x)\cos^n(x)\,dx$

We consider integrals of the form

$$\int \sin^m(x)\cos^n(x)\,dx,$$

where m,n are nonnegative integers. Our strategy for evaluating these integrals is to use the identity $\cos^2(x)+\sin^2(x)=1$ to convert high powers of one trigonometric function into the other, leaving a single sine or cosine term in the integrand. This is summarized below.

1. If m is odd, then $m=2k+1$ for some integer k. Rewrite

   $$\sin^m(x)=\sin^{2k+1}(x)=\sin^{2k}(x)\sin(x)=(\sin^2(x))^k\sin(x)=(1-\cos^2(x))^k\sin(x).$$

   Then
   $$\int \sin^m(x)\cos^n(x)\,dx=\int (1-\cos^2(x))^k\sin(x)\cos^n(x)\,dx=-\int (1-u^2)^k u^n\,du,$$
   where $u=\cos(x)$ and $du=-\sin(x)\,dx$.

2. If n is odd, then using substitutions similar to that outlined above we have

   $$\int \sin^m(x)\cos^n(x)\,dx=\int u^m(1-u^2)^k\,du,$$

   where $u=\sin(x)$ and $du=\cos(x)\,dx$.

3. If both m and n are even, use Theorem 5.12 to reduce the degree of the integrand. Expand the result and apply (1)-(3) again.

Let us check out how this all works in the following examples.

Example 12.18

Evaluate

$$\int \sin^5(x)\cos^9(x)\,dx.$$

Solution

The powers of both the sine and cosine terms are odd, therefore we can apply the techniques above to either power. We choose to work with the power of the cosine term.

We rewrite $\cos^9(x)$ as

$$\begin{aligned}\cos^9(x)&=\cos^8(x)\cos(x)\\&=(\cos^2(x))^4\cos(x)\\&=(1-\sin^2(x))^4\cos(x)\\&=(1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x))\cos x.\end{aligned}$$

We rewrite the integral as

$$\int \sin^5(x)\cos^9(x)\,dx=\int \sin^5(x)\big(1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x)\big)\cos(x)\,dx.$$

Now substitute using $u = \sin(x)$ and $du = \cos(x)\ dx$ to arrive at the following integral

$$\int u^5(1-4u^2+6u^4-4u^6+u^8)\ du\,,$$

which can then be integrated:

$$\begin{aligned}\int u^5(1-4u^2+6u^4-4u^6+u^8)\ du &= \int \left(u^5-4u^7+6u^9-4u^{11}+u^{13}\right) du \\ &= \frac{1}{6}u^6-\frac{1}{2}u^8+\frac{3}{5}u^{10}-\frac{1}{3}u^{12}+\frac{1}{14}u^{14}+C \\ &= \frac{1}{6}\sin^6(x)-\frac{1}{2}\sin^8(x)+\frac{3}{5}\sin^{10}(x)+\ldots \\ &\qquad -\frac{1}{3}\sin^{12}(x)+\frac{1}{14}\sin^{14}(x)+C.\end{aligned}$$

The work we are doing here can be a bit tedious, but the skills developed (problem solving, algebraic manipulation, etc.) are important. Nowadays problems of this sort are often solved using a computer algebra system. Mathematica, for instance, integrates $\int \sin^5(x)\cos^9(x)\ dx$ as

$$g(x) = -\frac{45\cos(2x)}{16384}-\frac{5\cos(4x)}{8192}+\frac{19\cos(6x)}{49152}+\frac{\cos(8x)}{4096}-\frac{\cos(10x)}{81920}-\frac{\cos(12x)}{24576}-\frac{\cos(14x)}{114688},$$

which clearly has a different form than our answer in Example 12.18, which we now refer to as $f(x)$. Figure 12.12 shows a graph of f and g; they are clearly not equal, but they differ only by a constant. That is $g(x) = f(x)+C$ for some constant C. So we have two different antiderivatives of the same function, meaning both answers are correct.

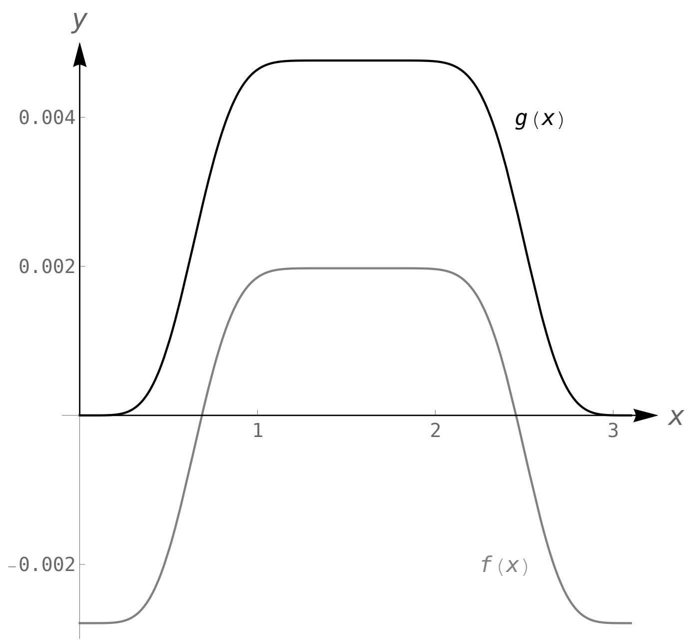


Figure 12.12: A plot of $f(x)$ and $g(x)$ from Example 12.18.

Example 12.19

Evaluate

$$\int \cos^4(x)\sin^2(x)\ dx.$$

Solution

The powers of sine and cosine are both even, so we employ the power–reducing formulas and algebra as follows.

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\ dx &= \int \left(\frac{1+\cos(2x)}{2}\right)^2 \left(\frac{1-\cos(2x)}{2}\right)\ dx\\ &= \int \frac{1+2\cos(2x)+\cos^2(2x)}{4}\cdot\frac{1-\cos(2x)}{2}\ dx\\ &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\ dx\end{aligned}$$

The $\cos(2x)$ term is easy to integrate. The $\cos^2(2x)$ term is another trigonometric integral with an even power, requiring the power–reducing formula again. The $\cos^3(2x)$ term is a cosine function with an odd power, requiring a substitution as done before. We integrate each in turn below.

$$\int \cos(2x)\ dx = \frac{1}{2}\sin(2x)+C.$$

$$\int \cos^2(2x)\ dx = \int \frac{1+\cos(4x)}{2}\ dx = \frac{1}{2}\left(x+\frac{1}{4}\sin(4x)\right)+C.$$

Finally, we rewrite $\cos^3(2x)$ as

$$\cos^3(2x) = \cos^2(2x)\cos(2x) = \big(1-\sin^2(2x)\big)\cos(2x).$$

Letting $u=\sin(2x)$, we have $du = 2\cos(2x)\ dx$, hence

$$\begin{aligned}\int \cos^3(2x)\ dx &= \int \big(1-\sin^2(2x)\big)\cos(2x)\ dx\\ &= \int \frac{1}{2}(1-u^2)\ du\\ &= \frac{1}{2}\left(u-\frac{1}{3}u^3\right)+C\\ &= \frac{1}{2}\left(\sin(2x)-\frac{1}{3}\sin^3(2x)\right)+C\,.\end{aligned}$$

Putting all the pieces together, we have

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\ dx &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\ dx\\ &= \frac{1}{8}\left[x+\frac{1}{2}\sin(2x)-\frac{1}{2}\left(x+\frac{1}{4}\sin(4x)\right)-\frac{1}{2}\left(\sin(2x)-\frac{1}{3}\sin^3(2x)\right)\right]+C\end{aligned}$$

$$= \frac{1}{8}\left[\frac{1}{2}x - \frac{1}{8}\sin(4x) + \frac{1}{6}\sin^3(2x)\right] + C.$$

12.4.3.2 Integrals of products of sines and cosines of differing period

Integrals of the form

$$\int \sin(mx)\sin(nx)\ dx, \quad \int \cos(mx)\cos(nx)\ dx \quad \text{and} \quad \int \sin(mx)\cos(nx)\ dx$$

are best approached by first applying the product to sum formulas (Theorem 5.13).

Example 12.20
Evaluate

$$\int \sin(5x)\cos(2x)\ dx.$$

Solution

The application of the appropriate Simpson formula and subsequent integration are straightforward:

$$\begin{aligned}\int \sin(5x)\cos(2x)\ dx &= \int \frac{1}{2}\left[\sin(3x) + \sin(7x)\right] dx \\ &= -\frac{1}{6}\cos(3x) - \frac{1}{14}\cos(7x) + C\end{aligned}$$

12.4.3.3 Integrals of the form $\int \tan^m(x)\sec^n(x)\ dx$.

When evaluating integrals of the form $\int \sin^m(x)\cos^n(x)\ dx$, the Pythagorean theorem allowed us to convert even powers of sine into even powers of cosine, and vise–versa. The same basic strategy applies to integrals of the form $\int \tan^m(x)\sec^n(x)\ dx$, albeit a bit more nuanced.

Basically, if the integrand can be manipulated to separate a $\sec^2(x)$ term with the remaining secant power even, or if a $\sec(x)\tan(x)$ term can be separated with the remaining $\tan(x)$ power even, the Pythagorean theorem can be employed, leading to a simple substitution. This strategy is outlined below.

1. If n is even, then $n = 2k$ for some integer k. Rewrite $\sec^n x$ as

$$\sec^n(x) = \sec^{2k}(x) = \sec^{2k-2}(x)\sec^2(x) = (1 + \tan^2(x))^{k-1}\sec^2(x).$$

   Then

$$\int \tan^m(x)\sec^n(x)\ dx = \int \tan^m(x)(1 + \tan^2(x))^{k-1}\sec^2(x)\ dx = \int u^m(1 + u^2)^{k-1}\ du,$$

   where $u = \tan(x)$ and $du = \sec^2(x)\ dx$.

2. If m is odd, then $m = 2k + 1$ for some integer k. Rewrite $\tan^m(x)\sec^n(x)$ as

$$\tan^m(x)\sec^n(x) \quad = \quad \tan^{2k+1}(x)\sec^n(x) = \tan^{2k}(x)\sec^{n-1}(x)\sec(x)\tan(x)$$

$$= \quad (\sec^2(x)-1)^k \sec^{n-1}(x)\sec(x)\tan(x).$$

Then

$$\int \tan^m(x)\sec^n(x)\,dx = \int (\sec^2(x)-1)^k \sec^{n-1}(x)\sec(x)\tan(x)\,dx = \int (u^2-1)^k u^{n-1}\,du,$$

where $u = \sec(x)$ and $du = \sec(x)\tan(x)\,dx$.

3. If n is odd and m is even, then $m = 2k$ for some integer k. Convert $\tan^m(x)$ to $(\sec^2(x)-1)^k$. Expand the new integrand and use Integration By Parts, with $dv = \sec^2(x)\,dx$.

4. If m is even and $n = 0$, rewrite $\tan^m(x)$ as

$$\tan^m(x) = \tan^{m-2}(x)\tan^2(x) = \tan^{m-2}(x)(\sec^2(x)-1) = \tan^{m-2}(x)\sec^2(x) - \tan^{m-2}(x).$$

So

$$\int \tan^m(x)\,dx = \underbrace{\int \tan^{m-2}(x)\sec^2(x)\,dx}_{\text{apply rule \#1}} - \underbrace{\int \tan^{m-2}(x)\,dx}_{\text{apply rule \#4 again}}.$$

Example 12.21

Evaluate the following indefinite integrals:

1. $\displaystyle\int \tan^2(x)\sec^6(x)\,dx$,
2. $\displaystyle\int \tan^6(x)\,dx$.

Solution

1. Since the power of secant is even, we use rule #1 above and pull out a $\sec^2(x)$ in the integrand. We convert the remaining powers of secant into powers of tangent.

$$\begin{aligned}\int \tan^2(x)\sec^6(x)\,dx &= \int \tan^2(x)\sec^4(x)\sec^2(x)\,dx \\ &= \int \tan^2(x)\big(1+\tan^2(x)\big)^2\sec^2(x)\,dx\end{aligned}$$

Now substitute, with $u = \tan(x)$, with $du = \sec^2(x)\,dx$:

$$= \int u^2\big(1+u^2\big)^2\,du.$$

We leave the integration and subsequent substitution to the reader. The final answer is

$$= \frac{1}{3}\tan^3(x) + \frac{2}{5}\tan^5(x) + \frac{1}{7}\tan^7(x) + C.$$

2. We employ rule #4 of the workflow outlined above.

$$\int \tan^6(x)\,dx = \int \tan^4(x)\tan^2(x)\,dx$$

$$= \int \tan^4(x)\left(\sec^2(x) - 1\right)\,dx$$

$$= \int \tan^4(x)\sec^2(x)\,dx - \int \tan^4(x)\,dx$$

Integrate the first integral with substitution, $u = \tan(x)$; integrate the second by employing rule #4 again.

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\tan^2(x)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\left(\sec^2(x) - 1\right)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\sec^2(x)\,dx + \int \tan^2(x)\,dx$$

Again, use substitution for the first integral and rule #4 for the second.

$$= \frac{1}{5}\tan^5(x) - \frac{1}{3}\tan^3(x) + \int \left(\sec^2(x) - 1\right)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \frac{1}{3}\tan^3(x) + \tan(x) - x + C$$

These latter examples were admittedly long, with repeated applications of the same rule. Try to not be overwhelmed by the length of the problem, but rather admire how robust this solution method is. A trigonometric function of a high power can be systematically reduced to trigonometric functions of lower powers until all antiderivatives can be computed.

12.4.4 Trigonometric substitution

We have since learned a number of integration techniques, yet we are still unable to evaluate an integral like

$$\int_{-3}^{3} \sqrt{9 - x^2}\,dx. \tag{12.21}$$

without resorting to a geometric interpretation. This section introduces **trigonometric substitution** (*goniometrische substitutie*), a method of integration that fills this gap in our integration skill. This technique works on the same principle as substitution, by setting set $x = f(\theta)$, where f is a trigonometric function, and then replacing x with $f(\theta)$.

For what concerns the integral given by Equation (12.21), we begin by noting that $9\sin^2(\theta) + 9\cos^2(\theta) = 9$, and hence $9\cos^2(\theta) = 9 - 9\sin^2(\theta)$. If we let $x = 3\sin(\theta)$, then $9 - x^2 = 9 - 9\sin^2(\theta) = 9\cos^2(\theta)$.

Setting $x = 3\sin(\theta)$ gives $dx = 3\cos(\theta)\,d\theta$. We are almost ready to substitute. We also wish to change our bounds of integration. The bound $x = -3$ corresponds to $\theta = -\pi/2$. Likewise, the bound of $x = 3$ is replaced by the bound $\theta = \pi/2$. Thus

$$\int_{-3}^{3} \sqrt{9 - x^2}\,dx = \int_{-\pi/2}^{\pi/2} \sqrt{9 - 9\sin^2(\theta)}(3\cos(\theta))\,d\theta$$

$$= \int_{-\pi/2}^{\pi/2} 3\sqrt{9\cos^2(\theta)}\cos(\theta)\; d\theta$$

$$= \int_{-\pi/2}^{\pi/2} 3\left|3\cos(\theta)\right|\cos(\theta)\; d\theta.$$

On $[-\pi/2, \pi/2]$, $\cos\theta$ is always positive, so we can drop the absolute value bars, then employ a power–reducing formula:

$$= \int_{-\pi/2}^{\pi/2} 9\cos^2(\theta)\; d\theta$$

$$= \int_{-\pi/2}^{\pi/2} \frac{9}{2}\big(1+\cos(2\theta)\big)\; d\theta$$

$$= \frac{9}{2}\left(\theta + \frac{1}{2}\sin(2\theta)\right)\Bigg|_{-\pi/2}^{\pi/2} = \frac{9}{2}\pi.$$

This matches our answer in Example 12.3.

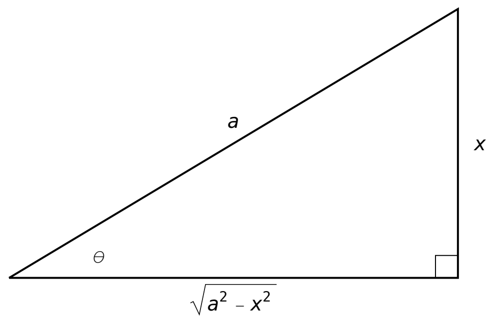
Trigonometric substitution excels when dealing with integrands that contain $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ and $\sqrt{x^2+a^2}$. The following outlines the procedure for each case. Each right triangle acts as a reference to help us understand the relationships between x and θ.

(a) For integrands containing $\sqrt{a^2-x^2}$:

Let $x = a\sin(\theta)$, then $dx = a\cos(\theta)\; d\theta$.

Thus $\theta = \arcsin(x/a)$, for $-\pi/2 \le \theta \le \pi/2$.

On this interval, $\cos(\theta) \ge 0$, so $\sqrt{a^2-x^2} = a\cos(\theta)$.

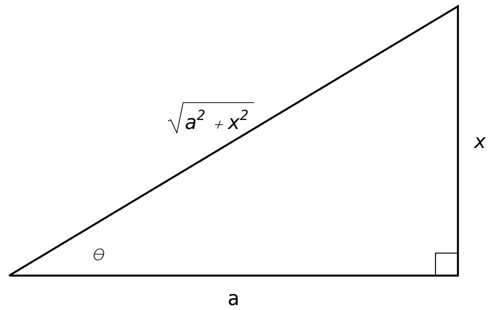


(b) For integrands containing $\sqrt{x^2+a^2}$:

Let $x = a\tan(\theta)$, then $dx = a\sec^2(\theta)\; d\theta$.

Thus $\theta = \arctan(x/a)$, for $-\pi/2 < \theta < \pi/2$.

On this interval, $\sec(\theta) > 0$, so $\sqrt{x^2+a^2} = a\sec(\theta)$.

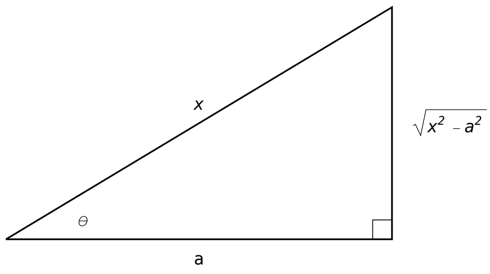


(c) For integrands containing $\sqrt{x^2-a^2}$:

Let $x = a\sec(\theta)$, then $dx = a\sec(\theta)\tan(\theta)\; d\theta$.

Thus $\theta = \operatorname{arcsec}(x/a)$. If $x/a \geq 1$, then $0 \leq \theta < \pi/2$; if $x/a \leq -1$, then $\pi/2 < \theta \leq \pi$.

We restrict our work to where $x \geq a$, so $x/a \geq 1$, and $0 \leq \theta < \pi/2$. On this interval, $\tan\theta \geq 0$, so $\sqrt{x^2-a^2} = a\tan(\theta)$.



Example 12.22

Evaluate

$$\int \sqrt{4x^2-1}\; dx.$$

Solution

We start by rewriting the integrand so that it looks like $\sqrt{x^2-a^2}$ for some value of a:

$$\begin{aligned}\sqrt{4x^2-1} &= \sqrt{4\left(x^2-\frac{1}{4}\right)}\\ &= 2\sqrt{x^2-\left(\frac{1}{2}\right)^2}.\end{aligned}$$

So we have $a = 1/2$, and following rule (c) from the above workflow, we set $x = \sec(\theta)/2$, and hence $dx = \sec(\theta)\tan(\theta)/2\; d\theta$. We now rewrite the integral with these substitutions:

$$\begin{aligned}\int \sqrt{4x^2-1}\; dx &= \int 2\sqrt{x^2-\left(\frac{1}{2}\right)^2}\; dx\\ &= \int 2\sqrt{\frac{1}{4}\sec^2(\theta)-\frac{1}{4}}\left(\frac{1}{2}\sec(\theta)\tan(\theta)\right)\; d\theta\\ &= \int \sqrt{\frac{1}{4}(\sec^2(\theta)-1)}\Big(\sec(\theta)\tan(\theta)\Big)\; d\theta\\ &= \int \sqrt{\frac{1}{4}\tan^2(\theta)}\Big(\sec\theta\tan(\theta)\Big)\; d\theta\\ &= \int \frac{1}{2}\tan^2(\theta)\sec(\theta)\; d\theta\\ &= \frac{1}{2}\int \Big(\sec^2(\theta)-1\Big)\sec(\theta)\; d\theta\\ &= \frac{1}{2}\int \big(\sec^3(\theta)-\sec(\theta)\big)\; d\theta.\end{aligned}$$

We can now integrate $\sec^3(\theta)$ using integration by parts with $dv = \sec^2(\theta)$ and $u = \sec(\theta)$, finding its antiderivatives to be

$$\int \sec^3(\theta)\, d\theta = \frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\big|\sec(\theta)+\tan(\theta)\big| \Big) + C.$$

Thus

$$\begin{aligned}\int \sqrt{4x^2-1}\, dx &= \frac{1}{2}\int \left(\sec^3(\theta) - \sec(\theta)\right) d\theta \\ &= \frac{1}{2}\left(\frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\big|\sec(\theta)+\tan(\theta)\big| \Big) - \ln\big|\sec(\theta)+\tan(\theta)\big|\right) + C \\ &= \frac{1}{4}\left(\sec(\theta)\tan(\theta) - \ln\big|\sec(\theta)+\tan(\theta)\big|\right) + C.\end{aligned}$$

We are not yet done. Our original integral is given in terms of x, whereas our final answer, as given, is in terms of θ. We need to rewrite our answer in terms of x. With $a = 1/2$, and $x = \sec(\theta)/2$, the reference triangle in rule (c) of the above workflow shows that

$$\tan\theta = \sqrt{x^2 - \frac{1}{4}}\Big/\frac{1}{2} = 2\sqrt{x^2-\frac{1}{4}} \quad \text{and} \quad \sec(\theta) = 2x.$$

Thus

$$\frac{1}{4}\Big(\sec(\theta)\tan(\theta) - \ln\big|\, \sec(\theta)+\tan(\theta)\big| \Big) + C$$

becomes

$$\frac{1}{4}\left(2x\, 2\sqrt{x^2-\frac{1}{4}} - \ln\left|2x + 2\sqrt{x^2-\frac{1}{4}}\right|\right) + C.$$

The final answer hence is:

$$\int \sqrt{4x^2-1}\, dx = \frac{1}{4}\left(4x\sqrt{x^2-\frac{1}{4}} - \ln\left|2x + 2\sqrt{x^2-\frac{1}{4}}\right|\right) + C.$$

It is important to realize that trigonometric substitution can be applied in many situations, even those not of the form $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ or $\sqrt{x^2+a^2}$. This is illustrated in the following example.

Example 12.23

Evaluate

$$\int \frac{1}{(x^2+6x+10)^2}\, dx.$$

Solution

We start by completing the square, then make the substitution $u = x+3$, followed by the trigonometric substitution of $u = \tan(\theta)$:

$$\int \frac{1}{(x^2+6x+10)^2}\, dx = \int \frac{1}{\big((x+3)^2+1\big)^2}\, dx = \int \frac{1}{(u^2+1)^2}\, du.$$

Now make the substitution $u = \tan(\theta)$, $du = \sec^2(\theta)\ d\theta$:

$$\begin{aligned}\int \frac{1}{(u^2+1)^2}\ du &= \int \frac{1}{(\tan^2(\theta)+1)^2} \sec^2(\theta)\ d\theta \\ &= \int \frac{1}{(\sec^2(\theta))^2} \sec^2(\theta)\ d\theta \\ &= \int \cos^2(\theta)\ d\theta.\end{aligned}$$

Applying a power reducing formula, we have

$$\begin{aligned}\int \cos^2(\theta)\ d\theta &= \int \left(\frac{1}{2} + \frac{1}{2}\cos(2\theta)\right)\ d\theta \\ &= \frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + C. \qquad (12.22)\end{aligned}$$

We need to return to the variable x. As $u = \tan(\theta)$, $\theta = \arctan(u)$. Using the identity $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ and using the reference triangle found in rule (b) of the workflow above, we have

$$\frac{1}{4}\sin(2\theta) = \frac{1}{2}\frac{u}{\sqrt{u^2+1}}\frac{1}{\sqrt{u^2+1}} = \frac{1}{2}\frac{u}{u^2+1}.$$

Finally, we return to x with the substitution $u = x+3$. We start with the expression in Equation (12.22):

$$\begin{aligned}\frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + C &= \frac{1}{2}\arctan(u) + \frac{1}{2}\frac{u}{u^2+1} + C \\ &= \frac{1}{2}\arctan(x+3) + \frac{x+3}{2(x^2+6x+10)} + C.\end{aligned}$$

Stating our final result in one line:

$$\int \frac{1}{(x^2+6x+10)^2}\ dx = \frac{1}{2}\arctan(x+3) + \frac{x+3}{2(x^2+6x+10)} + C.$$

Finally, it should be mentioned that given a definite integral that can be evaluated using trigonometric substitution, we could first evaluate the corresponding indefinite integral and then evaluate using the original bounds. It is much more straightforward, though, to change the bounds as we substitute.

12.4.5 Partial fraction decomposition

Here we investigate the antiderivatives of rational functions. Recall that rational functions are functions of the form $f(x) = \frac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are polynomials and $q(x) \neq 0$.

Consider the integral

$$\int \frac{1}{x^2-1}\ dx.$$

We do not have a simple formula for this. It can be evaluated using trigonometric substitution, but note how the integral is easy to evaluate once we realize:

$$\frac{1}{x^2-1} = \frac{1/2}{x-1} - \frac{1/2}{x+1}.$$

Thus

$$\int \frac{1}{x^2-1}\,dx = \int \frac{1/2}{x-1}\,dx - \int \frac{1/2}{x+1}\,dx$$
$$= \frac{1}{2}\ln|x-1| - \frac{1}{2}\ln|x+1| + C.$$

Here, we will learn how to decompose fractions like

$$\frac{1}{x^2-1}.$$

We start with a rational function

$$f(x) = \frac{p(x)}{q(x)},$$

where p and q do not have any common factors and the degree of p is less than the degree of q. It can be shown that any polynomial, and hence q, can be factored into a product of real linear and irreducible quadratic terms. The following workflow states how to **decompose a rational function into partial fractions** (*splitsing in partieelbreuken*) as a sum of rational functions whose denominators are all of lower degree than q.

1. **Linear Terms:** Let $(x-a)$ divide $q(x)$, where $(x-a)^n$ is the highest power of $(x-a)$ that divides $q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{A_1}{(x-a)} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_n}{(x-a)^n}.$$

2. **Quadratic Terms:** Let (x^2+bx+c) divide $q(x)$, where $(x^2+bx+c)^n$ is the highest power of (x^2+bx+c) that divides $q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{B_1x+C_1}{x^2+bx+c} + \frac{B_2x+C_2}{(x^2+bx+c)^2} + \cdots + \frac{B_nx+C_n}{(x^2+bx+c)^n}.$$

To find the coefficients A_i, B_i and C_i:

1. Multiply all fractions by $q(x)$, clearing the denominators. Collect like terms.
2. Equate the resulting coefficients of the powers of x and solve the resulting system of linear equations.

Example 12.24

Perform the partial fraction decomposition of

$$\frac{1}{x^2-1}.$$

Solution

The denominator factors into two linear terms: $x^2-1=(x-1)(x+1)$. Thus

$$\frac{1}{x^2-1} = \frac{A}{x-1} + \frac{B}{x+1}.$$

To solve for A and B, first multiply through by $x^2-1=(x-1)(x+1)$:

$$\begin{aligned} 1 &= \frac{A(x-1)(x+1)}{x-1}+\frac{B(x-1)(x+1)}{x+1} \\ &= A(x+1)+B(x-1) \\ &= Ax+A+Bx-B \\ &= (A+B)x+(A-B). \end{aligned}$$

The next step is key. Note the equality we have:

$$1=(A+B)x+(A-B).$$

For clarity's sake, rewrite the left hand side as

$$0x+1=(A+B)x+(A-B).$$

On the left, the coefficient of the x term is 0; on the right, it is $(A+B)$. Since both sides are equal, we must have that $0=A+B$.

Likewise, on the left, we have a constant term of 1; on the right, the constant term is $(A-B)$. Therefore we have $1=A-B$.

We have two linear equations with two unknowns. This one is easy to solve by hand, leading to

$$\begin{cases} A+B=0 \\ A-B=1 \end{cases} \quad \Rightarrow \quad \begin{cases} A=1/2 \\ B=-1/2. \end{cases}$$

Thus

$$\frac{1}{x^2-1}=\frac{1/2}{x-1}-\frac{1/2}{x+1}.$$

Clearly, it can become rather tedious to do a a partial fraction decomposition by hand if one is confronted with a more complex rational fraction. Luckily, we can resort in such cases to Mathematica, which can accomplish this with the command **Apart**. For instance, for what concerns the rational function in Example (12.24), we should proceed as follows.

```
In[21]:= Apart[1/(x^2 - 1), x]
```

The second argument of the command **Apart** is nothing but the variable at stake.

Out[21]= $\frac{1}{2\ (-1+x)}-\frac{1}{2\ (1+x)}$

Example 12.25

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{(x-1)(x+2)^2}\,dx,$
2. $\displaystyle\int \frac{x^3}{(x-5)(x+3)}\,dx,$
3. $\displaystyle\int \frac{2+\sin(x)}{3+\cos(x)}\,dx.$

Solution

1. We decompose the integrand as follows:

$$\frac{1}{(x-1)(x+2)^2} = \frac{A}{x-1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}.$$

To solve for A, B and C, we multiply both sides by $(x-1)(x+2)^2$ and collect like terms:

$$\begin{aligned} 1 &= A(x+2)^2 + B(x-1)(x+2) + C(x-1) \\ &= Ax^2 + 4Ax + 4A + Bx^2 + Bx - 2B + Cx - C \\ &= (A+B)x^2 + (4A+B+C)x + (4A-2B-C). \end{aligned} \tag{12.23}$$

We have

$$0x^2 + 0x + 1 = (A+B)x^2 + (4A+B+C)x + (4A-2B-C),$$

leading to the equations

$$\left\{ \begin{array}{rcl} A+B &=& 0 \\ 4A+B+C &=& 0 \\ 4A-2B-C &=& 1 \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{rcl} A &=& \frac{1}{9} \\ B &=& -\frac{1}{9} \\ C &=& -\frac{1}{3} \end{array} \right.$$

Thus

$$\int \frac{1}{(x-1)(x+2)^2}\,dx = \int \frac{1/9}{x-1}\,dx + \int \frac{-1/9}{x+2}\,dx + \int \frac{-1/3}{(x+2)^2}\,dx.$$

Each can be integrated with a simple substitution with $u = x-1$ or $u = x+2$. The end result is

$$\int \frac{1}{(x-1)(x+2)^2}\,dx = \frac{1}{9}\ln|x-1| - \frac{1}{9}\ln|x+2| + \frac{1}{3(x+2)} + C.$$

2. Since the degree of the numerator is now higher than the one of the denominator, we begin by using polynomial division to reduce the degree of the numerator (see Section 4.1). Doing so, we arrive at

$$\frac{x^3}{(x-5)(x+3)} = x + 2 + \frac{19x+30}{(x-5)(x+3)}.$$

Consequently, we can rewrite the new rational function as:

$$\frac{19x+30}{(x-5)(x+3)} = \frac{A}{x-5} + \frac{B}{x+3},$$

for appropriate values of A and B. Clearing denominators, we have

$$\begin{aligned} 19x+30 &= A(x+3) + B(x-5) \\ &= (A+B)x + (3A-5B). \end{aligned}$$

This implies that:

$$\begin{cases} 19 = A + B \\ 30 = 3A - 5B. \end{cases} \qquad \Leftrightarrow \qquad \begin{cases} A = \dfrac{125}{8} \\ B = \dfrac{27}{8}. \end{cases}$$

We can now integrate:

$$\begin{aligned} \int \frac{x^3}{(x-5)(x+3)}\, dx &= \int \left(x + 2 + \frac{125/8}{x-5} + \frac{27/8}{x+3} \right) dx \\ &= \frac{x^2}{2} + 2x + \frac{125}{8} \ln|x-5| + \frac{27}{8} \ln|x+3| + C. \end{aligned}$$

3. We observe that we are confronted with a rational function of trigonometric functions, so we first of all resort to the Weierstrass substitution. This leads to the following integral

$$2 \int \frac{t^2 + t + 1}{(t^2+2)(t^2+1)}\, dt\,,$$

which can be finished off using partial fraction decomposition. In this way, we get

$$2 \int \frac{t^2 + t + 1}{(t^2+2)(t^2+1)}\, dt = 2 \int \frac{1-t}{t^2+2}\, dt + \int \frac{t}{t^2+1}\, dt\,.$$

Hence, we arrive at

$$2 \int \frac{t^2 + t + 1}{(t^2+2)(t^2+1)}\, dt = \frac{1}{\sqrt{2}} \arctan\left(\frac{t}{\sqrt{2}}\right) - \frac{1}{2} \ln\left(t^2+2\right) + \frac{1}{2} \ln\left(t^2+1\right) + C\,,$$

where $t = \tan(x/2)$.

We conclude our discussion of partial fraction decomposition with a final example that combines several of the techniques we encountered earlier in this section.

Example 12.26

Evaluate

$$\int \frac{7x^2 + 31x + 54}{(x+1)(x^2+6x+11)}\, dx.$$

Solution

The degree of the numerator is less than the degree of the denominator, so we have:

$$\frac{7x^2 + 31x + 54}{(x+1)(x^2+6x+11)} = \frac{A}{x+1} + \frac{Bx + C}{x^2+6x+11}.$$

Now clear the denominators.

$$\begin{aligned} 7x^2 + 31x + 54 &= A(x^2 + 6x + 11) + (Bx + C)(x+1) \\ &= (A+B)x^2 + (6A + B + C)x + (11A + C). \end{aligned}$$

This implies that:

$$\begin{cases} 7 = A + B \\ 31 = 6A + B + C \\ 54 = 11A + C. \end{cases} \quad \Leftrightarrow \quad \begin{cases} A = 5 \\ B = 2 \\ C = -1. \end{cases}$$

Thus

$$\int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\,dx = \int \left(\frac{5}{x+1} + \frac{2x-1}{x^2+6x+11}\right)\,dx.$$

The first term of this new integrand is easy to evaluate; it leads to a $5\ln|x+1|$ term. The second term is not hard, but takes several steps and uses substitution techniques.

The integrand $\dfrac{2x-1}{x^2+6x+11}$ has a quadratic in the denominator and a linear term in the numerator. This leads us to try substitution. Let $u = x^2+6x+11$, so $du = (2x+6)\,dx$. The numerator is $2x-1$, not $2x+6$, but we can get a $2x+6$ term in the numerator by adding 0 in the form of "$7-7$."

$$\begin{aligned} \frac{2x-1}{x^2+6x+11} &= \frac{2x-1+7-7}{x^2+6x+11} \\ &= \frac{2x+6}{x^2+6x+11} - \frac{7}{x^2+6x+11}. \end{aligned}$$

We can now integrate the first term with substitution, leading to a $\ln|x^2+6x+11|$ term. The final term can be integrated using arctangent. First, complete the square in the denominator:

$$\frac{7}{x^2+6x+11} = \frac{7}{(x+3)^2+2}.$$

An antiderivative of the latter term can be found using Equation (12.13) and substitution:

$$\int \frac{7}{x^2+6x+11}\,dx = \frac{7}{\sqrt{2}}\arctan\left(\frac{x+3}{\sqrt{2}}\right) + C.$$

Let's start at the beginning and put all of the steps together.

$$\begin{aligned} \int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\,dx &= \int \left(\frac{5}{x+1} + \frac{2x-1}{x^2+6x+11}\right)\,dx \\ &= \int \frac{5}{x+1}\,dx + \int \frac{2x+6}{x^2+6x+11}\,dx - \int \frac{7}{x^2+6x+11}\,dx \\ &= 5\ln|x+1| + \ln\left|x^2+6x+11\right| - \frac{7}{\sqrt{2}}\arctan\left(\frac{x+3}{\sqrt{2}}\right) + C \end{aligned}$$

It is important to remember that one is not expected to see the final answer immediately after seeing the problem. Rather, given the initial problem, we break it down into smaller problems that are easier to solve. The final answer is a combination of the answers of the smaller problems.

Partial fraction decomposition is an important tool when dealing with rational functions. Note that at its heart, it is a technique of algebra, not calculus, as we are rewriting a fraction in a new form. Still, it is very useful in the realm of calculus as it lets us evaluate a certain set of complicated integrals.

Integral equations

In Chapter 9, we encountered differential equations, which are equations that relate some function with its derivatives. Likewise, we can formulate integral equations, which are equations in which an unknown function appears under an integral sign. Consider, for instance, the following integral equation:

$$f(x) = e^{-x} - \frac{1}{2} + \frac{1}{2}e^{-x-1} + \frac{1}{2}\int_0^1 (x+1)e^{-xy} f(y)\, dy.$$

Its solution is $f(x) = e^{-x}$, which can verified easily.

Just are the differential equations, integral equations are omnipresent in physics and engineering. For instance, Maxwell's equations of electromagnetism can be formulated in integral form.

12.5 Improper Integration

Consider the following definite integrals:

- $\displaystyle\int_0^{100} \frac{1}{1+x^2}\, dx \approx 1.5608,$
- $\displaystyle\int_0^{1000} \frac{1}{1+x^2}\, dx \approx 1.5698,$
- $\displaystyle\int_0^{10,000} \frac{1}{1+x^2}\, dx \approx 1.5707.$

Notice how the integrand is $1/(1+x^2)$ in each integral. It is sketched in Figure 12.13. As the upper bound gets larger, one would expect the area under the curve would also grow. While the definite integrals do increase in value as the upper bound grows, they are not increasing by much. In fact, consider:

$$\int_0^b \frac{1}{1+x^2}\, dx = \arctan(x)\Big|_0^b = \arctan(b) - \arctan(0) = \arctan(b).$$

As $b \to +\infty$, $\arctan(b) \to \pi/2$. Therefore it seems that as the upper bound b grows, the value of the concerned definite integral approaches $\pi/2 \approx 1.5708$. This should strike the reader as being a bit amazing: even though the curve extends to infinity, it has a finite amount of area underneath it.

When we defined the definite integral $\int_a^b f(x)\, dx$ in Definition 12.2, we made two stipulations:

1. The interval over which we integrated, $[a,b]$, was a finite interval, and
2. The function $f(x)$ was continuous on $[a,b]$ (ensuring that the range of f was finite).

In this section we consider integrals where one or both of the above conditions do not hold. Such integrals are called **improper integrals** (*oneigenlijke integraal*)

12.5.1 Improper integrals with infinite bounds

We start with a definition of Improper integrals with infinite bounds.

Definitie 12.6 (Improper integrals with infinite bounds)

1. Let f be a continuous function on $[a, +\infty[$. Define

$$\int_a^{+\infty} f(x)\, dx \quad \text{to be} \quad \lim_{b\to+\infty} \int_a^b f(x)\, dx.$$

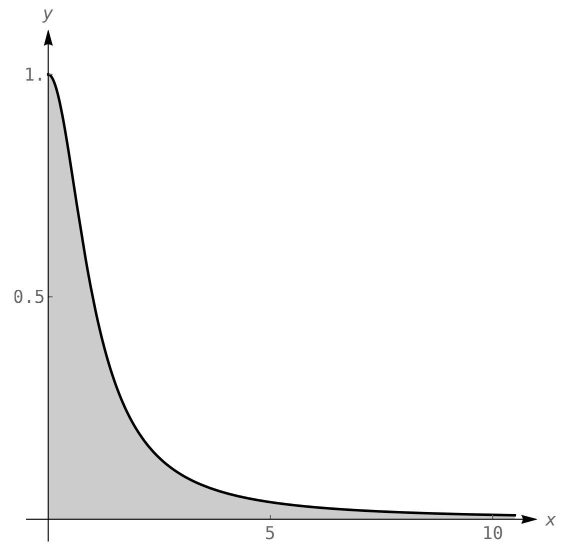


Figure 12.13: Graphing $f(x) = \dfrac{1}{1+x^2}$.

2. Let f be a continuous function on $]-\infty, b]$. Define

$$\int_{-\infty}^{b} f(x)\, dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{b} f(x)\, dx.$$

3. Let f be a continuous function on $\mathbb{R}$. Let c be any real number; define

$$\int_{-\infty}^{+\infty} f(x)\, dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{c} f(x)\, dx \;+\; \lim_{b\to+\infty} \int_{c}^{b} f(x)\, dx.$$

An improper integral is said to converge if its corresponding limit exists (is finite); otherwise, it diverges. The improper integral in part 3 converges if and only if both of its limits exist.

Example 12.27

Evaluate the following improper integrals:

1. $\displaystyle\int_{1}^{+\infty} \frac{1}{x^2}\, dx$,
2. $\displaystyle\int_{1}^{+\infty} \frac{1}{x}\, dx$,
3. $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\, dx$.

Solution

1.

$$[t] \int_{1}^{+\infty} \frac{1}{x^2}\, dx = \lim_{b\to+\infty} \int_{1}^{b} \frac{1}{x^2}\, dx = \lim_{b\to+\infty} \left.\frac{-1}{x}\right|_{1}^{b} \tag{12.24}$$

$$= \lim_{b\to+\infty} \frac{-1}{b} + 1 = 1. \tag{12.25}$$

A graph of the area defined by this integral is given in Figure 12.14(a). In Mathematica, this result can be checked as follows:

```
In[22]:= Integrate[1/x^2, x, 1, +Infinity]

Out[22]= 1
```

2.
$$\begin{aligned}\int_1^\infty \frac{1}{x}\,dx &= \lim_{b\to+\infty}\int_1^b \frac{1}{x}\,dx\\ &= \lim_{b\to+\infty} \ln|x|\Big|_1^b\\ &= \lim_{b\to+\infty} \ln(b)\\ &= +\infty.\end{aligned}$$

The limit does not exist, hence the concerned improper integral diverges. Compare the graphs in Figures 12.14(a) and 12.14(b); notice how the graph of $f(x) = 1/x$ is noticeably larger. This difference is enough to cause the improper integral to diverge.

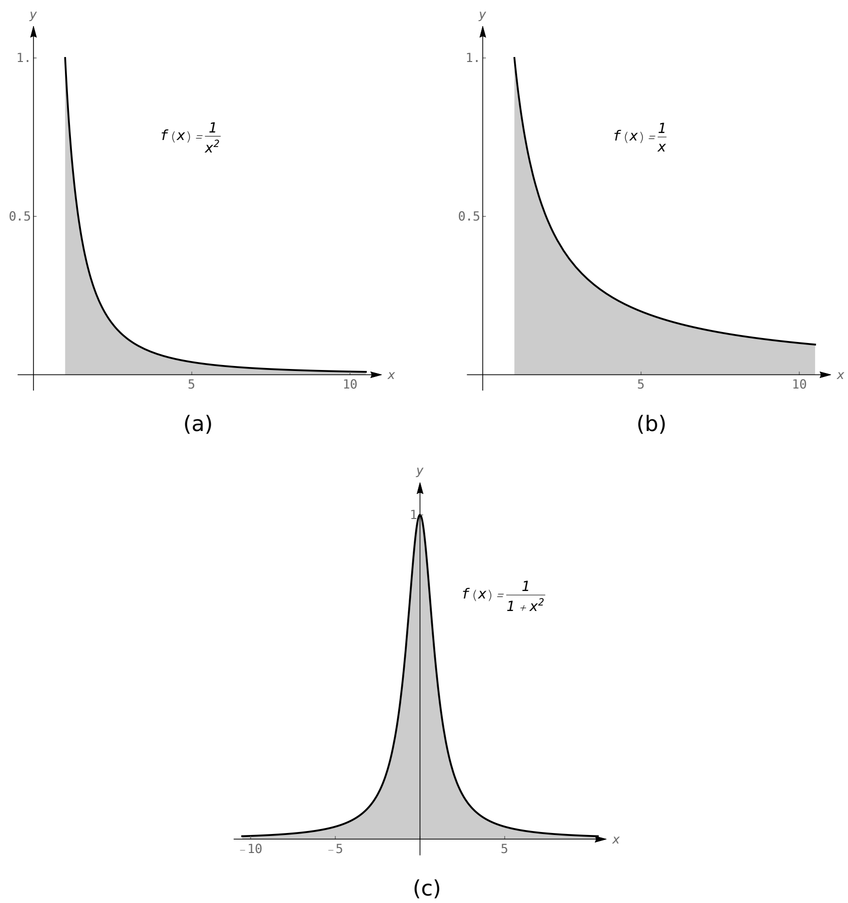


Figure 12.14: A graph of $f(x) = \frac{1}{x^2}$ (a), $f(x) = \frac{1}{x}$ (b) and $f(x) = \frac{1}{1+x^2}$ (c) in Example 12.27.

3. We will need to break this into two improper integrals and choose a value of c as in part 3 of Definition 12.6. Any value of c is fine; we choose $c = 0$.

$$
\begin{aligned}
\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx &= \lim_{a\to-\infty}\int_a^0 \frac{1}{1+x^2}\,dx + \lim_{b\to+\infty}\int_0^b \frac{1}{1+x^2}\,dx \\
&= \lim_{a\to-\infty} \arctan(x)\Big|_a^0 + \lim_{b\to+\infty} \arctan(x)\Big|_0^b \\
&= \lim_{a\to-\infty}\big(\arctan(0)-\arctan(a)\big) + \lim_{b\to+\infty}\big(\arctan(b)-\arctan(0)\big) \\
&= \left(0-\frac{-\pi}{2}\right)+\left(\frac{\pi}{2}-0\right)
\end{aligned}
$$

Each limit exists, hence the original integral converges and has value:

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx = \pi.$$

A graph of the area defined by this integral is given in Figure 12.14(c).

Note that it is not uncommon for the limits resulting from improper integrals to need l'Hôpital's rule.

12.5.2 Improper integrals with infinite range

We have just considered definite integrals where the interval of integration was infinite. We now consider another type of improper integration, where the range of the integrand is infinite.

Definitie 12.7 (Improper integrals with infinite range)
Let $f(x)$ be a continuous function on $[a,b]$ except at c, $a \le c \le b$, where $x=c$ is a vertical asymptote of f. Define

$$\int_a^b f(x)\,dx = \lim_{\substack{t\to c \\ <}} \int_a^t f(x)\,dx + \lim_{\substack{t\to c \\ >}} \int_t^b f(x)\,dx.$$

Example 12.28

Evaluate the following improper integrals:

1. $\displaystyle\int_0^1 \frac{1}{\sqrt{x}}\,dx$,
2. $\displaystyle\int_{-1}^1 \frac{1}{x^2}\,dx$.

Solution

1. A graph of $f(x)=1/\sqrt{x}$ is given in Figure 12.15(a). Notice that f has a vertical asymptote at $x=0$; in some sense, we are trying to compute the area of a region that has no top. Could

this have a finite value?

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = \lim_{a \xrightarrow[>]{} 0} \int_a^1 \frac{1}{\sqrt{x}}\,dx$$
$$= \lim_{a \xrightarrow[>]{} 0} 2\sqrt{x}\Big|_a^1$$
$$= \lim_{a \xrightarrow[>]{} 0} 2\left(\sqrt{1}-\sqrt{a}\right)$$
$$= 2$$

It turns out that the region does have a finite area even though it has no upper bound.

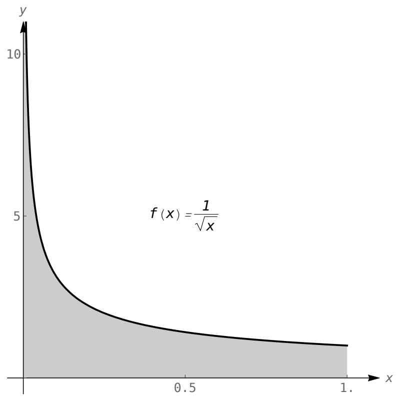
2. The function $f(x) = 1/x^2$ has a vertical asymptote at $x = 0$, as shown in Figure 12.15(b), so this integral is an improper integral. Let's eschew using limits for a moment and proceed without recognizing the improper nature of the integral. This leads to:

$$\int_{-1}^1 \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_{-1}^1$$
$$= -1-(1)$$
$$= -2.\ (!)$$

Clearly the area in question is above the x-axis, yet the area is supposedly negative! Why does our answer not match our intuition? To answer this, evaluate the integral using Definition 12.7.

$$\int_{-1}^1 \frac{1}{x^2}\,dx = \lim_{t \xrightarrow[<]{} 0} \int_{-1}^t \frac{1}{x^2}\,dx + \lim_{t \xrightarrow[>]{} 0} \int_t^1 \frac{1}{x^2}\,dx$$
$$= \lim_{t \xrightarrow[<]{} 0} \left(-\frac{1}{x}\right)\Big|_{-1}^t + \lim_{t \xrightarrow[>]{} 0} \left(-\frac{1}{x}\right)\Big|_t^1$$
$$= \lim_{t \xrightarrow[<]{} 0} \left(-\frac{1}{t}-1\right) + \lim_{t \xrightarrow[>]{} 0} \left(-1+\frac{1}{t}\right)$$
$$= \Big(+\infty-1\Big) + \Big(-1+\infty\Big)$$

Neither limit converges hence the original improper integral diverges. The nonsensical answer we obtained by ignoring the improper nature of the integral is just that: nonsensical.



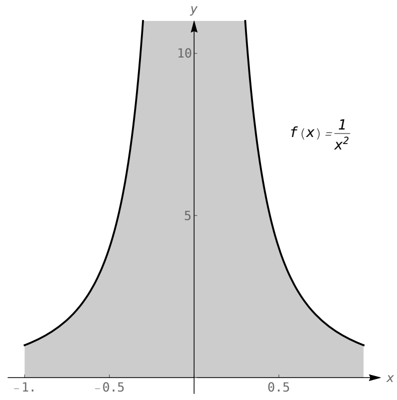


Figure 12.15: A graph of $f(x) = \frac{1}{\sqrt{x}}$ (a) and $f(x) = \frac{1}{x^2}$ (b) in Example 12.28.

12.5.3 Convergence and divergence

Oftentimes we are interested in knowing simply whether or not an improper integral converges, and not necessarily the value of a convergent integral. We provide here several tools that help determine the **convergence** (*convergentie*) or **divergence** (*divergentie*) of improper integrals without integrating.

For instance, let us try to determine the values of p for which

$$\int_1^{+\infty} \frac{1}{x^p}\, dx$$

converges.

We begin by integrating and then evaluating the limit:

$$\begin{aligned}
\int_1^{+\infty} \frac{1}{x^p}\, dx &= \lim_{b\to+\infty} \int_1^{b} \frac{1}{x^p}\, dx \\
&= \lim_{b\to+\infty} \int_1^{b} x^{-p}\, dx \qquad \text{(assume } p \neq 1\text{)} \\
&= \lim_{b\to+\infty} \frac{1}{-p+1} x^{-p+1}\Big|_1^b \\
&= \lim_{b\to+\infty} \frac{1}{1-p}\big(b^{1-p} - 1^{1-p}\big).
\end{aligned}$$

When does this limit converge – i.e., when is this limit not ∞? This limit converges precisely when the power of b is less than 0: when $1-p < 0 \Rightarrow 1 < p$.

So, if $p > 1$, then $\int_1^{\infty} \frac{1}{x^p}\, dx$ converges. When $p < 1$ the improper integral diverges; we showed in Example 12.27 that when $p = 1$ the integral also diverges. Figure 12.16 graphs $y = 1/x$ with a dashed line, along with graphs of $y = 1/x^p$, $p < 1$, and $y = 1/x^q$, $q > 1$. Somehow the dashed line forms a dividing line between convergence and divergence.

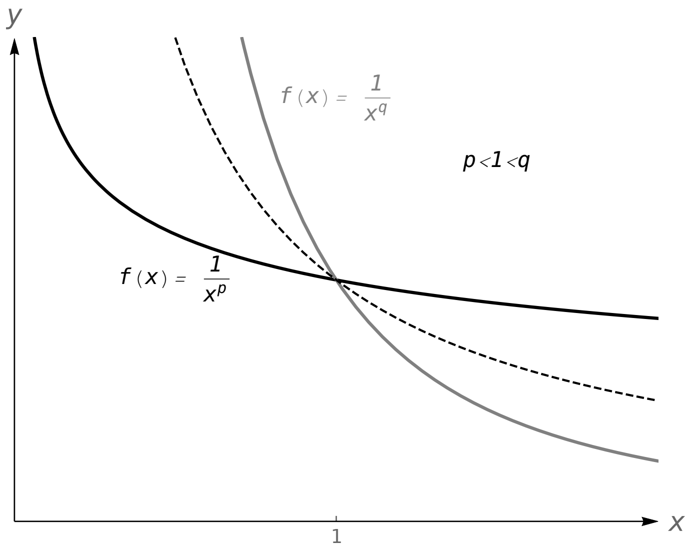


Figure 12.16: Plotting functions of the form $1/x^p$.

A similar result is proved in the exercises about improper integrals of the form

$$\int_0^1 \frac{1}{x^p}\,dx,$$

i.e. this improper integral converges when $p < 1$ and diverges when $p \geq 1$.

Note that we used the upper and lower bound of 1 just for convenience. It can be replaced by any a where $a > 0$.

A basic technique in determining convergence of improper integrals is to compare an integrand whose convergence is unknown to an integrand whose convergence is known. We often use integrands of the form $1/x^p$ to compare to as their convergence on certain intervals is known. This is described in the following theorem.

Theorem 12.9 (Direct comparison test for improper integrals)
Let f and g be continuous on $[a, +\infty[$ where $0 \leq f(x) \leq g(x)$ for all x in $[a, +\infty[$.

1. *If $\int_a^{+\infty} g(x)\,dx$ converges, then $\int_a^{+\infty} f(x)\,dx$ converges.*
2. *If $\int_a^{+\infty} f(x)\,dx$ diverges, then $\int_a^{+\infty} g(x)\,dx$ diverges.*

To prove Theorem 12.9, let us first of all prove the following theorem, which will need later on.

Theorem 12.10 (Limit of an increasing function bounded from above)
Let $F(t)$ be an increasing function on an interval $]\,a,+\infty\,[$. Assume there exists $M > 0$ such that $F(t) \leq M$ for all $t \in \,]\,a,+\infty\,[$. Then the following limit exists:

$$L = \lim_{t\to+\infty} F(t),$$

and $L \leq M$.

Proof Let S be the set of values of $F(t)$ on $]\,a,+\infty\,[$:

$$S = \{y \mid y = F(t),\ \forall t > a\}.$$

By assumption, S is bounded by M, that is, $y \leq M$ for all $y \in S$, so S has a least upper bound (supremum). Let $L = \sup(S)$. Then for all $\epsilon > 0$, $L-\epsilon$ is not an upper bound for S, so there exists some $y_0 > a$ such that $F(y_0) > L-\epsilon$. Since $F(t)$ is an increasing function, it follows that

$$L-\epsilon \leq F(y_0) \leq F(t) \leq L$$

for $t > y_0$. Therefore, $\left|L-F(t)\right| < \epsilon$ for $t > y_0$. Since ϵ is an arbitrary positive number, this is precisely what is needed to conclude that

$$L = \lim_{t\to+\infty} F(t).$$

□

Proof (of Theorem 12.9) Now to prove the first part of Theorem 12.9, consider the functions

$$G(t) = \int_a^t g(x)\,dx \qquad \text{and} \qquad F(t) = \int_a^t f(x)\,dx$$

They are defined for $t > a$. Since $f(x) \geq 0$ and $g(x) \geq 0$, both $F(t)$ and $G(t)$ are increasing. Furthermore, $f(x) \leq g(x)$ for all $x \geq a$ and therefore,

$$F(t) \leq G(t) \tag{12.26}$$

for all $t \geq a$.

Our assumption now is that the following improper integral converges:

$$M = \int_a^{+\infty} g(x)\,dx\,.$$

By definition, we have that $M = \lim\limits_{t\to+\infty} G(t)$. Since $G(t)$ is increasing, it holds that

$$G(t) \leq M$$

for all $t \geq a$, and it subsequently follows from Inequality (12.26) that

$$F(t) \leq M$$

for all $t \geq a$.

Since we have shown that $F(t)$ is increasing and bounded by M, we can conclude that $\lim\limits_{t\to+\infty} F(t)$ exists.

Since this limit is equal to the desired improper integral

$$\lim_{t\to+\infty} F(t) = \int_a^{+\infty} f(x)\,dx\,,$$

this concludes our proof of the first part. Moreover, the second part follows immediately. Indeed, assume that the first part is known to be true and that

$$\int_a^{+\infty} f(x)\,dx$$

diverges. Then

$$\int_a^{+\infty} g(x)\,dx$$

must diverge as well, for if it converged, the first part would imply that

$$\int_a^{+\infty} f(x)\,dx$$

converges. Similarly, the second part implies the first. □

There is also a counterpart of Theorem 12.9 for improper integrals with infinite range.

Theorem 12.11 (Direct comparison test for improper integrals with infinite range)
Let f and g be continuous on $[a, x_0[$ where $0 \le f(x) \le g(x)$ for all x in $[a, x_0[$.

1. *If $\displaystyle\int_a^{x_0} g(x)\,dx$ converges, then $\displaystyle\int_a^{x_0} f(x)\,dx$ converges.*
2. *If $\displaystyle\int_a^{x_0} f(x)\,dx$ diverges, then $\displaystyle\int_a^{x_0} g(x)\,dx$ diverges.*

Example 12.29

Determine the convergence of the following improper integrals.

1. $\displaystyle\int_1^{+\infty} e^{-x^2}\,dx$
2. $\displaystyle\int_3^{+\infty} \frac{1}{\sqrt{x^2-x}}\,dx$

Solution

1. The function $f(x) = e^{-x^2}$ does not have an antiderivative expressible in terms of elementary functions, so we cannot integrate directly. It is comparable to $g(x) = 1/x^2$, and as demonstrated in Figure 12.17(a), $e^{-x^2} < 1/x^2$ on $[1, +\infty[$. We know that $\int_1^{+\infty} x^{-2}\,dx$ converges, hence also the improper integral under consideration converges.

2. Note that for large values of x, we have

$$\frac{1}{\sqrt{x^2-x}} \approx \frac{1}{\sqrt{x^2}} = \frac{1}{x}.$$

We know that $\int_3^{+\infty} x^{-1}\, dx$ diverges, so we seek to compare the original integrand to $1/x$. It is easy to see that when $x > 0$, we have

$$x = \sqrt{x^2} > \sqrt{x^2-x} \quad \Leftrightarrow \quad \frac{1}{x} < \frac{1}{\sqrt{x^2-x}}.$$

Using Theorem 12.9, we conclude that since $\int_3^{+\infty} x^{-1}\, dx$ diverges, the concerned improper integral diverges as well. Figure 12.17(b) illustrates this.

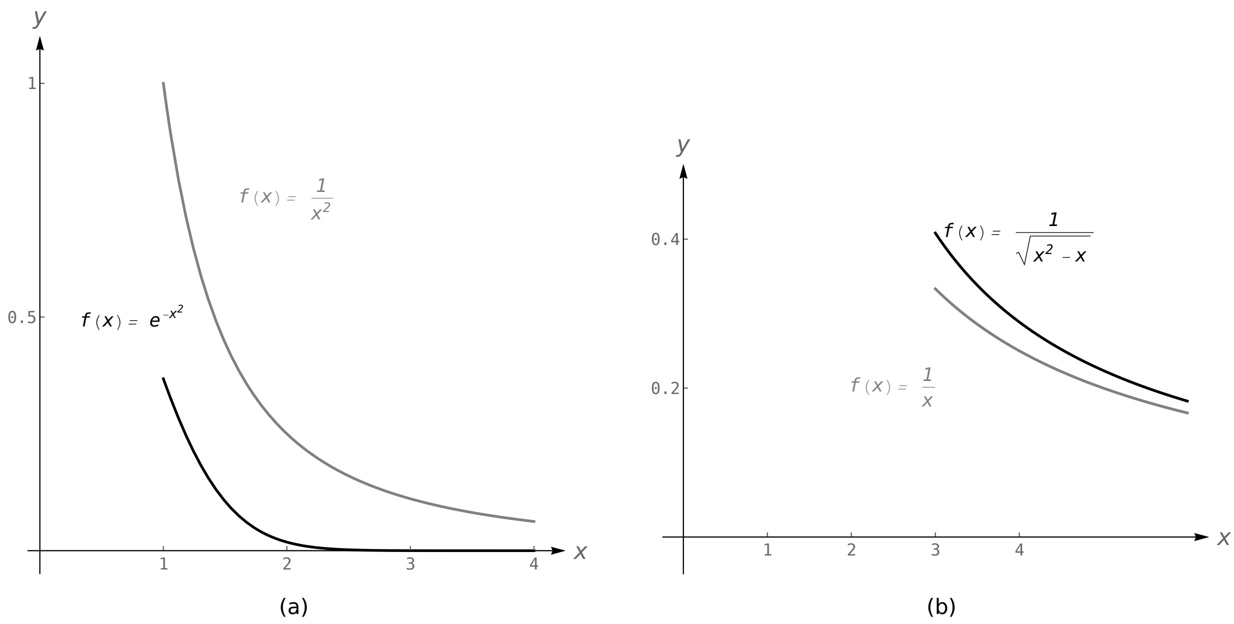


Figure 12.17: Graphs of $f(x) = e^{-x^2}$ and $f(x) = 1/x^2$ (a) and of $f(x) = 1/\sqrt{x^2-x}$ and $f(x) = 1/x$ (b) in Example 12.29.

Being able to compare unknown integrals to known integrals is very useful in determining convergence. However, some of our examples were a little too nice. For instance, it was convenient that $\frac{1}{x} < \frac{1}{\sqrt{x^2-x}}$, but what if the $-x$ were replaced with a $+2x+5$? That is, what can we say about the convergence of

$$\int_3^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\, dx?$$

We have

$$\frac{1}{x} > \frac{1}{\sqrt{x^2+2x+5}},$$

so we cannot use Theorem 12.9.

In cases like this (and many more) it is useful to employ the following theorem.

Theorem 12.12 (Limit comparison test for improper integrals)
Let f and g be continuous functions on $[a,+\infty[$ where $f(x)>0$ and $g(x)>0$ for all x. If

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)}=L, \qquad 0<L<+\infty,$$

then

$$\int_a^{+\infty} f(x)\,dx \text{ is convergent} \quad\Leftrightarrow\quad \int_a^{+\infty} g(x)\,dx \text{ is convergent},$$

and equivalently,

$$\int_a^{+\infty} f(x)\,dx \text{ is divergent} \quad\Leftrightarrow\quad \int_a^{+\infty} g(x)\,dx \text{ is divergent}.$$

Proof We assume that L exists and is a positive finite number, and that the limit from a to $+\infty$ of g converges; we will show that the limit from a to $+\infty$ of f converges as well.

The definition of the limit tells us that, given the number $\epsilon=L/2$, there exists some M such that

$$\frac{L}{2}=L-\epsilon<\frac{f(x)}{g(x)}<L+\epsilon=\frac{3L}{2}$$

whenever $x>M$. So, for those values of x, we have that

$$\frac{L}{2}g(x)<f(x)<\frac{3L}{2}g(x). \tag{12.27}$$

Let us now break the following integral in question into two pieces:

$$\int_a^{+\infty} f(x)\,dx=\int_a^{M} f(x)\,dx+\int_M^{+\infty} f(x)\,dx.$$

The first integral is of a continuous function on a closed, bounded interval, so we know that is finite. The convergence of the second integral is concluded by the following, which we can do because of Inequality (12.27):

$$\int_M^{+\infty} f(x)\,dx<\frac{3L}{2}\int_M^{+\infty} g(x)\,dx.$$

the last integral in this equation is given to converge (our assumption); therefore, by Theorem 12.9, the integral on the left converges as well. Hence, we conclude, as desired, that the integral of f converges.

Proving the other direction can be done similarly, or simply by observing that if $\lim_{x\to+\infty}\frac{f(x)}{g(x)}=k$ exists and is positive, then $\lim_{x\to+\infty}\frac{g(x)}{f(x)}=\frac{1}{k}$ must also exist and be positive. □

The limit comparison test can as well be given for unproper integrals with an infinite range.

Theorem 12.13 (Limit comparison test for improper integrals with infinite range)
Let f and g be continuous functions on $[a, x_0[$ where $f(x) > 0$ and $g(x) > 0$ for all x. If

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = L, \qquad 0 < L < \infty,$$

then

$$\int_a^{x_0} f(x)\, dx \text{ is convergent} \iff \int_a^{x_0} g(x)\, dx \text{ is convergent,}$$

and equivalently,

$$\int_a^{x_0} f(x)\, dx \text{ is divergent} \iff \int_a^{x_0} g(x)\, dx \text{ is divergent.}$$

Example 12.30

Determine the convergence of

$$\int_3^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\, dx.$$

Solution

As x gets large, the denominator of the integrand will begin to behave much like $y = x$. So we compare

$$\frac{1}{\sqrt{x^2+2x+5}}$$

to $1/x$ using Theorem 12.12:

$$\lim_{x \to +\infty} \frac{1/\sqrt{x^2+2x+5}}{1/x} = \lim_{x \to +\infty} \frac{x}{\sqrt{x^2+2x+5}}.$$

The immediate evaluation of this limit returns ∞/∞, an indeterminate form. Using l'Hôpital's rule seems appropriate, but in this situation, it does not lead to useful results.

The trouble is the square root function. To get rid of it, we employ the following fact: If $\lim_{x \to c} f(x) = L$, then $\lim_{x \to c} f(x)^2 = L^2$. So we consider now the limit

$$\lim_{x \to +\infty} \frac{x^2}{x^2+2x+5}.$$

This converges to 1, meaning the original limit also converged to 1. As x gets very large, the function $\dfrac{1}{\sqrt{x^2+2x+5}}$ looks very much like $\dfrac{1}{x}$. Since we know that

$$\int_3^{+\infty} \frac{1}{x}\, dx$$

diverges, by Theorem 12.12 we know that

$$\int_{3}^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\,dx$$

also diverges. Figure 12.18 graphs $f(x) = 1/\sqrt{x^2+2x+5}$ and $f(x) = 1/x$, illustrating that as x gets large, the functions become indistinguishable.

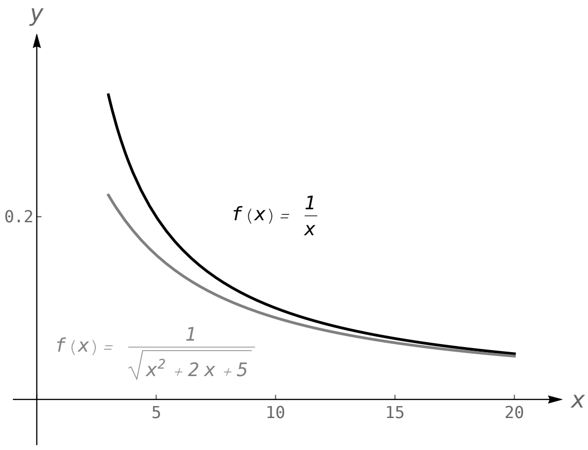


Figure 12.18: Graphing $f(x) = \frac{1}{\sqrt{x^2+2x+5}}$ and $f(x) = \frac{1}{x}$ in Example 12.30.

This chapter has explored many integration techniques. All of them effectively have one goal in common: rewrite the integrand in a new way so that the integration step is easier to see and implement. As stated before, integration is, in general, hard. It is easy to write a function whose antiderivative is impossible to write in terms of elementary functions, and even when a function does have an antiderivative expressible by elementary functions, it may be really hard to discover what it is. Mathematica, for instance, has approximately 1,000 pages of code dedicated to integration. Do not let this difficulty discourage you. There is great value in learning integration techniques, as they allow one to manipulate an integral in ways that can illuminate a concept for greater understanding. There is also great value in understanding the need for good numerical techniques. The next chapter stresses the uses of integration. We generally do not find antiderivatives for antiderivative's sake, but rather because they provide the solution to some type of problem. The following chapter introduces us to a number of different problems whose solution is provided by integration.

12.6 Exercises

12.6.1 Analytical exercises

Antiderivatives and (in)definite integration

Assignments 12.1 — Determine the area between the x-axis and the graph of the function:

$$f(x) = \begin{cases} x, & \text{als} \quad 0 < x \leq 1\,, \\ -2x+3, & \text{als} \quad 1 < x \leq 2\,, \\ -1, & \text{als} \quad 2 < x \leq 3\,, \\ 0, & \text{als} \quad x \leq 0 \vee x > 3\,. \end{cases}$$

Riemann sums

Assignments 12.2 — Consider a partition of the interval $[a,b]$ in n subintervals of equal width $\Delta x_i = (b-a)/n$. Determine the upper and lower Riemann sum for the given functions and n.

(a) $f(x) = x, \quad [0,2], \quad n = 8$

(b) $f(x) = \ln(x), \quad [1,2], \quad n = 5$

(c) $f(x) = \sin(x), \quad [0,\pi], \quad n = 6$

(d) $f(x) = \cos(x), \quad [0,2\pi], \quad n = 4$

(e) $f(x) = x^2, \quad [-3,3], \quad n = 6$

(f) $f(x) = \dfrac{1}{x}, \quad [1,9], \quad n = 4$

Assignments 12.3 — Express the given limit as a definite integral.

(a) $\displaystyle\lim_{n\to+\infty} \sum_{i=1}^{n} \frac{1}{n}\sqrt{\frac{i}{n}}$

(b) $\displaystyle\lim_{n\to+\infty} \sum_{i=1}^{n} \frac{1}{n}\sqrt{\frac{i-1}{n}}$

(c) $\displaystyle\lim_{n\to+\infty} \sum_{i=1}^{n} \frac{2}{n}\ln\left(1+\frac{2i}{n}\right)$

The fundamental theorem of calculus

Assignments 12.4 — If $a < b$ and f is continuous over $[a,b]$, prove that

$$\int_a^b \left(f(x) - \bar{f}\right) dx = 0,$$

with $\bar{f}$ being the mean of f.

Assignments 12.5 — Find the derivative of the functions below.

(a) $F(x) = \displaystyle\int_{2}^{x^3+x} \frac{1}{t}\,dt$

(b) $F(x) = \displaystyle\int_{x^3}^{0} t^3\,dt$

(c) $F(t) = \displaystyle\int_{-\pi}^{t} \frac{\cos(y)}{1+y^2}\,dy$

(d) $F(t) = \displaystyle\int_{t}^{3} \frac{\sin(x)}{x}\,dx$

(e) $F(x) = x^2 \displaystyle\int_{0}^{x^2} \frac{\sin(u)}{u}\,du$

(f) $F(\theta) = \displaystyle\int_{\sin(\theta)}^{\cos(\theta)} \frac{dx}{1-x^2}$

(g) $F(x) = 3x \displaystyle\int_{4}^{x^2} e^{-\sqrt{t}}\,dt$

(h) $F(x) = \displaystyle\int_{x}^{x^2} (t+2)\,dt$

(i) $F(x) = \displaystyle\int_{\ln(x)}^{e^x} \sin(t)\,dt$

Techniques of antidifferentiation

Assignments 12.6 — Find the following indefinite integrals.

(a) $\displaystyle\int \frac{e^x\sqrt{1-x^2}-1}{\sqrt{1-x^2}}\,dx$

(b) $\displaystyle\int \frac{2x+1}{4x^2+4x+3}\,dx$

(c) $\displaystyle\int \frac{\sin(x)}{\cos^6(x)}\,dx$

(d) $\displaystyle\int \cos^5(x)\,dx$

(e) $\displaystyle\int \frac{\sin(x)-\cos(x)}{\sin(x)+\cos(x)}\,dx$

(f) $\displaystyle\int \frac{dx}{\cos^2(x)\,\sqrt{1-4\tan^2(x)}}$

(g) $\displaystyle\int \frac{dx}{\left(\cos(x)+\sin(x)\right)^2}$

(h) $\displaystyle\int \ln\left(x+\sqrt{x^2+5}\right)\,dx$

(i) $\displaystyle\int \frac{2x-1}{x^2+x-6}\,dx$

(j) $\displaystyle\int \left(\frac{x-1}{x^2-5x+6}\right)^2\,dx$

(k) $\displaystyle\int \frac{x^2+1}{x^2+2x+2}\,dx$

(l) $\displaystyle\int \frac{x+1}{\left(x^2+1\right)^{3/2}}\,dx$

(m) $\displaystyle\int \frac{dx}{\sqrt{4x-x^2}}$

(n) $\displaystyle\int e^{2x}\sin(4x)\,dx$

(o) $\displaystyle\int \sin^4(x)\cos^2(x)\,dx$

(p) $\displaystyle\int \frac{\cos(x)}{2\cos^2(x)+\sin(x)-1}\,dx$

(q) $\displaystyle\int \frac{dx}{\sin^2(x)\cos^4(x)}$

(r) $\displaystyle\int \frac{dx}{\sinh(x)}$

(s) $\displaystyle\int \tanh^3(x)\,dx$

Assignments 12.7 — Find the following indefinite integrals.

(a) $\displaystyle\int \sin(x)\ \sinh(x)\,dx$

(b) $\displaystyle\int \frac{3x^2-4}{x^2+1}\,dx$

(c) $\displaystyle\int \frac{x^4}{x^3-8}\,dx$

(d) $\displaystyle\int \frac{dx}{x^4\sqrt{x^2-1}}$

(e) $\displaystyle\int \frac{3-4x}{\left(1-2\sqrt{x}\right)^2}\,dx$

(f) $\displaystyle\int \frac{dx}{\left(\tan(x)+1\right)\sin^2(x)}$

(g) $\displaystyle\int x\,e^{2x}\,dx$

(h) $\displaystyle\int \frac{5x}{\sqrt{x^4+1}}\,dx$

(i) $\displaystyle\int \sin\left(\frac{\pi}{4}-x\right)\sin\left(\frac{\pi}{4}+x\right)dx$

(j) $\displaystyle\int \frac{dx}{\sqrt[4]{5-x}+\sqrt{5-x}}$

Assignments 12.8 — Find the following indefinite integrals.

(a) $\displaystyle\int x^2\ln\left(\sqrt{1-x}\right)dx$

(b) $\displaystyle\int \frac{2x-1}{2x+3}\,dx$

(c) $\displaystyle\int \sin(2x)\cos(2x)\,dx$

(d) $\displaystyle\int \frac{dx}{e^x+1}$

(e) $\displaystyle\int \frac{dx}{x^2+x+1}$

(f) $\displaystyle\int \frac{2x+3}{\left(x^2+x+1\right)^2}\,dx$

(g) $\displaystyle\int \sqrt{\frac{a+x}{a-x}}\,dx$

(h) $\displaystyle\int \frac{x-2\sqrt{x-1}}{1+\sqrt[4]{x-1}}\,dx$

(i) $\displaystyle\int \arctan(\sqrt{x})\,dx$

(j) $\displaystyle\int x^5\left(1+x^3\right)^{1/2}\,dx$

(k) $\displaystyle\int \frac{dx}{\sin^3(x)\ \cos^5(x)}$

(l) $\displaystyle\int \frac{dx}{\sin^6(x)}$

(m) $\displaystyle\int \frac{dx}{\sqrt{1+e^x}}$

(n) $\displaystyle\int 2^x\cosh(x)\,dx$

(o) $\displaystyle\int \sin^4(x)\,dx$

(p) $\displaystyle\int \frac{dx}{1+\cos(x)+\sin(x)}$

Improper Integration

Assignments 12.9 — Investigate the convergence of the following improper integrals. Also give an explanation.

(a) $\displaystyle\int_0^{\frac{\pi}{2}} \sec(x)\, dx$

(b) $\displaystyle\int_0^{+\infty} e^{-x}\sin(x)\, dx$

(c) $\displaystyle\int_{-1}^{8} x^{-\frac{2}{3}}\, dx$

(d) $\displaystyle\int_{-2}^{1} \frac{1}{\sqrt{|x|}}\, dx$

(e) $\displaystyle\int_0^{+\infty} xe^{x^2}\, dx$

(f) $\displaystyle\int_{-1}^{1} \frac{1}{x^4}\, dx$

(g) $\displaystyle\int_0^{e^2} \big(1+\ln(x)\big)\, dx$

(h) $\displaystyle\int_0^{+\infty} \frac{x^2}{x^5+1}\, dx$

(i) $\displaystyle\int_0^{+\infty} \frac{dx}{1+\sqrt{x}}$

(j) $\displaystyle\int_0^{+\infty} \frac{dx}{\sqrt{x}+x^2}$

(k) $\displaystyle\int_{-1}^{1} \frac{e^x}{x+1}\, dx$

(l) $\displaystyle\int_2^{+\infty} \frac{x\sqrt{x}}{x^2-1}\, dx$

(m) $\displaystyle\int_0^{\frac{\pi}{2}} \frac{\cos(x)}{\sqrt{x}}\, dx$

(n) $\displaystyle\int_1^{+\infty} \frac{\sin(x)}{x^2}\, dx$

Assignments 12.10 — The shape of the spectral lines in magnetic resonance spectroscopy is often described by the Lorentz function

$$g(\omega) = \frac{1}{\pi}\cdot\frac{T}{1+T^2(\omega-\omega_0)^2},$$

with T and ω_0 constants. Evaluate

$$\int_{\omega_0}^{+\infty} g(\omega)\, d\omega\,.$$

Review exercises

Assignments 12.11 — Prove

$$I_n = \int_0^{\frac{\pi}{2}} \sin^n(x)\, dx = \left(\frac{n-1}{n}\right) I_{n-2}, \quad n \in \mathbb{N}\setminus\{0,1\}.$$

Evaluate $\displaystyle\int_0^{\frac{\pi}{2}} \sin^9(x)\, dx$.

Assignments 12.12 — Assume $n \in \mathbb{N}\backslash\{0,1\}$. Set up a formula for the definite integrals below.

(a) $I_n = \int\limits_0^{\pi} \sin^n(x)\,dx$

(b) $I_n = \int\limits_0^{\pi} \cos^n(x)\,dx$

(c) $I_n = \int\limits_0^{2\pi} \sin^n(x)\,dx$

(d) $I_n = \int\limits_0^{2\pi} \cos^n(x)\,dx$

Assignments 12.13 — Evaluate the definite integrals below.

(a) $\int\limits_4^9 \frac{1-\sqrt{x}}{1+\sqrt{x}}\,dx$

(b) $\int\limits_0^2 \sqrt{4-x^2}\,\frac{|x-1|}{x-1}\,dx$

Assignments 12.14 — The gamma function $\Gamma(x)$ can be defined by the improper integral:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}\,e^{-t}\,dt.$$

(a) Prove that the integral converges for $x > 0$.

(b) Prove by using integration by parts that $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$.

(c) Prove $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$

(d) Assume

$$\int_0^{+\infty} e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}.$$

Prove

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \qquad \text{en} \qquad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

Assignments 12.15 — The probability that a molecule with mass m in a gas at temperature T, has velocity v is given by the Maxwell-Boltzmann distribution

$$f(v) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} v^2 e^{-\frac{mv^2}{2kT}},$$

with k a constant.

Show that the average speed $\overline{v}$, given by $\overline{v} = \int_0^{+\infty} v f(v)\,dv$, equals $\overline{v} = \sqrt{\frac{8kT}{\pi m}}$.

Use the reduction formula.

$$I_n = \frac{n-1}{2a} I_{n-2} \qquad \text{met} \quad I_n = \int_0^{+\infty} e^{-ax^2} x^n\,dx.$$

12.6.2 Numerical integration

Finding an antiderivative is far from obvious. Although Sections 12.4 and 12.5 provided many integration techniques, still, in many cases these techniques are not useful. Because, for example, the antiderivative cannot be expressed in terms of elementary function(s). It becomes even more difficult if we do not have a closed-form function in the integrand, something that occurs constantly in practice. What do we do in these cases? We approximate the (definite) integral as a sum of computable areas (Section 12.2.1). This method is conceptually very simple, but it is tedious if we want to obtain an approximation with an acceptable accuracy. Therefore, here we will implement and study some numerical integration methods in Python.

12.6.2.1 The midpoint method

We can approximate a definite integral by summing the areas of a series of rectangles. For the definite integral

$$S = \int_a^b f(x)\, dx\,,$$

we obtain these n rectangles as follows:

- Subdivide the integration interval $[a,b]$ into a partition $\{x_1, x_2, \ldots, x_n, x_{n+1}\}$, where
$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b\,;$$
- the length of the i-th subinterval $[x_i, x_{i+1}]$, denoted by Δx, is the base of the i-th rectangle;
- the height of the i-th rectangle is determined by the left, right, or midpoint rule.

If Δx is positive and we apply the midpoint rule, we call the resulting method the **midpoint method**. The approximation $\hat{\boldsymbol{S}}$ of an integral S then follows from

$$\begin{aligned} S = \int_a^b f(x)\, dx &\approx \Delta x f\left(\frac{x_1 + x_2}{2}\right) + \Delta x f\left(\frac{x_2 + x_3}{2}\right) + \cdots + \Delta x f\left(\frac{x_n + x_{n+1}}{2}\right) \\ &\approx \Delta x \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right) = \hat{\boldsymbol{S}}\,. \end{aligned} \tag{12.28}$$

If we rewrite Equation (12.28) as

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f(m_i) = \Delta x \sum_{i=1}^{n} f\left(a + \frac{\Delta x}{2} + (i-1)\,\Delta x\right),$$

where $m_i = \frac{x_i + x_{i+1}}{2}$ is the midpoint of interval $[x_i, x_{i+1}]$, we can translate the midpoint method to executable Python code as follows.

```
def midpoint(f, interval, n=20, plotf = False):
        '''
```

```
	Midpoint method for the approximation of the integral
	of f on a given interval [a,b]
	Inputs:
		- f: integrand
		- interval: integration interval [a,b], given as a list
		- n: number of subintervals (default: 20)
		- plotf (optional): indicates whether the integrand should be plotted
(default: False)

	Output(s):
		- S_h: the approximated integral on [a,b] using the midpoint method
		* m_list: list with the midpoints of the n subintervals
		* fm_list: list with the function values of the midpoints of the n subintervals
		* deltax: width of the subintervals
	[*] these outputs are only returned if the function is plotted (plotf=True)
	'''
    # check if n is a strictly positive integer. If not, display an error message.
	if not (isinstance(n, int) and n>0):
	print('Error: n must be a strictly positive integer')
	return None

	# extract the values of a and b from the interval
	a = interval[0]
	b = interval[1]

	# calculate the width of the subintervals
	deltax = (b-a)/n

	# make a list with the midpoints of the n subintervals [m_1,...,m_n]
	# and a list with the function values [f(m_1),...,f(m_n)]
	m_list =  [a+deltax*i-deltax/2 for i in range(1,n+1)]
	fm_list = [f(m_i) for m_i in m_list]

	# calculate the approximated integral
	S_h = deltax*sum(fm_list)

	if plotf:
		return S_h, m_list, fm_list, deltax
	else:
		return S_h
```

The function **`plot_numerical_integration(method, f, interval, n=[1,100],interactive=True)`** makes a static or interactive plot of f and the approximation of the integral on the interval $[a, b]$ with a specified numerical integration method. In the interactive plot, the number of subintervals can be determined by means of a slider. The inputs of this function are defined as follows:

- method: numerical integration method (`'midpoint'` or `'trapezium'`)
- f: integrand
- interval: integration interval $[a, b]$
- n: number of subintervals
  - `interactive = True` : range of values $[n_{min}, n_{max}]$ (default = $[1, 100]$)
  - `interactive = False`: one value for n

Question 1.a Test the function **`midpoint`** for the definite integral from Example 12.4.

$$S_1 = \int_{a_1}^{b_1} f_1(x)\,dx = \int_0^4 (4x - x^2)\,dx.$$

```
def f_1(x):
    return 4*x-x**2

    # call to the function midpoint
    ...
```

Question 1.b Use the function **`plot_numerical_integration`** to plot $f_1(x)$ and the approximation of the corresponding integral on the interval $[0.4]$ with $n = 10$.

Now make an interactive plot of $f_1(x)$ and the approximation of the corresponding integral over the interval $[0,4]$ with $n \in [1,20]$.

Question 1.c Using the instruction(s) below, check the computing time for $n = 100, 10^4, 10^6, \ldots$. What is the influence of n?

```
%timeit midpoint(f=f_1, interval=[0,4], n=100)
```

12.6.2.2 Approximation error

To quantitatively check the accuracy of the numerical integration method, we can use the relative approximation error ϵ:

$$\epsilon = \frac{|S - \hat{\mathbf{S}}|}{|S|}.$$

To investigate the effect of the number of subintervals n on ϵ, we use the function **`plot_error(S, method, f, interval, n_range)`**. It plots the relative approximation error ϵ of a specified numerical integration method as a function of the number of subintervals n. The inputs of this function are defined as follows:

- `S`: exact value of the integral
- method: numerical integration method (`'midpoint'` or `'trapezium'`)
- `f`: integrand
- `interval`: integration interval $[a,b]$
- `n\_range`: range of values for n $[n_{min}, n_{max}]\,(default = [1,20])$

```
from teachingtools import plot_error
```

Question 2.a We want to test the function **`plot_error`** for the approximation of S_1 for n ranging from 1 to 20. For this, we need S_1. First calculate the exact value of S_1 and enter it below.

```
S_1 = ... # to be completed
```

Question 2.b Now plot with the function `plot_error` the relative approximation error (ϵ) of the integral of $f_1(x)$ for n ranging from 1 to 20.

```
# to be completed with call to the function plot_error
...
```

Question 2.c Implement the integrand of the following definite integrals and calculate their exact value (if possible):

- $S_2 = \int_{a_2}^{b_2} f_2(x)\,dx = \int_0^2 \sin(2x)\cos(2x)\,dx\,,$
- $S_3 = \int_{a_3}^{b_3} f_3(x)\,dx = \int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin\left(x^3\right)\,dx\,.$

What do you notice?

```
def f_2(x):
        return ... # to be completed
S_2 = ... # to be completed with the exact value of the integral of f2 on [0,2]


def f_3(x):
        return ... # to be completed
S_3 = ... # to be completed with the exact value of the integral of f3 on [-pi/4,pi/2]
```

Question 2.d For the integral of $f_2(x)$ and $f_3(x)$, calculate the approximation with the midpoint method, and plot the relative approximation error as a function of n (for n ranging from 1 to 20) with **`plot_error`**.

```
# numerical approximations for S_2 (no plot)

# errors for S_2

# numerical approximations for S_3 (no plot)

# errors for S_3 (if possible)
```

12.6.2.3 The trapezium method

The midpoint method approximates a definite integral as a sum of rectangular areas. The integrand in each subinterval $]x_i, x_{i+1}[$ is approximated by a constant function $f(m_i)$. However, this approximation is not accurate in subintervals where the function value changes significantly. This can be clearly seen in the plot of f_3.

An alternative to the midpoint method is the **trapezium method**, where the areas in the subintervals are approximated by - surprise - trapezia. This method approximates the integrand over the subinterval $[x_i, x_{i+1}]$ as a straight line going from $\left(x_i, f(x_i)\right)$ to $\left(x_{i+1}, f(x_{i+1})\right)$. The definite integral is then approximated as follows:

$$S = \int_a^b f(x)\,dx \quad \approx \quad \Delta x \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i+1})}{2} = \hat{\boldsymbol{S}}\,. \tag{12.29}$$

Question 3.a Implement the trapezium method by completing the function below where you find "...".

```
def trapezium(f, interval, n = 20, plotf = False):
    '''
    Trapezium method for the approximation of the integral
    of the function f on a given interval [a,b]
    Inputs:
        - f: integrand
        - interval: integration interval [a,b], given as a list
        - n: number of subintervals (default: 20)
        - plotf (optional): indicates whether the function
                            should be plotted (default: False)

    Output:
        - S_h: the approximated integral on [a,b] using the trapezium method
        * x_list: list with the boundaries of the n subintervals [x_1,...,x_n+1]
        * f_list: list with the function values of the boundaries of
                  the n subintervals [f(x_1),...,f(x_n+1)]
        * deltax: width of the subintervals
        * these outputs are only returned if the function
         is plotted (plotf = True)
    '''
    # Check if n is a strictly positive integer.
    # If not, display an error message.
    if not (isinstance(n, int) and n > 0):
        print('Error: n must be a strictly positive integer')
        return None

    # 1) Extract the values of a and b from the interval
    ...
    ...

    # 2) Calculate the width of the subintervals
    ...
    ...

    # 3) Create a vector with the (n+1) values of x [x_0, ..., x_n]
     # in the interval [a,b]
    ...

    # 4) Create a vector with the n+1 values of f [f(x_0), ..., f(x_n)]
     # of the n+1 x values in the interval [a, b]
    ...

    # 5) Calculate the approximated integral
    ...

    if plotf == True:
        return S_h, x_list, f_list, deltax
```

```
    else:
        return S_h
```

Question 3.b Compare the results of the midpoint and trapezium methods for f_1, f_2 and f_3. Do you notice any differences? Which method do you prefer and why?

```
% % approximation for S1 with midpoint

% % approximation for S1 with trapezium

% % approximation for S2 with midpoint

% % approximation for S2 with trapezium

% % approximation for S3 with midpoint

% % approximation for S3 with trapezium
```

Question 3.c Consider the following definite integral:

$$S_{tot} = \int_a^b f_t(x)\,dx = \int_a^b \big(f_s(x) + f_r(x)\big)\,dx.$$

The integrand $f_t(x)$ here is a sum of two functions, i.e. a signal function $f_s(x)$ and noise function $f_r(x)$, of which we do not know the mathematical form. We constantly encounter this in practice, for example when we use a measuring device on which dirt is deposited (after a while). In this case, x is the time, $f_s(x)$ is the quantity to be measured as a function of time and $f_r(x)$ is the disturbance of the signal due to dirt deposition as a function of time.

Import the functions f_s, f_r and f_t and check for both numerical methods the effect of the noise on the approximated integral of $f_t(x)$ on the interval $[0.20]$. Would you prefer one method over another?

12.6.2.4 Simpson's rule

The trapezium rule locally replaces the function with a straight line so that the areas in the subintervals can be approximated by trapezia. Intuitively, you expect the approximation to get better if the function is locally approximated by a nonlinear function that is still easy to integrate.

Simpson's method, also called Simpson's 1/3 method, approximates the integrand $f(x)$ by means of a quadratic curve or parabola

$$f(x) \mapsto P_2(x) = a_0 + a_1 x + a_2 x^2. \tag{12.30}$$

On the one hand, we now need three conditions to determine the three coefficients a_0, a_1 and a_2 in Eq. (12.30). On the other hand, we do not immediately know the formula for the area under a parabola. However, we can easily find it. To do this, we consider three points at a distance Δx from each other: $(0, y_i)$, $(\Delta x, y_{i+1})$ and $(2\Delta x, y_{i+2})$. The parabola $P_2(x)$ can then be identified by requiring that these three points should lie on it:

$$x = 0 \quad \rightarrow \quad P_2(0) = a_0 \overset{\text{require}}{=} y_i,$$

$$x = \Delta x \quad \rightarrow \quad P_2(\Delta x) = a_0 + a_1 \Delta x + a_2 \Delta x^2 \overset{\text{require}}{=} y_{i+1},$$
$$x = 2\Delta x \quad \rightarrow \quad P_2(2\Delta x) = a_0 + 2a_1 \Delta x + 4a_2 \Delta x^2 \overset{\text{require}}{=} y_{i+2}.$$

If we substitute $a_0 = y_i$ in the second and third equation, we obtain a linear 2 × 2 system:

$$\begin{aligned} \Delta x a_1 + \Delta x^2 a_2 &= y_{i+1} - y_i, \\ 2\Delta x a_1 + 4\Delta x^2 a_2 &= y_{i+2} - y_i, \end{aligned}$$

with solution

$$\begin{aligned} a_1 &= -\frac{1}{2\Delta x}\left(3y_i - 4y_{i+1} + y_{i+2}\right), \\ a_2 &= \frac{1}{2\Delta x^2}\left(y_i - 2y_{i+1} + y_{i+2}\right). \end{aligned}$$

This allows us to unambiguously determine the coefficients in Eq. (12.30).

Question 4.a Show analytically that the area under the parabola passing through the three considered points is given by

$$S_i = \int_0^{2h} P_2(x)\, dx = \frac{\Delta x}{3}\left(y_i + 4y_{i+1} + y_{i+2}\right).$$

Note that the area S_i only depends on the heights y_i, y_{i+1} and y_{i+2} and on Δx.

Simpson's method divides the integration interval in n subintervals of equal width $\Delta x = (b-a)/n$. We require that n is even and define $x_1 = a$ and $x_{i+1} = x_i + \Delta x$ for $i = 1, \ldots, n$, so that $x_{n+1} = b$.
The integral of the function $f(x)$ on the interval $[a, b]$ can now be approximated by

$$\begin{aligned} S = \int_a^b f(x)\, dx \quad &\approx \quad \sum_{i=0}^{\frac{n}{2}-1} \left(\int_{x_{2i}}^{x_{2i+2}} P_2(x)\, dx \right) \\ &= \sum_{i=0}^{\frac{n}{2}-1} S_{2i} \\ &= \sum_{i=0}^{\frac{n}{2}-1} \frac{\Delta x}{3}\left(y_i + 4y_{i+1} + y_{i+2}\right) = \hat{\boldsymbol{S}} \end{aligned}$$

The expanded form of the latter is easier to remember:

$$\hat{\boldsymbol{S}} = \frac{\Delta x}{3}\left(f(x_0) + 4 \sum_{\substack{i=1 \\ i \text{ odd}}}^{n-1} f(x_i) + 2 \sum_{\substack{i=2 \\ i \text{ even}}}^{n-2} f(x_i) + f(x_n) \right).$$

For example, if $n = 8$, the approximation is given by:

$$\hat{\boldsymbol{S}} = \frac{\Delta x}{3}\Big(f(x_0) + 4\big(f(x_1) + f(x_3) + f(x_5) + f(x_7)\big) + 2\big(f(x_2) + f(x_4) + f(x_6)\big) + f(x_8)\Big).$$

Question 4.b Implement Simpson's method by completing the function below where you find "...".

```
def simpson(f, interval, n):
```

```
    """Simpson's method for approximating the integral of f on an interval [a,b]
    Inputs:
    - f: integrand (function handle)
    - interval: integration interval [a, b] given as a 1x2 row vector
    - n: the number of subintervals
    Outputs:
    - Sh: the approximation of the integral on [a,b] using Simpson's method
    - If the value of n is odd, there will be an error message

      'The value of n should be even!' and the function will stop"""

    if n%2 != 0:
        raise ValueError("...")
    # Extract the start and end value a and b from the interval
    ...
    ...

    # Get the interval width h
    ...

    # calculate the approximated integral Sh based on the start and end value of the interval
    Sh = ..

    # calculate the approximation of the integral Sh using all start and end values from the
      partial intervals

    for i in range(...):
        if n%2 == 0:
            Sh += ...
        else:
            Sh += ...
    return Sh
```

Question 4.c Use the function **`simpson`** to approximate

$$\int_{\frac{\pi}{10}}^{\frac{\pi}{2}} \frac{\sin(x)}{x}\, dx$$

using 10 subintervals. Give your result up to 5 digits after the decimal point.

Engineering is achieving function while avoiding failure.

— Henry Petroski —

13

Applications of integration

This chapter employs the following technique to a variety of applications. Suppose the value Q of a quantity is to be calculated. We first approximate the value of Q using a Riemann sum, then find the exact value via a definite integral. This goes as follows.

1. Divide the quantity into n smaller subquantities of value Q_i.
2. Identify a variable x and function $f(x)$ such that each subquantity can be approximated with the product $f(c_i)\Delta x$, where Δx represents a small change in x. Thus $Q_i \approx f(c_i)\Delta x$. A sample approximation $f(c_i)\Delta x$ of Q_i is called a **differential element**.
3. Recognize that
$$Q = \sum_{i=1}^{n} Q_i \approx \sum_{i=1}^{n} f(c_i)\Delta x,$$
which is a Riemann Sum.
4. Taking the appropriate limit gives $Q = \int_a^b f(x)\, dx$.

13.1 Area between curves

13.1.1 Rectangular coordinates

Let Q be the area of a region bounded by continuous functions f and g. If we break the region into many subregions, we have an obvious equation:

Total area = sum of the areas of the subregions.

The issue to address next is how to systematically break a region into subregions. Consider Figure 13.1(a) where a region between two curves is shaded. While there are many ways to break this into subregions, one particularly efficient way is to slice it vertically, as shown in Figure 13.1(b), into n equally spaced slices.

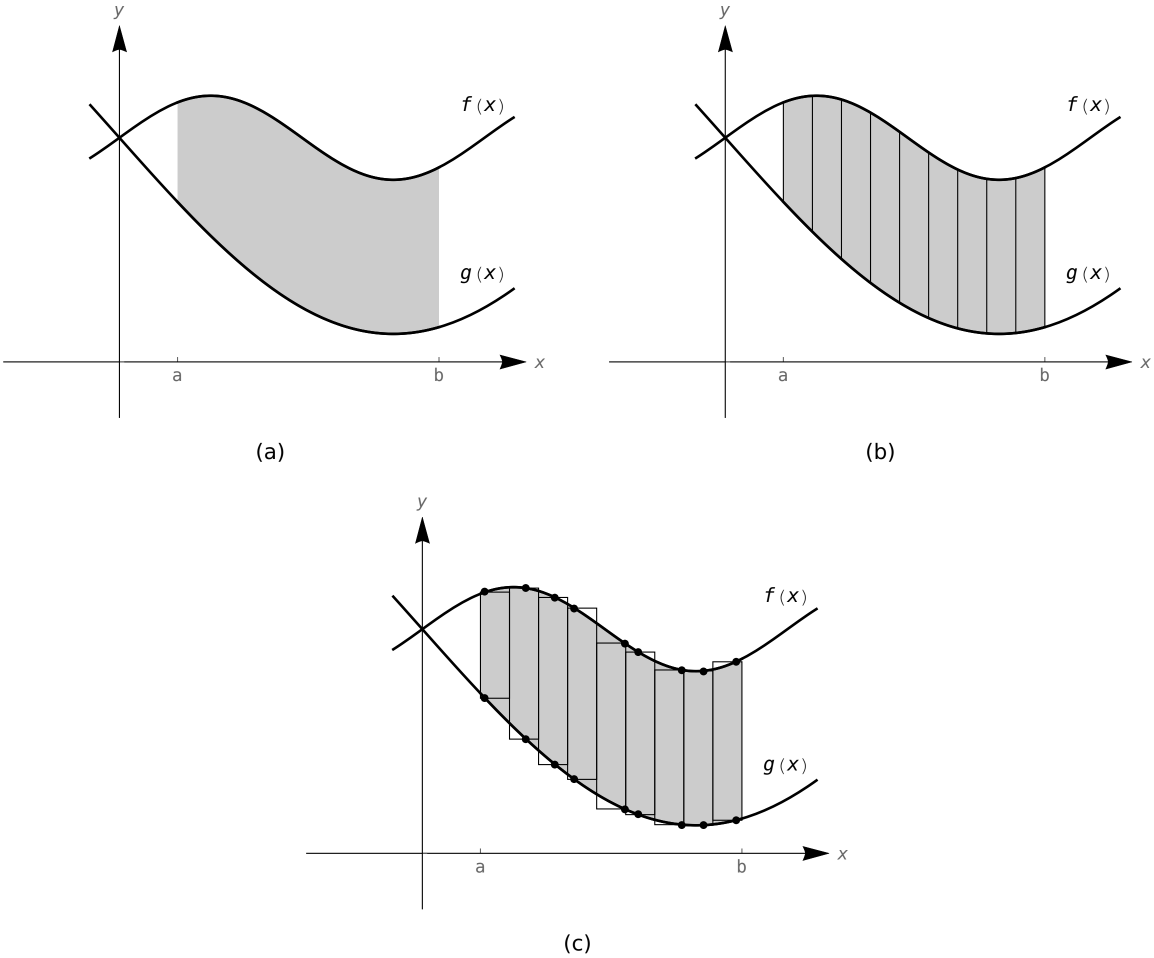


Figure 13.1: Subdividing a region into vertical slices and approximating the areas with rectangles.

We now approximate the area of a slice. Again, we have many options, but using a rectangle seems simplest. Picking any x-value c_i in the i^{th} slice, we set the height of the rectangle to be $f(c_i)-g(c_i)$, the difference of the corresponding y-values. The width of the rectangle is a small difference in x-values, which we represent with Δx. Figure 13.1(c) shows sample points c_i chosen in each subinterval and appropriate rectangles drawn. Each of these rectangles represents a differential element. Each slice has an area approximately equal to $\big(f(c_i)-g(c_i)\big)\Delta x$; hence, the total area is approximately the Riemann sum

$$Q=\sum_{i=1}^{n}\big(f(c_i)-g(c_i)\big)\Delta x.$$

Taking the limit as $n\to+\infty$ gives the exact area A as

$$\int_a^b \big(f(x)-g(x)\big)\,dx. \tag{13.1}$$

Example 13.1

Find the total area of the region enclosed by the functions $f(x)=-2x+5$ and $g(x)=x^3-7x^2+12x-3$ as shown in Figure 13.2.

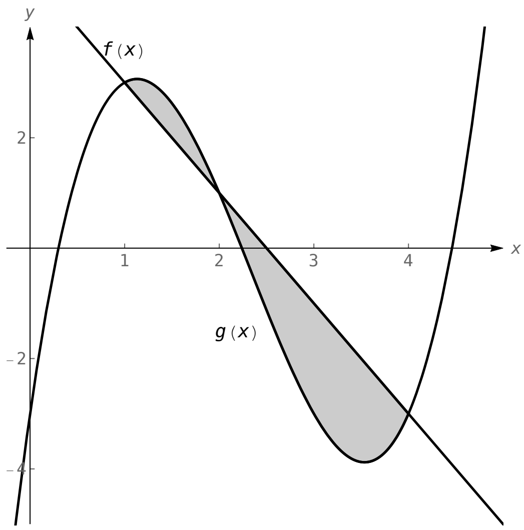


Figure 13.2: Graphing a region enclosed by two functions in Example 13.1.

Solution

A quick calculation shows that $f = g$ at $x = 1, 2$ and 4. One can proceed thoughtlessly by computing $\int_1^4 \big(f(x) - g(x)\big)\,dx$, but this ignores the fact that on $[1,2]$, $g(x) > f(x)$. Thus we compute the total area by breaking the interval $[1,4]$ into two subintervals, $[1,2]$ and $[2,4]$ and using the proper integrand in each.

$$\begin{aligned}\text{Total Area} &= \int_1^2 \big(g(x) - f(x)\big)\,dx + \int_2^4 \big(f(x) - g(x)\big)\,dx\\ &= \int_1^2 \big(x^3 - 7x^2 + 14x - 8\big)\,dx + \int_2^4 \big(-x^3 + 7x^2 - 14x + 8\big)\,dx\\ &= \frac{5}{12} + \frac{8}{3}\\ &= \frac{37}{12} = 3.083\ \text{units}^2.\end{aligned}$$

The previous example makes note that we are expecting area to be positive. When first learning about the definite integral, we interpreted it as signed area under the curve, allowing for negative area. That does not apply here; area is to be positive.

The previous example also demonstrates that we often have to break a given region into subregions. The following example shows another situation where this is applicable.

Example 13.2

Find the area of the region enclosed by the functions $y = \sqrt{x} + 2$, $y = -(x-1)^2 + 3$ and $y = 2$, as shown in Figure 13.3(a).

Solution

We give two approaches to this problem. In the first approach, we notice that the region's top is defined by two different curves. On $[0,1]$, the top function is $y=\sqrt{x}+2$; on $[1,2]$, the top function is $y=-(x-1)^2+3$.

Thus we compute the area as the sum of two integrals:

$$\begin{aligned} A &= \int_0^1 \Big(\big(\sqrt{x}+2\big)-2\Big)\,dx + \int_1^2 \Big(\big(-(x-1)^2+3\big)-2\Big)\,dx \\ &= \frac{2}{3}+\frac{2}{3} \\ &= \frac{4}{3}. \end{aligned}$$

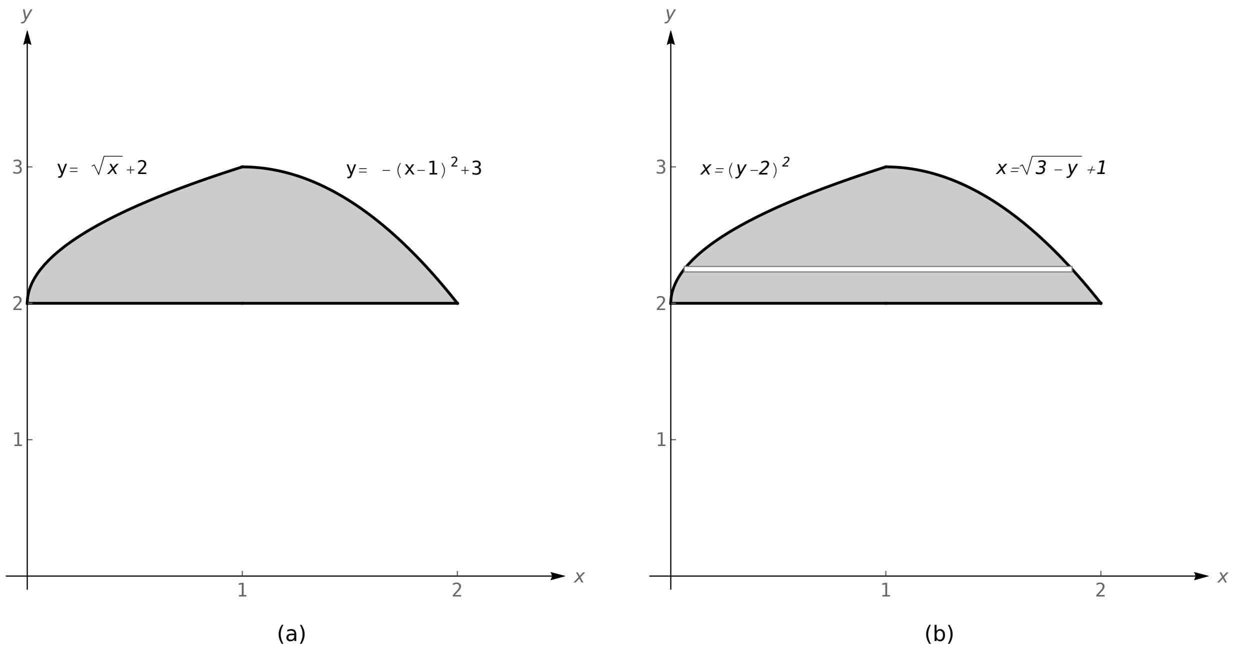


Figure 13.3: Graphing a region for Example 13.2 (a) and the region with boundaries relabelled as functions of y (b).

The second approach is clever and very useful in certain situations. We are used to viewing curves as functions of x; we input an x-value and a y-value is returned. Some curves can also be described as functions of y: input a y-value and an x-value is returned. We can rewrite the equations describing the boundary by solving for x:

$$y=\sqrt{x}+2 \quad\Rightarrow\quad x=(y-2)^2$$

$$y=-(x-1)^2+3 \quad\Rightarrow\quad x=\sqrt{3-y}+1.$$

Figure 13.3(b) shows the region with the boundaries relabelled. A differential element, a horizontal rectangle, is also pictured. The width of the rectangle is a small change in y: Δy. The height of the rectangle is a difference in x-values. The top x-value is the largest value, i.e., the rightmost.

The bottom x-value is the smaller, i.e., the leftmost. Therefore the height of the rectangle is

$$\left(\sqrt{3-y}+1\right)-(y-2)^2.$$

The area is found by integrating the above function with respect to y with the appropriate bounds. We determine these by considering the y-values the region occupies. It is bounded below by $y=2$, and bounded above by $y=3$. That is, both the top and bottom functions exist on the y interval $[2,3]$. Thus

$$\begin{aligned} A &= \int_2^3 \left(\sqrt{3-y}+1-(y-2)^2\right)\,dy \\ &= \left(-\frac{2}{3}(3-y)^{3/2}+y-\frac{1}{3}(y-2)^3\right)\Big|_2^3 \\ &= \frac{4}{3}. \end{aligned}$$

While we have focused on producing exact answers, we are also able to make approximations. The integrand in Equation (13.1) is a distance (top minus bottom); integrating this distance function gives an area. By taking discrete measurements of distance, we can approximate an area using numerical integration techniques developed in Chapter 12.

In the next section we apply our applications–of–integration techniques to finding the volumes of certain solids.

13.1.2 Polar coordinates

When using polar coordinates, the equations $\theta=\alpha$ form lines through the origin and $r=c$ form circles centred at the origin, respectively, and combinations of these curves form sectors of circles. It is then somewhat natural to calculate the area of regions defined by polar functions by first approximating with sectors of circles.

Consider Figure 13.4(a) where a region defined by $r=f(\theta)$ on $[\alpha,\beta]$ is given. Note how the sides of the region are the lines $\theta=\alpha$ and $\theta=\beta$, whereas in rectangular coordinates the sides of regions were often the vertical lines $x=a$ and $x=b$.

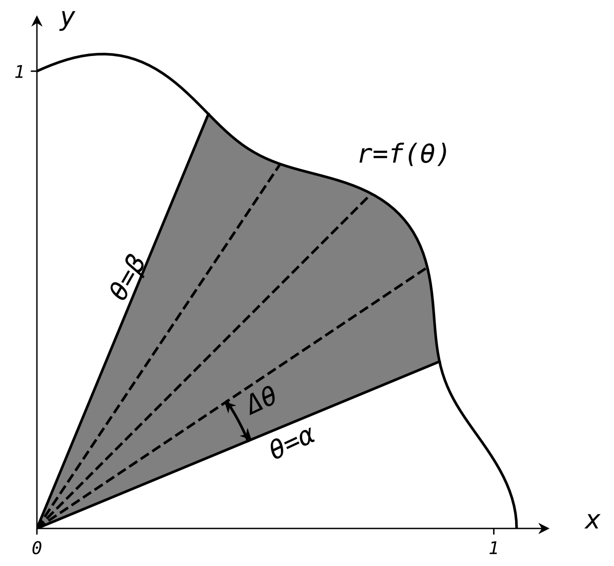
Partition the interval $[\alpha,\beta]$ into n equally spaced subintervals as $\alpha=\theta_1<\theta_2<\cdots<\theta_{n+1}=\beta$. The length of each subinterval is $\Delta\theta=(\beta-\alpha)/n$, representing a small change in angle. The area of the region defined by the i^{th} subinterval $[\theta_i,\theta_{i+1}]$ can be approximated with a sector of a circle with radius $f(c_i)$, for some c_i in $[\theta_i,\theta_{i+1}]$. The area of this sector is $\frac{1}{2}f(c_i)^2\Delta\theta$. This is shown in Figure 13.4(b), where $[\alpha,\beta]$ has been divided into 4 subintervals. We approximate the area of the whole region by summing the areas of all sectors:

$$\text{Area} \approx \sum_{i=1}^{n}\frac{1}{2}f(c_i)^2\Delta\theta.$$

This is a Riemann sum. By taking the limit of the sum as $n\to+\infty$, we find the exact area of the region in the form of a definite integral:

$$A \;=\; \frac{1}{2}\int_\alpha^\beta f(\theta)^2\,d\theta \;=\; \frac{1}{2}\int_\alpha^\beta r^2\,d\theta. \tag{13.2}$$

By having $0 \leq \beta - \alpha \leq 2\pi$, we ensure that the region does not overlap itself, which would give a result that does not correspond directly to the area.



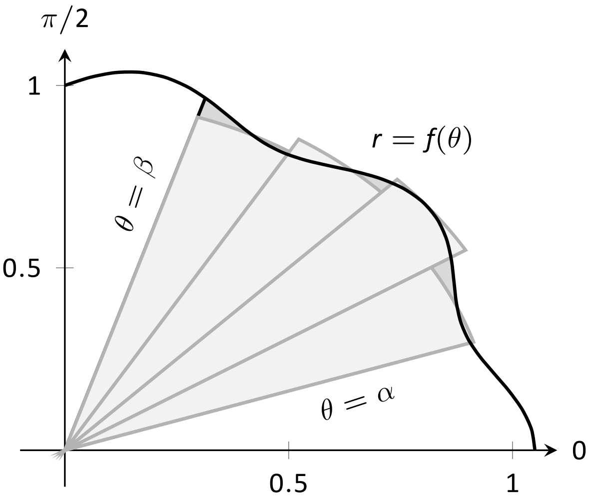


Figure 13.4: Computing the area of a polar region.

Example 13.3

Find the area of the cardioid $r = 1 + \cos(\theta)$ bound between $\theta = \pi/6$ and $\theta = \pi/3$, as shown in Figure 13.5.

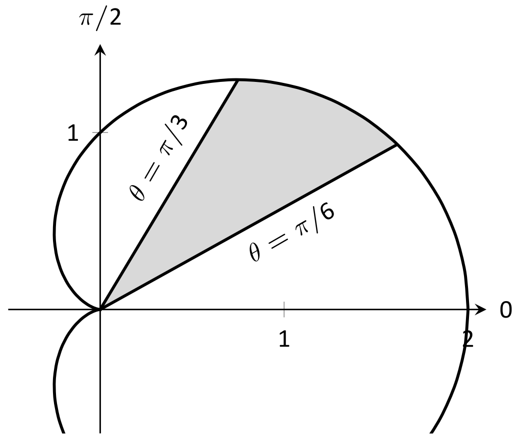


Figure 13.5: Finding the area of the shaded region of a cardioid in Example 13.3.

Solution

This is a direct application of Equation (13.2).

$$\text{Area} = \frac{1}{2}\int_{\pi/6}^{\pi/3} (1+\cos(\theta))^2 \, d\theta$$

$$= \frac{1}{2}\int_{\pi/6}^{\pi/3} \left(1+2\cos(\theta)+\cos^2(\theta)\right)\, d\theta$$

$$= \frac{1}{2}\left(\theta+2\sin(\theta)+\frac{1}{2}\theta+\frac{1}{4}\sin(2\theta)\right)\Bigg|_{\pi/6}^{\pi/3}$$

$$= \frac{1}{8}\left(\pi+4\sqrt{3}-4\right) \approx 0.7587$$

We may of course also determine the region enclosed between two polar curves. Consider for that purpose the shaded region shown in Figure 13.6. We can find the area of this region by computing the area bounded by $r_2 = f_2(\theta)$ and subtracting the area bounded by $r_1 = f_1(\theta)$ on $[\alpha, \beta]$. Thus

$$\text{Area} \;=\; \frac{1}{2}\int_{\alpha}^{\beta} r_2^2\, d\theta - \frac{1}{2}\int_{\alpha}^{\beta} r_1^2\, d\theta = \frac{1}{2}\int_{\alpha}^{\beta} \left(r_2^2 - r_1^2\right) d\theta. \tag{13.3}$$

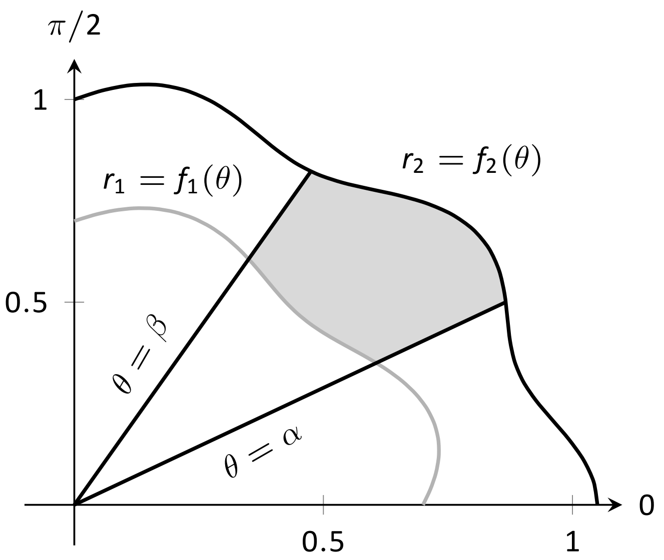


Figure 13.6: Illustrating area bound between two polar curves.

Example 13.4

Find the area bounded between the polar curves $r = 1$ and $r = 2\cos(2\theta)$, as shown in Figure 13.7(a).

Solution

We need to find the point of intersection between the two curves. Setting the two functions equal to each other, we have

$$2\cos(2\theta) = 1 \quad \Leftrightarrow \quad \cos(2\theta) = \frac{1}{2} \quad \Rightarrow \quad 2\theta = \frac{\pi}{3} \quad \Rightarrow \quad \theta = \frac{\pi}{6}.$$

In Figure 13.7(b), we zoom in on the region and note that it is not really bounded between two polar curves, but rather by two polar curves, along with $\theta = 0$. The dashed line breaks the region into its component parts. Below the dashed line, the region is defined by $r = 1$, $\theta = 0$ and $\theta = \pi/6$. Above the dashed line the region is bounded by $r = 2\cos(2\theta)$ and $\theta = \pi/6$. Since we have two separate regions, we find the area using two separate integrals.

Call the area below the dashed line A_1 and the area above the dashed line A_2. They are determined by the following integrals:

$$A_1 = \frac{1}{2}\int_0^{\pi/6} (1)^2\, d\theta \qquad \text{and} \qquad A_2 = \frac{1}{2}\int_{\pi/6}^{\pi/4} \big(2\cos(2\theta)\big)^2\, d\theta.$$

The upper bound of the integral for A_2 is $\pi/4$ as $r = 2\cos(2\theta)$ is at the pole when $\theta = \pi/4$. We omit the integration details and let the reader verify that $A_1 = \pi/12$ and $A_2 = \pi/12 - \sqrt{3}/8$; the total area is $A = \pi/6 - \sqrt{3}/8$.

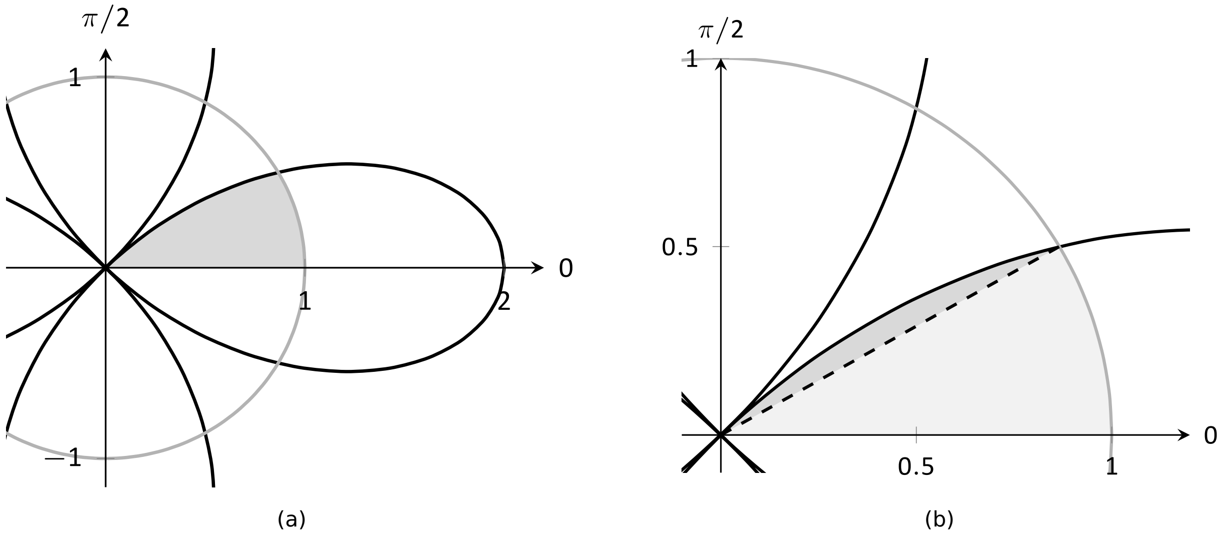


Figure 13.7: Graphing the region bounded by the functions in Example 13.4.

The error function

The error function is an example of a non-elementary function that contains an integral in its definition. More precisely, it is defined as:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}}\int_0^{x} e^{-t^2}\, dt.$$

It is of sigmoid shape and occurs in probability and statistic (Figure 13.8). There, for nonnegative values of x, it has the following interpretation: for a random variable Y that is normally distributed with mean 0 and variance 1/2, $\operatorname{erf}(x)$ describes the probability of Y falling in the range $[-x, x]$.

The error function

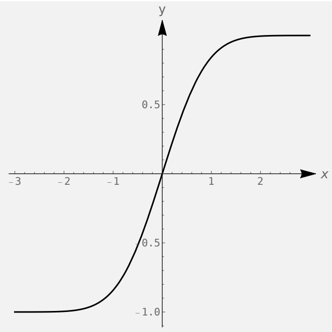


Figure 13.8: A graph of the error function.

13.2 Volume by cross-sectional area

13.2.1 Volumes by slicing

The volume of a general right cylinder, as shown in Figure 13.9, is

$$\text{Area of the base} \times \text{height}.$$

We can use this fact as the building block in finding volumes of a variety of shapes.

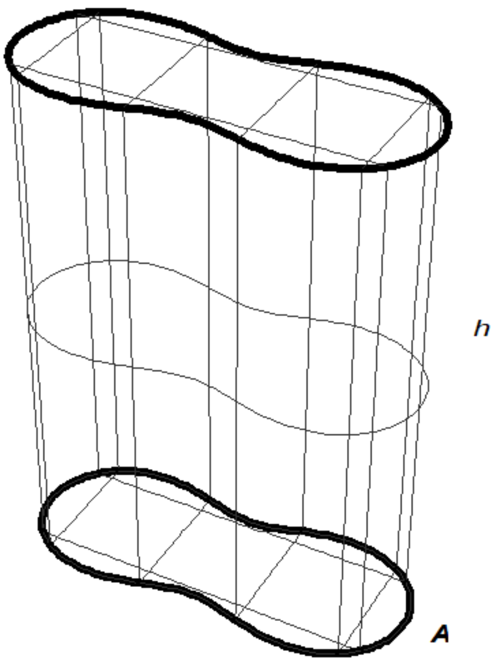


Figure 13.9: The volume of a general right cylinder.

Given an arbitrary solid, we can approximate its volume by cutting it into n thin slices. When the slices are thin, each slice can be approximated well by a general right cylinder. Thus the volume of each slice is approximately its cross-sectional area × thickness. These slices are the differential elements.

By orienting a solid along the x-axis, we can let $A(x_i)$ represent the cross-sectional area of the i^{th} slice, and let Δx_i represent the thickness of this slice. The total volume of the solid is approximately:

$$\begin{aligned}\text{Volume} &\approx \sum_{i=1}^{n}\Big[\text{Area} \ \times \ \text{thickness}\Big]\\ &= \sum_{i=1}^{n} A(x_i)\Delta x_i.\end{aligned}$$

Recognize that this is a Riemann sum. By taking a limit as the thickness of the slices goes to 0 we can find the volume exactly.

Theorem 13.1 (Volume by cross-sectional area)
The volume V of a solid, oriented along the x-axis with cross-sectional area $A(x)$ from $x = a$ to $x = b$, is

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

Example 13.5

Find the volume of a pyramid with a square base of side length 10 cm and a height of 5 cm.

Solution

There are many ways to orient the pyramid along the x-axis; Figure 13.10(a) gives one such way, with the pointed top of the pyramid at the origin and the x-axis going through the centre of the base.

Each cross section of the pyramid is a square; this is a sample differential element. To determine its area $A(x)$, we need to determine the side lengths of the square. When $x = 5$, the square has side length 10; when $x = 0$, the square has side length 0. Since the edges of the pyramid are lines, it is easy to figure that each cross-sectional square has side length $2x$, giving $A(x) = (2x)^2 = 4x^2$.

If one were to cut a slice out of the pyramid at $x = 3$, as shown in Figure 13.10(b), one would have a shape with square bottom and top with sloped sides. If the slice were thin, both the bottom and top squares would have sides lengths of about 6, and thus the cross–sectional area of the bottom and top would be about 36cm^2. Letting Δx_i represent the thickness of the slice, the volume of this slice would then be about $36\Delta x_i\ \text{cm}^3$.

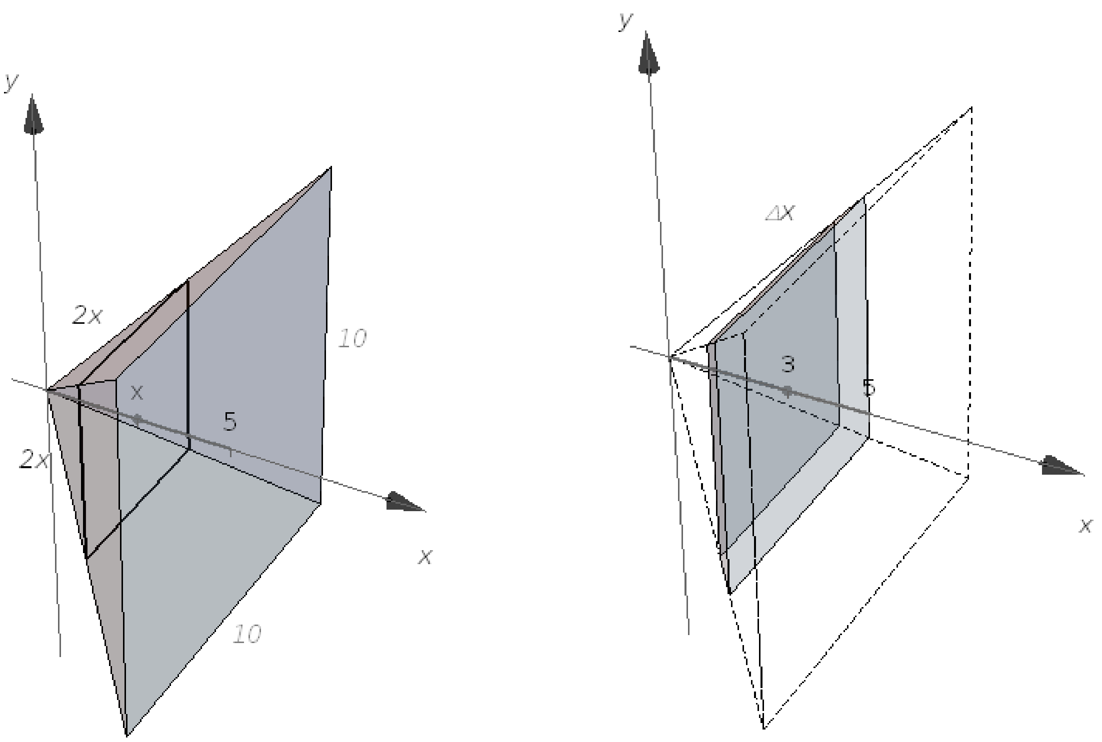
Cutting the pyramid into n slices divides the total volume into n equally–spaced smaller pieces, each with volume $(2x_i)^2\Delta x$, where x_i is the approximate location of the slice along the x-axis and Δx represents the thickness of each slice. One can approximate the total volume of the pyramid by summing up the volumes of these slices:

$$\text{Approximate volume } = \sum_{i=1}^{n} (2x_i)^2 \Delta x.$$

Taking the limit as $n \to +\infty$ gives the actual volume of the pyramid; recognizing this sum as a Riemann sum allows us to find the exact answer using a definite integral, matching the definite integral given by Theorem 13.1.

We have

$$\begin{aligned} V &= \lim_{n\to+\infty} \sum_{i=1}^{n} (2x_i)^2 \Delta x \\ &= \int_0^5 4x^2\, dx \\ &= \frac{4}{3}x^3 \Big|_0^5 \\ &= \frac{500}{3}\ \text{in}^3 \approx 166.67\ \text{cm}^3. \end{aligned}$$



(a) (b)

Figure 13.10: Orienting a pyramid along the x-axis (a) and cutting a slice in it at $x = 3$ (b) in Example 13.5.

13.2.2 Solids of revolution

An important special case of Theorem 13.1 is when the solid is a **solid of revolution** (*omwentelingslichaam*), that is, when the solid is formed by rotating a shape about an axis.

Start with a function $y = f(x)$ from $x = a$ to $x = b$. Revolving this curve about a horizontal axis creates a three-dimensional solid whose cross sections are disks (thin circles). Let $R(x)$ represent the radius of the cross-sectional disk at x; the area of this disk is $\pi R(x)^2$. Applying Theorem 13.1 gives the disk method.

More precisely, let a solid be formed by revolving the curve $y = f(x)$ from $x = a$ to $x = b$ about a horizontal axis, and let $R(x)$ be the radius of the cross-sectional disk at x. The volume of the resulting solid is

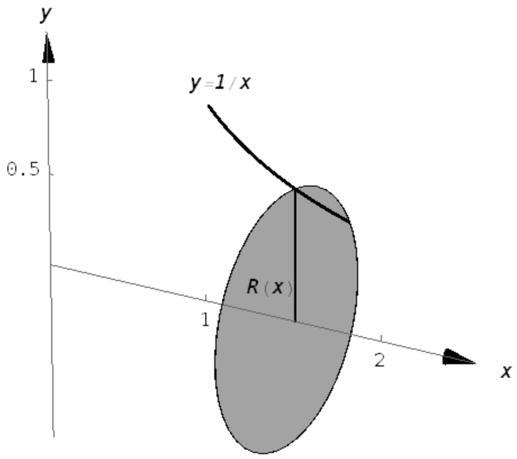
$$V = \pi \int_a^b R(x)^2 \, dx. \tag{13.4}$$

Example 13.6

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, about the x-axis.

Solution

A sketch can help us understand this problem. In Figure 13.11(a) the curve $y = 1/x$ is sketched along with the differential element – a disk – at x with radius $R(x) = 1/x$. In Figure 13.11(b) the whole solid is pictured, along with the differential element.



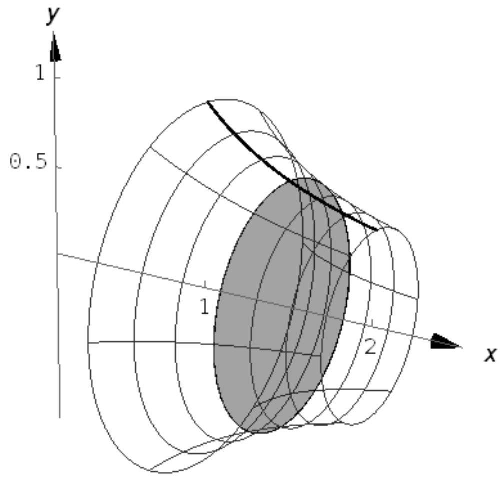


Figure 13.11: Sketching a solid in Example 13.6.

The volume of the differential element shown in part (a) of the figure is approximately $\pi R(x_i)^2\Delta x$, where $R(x_i)$ is the radius of the disk shown and Δx is the thickness of that slice. The radius $R(x_i)$ is the distance from the x-axis to the curve, hence $R(x_i) = 1/x_i$.

Slicing the solid into n equally–spaced slices, we can approximate the total volume by adding up the approximate volume of each slice:

$$\text{Approximate volume } = \sum_{i=1}^{n} \pi\left(\frac{1}{x_i}\right)^2 \Delta x.$$

Taking the limit of the above sum as $n \to +\infty$ gives the actual volume; recognizing this sum as a Riemann sum allows us to evaluate the limit with a definite integral, which matches Equation (13.4):

$$\begin{aligned} V &= \lim_{n\to+\infty} \sum_{i=1}^{n} \pi\left(\frac{1}{x_i}\right)^2 \Delta x \\ &= \pi\int_1^2 \left(\frac{1}{x}\right)^2\, dx \\ &= \pi\int_1^2 \frac{1}{x^2}\, dx \\ &= \pi\left[-\frac{1}{x}\right]\Bigg|_1^2 \\ &= \pi\left[-\frac{1}{2}-(-1)\right] \end{aligned}$$

$$= \frac{\pi}{2} \text{ units}^3.$$

While Equation (13.4) is given in terms of functions of x, the principle involved can be as well applied to functions of y when the axis of rotation is vertical, not horizontal. We demonstrate this in the next example.

Example 13.7

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, about the y-axis.

Solution

Since the axis of rotation is vertical, we need to convert the function into a function of y and convert the x-bounds to y-bounds. Since $y = 1/x$ defines the curve, we rewrite it as $x = 1/y$. The bound $x = 1$ corresponds to the y-bound $y = 1$, and the bound $x = 2$ corresponds to the y-bound $y = 1/2$. Thus we are rotating the curve $x = 1/y$, from $y = 1/2$ to $y = 1$ about the y-axis to form a solid. The curve and sample differential element are sketched in Figure 13.12(a), with a full sketch of the solid in Figure 13.12(b).

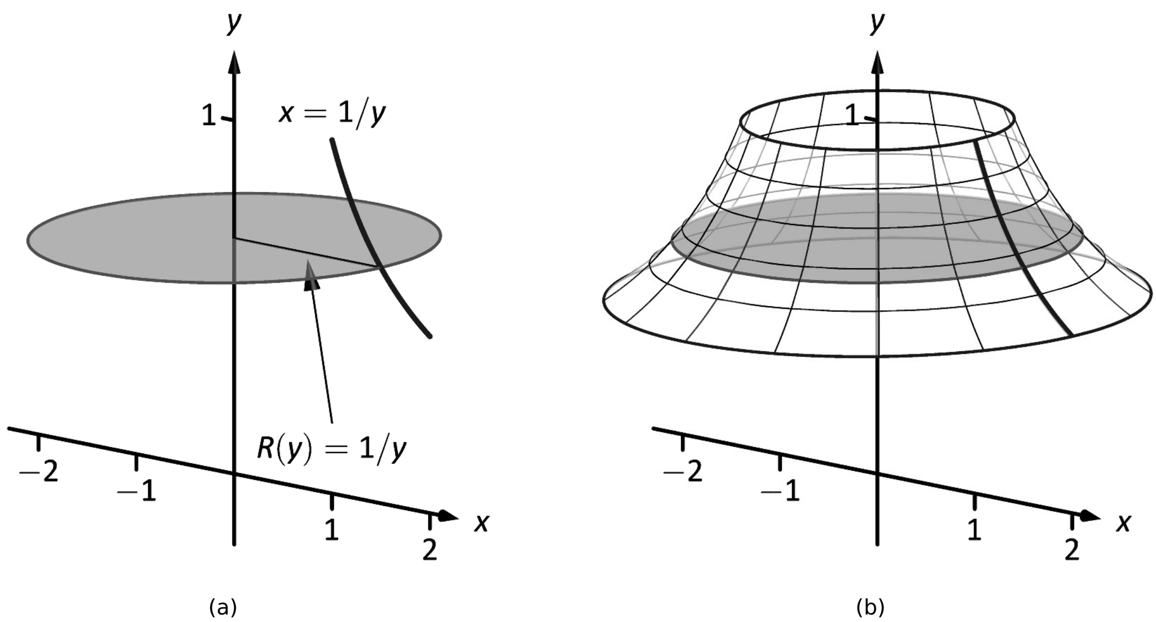


Figure 13.12: Sketching a solid in Example 13.7.

We integrate to find the volume:

$$V = \pi \int_{1/2}^{1} \frac{1}{y^2}\, dy$$
$$= -\frac{\pi}{y}\Big|_{1/2}^{1}$$
$$= \pi \text{ units}^3.$$

We can also compute the volume of solids of revolution that have a hole in the centre. The general principle is simple: compute the volume of the solid irrespective of the hole, then subtract the volume of the hole. If the outside radius of the solid is $R(x)$ and the inside radius (defining the hole) is $r(x)$, then the volume is

$$V = \pi \int_a^b R(x)^2 \, dx - \pi \int_a^b r(x)^2 \, dx = \pi \int_a^b \left(R(x)^2 - r(x)^2 \right) \, dx.$$

One can generate a solid of revolution with a hole in the middle by revolving a region about an axis. Consider Figure 13.13(a), where a region is sketched along with a dashed, horizontal axis of rotation. By rotating the region about the axis, a solid is formed as sketched in Figure 13.13(b). The outside of the solid has radius $R(x)$, whereas the inside has radius $r(x)$. Each cross section of this solid will be a washer (a disk with a hole in the center) as sketched in Figure 13.13(c). This leads us to the washer method.

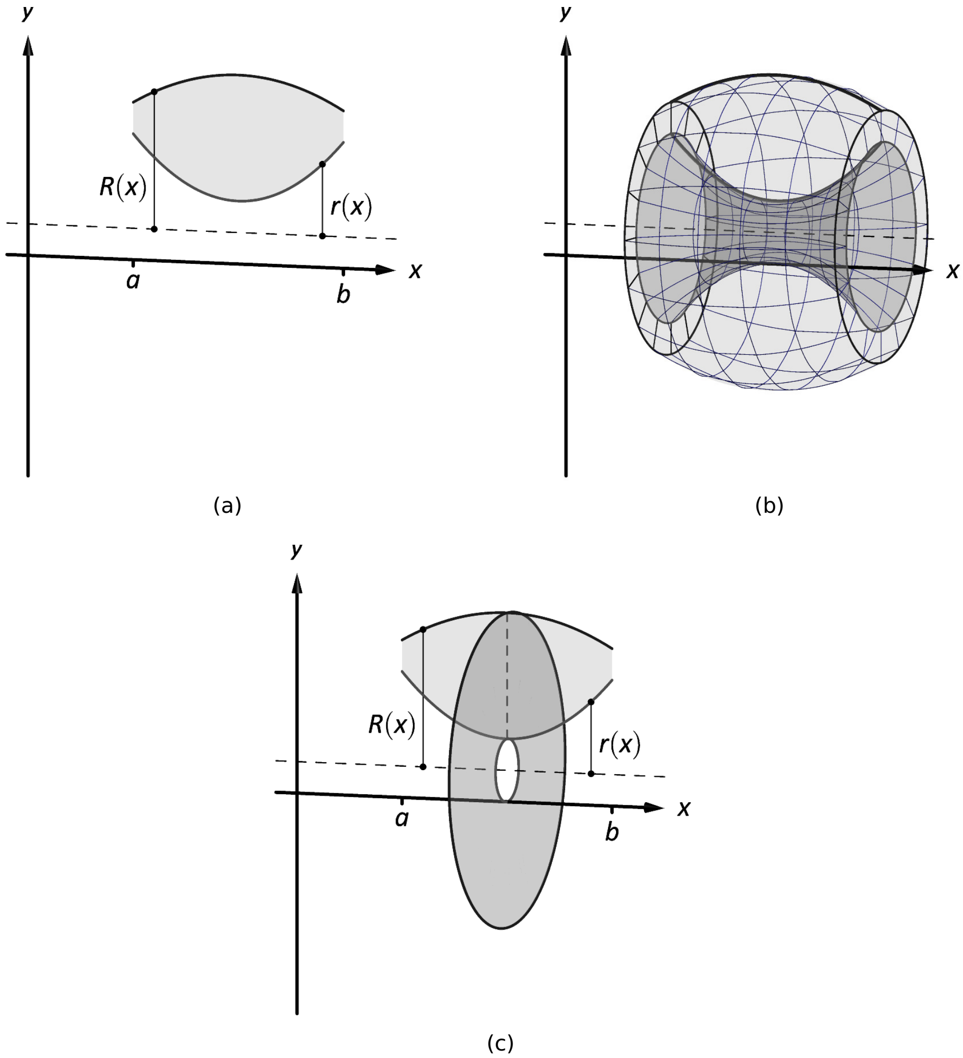


Figure 13.13: Establishing the washer method.

Let a region bounded by $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$ be rotated about a horizontal axis that does not intersect the region, forming a solid. Each cross section at x will be a washer with outside radius $R(x)$ and inside radius $r(x)$. The volume of the solid is

$$V = \pi \int_a^b \left(R(x)^2 - r(x)^2 \right) \, dx. \qquad (13.5)$$

Obviously, the disk method is just a special case of the washer method with an inside radius of $r(x) = 0$.

Example 13.8

Find the volume of the solid formed by rotating the region bounded by $y = x^2 - 2x + 2$ and $y = 2x - 1$ about the x-axis.

Solution

A sketch of the region will help, as given in Figure 13.10(a). Rotating about the x-axis will produce cross sections in the shape of washers, as shown in Figure 13.14(b); the complete solid is shown in Figure 13.14(c). The outside radius of this washer is $R(x) = 2x - 1$; the inside radius is $r(x) = x^2 - 2x + 2$. As the region is bounded from $x = 1$ to $x = 3$, we integrate as follows to compute the volume.

$$\begin{aligned} V &= \pi \int_1^3 \Big((2x-1)^2 - (x^2 - 2x + 2)^2 \Big)\, dx \\ &= \pi \int_1^3 \left(-x^4 + 4x^3 - 4x^2 + 4x - 3 \right) dx \\ &= \pi \Big[-\frac{1}{5}x^5 + x^4 - \frac{4}{3}x^3 + 2x^2 - 3x \Big]\Big|_1^3 \\ &= \frac{104}{15}\pi \text{ units}^3 \approx 21.78 \text{ units}^3. \end{aligned}$$

When rotating about a vertical axis, the outside and inside radius functions must be functions of y.

Example 13.9

Find the volume of the solid formed by rotating the triangular region with vertices at $(1,1)$, $(2,1)$ and $(2,3)$ about the y-axis.

Solution

The triangular region is sketched in Figure 13.15(a); the differential element is sketched in Figure 13.15(b) and the full solid is drawn in Figure 13.15(c). They help us establish the outside and inside radii. Since the axis of rotation is vertical, each radius is a function of y.

The outside radius $R(y)$ is formed by the line connecting $(2,1)$ and $(2,3)$; it is a constant function, because $R(y) = 2$. The inside radius is formed by the line connecting $(1,1)$ and $(2,3)$. The equation of this line is $y = 2x - 1$, but we need to refer to it as a function of y. Solving for x gives $r(y) = \frac{1}{2}(y+1)$.

We integrate over the y-bounds of $y = 1$ to $y = 3$. Thus the volume is

$$V = \pi \int_1^3 \left(2^2 - \left(\frac{1}{2}(y+1) \right)^2 \right) dy$$

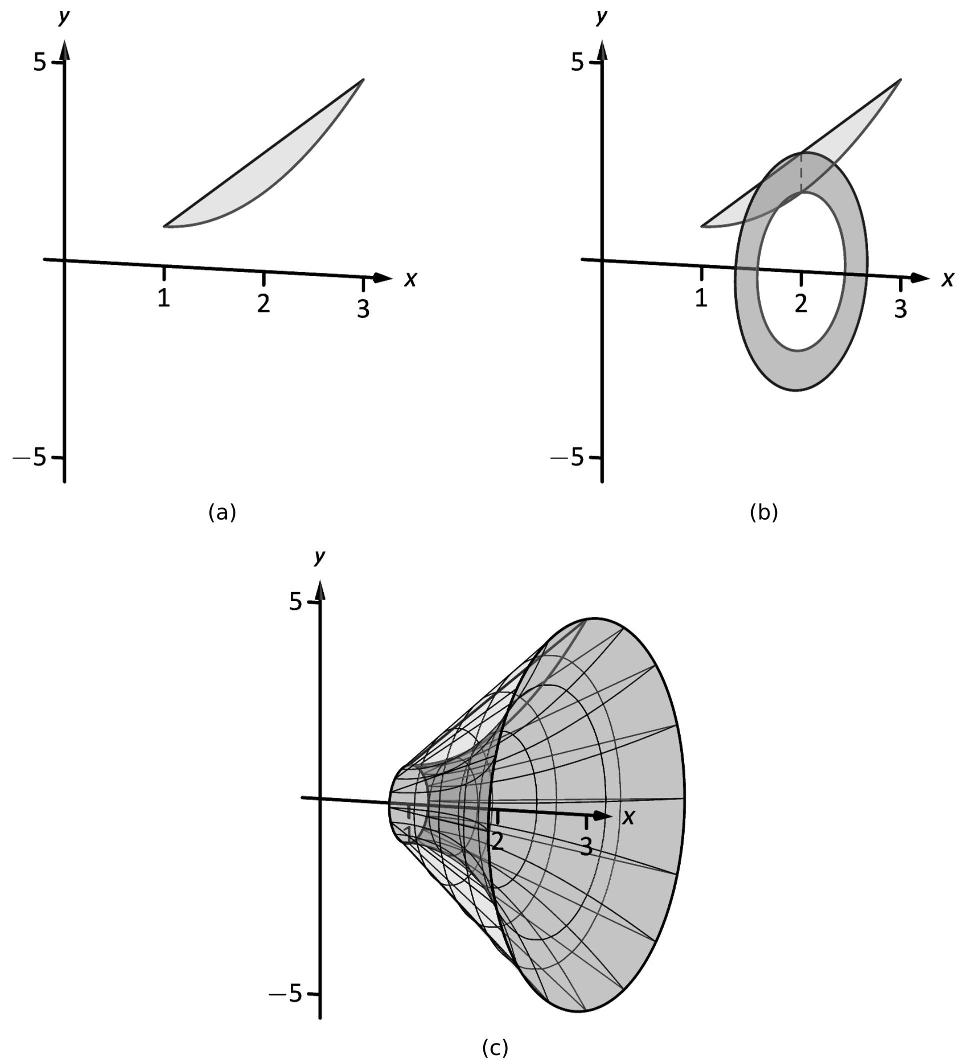


Figure 13.14: Sketching the differential element and solid in Example 13.8.

$$
\begin{aligned}
&= \pi \int_1^3 \left(-\frac{1}{4}y^2 - \frac{1}{2}y + \frac{15}{4} \right) dy \\
&= \pi \Big[-\frac{1}{12}y^3 - \frac{1}{4}y^2 + \frac{15}{4}y \Big]\Big|_1^3 \\
&= \frac{10}{3}\pi \text{units}^3 \approx 10.47 \text{ units}^3.
\end{aligned}
$$

The ultimate goal of this section is not to compute volumes of solids. That can be useful, but what is more useful is the understanding of this basic principle of integral calculus: to find the exact value of some quantity,

- we start with an approximation (in this section, slice the solid and approximate the volume of each slice),
- then make the approximation better by refining our original approximation (i.e., use more slices),
- then use limits to establish a definite integral which gives the exact value.

We practice this principle in the next section where we find volumes by slicing solids in a different way.

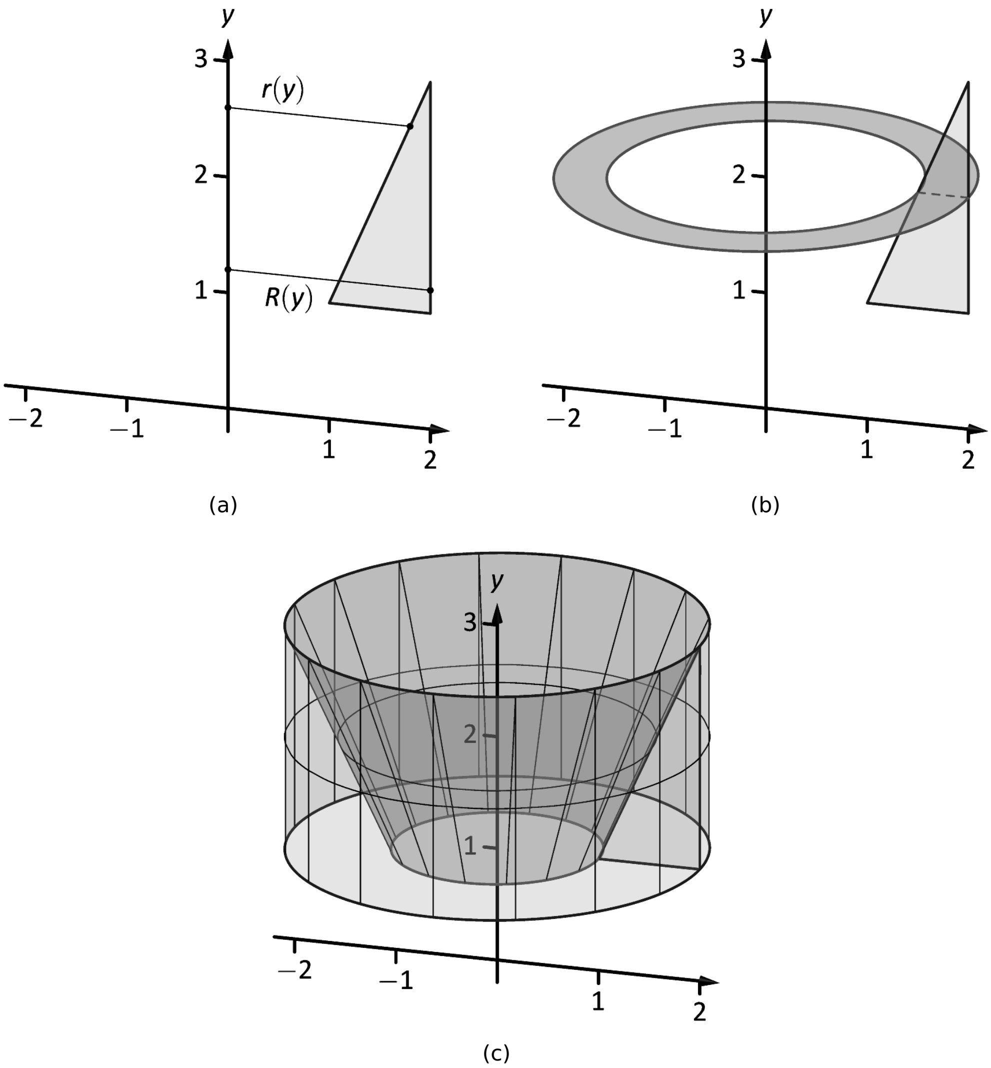


Figure 13.15: Sketching the differential element and solid in Example 13.9.

13.3 The shell method

This section develops another method of computing volume, the **shell method.** Instead of slicing the solid perpendicular to the axis of rotation creating cross-sections, we now slice it parallel to the axis of rotation, creating shells.

Consider Figure 13.16(a), where the region is rotated about the y-axis forming the solid shown in Figure 13.16(b). A small slice of the region is drawn in Figure 13.16(a), parallel to the axis of rotation. When the region is rotated, this thin slice forms a cylindrical shell, as pictured in Figure 13.16(c). The previous section approximated a solid with lots of thin disks (or washers); we now approximate a solid with many thin cylindrical shells.

To compute the volume of one shell, first consider the paper label on a soup can with radius r and height h. What is the area of this label? A simple way of determining this is to cut the label and lay it out flat, forming a rectangle with height h and length $2\pi r$. Thus the area is $A = 2\pi rh$; see Figure 13.17(a). Do a similar process with a cylindrical shell, with height h, thickness Δx, and approximate radius r. Cutting the shell and laying it flat forms a rectangular solid with length $2\pi r$, height h and depth Δx. Thus the volume is $V \approx 2\pi rh\Delta x$; see Figure 13.17(b). We say approximately since our radius was an approximation.

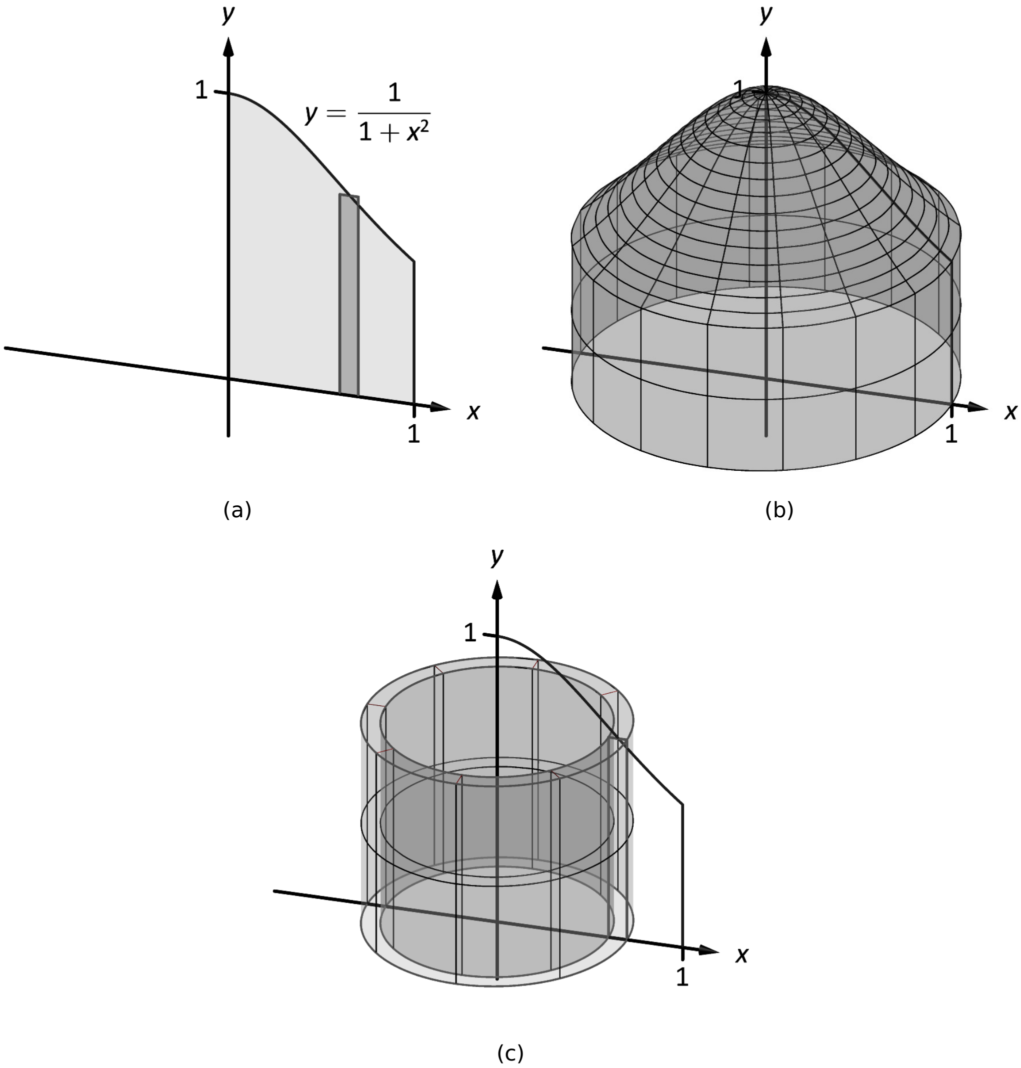


Figure 13.16: The shell method.

By breaking the solid into n cylindrical shells, we can approximate the volume of the solid as

$$V \approx \sum_{i=1}^{n} 2\pi r_i h_i \Delta x_i,$$

where r_i, h_i and Δx_i are the radius, height and thickness of the i^{th} shell, respectively. This is a Riemann sum. Taking a limit as the thickness of the shells approaches 0 leads to a definite integral. So we arrive at the following.

Let a solid be formed by revolving a region R, bounded by $x = a$ and $x = b$, about a vertical axis. Let $r(x)$ represent the distance from the axis of rotation to x (i.e., the radius of a sample shell) and let $h(x)$ represent the height of the solid at x (i.e., the height of the shell). The volume of the solid is

$$V = 2\pi \int_a^b r(x)h(x)\ dx. \tag{13.6}$$

There are two special cases:

1. When the region R is bounded above by $y = f(x)$ and below by $y = g(x)$, then $h(x) = f(x) - g(x)$.
2. When the axis of rotation is the y-axis (i.e., $x = 0$) then $r(x) = x$.

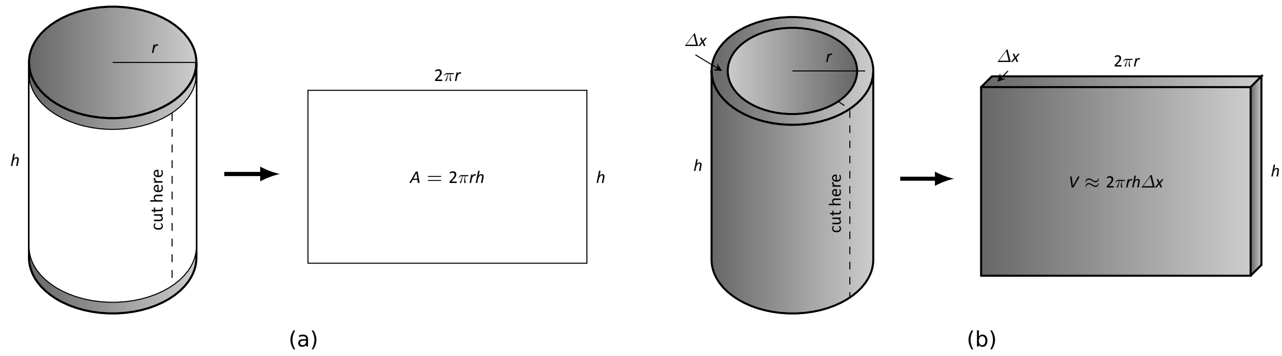


Figure 13.17: Determining the volume of a thin cylindrical shell.

Let us practice using this method.

Example 13.10

Find the volume of the solid formed by rotating the region bounded by $y = 0$, $y = 1/(1+x^2)$, $x = 0$ and $x = 1$ about the y-axis.

Solution

This is the region used to introduce the shell method in Figure 13.16(a), but is sketched again in Figure 13.18 for closer reference. A line is drawn in the region parallel to the axis of rotation representing a shell that will be carved out as the region is rotated about the y-axis. This is the differential element.

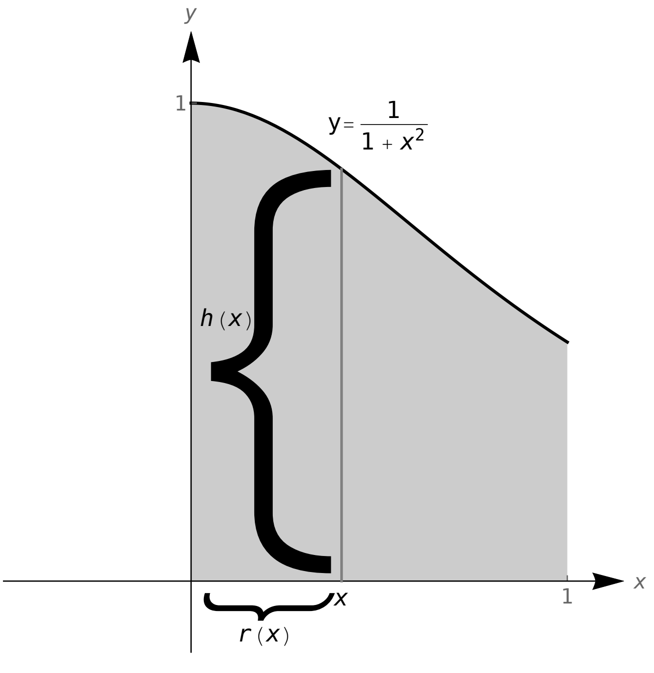


Figure 13.18: Graphing a region in Example 13.10.

The distance this line is from the axis of rotation determines $r(x)$; as the distance from x to the y-axis is x, we have $r(x) = x$. The height of this line determines $h(x)$; the top of the line is at $y = 1/(1+x^2)$, whereas the bottom of the line is at $y = 0$. Thus $h(x) = 1/(1+x^2) - 0 = 1/(1+x^2)$.

The region is bounded from $x = 0$ to $x = 1$, so the volume is

$$V = 2\pi \int_0^1 \frac{x}{1+x^2}\, dx.$$

This requires substitution. Let $u = 1+x^2$, so $du = 2x\ dx$. We also change the bounds: $u(0) = 1$ and $u(1) = 2$. Thus we have:

$$\begin{aligned} &= \pi \int_1^2 \frac{1}{u}\, du \\ &= \pi \ln(u)\Big|_1^2 \\ &= \pi \ln(2) \approx 2.178 \text{ units}^3. \end{aligned}$$

Note that in order to find this volume using the disk method, two integrals would be needed to account for the regions above and below $y = 1/2$.

With the shell method, nothing special needs to be accounted for to compute the volume of a solid that has a hole in the middle, as demonstrated next.

Example 13.11

Find the volume of the solid formed by rotating the triangular region determined by the points $(0,1)$, $(1,1)$ and $(1,3)$ about the line $x = 3$.

Solution

The region is sketched in Figure 13.19(a) along with the differential element, a line within the region parallel to the axis of rotation. In Figure 13.19(b), we see the shell traced out by the differential element, and in Figure 13.19(c) the whole solid is shown.

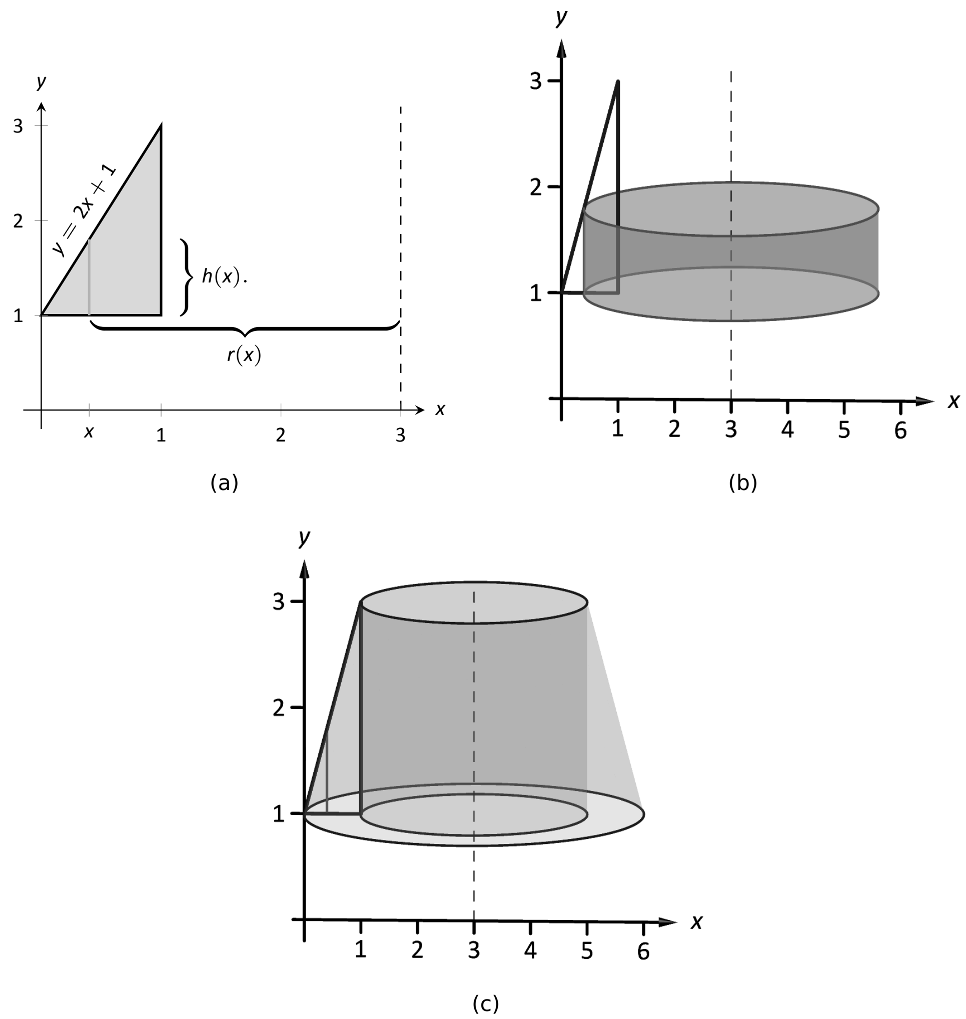


Figure 13.19: Graphing a region in Example 13.11.

The height of the differential element is the distance from $y = 1$ to $y = 2x+1$, the line that connects the points $(0,1)$ and $(1,3)$. Thus $h(x) = 2x+1-1 = 2x$. The radius of the shell formed by the differential element is the distance from x to $x = 3$; that is, it is $r(x) = 3-x$. The x-bounds of the region are $x = 0$ to $x = 1$, giving

$$\begin{aligned} V &= 2\pi \int_0^1 (3-x)(2x)\, dx \\ &= 2\pi \int_0^1 \left(6x-2x^2\right) dx \\ &= 2\pi\left[3x^2 - \frac{2}{3}x^3\right]\Big|_0^1 \\ &= \frac{14}{3}\pi \text{ units}^3 \approx 14.66 \text{ units}^3. \end{aligned}$$

When revolving a region about a horizontal axis, we must consider the radius and height functions in terms of y, not x.

Example 13.12

Find the volume of the solid formed by rotating the region given in Example 13.11 about the x-axis.

Solution

The region is sketched in Figure 13.20(a) with a sample differential element. In Figure 13.20(b) the shell formed by the differential element is drawn, and the solid is sketched in Figure 13.20(c).

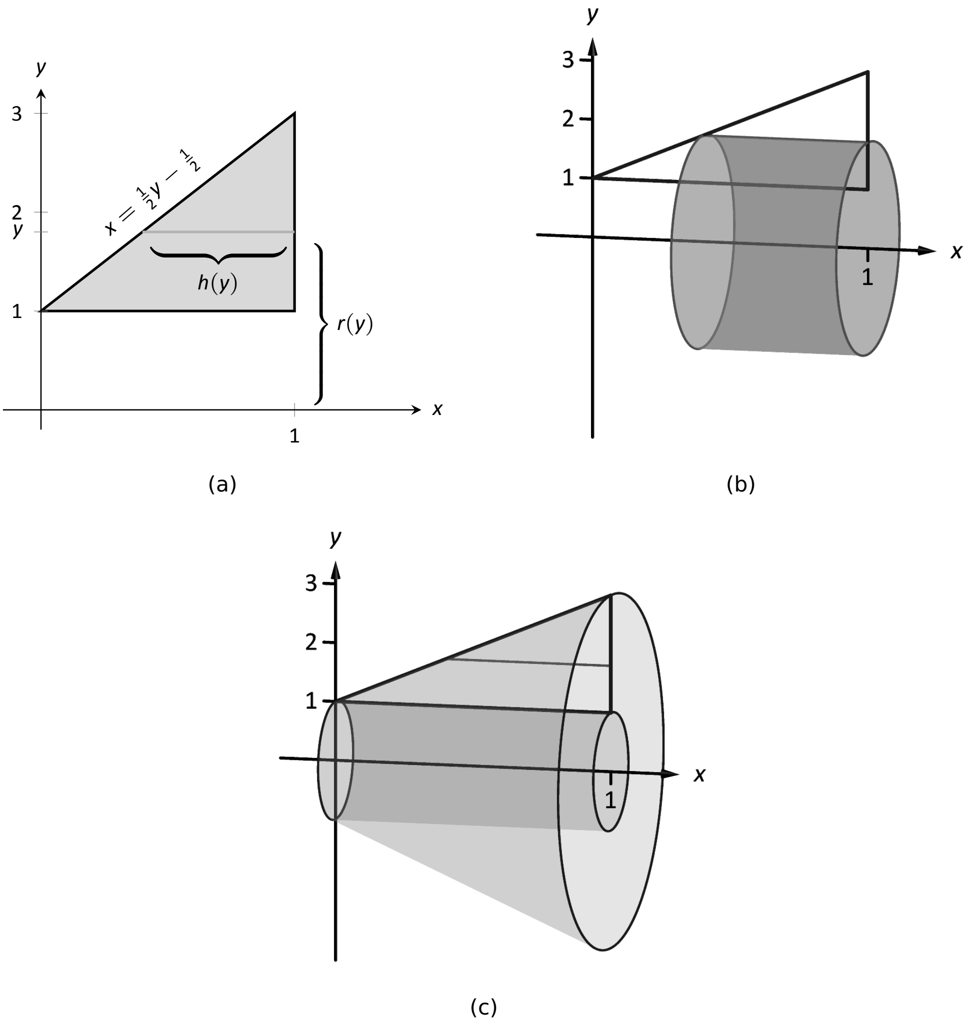


Figure 13.20: Graphing a region in Example 13.12.

The height of the differential element is an x-distance, between $x = y/2 - 1/2$ and $x = 1$. Thus

$$h(y) = 1 - \left(\frac{1}{2}y - \frac{1}{2}\right) = -\frac{1}{2}y + \frac{3}{2}.$$

The radius is the distance from y to the x-axis, so $r(y) = y$. The y bounds of the region are $y = 1$ and $y = 3$, leading to the integral

$$V = 2\pi \int_1^3 \left[y\left(-\frac{1}{2}y + \frac{3}{2}\right)\right]\, dy$$

$$= 2\pi \int_1^3 \left[-\frac{1}{2}y^2 + \frac{3}{2}y \right] dy$$

$$= 2\pi \left[-\frac{1}{6}y^3 + \frac{3}{4}y^2 \right] \Bigg|_1^3$$

$$= 2\pi \left[\frac{9}{4} - \frac{7}{12} \right]$$

$$= \frac{10}{3}\pi \text{ units}^3 \approx 10.472 \text{ units}^3.$$

We end this section with a table summarizing the usage of the washer and shell Methods.

	Washer method	Shell method
Horizontal axis	$\pi \int_a^b \left(R(x)^2 - r(x)^2 \right) dx$	$2\pi \int_c^d r(y)h(y)\, dy$
Vertical axis	$\pi \int_c^d \left(R(y)^2 - r(y)^2 \right) dy$	$2\pi \int_a^b r(x)h(x)\, dx$

In this section, we approximate the volume of a solid by cutting it into thin cylindrical shells. By summing up the volumes of each shell, we get an approximation of the volume. By taking a limit as the number of equally spaced shells goes to infinity, our summation can be evaluated as a definite integral, giving the exact value.

We use this same principle again in the next section, where we find the length of curves in the plane.

13.4 Arc length

13.4.1 Rectangular coordinates

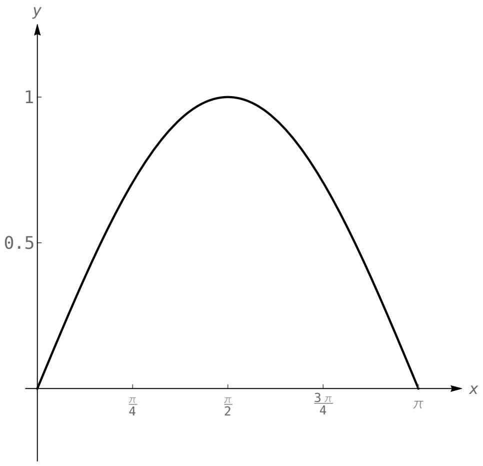
In this section, we address the question: Given a curve, what is its length? This is often referred to as **arc length** (*booglengte*).

Consider the graph of $y = \sin(x)$ on $[0, \pi]$ given in Figure 13.21(a). How long is this curve? That is, if we were to use a piece of string to exactly match the shape of this curve, how long would the string be?

As we have done in the past, we start by approximating; later, we will refine our answer using limits to get an exact solution.

The length of straight–line segments is easy to compute using the distance formula. We can approximate the length of the given curve by approximating the curve with straight lines and measuring their lengths.

In Figure 13.21(b), the curve $y = \sin(x)$ has been approximated with 4 line segments, i.e. the interval $[0, \pi]$ has been divided into 4 equally–lengthed subintervals. It is clear that these four line segments approximate $y = \sin(x)$ very well on the first and last subinterval, though not so well in the middle.



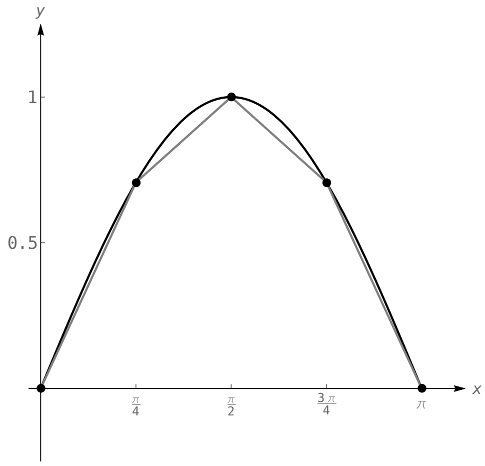


Figure 13.21: Graphing $y = \sin(x)$ on $[0, \pi]$ (a) and approximating the curve with line segments (b).

Regardless, the sum of the lengths of the line segments is 3.79, so we approximate the arc length of $y = \sin(x)$ on $[0, \pi]$ to be 3.79.

In general, we can approximate the arc length of $y = f(x)$ on $[a, b]$ in the following manner. Let $a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b$ be a partition of $[a, b]$ into n subintervals. Let Δx_i represent the length of the i^{th} subinterval $[x_i, x_{i+1}]$. Figure 13.22 zooms in on the i^{th} subinterval where $y = f(x)$ is approximated by a straight line segment. The dashed lines show that we can view this line segment as the hypotenuse of a right triangle whose sides have length Δx_i and Δy_i. Using the Pythagorean theorem, the length of this line segment is $\sqrt{\Delta x_i^2 + \Delta y_i^2}$. Summing over all subintervals gives an arc length approximation

$$L \approx \sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2}.$$

As shown here, this is not a Riemann sum. While we could conclude that taking a limit as the subinterval length goes to zero gives the exact arc length, we would not be able to compute the answer with a definite integral. We need first to do a little algebra.

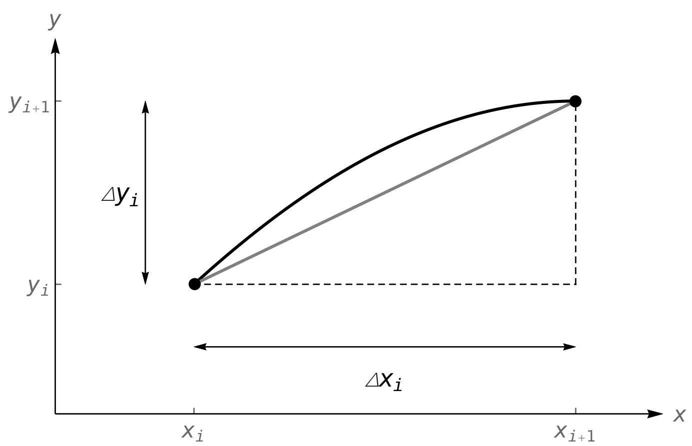


Figure 13.22: Zooming in on the i^{th} subinterval $[x_i, x_{i+1}]$ of a partition of $[a, b]$.

In the above expression factor out a Δx_i^2 term:

$$\sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{i=1}^{n} \sqrt{\Delta x_i^2 \left(1 + \frac{\Delta y_i^2}{\Delta x_i^2}\right)}.$$

Now pull the Δx_i^2 term out of the square root:

$$\sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{i=1}^{n} \sqrt{1 + \frac{\Delta y_i^2}{\Delta x_i^2}}\ \Delta x_i.$$

This is nearly a Riemann sum. Consider the $\Delta y_i^2/\Delta x_i^2$ term. The expression $\Delta y_i/\Delta x_i$ measures the (change in y)/(change in x) of f on the i^{th} subinterval. The mean value theorem of differentiation (Theorem 10.4) states that there is a c_i in the i^{th} subinterval where $f'(c_i) = \Delta y_i/\Delta x_i$. Thus we can rewrite our above expression as:

$$\sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{i=1}^{n} \sqrt{1 + f'(c_i)^2}\ \Delta x_i.$$

This is a Riemann sum. As long as f' is continuous, we can invoke Theorem 12.4 and conclude

$$L = \int_a^b \sqrt{1 + f'(x)^2}\ dx.$$

This result is summarized in the following theorem.

Theorem 13.2 (Arc length)
Let f be differentiable on $[a, b]$, where f' is also continuous on $[a, b]$. Then the arc length of f from $x = a$ to $x = b$ is

$$L = \int_a^b \sqrt{1 + f'(x)^2}\ dx. \tag{13.7}$$

The theorem also requires that f' is continuous on $[a, b]$; while examples are arcane, it is possible for f to be differentiable yet f' is not continuous.

As the integrand contains a square root, it is often difficult to use Equation (13.7) to find the length exactly. When exact answers are difficult to come by, we resort to using numerical methods.

Example 13.13

Find the arc length of $f(x) = x^{3/2}$ from $x = 0$ to $x = 4$.

Solution

We find $f'(x) = 3x^{1/2}/2$; note that on $[0, 4]$, f is differentiable and f' is also continuous. Using Equation (13.7), we find the arc length L as

$$L = \int_0^4 \sqrt{1 + \left(\frac{3}{2}x^{1/2}\right)^2}\ dx$$

$$= \int_0^4 \sqrt{1 + \frac{9}{4}x}\ dx$$

$$= \frac{2}{3} \cdot \frac{4}{9} \cdot \left(1 + \frac{9}{4}x\right)^{3/2} \Big|_0^4$$

$$= \frac{8}{27}\left(10^{3/2} - 1\right) \approx 9.07\text{units.}$$

A graph of f is given in Figure 13.23.

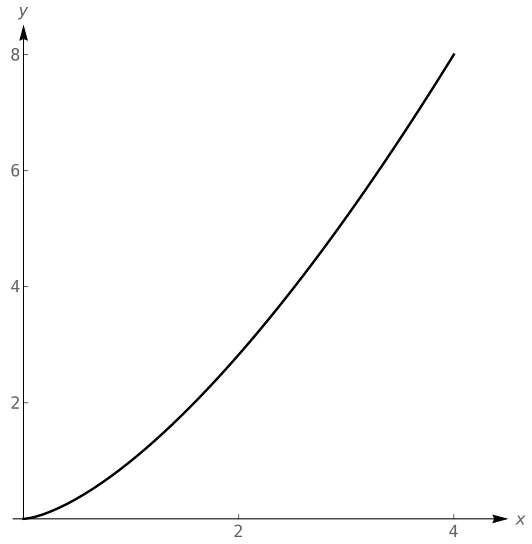


Figure 13.23: A graph of $f(x) = x^{3/2}$ from Example 13.13.

We conclude with one example where it is not possible to find an exact answer.

Example 13.14

Find the length of the sine curve from $x = 0$ to $x = \pi$.

Solution

The setup is straightforward: $f(x) = \sin(x)$ and $f'(x) = \cos(x)$. Thus

$$L = \int_0^\pi \sqrt{1 + \cos^2(x)}\ dx.$$

This integral cannot be evaluated in terms of elementary functions so we have to approximate it with one of the methods studied in Chapter 12. Doing this leads us to $L \approx 3.8202$.

13.4.2 Parametric and polar equations

When we are faced with a curve described by parametric equations, we can convert Equation (13.7) to such a context. Letting $x = f(t)$ and $y = g(t)$, we know that $dy/dx = g'(t)/f'(t)$. It will also be useful to calculate the differential of x:

$$dx = f'(t)dt \qquad \Rightarrow \qquad dt = \frac{1}{f'(t)}\,dx.$$

Starting with the arc length formula given by equation (13.7), consider:

$$\begin{aligned}
L &= \int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx \\
&= \int_a^b \sqrt{1+\frac{g'(t)^2}{f'(t)^2}}\,dx. \\
&= \int_a^b \sqrt{f'(t)^2+g'(t)^2}\,\underbrace{\frac{1}{f'(t)}\,dx}_{=dt} \\
&= \int_{t_1}^{t_2} \sqrt{f'(t)^2+g'(t)^2}\,dt.
\end{aligned} \tag{13.8}$$

Note the new bounds. They are found by solving $a = f(t)$ and $b = f(t)$ for t, and subsequently choosing t_1 such that $t_1 = \min(t_a, t_b)$, where t_a and t_b are the solutions of $a = f(t)$ and $b = f(t)$, respectively. Likewise, t_2 should be chosen such that $t_2 = \max(t_a, t_b)$.

Example 13.15

Find the arc length of the circle parametrized by $x = 3\cos(t)$, $y = 3\sin(t)$ on $[0, 3\pi/2]$.

Solution

By direct application of Equation (13.8), we have

$$L = \int_0^{3\pi/2} \sqrt{(-3\sin(t))^2 + (3\cos(t))^2}\,dt.$$

Then apply the Pythagorean theorem:

$$= \int_0^{3\pi/2} 3\,dt = \frac{9\pi}{2}.$$

This should make sense; we know from geometry that the circumference of a circle with radius 3 is 6π; since we are finding the arc length of $3/4$ of a circle, the arc length is $3/4 \cdot 6\pi = 9\pi/2$.

As mentioned above, care should be taken when setting the limits of integration for curves defined by means of parametric equations.

Example 13.16

Find the arc length of the astroid parametrized by $x = c\cos^3(t)$, $y = c\sin^3(t)$ (Figure 13.24).

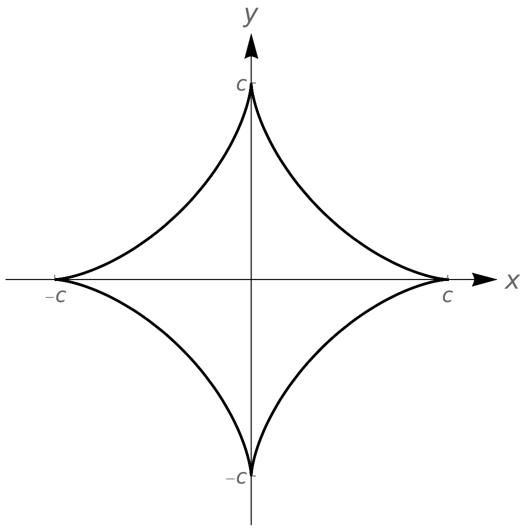


Figure 13.24: A graph of the astroid from Example 13.16.

Solution

By direct application of Equation (13.8), we have

$$\begin{aligned} L &= \int_{t_1}^{t_2} \sqrt{\left(-3c\cos^2(t)\sin(t)\right)^2 + \left(3c\sin^2(t)\cos(t)\right)^2}\, dt \\ &= \int_{t_1}^{t_2} 3c\sqrt{\sin^2(t)\cos^2(t)\left(\sin^2(t)+\cos^2(t)\right)}\, dt \\ &= \int_{t_1}^{t_2} \left|3c\sin(t)\cos(t)\right|\, dt \\ &= \int_{t_1}^{t_2} \frac{3c}{2}\left|\sin(2t)\right|. \end{aligned}$$

To find the limits of integration, we note that the arc length we are looking for is four times the length of one arc of the astroid. For what concerns the arc in the first quadrant, we would, when working with cartesian coordinates, vary x from $a = 0$ to $b = c$, where t is $\pi/2$ and 0, respectively. So, when integrating with respect to t the lower limit of integration should become $t_1 = 0$ and the upper limit $t_2 = \pi/2$. Taking into account these details, we get the following

$$L = 4\int_0^{\pi/2} \frac{3c}{2}\sin(2t) = 6c.$$

As we have already considered the arc length of curves defined by rectangular and parametric equations, we now consider it as well in the context of polar equations. Recall that the arc length L of the graph defined by the parametric equations $x = f(t)$, $y = g(t)$ on $[a, b]$ is

$$L = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2}\, dt. \tag{13.9}$$

Now consider the polar function $r = f(\theta)$. We again use the identities $x = f(\theta)\cos(\theta)$ and $y = f(\theta)\sin(\theta)$ to create parametric equations based on the polar function. We compute $x'(\theta)$ and $y'(\theta)$ as done before when computing $\frac{dy}{dx}$, then apply Equation (13.9).

The expression $x'(\theta)^2 + y'(\theta)^2$ can be simplified a great deal; we leave this as an exercise and state that

$$x'(\theta)^2 + y'(\theta)^2 = f'(\theta)^2 + f(\theta)^2.$$

This leads us to the arc length formula.

So, let $r = f(\theta)$ be a polar function with f' continuous on $[\alpha, \beta]$, on which the graph traces itself only once. The arc length L of the graph on $[\alpha, \beta]$ is

$$L = \int_\alpha^\beta \sqrt{f'(\theta)^2 + f(\theta)^2}\, d\theta = \int_\alpha^\beta \sqrt{(r')^2 + r^2}\, d\theta. \tag{13.10}$$

Again, care should be taken when setting the limits of integration for curves defined by means of polar equations.

13.5 Surface area

13.5.1 Rectangular coordinates

We have already seen how a curve $y = f(x)$ on $[a, b]$ can be revolved about an axis to form a solid. Instead of computing its volume, we now consider its surface area.

We begin as we have in the previous sections: we partition the interval $[a, b]$ with n subintervals, where the i^{th} subinterval is $[x_i, x_{i+1}]$. On each subinterval, we can approximate the curve $y = f(x)$ with a straight line that connects $f(x_i)$ and $f(x_{i+1})$ as shown in Figure 13.25(a). Revolving this line segment about the x-axis creates part of a cone (called a frustum of a cone) as shown in Figure 13.25(b). The surface area of a frustum of a cone is

$$2\pi \cdot \text{ length } \cdot \text{average of the two radii } R \text{ and } r.$$

The length is given by L_i. More precisely, we use the material just covered by arc length to state that

$$L \approx \sqrt{1 + f'(c_i)^2}\Delta x_i$$

for some c_i in the i^{th} subinterval. The radii are just the function evaluated at the endpoints of the interval. That is,

$$R_i = f(x_{i+1}) \quad \text{and} \quad r_i = f(x_i).$$

Thus the surface area of this sample frustum of the cone is approximately

$$2\pi \frac{f(x_i) + f(x_{i+1})}{2} \sqrt{1 + f'(c_i)^2}\Delta x_i.$$

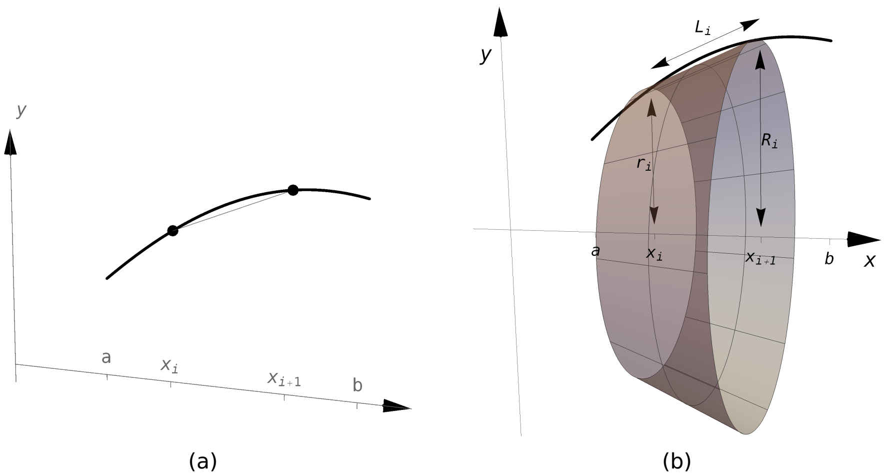


Figure 13.25: Establishing the formula for surface area.

Since f is a continuous function, the intermediate value theorem states there is some d_i in $[x_i, x_{i+1}]$ such that $f(d_i) = \frac{f(x_i)+f(x_{i+1})}{2}$; we can use this to rewrite the above equation as

$$2\pi f(d_i)\sqrt{1+f'(c_i)^2}\Delta x_i.$$

Summing over all the subintervals we get the total surface area to be approximately

$$SA \approx \sum_{i=1}^{n} 2\pi f(d_i)\sqrt{1+f'(c_i)^2}\Delta x_i,$$

which is a Riemann sum. Taking the limit as the subinterval lengths go to zero gives us the exact surface area, given in the following theorem.

Theorem 13.3 (Surface area of a solid of revolution using rectangular coordinates)
Let f be differentiable on $[a, b]$, where f' is also continuous on $[a, b]$.

1. *The surface area of the solid formed by revolving the graph of $y = f(x)$, where $f(x) \geq 0$, about the x-axis is*
$$SA = 2\pi \int_a^b f(x)\sqrt{1+f'(x)^2}\ dx.$$

2. *The surface area of the solid formed by revolving the graph of $y = f(x)$ about the y-axis, where $a, b \geq 0$, is*
$$SA = 2\pi \int_a^b x\sqrt{1+f'(x)^2}\ dx.$$

When revolving $y = f(x)$ about the y-axis, the radii of the resulting frustum are x_i and x_{i+1}; their average value is simply the midpoint of the interval. In the limit, this midpoint is just x. This gives the second part of Theorem 13.3.

Example 13.17

Find the surface area of the solid formed by revolving $y = \sin(x)$ on $[0, \pi]$ about the x-axis, as shown in Figure 13.26.

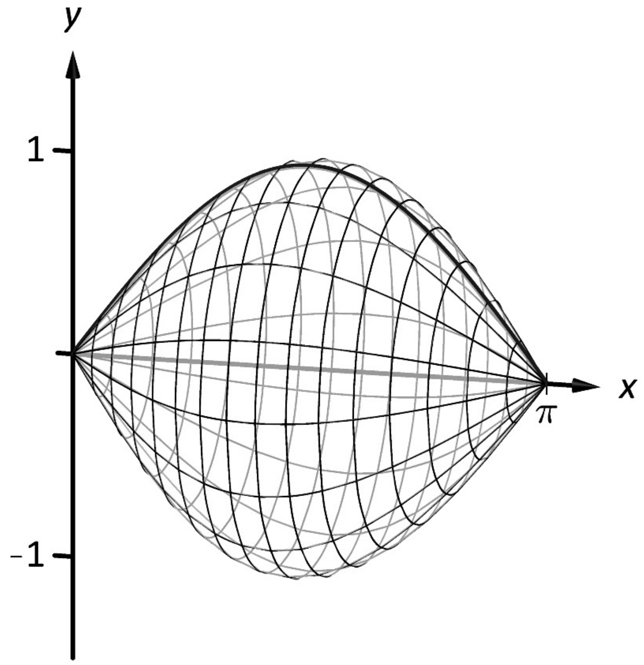


Figure 13.26: Revolving $y = \sin(x)$ on $[0, \pi]$ about the x-axis.

Solution

The setup is relatively straightforward. Using Theorem 13.3, we have the surface area SA is:

$$\begin{aligned} SA &= 2\pi \int_0^\pi \sin(x)\sqrt{1+\cos^2(x)}\,dx \\ &= -2\pi\frac{1}{2}\left(\operatorname{arsinh}(\cos(x)) + \cos(x)\sqrt{1+\cos^2(x)}\right)\Big|_0^\pi \\ &= 2\pi\left(\sqrt{2} + \operatorname{arsinh}(1)\right) \approx 14.42 \text{ units}^2. \end{aligned}$$

The integration step above is nontrivial uses trigonometric substitution.

It is interesting to see that the surface area of a solid, whose shape is defined by a trigonometric function, involves both a square root and an inverse hyperbolic trigonometric function.

Example 13.18

Find the surface area of the solid formed by revolving the curve $y = x^2$ on $[0, 1]$ about the x-axis and the y-axis.

Solution

About the x-axis: the integral is straightforward to setup:

$$SA = 2\pi \int_0^1 x^2\sqrt{1+(2x)^2}\,dx.$$

Like the integral in Example 13.17, this requires trigonometric substitution.

$$= \frac{\pi}{32}\left(2(8x^3+x)\sqrt{1+4x^2} - \operatorname{arsinh}(2x)\right)\Big|_0^1$$

$$= \frac{\pi}{32}\left(18\sqrt{5} - \text{arsinh}(2)\right)$$
$$\approx 3.81 \text{ units}^2.$$

The solid formed by revolving $y = x^2$ about the x-axis is graphed in Figure 13.27(a).

About the y-axis: since we are revolving about the y-axis, the radius of the solid is not $f(x)$ but rather x. Thus the integral to compute the surface area is:

$$SA = 2\pi \int_0^1 x\sqrt{1+(2x)^2}\, dx.$$

This integral can be solved using substitution. Set $u = 1 + 4x^2$; the new bounds are $u = 1$ to $u = 5$. We then have

$$= \frac{\pi}{4}\int_1^5 \sqrt{u}\, du$$
$$= \frac{\pi}{6}\left(5\sqrt{5} - 1\right)$$
$$\approx 5.33 \text{ units}^2.$$

The solid formed by revolving $y = x^2$ about the y-axis is graphed in Figure 13.27(b).

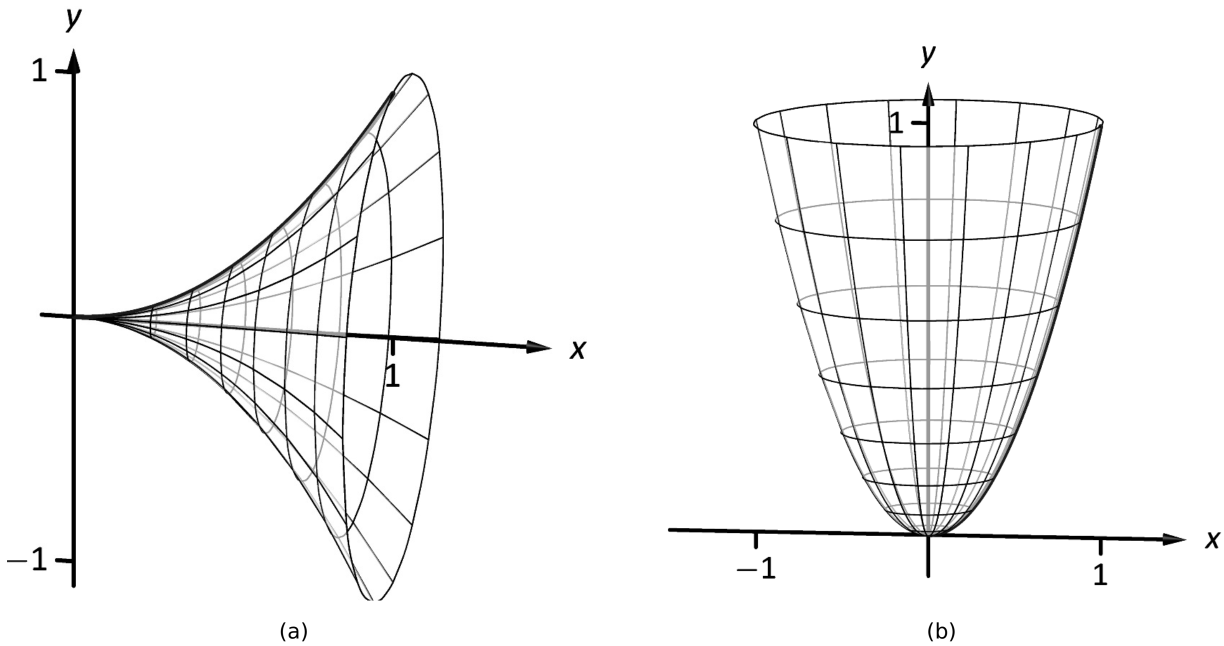


Figure 13.27: The solids used in Example 13.18.

We conclude this section with a famous mathematical paradox.

Example 13.19

Consider the solid formed by revolving $y = 1/x$ about the x-axis on $[1, +\infty[$. Find the volume and surface area of this solid. This shape, as graphed in Figure 13.28, is known as “Gabriel’s Horn” since it looks like a very long horn that only a supernatural person, such as an angel, could play.

Solution

To compute the volume it is natural to use the disk method. We have:

$$\begin{aligned} V &= \pi \int_1^{+\infty} \frac{1}{x^2}\,dx \\ &= \lim_{b\to+\infty} \pi \int_1^b \frac{1}{x^2}\,dx \\ &= \lim_{b\to+\infty} \pi\left(1-\frac{1}{b}\right) \\ &= \pi \text{ units}^3. \end{aligned}$$

Gabriel's Horn has a finite volume of π cubic units. Since we have already seen that regions with infinite length can have a finite area, this is not too difficult to accept.

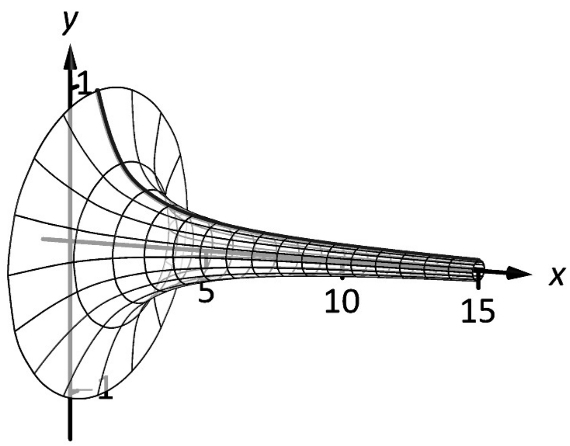


Figure 13.28: A graph of Gabriel's Horn.

We now consider its surface area. The integral is straightforward to setup:

$$SA = 2\pi \int_1^{+\infty} \frac{1}{x}\sqrt{1+\frac{1}{x^4}}\,dx.$$

Integrating this expression is not trivial. We can, however, compare it to other improper integrals. Since $1 < \sqrt{1+1/x^4}$ on $[1,+\infty[$, we can state that

$$2\pi \int_1^{+\infty} \frac{1}{x}\,dx < 2\pi \int_1^{+\infty} \frac{1}{x}\sqrt{1+\frac{1}{x^4}}\,dx.$$

The improper integral on the left diverges. Since the integral on the right is larger, we conclude it also diverges, meaning Gabriel's Horn has infinite surface area.

Hence the paradox: we can fill Gabriel's Horn with a finite amount of paint, but since it has infinite surface area, we can never paint it.

13.5.2 Parametric and polar equations

When dealing with a plane curve described by parametric equations, we can adapt the formula found in Theorem 13.3 in a similar way as done to produce the formula for arc length done before.

Theorem 13.4 (Surface area of a solid of revolution using parametric equations)
Consider the graph of the parametric equations $x = f(t)$ and $y = g(t)$, where f' and g' are continuous on an open interval I containing t_1 and t_2 on which the graph does not cross itself.

1. *The surface area of the solid formed by revolving the graph about the x-axis is (where $g(t) \geq 0$ on $[t_1, t_2]$):*
$$SA = 2\pi \int_{t_1}^{t_2} g(t)\sqrt{f'(t)^2 + g'(t)^2}\, dt.$$
2. *The surface area of the solid formed by revolving the graph about the y-axis is (where $f(t) \geq 0$ on $[t_1, t_2]$):*
$$SA = 2\pi \int_{t_1}^{t_2} f(t)\sqrt{f'(t)^2 + g'(t)^2}\, dt.$$

Example 13.20

Consider the teardrop shape formed by the parametric equations $x = t(t^2 - 1)$, $y = t^2 - 1$. Find the surface area if this shape is rotated about the x-axis, as shown in Figure 13.29.

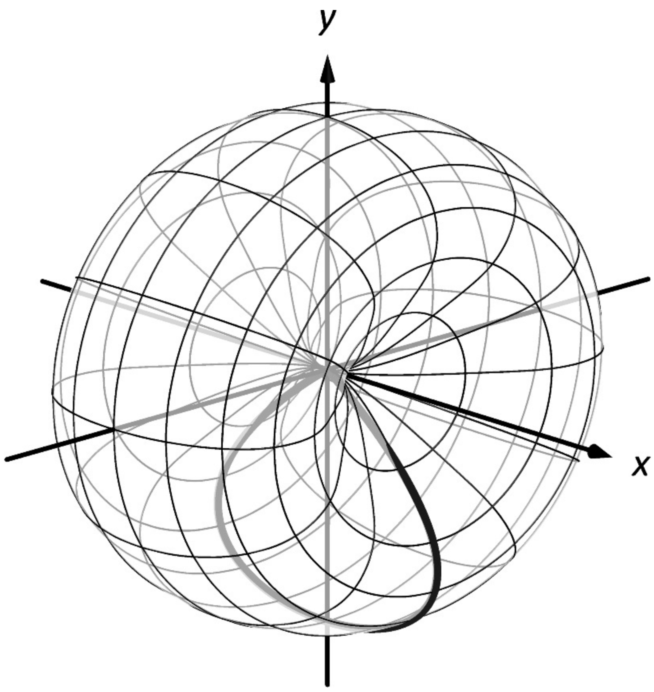


Figure 13.29: Rotating a teardrop shape about the x-axis in Example 13.20.

Solution

The teardrop shape is formed between $t=-1$ and $t=1$. Using Theorem 13.4, we see we need for $g(t)\geq 0$ on $[-1,1]$, and this is not the case. To fix this, we simplify replace $g(t)$ with $-g(t)$, which flips the whole graph about the x-axis. The surface area is:

$$SA = 2\pi\int_{-1}^{1}(1-t^2)\sqrt{(3t^2-1)^2+(2t)^2}\,dt$$
$$= 2\pi\int_{-1}^{1}(1-t^2)\sqrt{9t^4-2t^2+1}\,dt.$$

Once again we arrive at an integral that we cannot compute in terms of elementary functions. Using the midpoint rule with $n=20$, we find the area to be approximately $S=9.44$.

When dealing with polar equations, we may resort to the following theorem to find surface areas of solids of revolution.

Theorem 13.5 (Surface area of a solid of revolution using polar equations)
Consider the graph of the polar equation $r=f(\theta)$, where f' is continuous on $[\alpha,\beta]$, on which the graph does not cross itself.

1. *The surface area of the solid formed by revolving the graph about the initial ray ($\theta=0$) is:*

$$SA = 2\pi\int_{\alpha}^{\beta} f(\theta)\sin(\theta)\sqrt{f'(\theta)^2+f(\theta)^2}\,d\theta.$$

2. *The surface area of the solid formed by revolving the graph about the line $\theta=\pi/2$ is:*

$$SA = 2\pi\int_{\alpha}^{\beta} f(\theta)\cos(\theta)\sqrt{f'(\theta)^2+f(\theta)^2}\,d\theta.$$

Example 13.21

Find the surface area formed by revolving one petal of the rose curve $r=\cos(2\theta)$ about its central axis (see Figure 13.30(a)).

Solution

We choose, as implied by the figure, to revolve the portion of the curve that lies on $[0,\pi/4]$ about the initial ray. Using Theorem 13.5 and the fact that $f'(\theta)=-2\sin(2\theta)$, we have

$$SA = 2\pi\int_{0}^{\pi/4}\cos(2\theta)\sin(\theta)\sqrt{\big(-2\sin(2\theta)\big)^2+\big(\cos(2\theta)\big)^2}\,d\theta$$
$$\approx 1.36707.$$

The integral is another that cannot be evaluated in terms of elementary functions. The midpoint's rule, with $n=4$, approximates the value at 1.37.

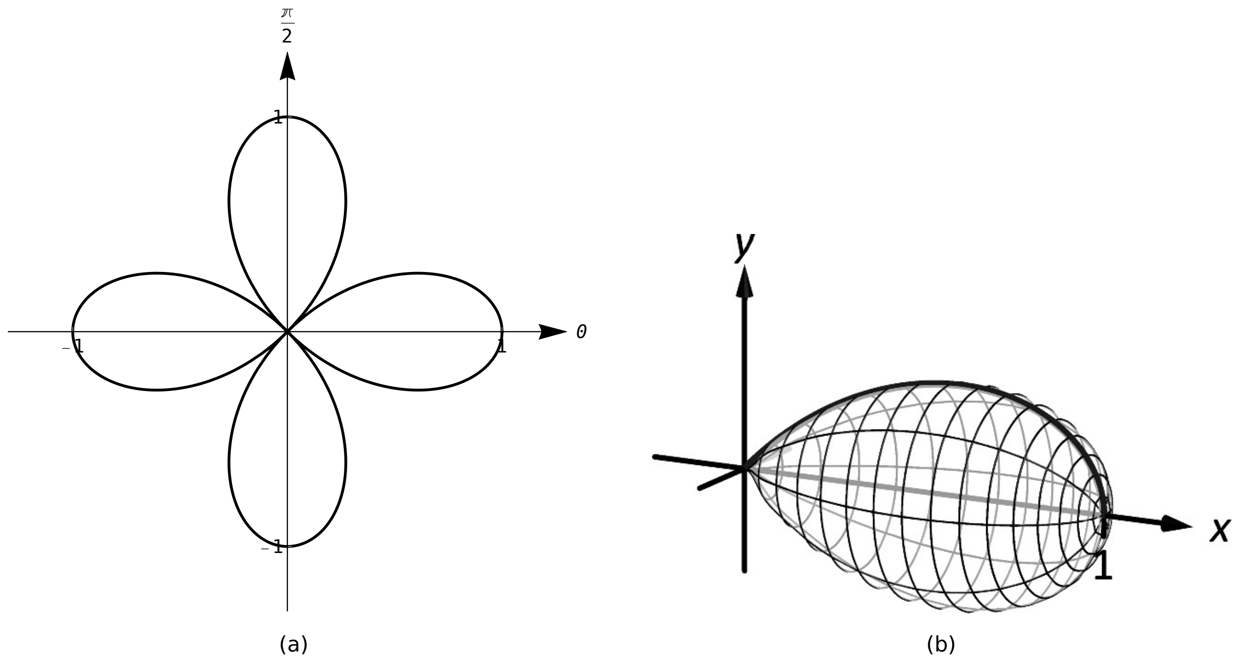


Figure 13.30: Finding the surface area of a rose–curve petal that is revolved about its central axis.

13.6 Work

Work is the scientific term used to describe the action of a force which moves an object. When a constant force F is applied to move an object a distance d, the amount of work performed is $W = F \cdot d$. The SI unit of force is the Newton, (kg·m/s^2), and the SI unit of distance is a meter (m). The fundamental unit of work is one Newton–meter, or a joule (J). That is, applying a force of one Newton for one meter performs one joule of work.

When force is constant, the measurement of work is straightforward. For instance, lifting an object with of force of 200 N for 5 m gives rise $200 \cdot 5 = 1000$ J of work. What if the force applied is variable? For instance, imagine a climber pulling a 200 m rope up a vertical face. The rope becomes lighter as more is pulled in, requiring less force and hence the climber performs less work.

In general, let $F(x)$ be a force function on an interval $[a, b]$. We want to measure the amount of work done applying the force F from $x = a$ to $x = b$. We can approximate the amount of work being done by partitioning $[a, b]$ into subintervals $a = x_1 < x_2 < \cdots < x_{n+1} = b$ and assuming that F is constant on each subinterval. Let c_i be a value in the i^{th} subinterval $[x_i, x_{i+1}]$. Then the work done on this interval is approximately $W_i \approx F(c_i)(x_{i+1} - x_i) = F(c_i)\Delta x_i$, a constant force × the distance over which it is applied. The total work is

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

This, of course, is a Riemann sum. Taking a limit as the subinterval lengths go to zero gives an exact value of work which can be evaluated through a definite integral.

So, if we let $F(x)$ be a continuous function on $[a, b]$ describing the amount of force being applied to an object in the direction of travel from distance $x = a$ to distance $x = b$, then the total work W done on $[a, b]$ is

$$W = \int_a^b F(x)\, dx. \qquad (13.11)$$

Example 13.22

A 60m climbing rope is hanging over the side of a tall cliff.

1. How much work is performed in pulling the rope up to the top, where the rope has a mass of 66g/m?
2. At what point is exactly half the work performed?

Solution

1. We need to create a force function $F(x)$ on the interval $[0, 60]$. To do so, we must first decide what x is measuring: it is the length of the rope still hanging or is it the amount of rope pulled in? As long as we are consistent, either approach is fine. We adopt for this example the convention that x is the amount of rope pulled in. This seems to match intuition better; pulling up the first 10 meters of rope involves $x = 0$ to $x = 10$ instead of $x = 60$ to $x = 50$.

   As x is the amount of rope pulled in, the amount of rope still hanging is $60 - x$. This length of rope has a mass of 66 g/m, or 0.066 kg/m. The mass of the rope still hanging is $0.066(60 - x)$ kg; multiplying this mass by the acceleration of gravity, 9.8 m/s^2, gives our variable force function

$$F(x) = (9.8)(0.066)(60 - x) = 0.6468(60 - x).$$

   Thus the total work performed in pulling up the rope is

$$W = \int_0^{60} 0.6468(60 - x)\, dx = 1164.24 \text{ J}.$$

   By comparison, consider the work done in lifting the entire rope 60 meters. The rope weighs $60 \times 0.066 \times 9.8 = 38.808$ N, so the work applying this force for 60 meters is $60 \times 38.808 = 2328.48$ J. This is exactly twice the work calculated before.

2. we know the total work performed is 1164.24 J. We want to find a height h such that the work in pulling the rope from a height of $x = 0$ to a height of $x = h$ is 582.12, half the total work. Thus we want to solve the equation

$$\int_0^{h} 0.6468(60 - x)\, dx = 582.12$$

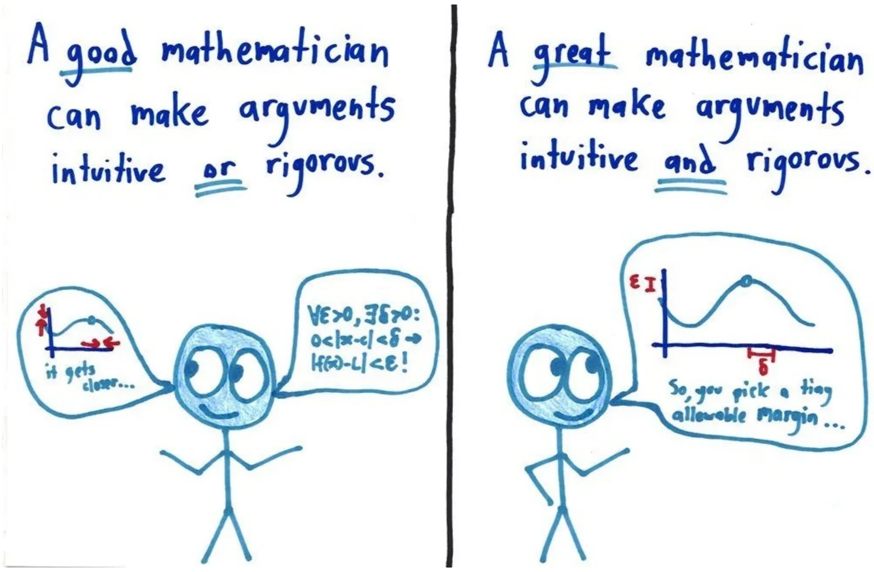
   for h.

$$\begin{aligned} \int_0^h 0.6468(60-x)\, dx &= 582.12 \\ \Leftrightarrow \quad \left(38.808x - 0.3234x^2\right)\Big|_0^h &= 582.12 \\ \Leftrightarrow \quad 38.808h - 0.3234h^2 &= 582.12 \\ \Leftrightarrow \quad -0.3234h^2 + 38.808h - 582.12 &= 0 \end{aligned}$$

   Apply the quadratic formula:

$$h = 17.57 \quad \text{and} \quad h = 102.43\,.$$

As the rope is only 60m long, the only sensible answer is $h = 17.57$. Thus about half the work is done pulling up the first 17.5m the other half of the work is done pulling up the remaining 42.43m.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

13.7 Exercises

Area between curves

Assignment 13.1 — Sketch the regions below and find their area.

(a) the area bounded by $y^2 = 4x$ and $y = 2x - 4$

(b) the area bounded by $x = 4 - y^2$ and the y-axis

(c) the smallest part within $x^2 + y^2 = 25$, cut off by $x = 3$

(d) the region enclosed between $y = 4x - x^2$, $y = 4 - x$ and the y-axis

(e) the region enclosed between $y = 6x - x^2$ and $y = x^2 - 2x$

(f) the region enclosed between $x^2 + y^2 = 12$ and $y^2 = x$

(g) the region enclosed between $y = 0$ and $y = \cos^2(x)$ for $x \in [0, 2\pi]$

(h) the region enclosed between $y = \ln(2x)$ and $y = \ln(x)$ for $x \in [1, e]$

(i) the region enclosed between $y = \cosh(x)$ and $y = \sinh(x)$ for $x \in [0, 1]$

(j) the region enclosed between $y = \sin(x)$, $x + y + \dfrac{\pi}{2} + 1 = 0$ and the x-axis

Assignment 13.2 — Find the areas of the regions below.

(a) the area inside the loop of the curve $\begin{cases} x = t^2 \\ y = t - \dfrac{t^3}{9} \end{cases}$

(b) the region enclosed by the curve $\begin{cases} x = 3 + \cos(\theta) \\ y = 4\sin(\theta) \end{cases}$

(c) the region enclosed by the curve $\begin{cases} x = 3\sin(2t) \\ y = 2\cos(t) \end{cases}$

(d) The region enclosed by the cardioid $r(\theta) = a(1 + \cos(\theta))$

(e) The region enclosed by one loop of the curve $r(\theta) = 4\cos(2\theta)$

(f) The region enclosed by the astroid

$$\begin{cases} x = c\cos^3(t) \\ y = c\sin^3(t). \end{cases}$$

(g) The region located between the first and second loops of Archimedes' spiral with polar equation $r(\theta) = a\theta$.

(h) The region located outside the curve $r = a$ and inside $r(\theta) = 2a\sin(3\theta)$

(i) The common region enclosed by the curves $\sqrt{3}r(\theta) = 1 + \sin(\theta)$ and $r(\theta) = \cos(\theta)$

(j) The region enclosed by the curve $(x^2 + y^2)^2 = ay(3x^2 - y^2)$.
Show that this curve is given by the polar equation $r = a\sin(3\theta)$ (see Figure 11.1.3).

(k) the region enclosed by the first loop of the logarithmic spiral $r(\theta) = 3e^{2\theta}$.

Volume by cross-sectional area and The shell method

Assignment 13.3 — Using the most efficient method to find the volume of the body of revolution obtained by rotating the given region about the given axis.

(a) the region in the first quadrant bounded by $y^2 = 8x$ en $x = 2$ about the x-axis

(b) the region bounded by $y^2 = 8x$ and $x = 2$ about the y-axis

(c) the region bounded by $y = x^2$, $y = \sqrt{x}$, $x = 0$ and $x = 1$ about the x-axis

(d) the region bounded by $y^2 = 8x$ and $x = 2$ about the line $x = 2$

(e) the region bounded by $y = x$ and $x = 4y - y^2$ about (a) the x-axis and (b) the y-axis

(f) the region inside $y = 4x - x^2$, cut off by the x-axis about $y = 6$

Assignment 13.4 — Find the volume of the body of revolution obtained by rotating the given region about the given axis.

(a) $y = 2x$ about $x = 3$ for $y \in [0, 6]$

(b) the region enclosed by $y^2 = x^2(1 - x^2)$ about (a) the x-axis and (b) the y-axis

(c) the area between the first loop of the cycloid $\begin{cases} x = \theta - \sin(\theta) \\ y = 1 - \cos(\theta) \end{cases}$ and the x-as about (a) the y-axis and (b) the line $y = 2$

(d) the region for which $0 \leq y \leq 1 - x^2$ about the line $y = 1$

(e) the region above $x - 2y + 5 = 0$ and inside $x^2 + y^2 = 25$ about the x-axis

(f) the upper half of the astroid $\begin{cases} x = c\cos^3(t) \\ y = c\sin^3(t) \end{cases}$ about the x-axis

(g) the region enclosed by $r(\theta) = 4\cos^2(\theta)$ about the polar axis ($\theta = 0$)

(h) a circular disk about one of the tangents

(i) the cissoid $y^2 = \dfrac{x^3}{2a - x}$ about the asymptote $x = 2a$. the graph of the cissoid with $a = 3$ is given in Figure 13.31.

(j) the cardioid $r(\theta) = a(1 + \cos(\theta))$ with $\theta \in [0, 2\pi]$ about the polar axis

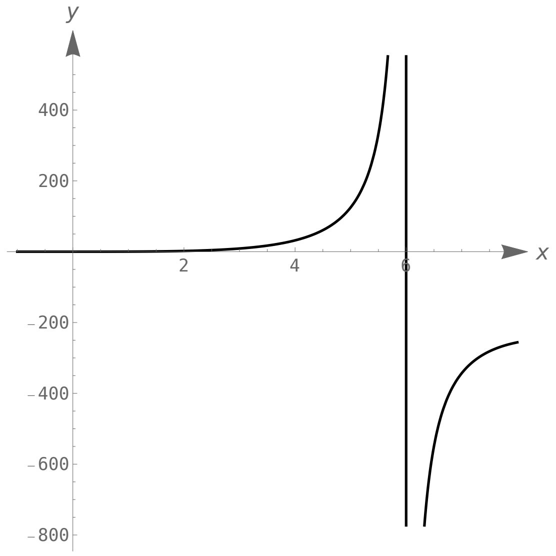


Figure 13.31: The graph of the cissoid with $a = 3$ from Exercise 13.4.

Arc length

Assignment 13.5 — Find the arc length of (the part of) the given curve.

(a) the astroid $x^{2/3} + y^{2/3} = a^{2/3}$

(b) the circle $r = 2\sin(\theta) + 4\cos(\theta)$

(c) the cardioid $r(\theta) = a(1 + \cos(\theta))$ with $0 \leq \theta \leq 2\pi$

(d) the part of $x = \ln\left(\dfrac{1}{\cos(y)}\right)$ between $y = 0$ and $y = \dfrac{\pi}{3}$

(e) the curve $r(\theta) = \dfrac{1}{\theta}$ with $\theta \in \left[\dfrac{1}{2}, 2\right]$

(f) the curve $r(\theta) = \dfrac{a}{\cos^2\left(\frac{\theta}{2}\right)}$ with $-\dfrac{\pi}{2} \leq \theta \leq \dfrac{\pi}{2}$

(g) the curve $\begin{cases} x = 2\cos(t) - \cos(2t) \\ y = 2\sin(t) - \sin(2t) \end{cases}$

(h) the part of $y^2 = x^3$ between the origin and the point with abscissa 4

(i) the closed part of $9y^2 = x(x-3)^2$

(j) the arc given by $\theta(r) = \dfrac{1}{2}\left(r + \dfrac{1}{r}\right)$ with $r \in [1, 3]$

Surface area

Assignment 13.6 — Find the surface area of the body of revolution obtained by rotating the given region about the given axis.

(a) the area under $y = \sin(x)$ about the x-axis for $x \in [0, \pi]$

(b) the ellipsoid $\dfrac{x^2}{16} + \dfrac{y^2}{4} = 1$ about the x-axis

(c) the line $y = 2x$ about the line $x = 3$ for $x \in \mathbb{R}^+$.

(d) the region between $x = y^3$, $y = 0$ and $y = 1$ about (a) the y-axis and (b) the line $x = 1$ (do not evaluate the integral)

(e) one loop of $8y^2 = x^2 - x^4$ about the x-axis

(f) the region enclosed by the astroid $\begin{cases} x = \cos^3(t) \\ y = \sin^3(t) \end{cases}$ and the x-axis about (a) the x-axis and (b) the line $y = -1$

(g) $r^2 = a^2 \cos(2\theta)$ about the polar axis

(h) the first loop of the cycloid $\begin{cases} x = a(\theta - \sin(\theta)) \\ y = a(1 - \cos(\theta)) \end{cases}$ about (a) the x-axis and (b) the line $x = a\pi$

(i) the cardioid $r(\theta) = a(1 + \cos(\theta))$ about the polar axis

Review exercises

Assignment 13.7 — A cherry floats in a cocktail glass. The glass has the shape of a sphere with diameter 8 cm. The idealized cherry is spherical and has a diameter of 2 cm. The glass is filled to $3/2$ cm from its border with Kir and the top of the cherry is located 1 cm from the rim of the glass.

(a) How much Kir does the glass contain? Tip: the glass is created by rotating $x = f(y)$ about the y-axis. You can model the cherry by rotating $x = g(y)$ about the y-axis.

(b) Find the area of the part of the cherry that extends above the liquid surface.

Mathematics is about giving the same name to different things.

— Henri Poincaré —

14

Sequences and series

This chapter introduces sequences and series, important mathematical constructions that are useful when solving a large variety of mathematical problems. The content of this chapter is considerably different from the content of the chapters before it. While the material we learn here definitely falls under the scope of calculus, we will make very little use of derivatives or integrals. Limits are extremely important, though, especially limits that involve infinity.

One of the problems addressed by this chapter is this: suppose we know information about a function and its derivatives at a point, such as $f(1) = 3$, $f'(1) = 1$, $f''(1) = -2$, $f'''(1) = 7$, and so on. What can I say about $f(x)$ itself? Is there any reasonable approximation of the value of $f(2)$? The topic of Taylor Series addresses this problem, and allows us to make excellent approximations of functions when limited knowledge of the function is available.

14.1 Sequences

14.1.1 Definition

In mathematics, we use the word **sequence** (*rij*) to refer to an ordered set of numbers, i.e., a set of numbers that occur one after the other.

For instance, the numbers 2, 4, 6, 8, . . . , form a sequence. The **order** (*ordening*) is important; the first number is 2, the second is 4, etc. It seems natural to seek a formula that describes a given sequence, and often this can be done. For instance, the sequence above could be described by the function $a(n) = 2n$, for the values of $n = 1, 2, \ldots$ To find the 10^{th} term in the sequence, we would compute $a(10)$. This leads us to the following, formal definition of a sequence.

Definitie 14.1 (Sequence)
A **sequence** (*rij*) is a function $a(n)$ such that $\operatorname{dom} a = \mathbb{N}_0$. The range of a sequence is the set of all distinct values of $a(n)$.
The terms of a sequence are the values $a(1)$, $a(2)$, ..., or equivalently, a_1, a_2, ..., where the subscript i refers to the **index** or **rank** of a_i.
A sequence $a(n)$ is often denoted as $\{a_n\}$.

Example 14.1

Find the n^{th} term of the following sequences, i.e., find a function that describes each of the given sequences.

1. 2, 5, 8, 11, 14, ...
2. 2, −5, 10, −17, 26, −37, ...

Solution

We should first note that there is never exactly one function that describes a finite set of numbers as a sequence. There are many sequences that start with 2, then 5, as our first example does. We are looking for a simple formula that describes the terms given, knowing there is possibly more than one answer.

1. Note how each term is 3 more than the previous one. This implies a linear function would be appropriate: $a(n) = a_n = 3n + b$ for some appropriate value of b. As we want $a_1 = 2$, we set $b = -1$. Thus $a_n = 3n - 1$.
2. First notice how the sign changes from term to term. This is most commonly accomplished by multiplying the terms by either $(-1)^n$ or $(-1)^{n+1}$. Using $(-1)^n$ multiplies the odd terms by (-1); using $(-1)^{n+1}$ multiplies the even terms by (-1). As this sequence has negative even terms, we will multiply by $(-1)^{n+1}$.

   After this, we might feel a bit stuck as to how to proceed. At this point, we are just looking for a pattern of some sort: what do the numbers 2, 5, 10, 17, etc., have in common? There are many correct answers, but the one that we will use here is that each is one more than a perfect square. That is, $2 = 1^2 + 1$, $5 = 2^2 + 1$, $10 = 3^2 + 1$, etc. Thus our formula is

$$a_n = (-1)^{n+1}(n^2 + 1).$$

14.1.2 Limits of sequences

14.1.2.1 Definition

A common mathematical endeavour is to create a new mathematical object, such as a sequence, and then apply previously known mathematics to the new object. We do so here. The fundamental concept of calculus is the limit, so we will investigate what it means to find the limit of a sequence.

Definitie 14.2 (Limit of a sequence)
Let $\{a_n\}$ be a sequence and let L be a real number. Given any $\varepsilon > 0$, if an m can be found such that $|a_n - L| < \varepsilon$ for all $n > m$, then we say the **limit of** $\{a_n\}$, as n approaches infinity, is L, denoted

$$\lim_{n\to+\infty} a_n = L.$$

If $\lim\limits_{n\to+\infty} a_n$ exists, we say the sequence converges; otherwise, the sequence diverges.

This definition states, informally, that if the limit of a sequence is L, then if you go far enough out along the sequence, all subsequent terms will be really close to L. This definition is reminiscent of the ε–δ definitions of Chapter 8. In that chapter we developed other tools to evaluate limits apart from the formal definition; we do so here as well.

Theorem 14.1 (Limit of a sequence)
Let $\{a_n\}$ be a sequence and let $f(x)$ be a function whose domain contains the strictly positive real numbers where $f(n) = a_n$ for all n in $\mathbb{N}_0$.

If

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

then

$$\lim_{n\to+\infty} a_n = L.$$

Theorem 14.1 allows us, in certain cases, to apply the tools developed in Chapter 8 to limits of sequences. Note two things not stated by the theorem:

1. If $\lim\limits_{x\to+\infty} f(x)$ does not exist, we cannot conclude that $\lim\limits_{n\to+\infty} a_n$ does not exist. It may, or may not, exist. For instance, we can define a sequence $\{a_n\} = \{\cos(2\pi n)\}$. Let $f(x) = \cos(2\pi x)$. Since the cosine function oscillates over the real numbers, the limit $\lim\limits_{x\to+\infty} f(x)$ does not exist. However, for every positive integer n, $\cos(2\pi n) = 1$, so $\lim\limits_{n\to+\infty} a_n = 1$.
2. If we cannot find a function $f(x)$ whose domain contains the positive real numbers where $f(n) = a_n$ for all n in $\mathbb{N}_1$, we cannot conclude $\lim\limits_{n\to+\infty} a_n$ does not exist. It may, or may not, exist.

Given the link between the limit of a sequence and the limit of its associated function as pointed out by Theorem 14.1, it is intuitive to understand that there must be also a squeeze theorem for sequences, just as we have a squeeze theorem for limits (Theorem 8.5). This is confirmed by the following theorem.

Theorem 14.2 (Squeeze theorem for sequences)
If $a_n \leq c_n \leq b_n$ for all $n > N$ for some N and it holds that $\lim\limits_{n\to+\infty} a_n = \lim\limits_{n\to+\infty} b_n = L$, then $\lim\limits_{n\to+\infty} c_n = L$.

Example 14.2

Determine the convergence or divergence of the following sequences.

1. $\{a_n\} = \left\{\dfrac{3n^2 - 2n + 1}{n^2 - 1000}\right\}$
2. $\{a_n\} = \{\cos(n)\}$
3. $\{a_n\} = \left\{\dfrac{(-1)^n}{n}\right\}$

Solution

1. By factoring out the highest power of x, we can state that

$$\lim_{x\to+\infty} \frac{3x^2-2x+1}{x^2-1000} = 3.$$

Thus the sequence $\{a_n\}$ converges, and its limit is 3. A scatter plot of every 5 values of a_n is given in Figure 14.1(a). The values of a_n vary widely near $n=30$, ranging from about -73 to 125, but as n grows, the values approach 3.

2. The limit $\lim_{x\to+\infty}\cos(x)$ does not exist as $\cos(x)$ oscillates. Thus we cannot apply Theorem 14.1. The fact that the cosine function oscillates strongly hints that $\cos(n)$, when n is restricted to $\mathbb{N}$, will also oscillate. Figure 14.1(b), where the sequence is plotted, implies that this is true. Because only discrete values of cosine are plotted, it does not bear strong resemblance to the familiar cosine wave.

3. We cannot actually apply Theorem 14.1 here, as the function $f(x) = (-1)^x/x$ is not well defined. What does $(-1)^{\sqrt{2}}$ mean? In fact, there is an answer, but it involves complex analysis, beyond the scope of this text. Instead, we invoke the definition of the limit of a sequence. By looking at the plot in Figure 14.1(c), we would like to conclude that the sequence converges to $L=0$. Let $\varepsilon>0$ be given. We can find a natural number m such that $1/m<\varepsilon$. Let $n>m$, and consider $|a_n-L|$:

$$\begin{aligned} |a_n-L| &= \left|\frac{(-1)^n}{n}-0\right| \\ &= \frac{1}{n} \\ &< \frac{1}{m} \quad \text{(since } n>m\text{)} \\ &< \varepsilon. \end{aligned}$$

We have shown that by picking m large enough, we can ensure that a_n is arbitrarily close to our limit, $L=0$, hence by the definition of the limit of a sequence, we can say $\lim_{n\to+\infty} a_n = 0$.

In the previous example we used the definition of the limit of a sequence to determine the convergence of a sequence as we could not apply Theorem 14.1. In general, we like to avoid invoking the definition of a limit, and the following theorem gives us tool that we could use in that example instead.

Theorem 14.3 (Absolute value theorem)
Let $\{a_n\}$ be a sequence. If $\lim_{n\to+\infty}|a_n| = 0$, then $\lim_{n\to+\infty} a_n = 0$.

Proof The main thing to this proof is to note that

$$-|a_n| \le a_n \le |a_n|$$

and

$$\lim_{n\to+\infty}(-|a_n|) = -\lim_{n\to+\infty}|a_n| = 0\,.$$

We then have $\lim_{n\to+\infty}(-|a_n|) = \lim_{n\to+\infty}(|a_n|) = 0$ and so by the squeeze theorem for sequences we must also have that $\lim_{n\to+\infty} a_n = 0$. □

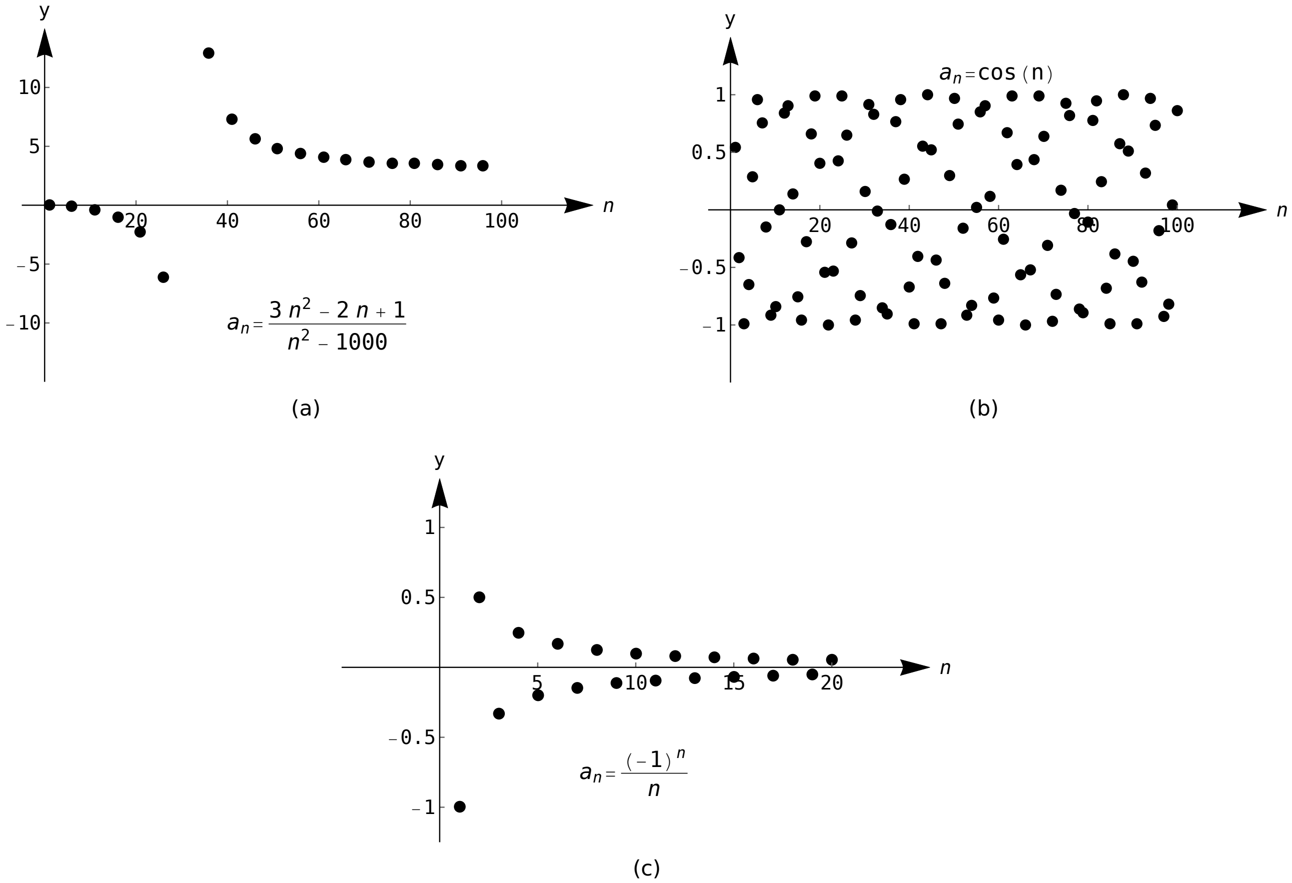


Figure 14.1: Scatter plots of the sequences in Example 14.2.

Example 14.3

Determine the convergence or divergence of the following sequences.

1. $\{a_n\} = \left\{ \dfrac{(-1)^n}{n} \right\}$

2. $\{a_n\} = \left\{ \dfrac{(-1)^n(n+1)}{n} \right\}$

Solution

1. This appeared in Example 14.2. We want to apply Theorem 14.3, so consider the limit of $\{|a_n|\}$:

$$\begin{aligned} \lim_{n\to+\infty} |a_n| &= \lim_{n\to+\infty} \left| \frac{(-1)^n}{n} \right| \\ &= \lim_{n\to+\infty} \frac{1}{n} \\ &= 0. \end{aligned}$$

Since this limit is 0, we can apply Theorem 14.3 and state that $\lim\limits_{n\to+\infty} a_n = 0$.

2. Because of the alternating nature of this sequence, we cannot simply look at the limit

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

We can try to apply the techniques of Theorem 14.3:

$$\lim_{n\to+\infty} |a_n| = \lim_{n\to+\infty} \left|\frac{(-1)^n(n+1)}{n}\right|$$
$$= \lim_{n\to+\infty} \frac{n+1}{n} = 1\,.$$

We have concluded that when we ignore the alternating sign, the sequence approaches 1. This means we cannot apply Theorem 14.3; it states the the limit must be 0 in order to conclude anything.

Since we know that the signs of the terms alternate and we know that the limit of $|a_n|$ is 1, we know that as n approaches infinity, the terms will alternate between values close to 1 and -1, meaning the sequence diverges. A plot of this sequence is given in Figure 14.2.

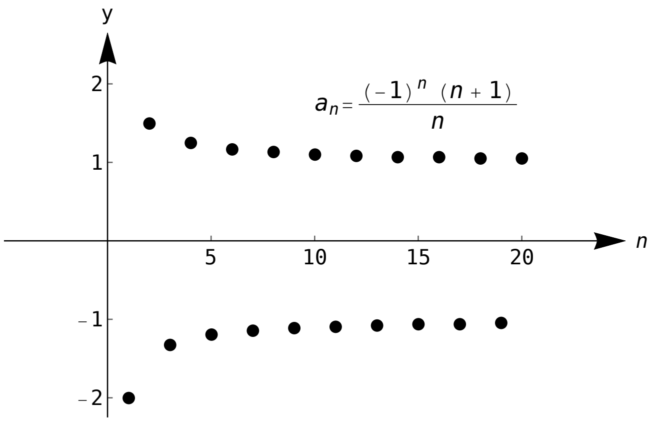


Figure 14.2: A plot of a sequence in Example 14.3, part 2.

14.1.2.2 Properties

We continue our study of the limits of sequences by considering some of the properties of these limits. These follow easily from the properties of limits of functions discussed in Chapter 8.

Let $\{a_n\}$ and $\{b_n\}$ be sequences such that $\lim_{n\to+\infty} a_n = L$, $\lim_{n\to+\infty} b_n = K$, and let c be a real number. The following properties hold

1. $\lim_{n\to+\infty}(a_n \pm b_n) = L \pm K$
2. $\lim_{n\to+\infty}(a_n \cdot b_n) = L \cdot K$
3. $\lim_{n\to+\infty}\left(\frac{a_n}{b_n}\right) = \frac{L}{K}$, $K \neq 0$
4. $\lim_{n\to+\infty} c \cdot a_n = c \cdot L$

For instance, let

$$\{a_n\} = \left\{\frac{n+1}{n^2}\right\},$$

and

$$\{b_n\} = \left\{\left(1+\frac{1}{n}\right)^n\right\},$$

with $\lim\limits_{n\to+\infty} a_n = 0$ and $\lim\limits_{n\to+\infty} b_n = e$, respectively. Then, $\lim\limits_{n\to+\infty}(a_n + b_n) = e$. Similarly, $\lim\limits_{n\to+\infty} 1000a_n = 1000\cdot 0 = 0$.

14.1.2.3 Behaviour

Let us start with some definitions describing properties of the range of a sequence.

Definitie 14.3 (Bounded and unbounded sequences)
A sequence $\{a_n\}$ is said to be **bounded** (*begrensd*) if there exist real numbers m and M such that $m \le a_n \le M$ for all n in $\mathbb{N}_0$.
A sequence $\{a_n\}$ is said to be **unbounded** (*onbegrensd*) if it is not bounded.
A sequence $\{a_n\}$ is said to be **bounded above** (*naar boven begrensd*) if there exists an M such that $a_n \le M$ for all n in $\mathbb{N}_0$; it is **bounded below** (*naar beneden begrensd*) if there exists an m such that $m \le a_n$ for all n in $\mathbb{N}_0$.

It follows from this definition that an unbounded sequence may be bounded above or bounded below; a sequence that is both bounded above and below is simply a bounded sequence. Alternatively, using absolute values, one may also say that a sequence $\{a_n\}$ is bounded if there exists $M \ge 0$ such that $|a_n| \le M$ for all $n \in \mathbb{N}_0$.

Example 14.4

Determine the boundedness of the following sequences.

1. $\{a_n\} = \left\{\frac{1}{n}\right\}$
2. $\{a_n\} = \{2^n\}$

Solution

1. The terms of this sequence are always positive but are decreasing, so we have $0 < a_n < 2$ for all n. Thus this sequence is bounded. Figure 14.3(a) illustrates this.
2. The terms of this sequence obviously grow without bound. However, it is also true that these terms are all positive, meaning $0 < a_n$. Thus we can say the sequence is unbounded, but also bounded below. Figure 14.3(b) illustrates this.

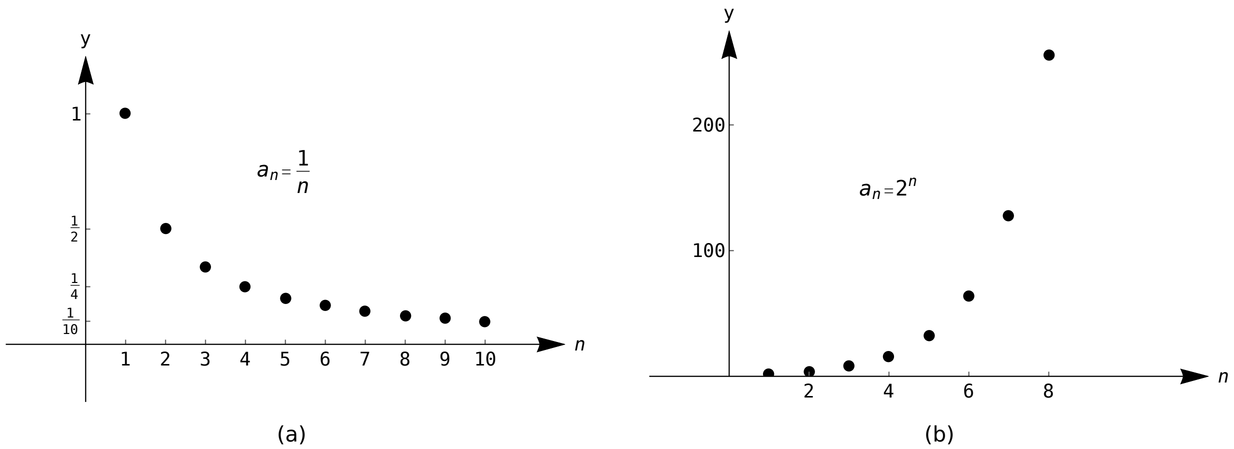


Figure 14.3: A plot of $\{a_n\} = \{1/n\}$ and $\{a_n\} = \{2^n\}$ from Example 14.4.

The previous example produces some interesting concepts. First, we can recognize that the sequence $\{1/n\}$ converges to 0. This says, informally, that most of the terms of the sequence are really close to 0. This implies that the sequence is bounded, using the following logic. First, most terms are near 0, so we could find some sort of bound on these terms (using Definition 14.2, the bound is ε). That leaves a few terms that are not near 0 (i.e., a finite number of terms). A finite list of numbers is always bounded.

This logic implies that if a sequence converges, it must be bounded. This is indeed true, as stated by the following theorem.

Theorem 14.4 (Convergent sequences are bounded)
Let $\{a_n\}$ be a convergent sequence. Then $\{a_n\}$ is bounded.

Note that this theorem does not say that bounded sequences must converge, nor does it say that if a sequence does not converge, it is not bounded.

Proof To prove this theorem, let $\{a_n\}$ be a convergent sequence and let $L = \lim_{n\to+\infty} a_n$. From Definition 14.2 with $\varepsilon = 1$, there exists an $N \in \mathbb{N}$ such that $|a_n - L| < 1$ whenever $n > N$. Thus for $n > N$ the triangle equality implies that $|a_n| = |a_n - L + L| \leq |a_n - L| + |L| < 1 + |L|$. Now, let $M = \max\{|a_1|, |a_2|, \ldots, |a_N|, |L|+1\}$. Then, we have $|a_n| \leq M$ for all $n \in \mathbb{N}_0$. Consequently, $\{a_n\}$ is bounded. □

We saw already the sequence $\{b_n\} = \{(1+1/n)^n\}$, for it was stated that $\lim\limits_{n\to+\infty} b_n = e$. Even though it may be difficult to intuitively grasp the behaviour of this sequence, we now know immediately that it is bounded.

Another interesting concept to come out of Example 14.4 involves the sequence $\{1/n\}$. The terms of the sequence are decreasing. That is, that $a_{n+1} < a_n$ for all n. This is easy to show as follows. Clearly $n < n+1$. Taking reciprocals flips the inequality: $1/n > 1/(n+1)$. This is the same as $a_n > a_{n+1}$. Sequences that either steadily increase or decrease are important, so we give this property a name.

Definitie 14.4 (Monotonic sequence)

1. A sequence $\{a_n\}$ is **monotonically increasing** (*monotoon stijgend*) if $a_n \leq a_{n+1}$ for all n, i.e.,
$$a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n \leq a_{n+1} \cdots$$
2. A sequence $\{a_n\}$ is **monotonically decreasing** (*monotoon dalend*) if $a_n \geq a_{n+1}$ for all n,

i.e.,

$$a_1 \geq a_2 \geq a_3 \geq \cdots \geq a_n \geq a_{n+1} \cdots$$

3. A sequence is **monotonic** (*monotoon*) if it is monotonically increasing or monotonically decreasing.

It is sometimes useful to call a monotonically increasing sequence strictly increasing if $a_n < a_{n+1}$ for all n. A similar statement holds for strictly decreasing.

Example 14.5

Determine the monotonicity of the following sequences.

1. $\{a_n\} = \left\{\dfrac{n+1}{n}\right\}$
2. $\{a_n\} = \left\{\dfrac{n^2}{n!}\right\}$

Solution

In each of the following, we will examine $a_{n+1} - a_n$. If $a_{n+1} - a_n \geq 0$, we conclude that $a_n \leq a_{n+1}$ and hence the sequence is increasing. If $a_{n+1} - a_n \leq 0$, we conclude that $a_n \geq a_{n+1}$ and the sequence is decreasing.

1. $$\begin{aligned} a_{n+1} - a_n &= \frac{n+2}{n+1} - \frac{n+1}{n} \\ &= \frac{(n+2)(n) - (n+1)^2}{(n+1)n} \\ &= \frac{-1}{n(n+1)} < 0 \quad \text{for all } n. \end{aligned}$$

   Since $a_{n+1} - a_n < 0$ for all n, we conclude that the sequence is strictly decreasing. A scatter plot of this sequence is shown in Figure 14.4(a).

2. The plot in Figure 14.4(b) shows that the sequence is not monotonic, but it suggests that it is monotonically decreasing after the first term. We perform the usual analysis to confirm this.

   $$\begin{aligned} a_{n+1} - a_n &= \frac{(n+1)^2}{(n+1)!} - \frac{n^2}{n!} \\ &= \frac{(n+1)^2 - n^2(n+1)}{(n+1)!} \\ &= \frac{-n^3 + 2n + 1}{(n+1)!} \end{aligned}$$

   When $n = 1$, the above expression is greater than zero; for $n \geq 2$, the above expression is lower than zero. Thus this sequence is not monotonic, but it is strictly decreasing after the first term.

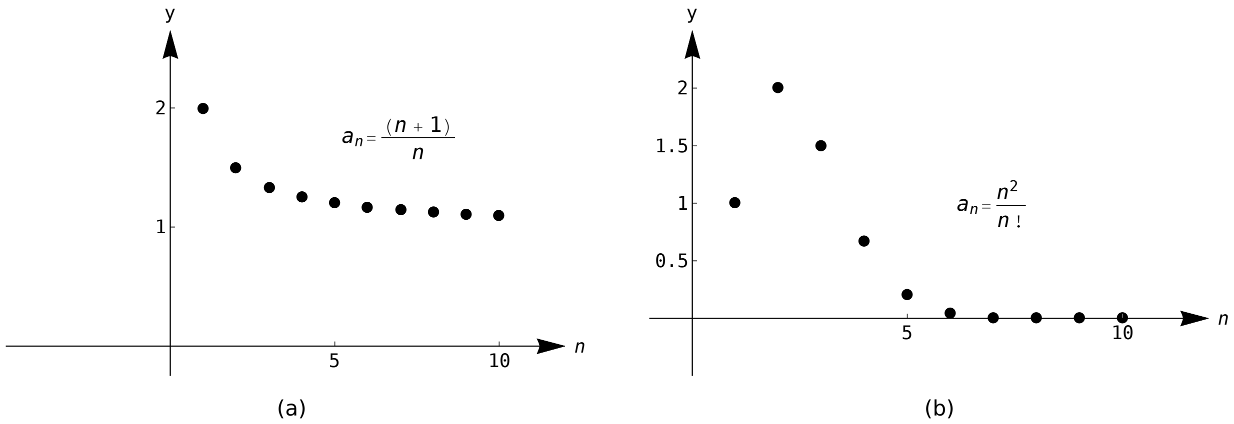


Figure 14.4: Plots of sequences in Example 14.5.

Knowing that a sequence is monotonic can be useful. Consider, for example, a sequence that is monotonically decreasing and is bounded below. We know the sequence is always getting smaller, but that there is a bound to how small it can become. This is enough to prove that the sequence will converge, as stated in the following theorem.

Theorem 14.5 (Monotone convergence theorem)
A monotone sequence of real numbers $\{a_n\}$ converges if and only if it is bounded.

Proof We already know that a convergent sequence is bounded. Hence we just need to show that a monotone and bounded sequence of real numbers is convergent. Let $\{a_n\}$ be such an increasing sequence bounded above. By assumption, $\{a_n\}$ is non-empty and bounded above. By the least-upper-bound property of real numbers, the supremum $c = \sup_n\{a_n\}$ exists and is finite. Now, for every $\varepsilon > 0$, $c-\varepsilon$ is not an upper bound. Thus there exists an integer N such that $a_N > c-\varepsilon$, since otherwise $c-\varepsilon$ is an upper bound of $\{a_n\}$, which contradicts to the definition of c. Then since $\{a_n\}$ is increasing, and c is its upper bound, we have

$$c-\varepsilon < a_N \leq a_n \leq c < c+\varepsilon$$

or equivalently

$$|a_n - c| < \varepsilon$$

for all $n \geq N$. Hence, by definition, the limit of $\{a_n\}$ is $\sup_n\{a_n\}$.

In the case when the sequence is decreasing, let $c = \inf_n\{a_n\}$ and proceed in a similar manner. □

Consider once again the sequence $\{a_n\} = \{1/n\}$. It is easy to show it is monotonically decreasing and that it is always positive (i.e., bounded below by 0). Therefore we can conclude by Theorem 14.5 that the sequence converges.

Sequences are a great source of mathematical inquiry. The On-Line Encyclopedia of Integer Sequences (OEIS) contains thousands of sequences and their formulae. Perusing this database quickly demonstrates that a single sequence can represent several different real life phenomena.

Interesting as this is, our interest actually lies elsewhere. We are more interested in the sum of a sequence. That is, given a sequence $\{a_n\}$, we are very interested in $a_1 + a_2 + a_3 + \cdots$. Of course, one might immediately think that thus adds up to infinity?". Many times, yes, but there are many important cases where the answer is no. This is the topic of series, which we begin to investigate in the next section.

Fibonacci numbers

The Fibonacci numbers make up an integer sequence, called the Fibonacci sequence, which is characterized by the fact that every number after the first two is the sum of the two preceding ones:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

Fibonacci sequences appear in biological settings, such as branching in trees, arrangement of leaves on a stem, the fruitlets of a pineapple, the flowering of artichoke, an uncurling fern and the arrangement of a pine cone, but there are also numerous poorly substantiated claims of Fibonacci numbers, relating to the breeding of rabbits, the seeds on a sunflower, the spirals of shells, and the curve of waves.

14.2 Infinite series

14.2.1 Definition

Given the sequence $\{a_n\} = \{1/2^n\} = 1/2,\ 1/4,\ 1/8,\ \ldots$, consider the following sums:

$$\begin{array}{ccccc} a_1 & = & 1/2 & = & 1/2 \\ a_1+a_2 & = & 1/2+1/4 & = & 3/4 \\ a_1+a_2+a_3 & = & 1/2+1/4+1/8 & = & 7/8 \\ a_1+a_2+a_3+a_4 & = & 1/2+1/4+1/8+1/16 & = & 15/16 \end{array}$$

In general, we can show that

$$a_1+a_2+a_3+\cdots+a_n = \frac{2^n-1}{2^n} = 1-\frac{1}{2^n}.$$

Let S_n be the sum of the first n terms of the sequence $\{1/2^n\}$. From the above, we see that $S_1 = 1/2$, $S_2 = 3/4$, etc. Our formula at the end shows that $S_n = 1-1/2^n$.

Now consider the following limit: $\lim\limits_{n\to+\infty} S_n = \lim\limits_{n\to+\infty}\left(1-1/2^n\right) = 1$. This limit can be interpreted as saying something amazing: the sum of all the terms of the sequence $\{1/2^n\}$ is 1. This example illustrates some interesting concepts that we explore in this section. We begin this exploration with some definitions.

Definitie 14.5 (Infinite series and partial sums)
Let $\{a_n\}$ be a sequence.

1. The sum $\displaystyle\sum_{n=1}^{+\infty} a_n$ is an **infinite series** (*oneindige reeks*) (or, simply series).
2. Let $S_n = \displaystyle\sum_{i=1}^{n} a_i$; the sequence $\{S_n\}$ is the **sequence of n^{th} partial sums of** $\{a_n\}$ (*partieelsom*).
3. If the sequence $\{S_n\}$ converges to L, we say the series $\displaystyle\sum_{n=1}^{+\infty} a_n$ **converges** (*convergeert*) to L, and we write $\displaystyle\sum_{n=1}^{+\infty} a_n = L$.

4. If the sequence $\{S_n\}$ **diverges** (*divergeert*), the series $\sum_{n=1}^{+\infty} a_n$ diverges.

We will explore a variety of series in this section. We start with two series that diverge, showing how we might discern divergence.

Example 14.6

1. Let $\{a_n\} = \{n^2\}$. Show $\sum_{n=1}^{+\infty} a_n$ diverges.
2. Let $\{b_n\} = \{(-1)^{n+1}\}$. Show $\sum_{n=1}^{+\infty} b_n$ diverges.

Solution

1. Consider S_n, the n^{th} partial sum.

$$\begin{aligned} S_n &= a_1 + a_2 + a_3 + \cdots + a_n \\ &= 1^2 + 2^2 + 3^2 \cdots + n^2. \end{aligned}$$

   By Theorem 2.1, this is

$$S_n = \frac{n(n+1)(2n+1)}{6}.$$

   Since $\lim_{n\to+\infty} S_n = +\infty$, we conclude that the series $\sum_{n=1}^{+\infty} n^2$ diverges. It is instructive to write

$$\sum_{n=1}^{+\infty} n^2 = +\infty$$

   for this tells us how the series diverges: it grows without bound. A scatter plot of the sequences $\{a_n\}$ and $\{S_n\}$ is given in Figure 14.5(a). The terms of $\{a_n\}$ are growing, so the terms of the partial sums $\{S_n\}$ are growing even faster, illustrating that the series diverges.

2. The sequence $\{b_n\}$ starts with 1, −1, 1, −1, ... Consider some of the partial sums S_n of $\{b_n\}$:

$$\begin{aligned} S_1 &= 1 \\ S_2 &= 0 \\ S_3 &= 1 \\ S_4 &= 0 \end{aligned}$$

   This pattern repeats; we find that $S_n = 1$ if n is odd and $S_n = 0$ if n is even. As $\{S_n\}$ oscillates, repeating 1, 0, 1, 0, ..., we conclude that $\lim_{n\to+\infty} S_n$ does not exist, hence the series under study diverges. A scatter plot of the sequence $\{b_n\}$ and the partial sums $\{S_n\}$ is given in Figure 14.5(b). When n is odd, $b_n = S_n$ so the marks for b_n are drawn oversized to show they coincide.

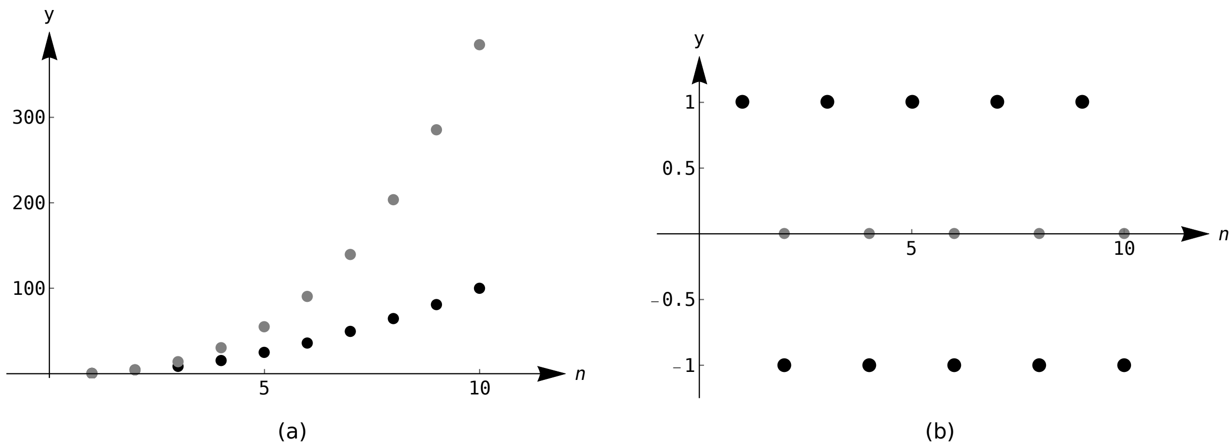


Figure 14.5: Scatter plots of a_n (black) and S_n (grey) relating to Example 14.6.

While it is important to recognize when a series diverges, we are generally more interested in series that converge. In this section we will only demonstrate a few general techniques for determining convergence; later sections will delve deeper into this topic.

14.2.2 Geometric and p-series

One important type of series is a geometric series.

Definitie 14.6 (Geometric series)
A **geometric series** (*meetkundige reeks*) is a series of the form

$$\sum_{n=0}^{+\infty} r^n = 1 + r + r^2 + r^3 + \cdots + r^n + \cdots$$

Note that the index starts at $n = 0$, not $n = 1$.

One reason geometric series are important is that they have nice convergence properties, as stated in the following theorem

Theorem 14.6 (Geometric series test)
Consider the geometric series

$$\sum_{n=0}^{+\infty} r^n.$$

1. *The n^{th} partial sum is:* $S_n = \dfrac{1-r^n}{1-r}$, $r \neq 1$.
2. *The series converges if and only if* $|r| < 1$. *When* $|r| < 1$,

$$\sum_{n=0}^{+\infty} r^n = \frac{1}{1-r}.$$

Proof Let us first of all construct the formula for the n-th partial sum. For $r \neq 1$, the sum of the first n

terms of a geometric series is

$$\begin{aligned} S_n &= 1+r+r^2+r^3+\cdots+r^{n-1}, \\ \Rightarrow \quad rS_n &= r+r^2+r^3+r^4+\cdots+r^n, \\ \Rightarrow \quad S_n - rS_n &= 1-r^n, \\ \Rightarrow \quad S_n(1-r) &= 1-r^n, \end{aligned}$$

so,

$$S_n = \frac{1-r^n}{1-r}$$

if $r \neq 1$.

Clearly, as n goes to infinity, the absolute value of r must be less than one for the series to converge. The sum then becomes

$$1+r+r^2+r^3+r^4+\cdots = \sum_{n=0}^{+\infty} r^n = \frac{a}{1-r},$$

for $|r| < 1$ because $\lim_{n\to+\infty} r^n = 0$ for $|r| < 1$. □

According to Theorem 14.6, the series of the introductory example, i.e.

$$\sum_{n=0}^{+\infty} \frac{1}{2^n} = \sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^n = 1 + \frac{1}{2} + \frac{1}{4} + \cdots$$

converges as $r = 1/2$, and

$$\sum_{n=0}^{+\infty} \frac{1}{2^n} = \frac{1}{1-1/2} = 2.$$

This concurs with our findings in the introductory example; while there we got a sum of 1, we skipped the first term of 1.

Example 14.7

Check the convergence of

$$\sum_{n=2}^{+\infty} \left(\frac{3}{4}\right)^n.$$

If it converges, find its sum.

Solution

Since $r = 3/4 < 1$, this series converges. By Theorem 14.6, we have that

$$\sum_{n=0}^{+\infty} \left(\frac{3}{4}\right)^n = \frac{1}{1-3/4} = 4.$$

However, note the subscript of the summation in the given series: we are to start with $n = 2$. Therefore we subtract the first two terms, giving:

$$\sum_{n=2}^{+\infty} \left(\frac{3}{4}\right)^n = 4 - 1 - \frac{3}{4} = \frac{9}{4}.$$

This is illustrated in Figure 14.6.

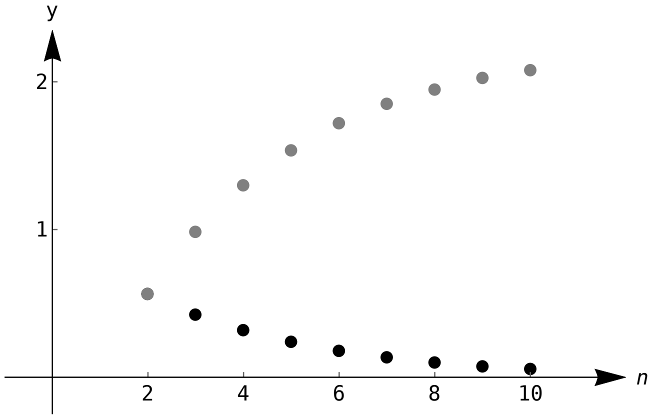


Figure 14.6: Scatter plots of a_n (black) and S_n (gray) relating to the series in Example 14.7.

Another important type of series is the p-series.

Definitie 14.7 (*p*-series)

1. A *p***–series** (*p-reeks*) is a series of the form

$$\sum_{n=1}^{+\infty} \frac{1}{n^p},$$

where $p > 0$.

2. A **general** *p***–series** is a series of the form

$$\sum_{n=1}^{+\infty} \frac{1}{(an+b)^p},$$

where $p > 0$ and a, b are real numbers.

Like geometric series, one of the nice things about p–series is that they have easy to determine convergence properties.

Theorem 14.7 (*p*–series test)

A general p–series

$$\sum_{n=1}^{+\infty} \frac{1}{(an+b)^p}$$

will converge if, and only if, $p > 1$.

Note that Theorem 14.7 assumes that $an + b \neq 0$ for all n. If $an + b = 0$ for some n, then of course the series does not converge regardless of p as not all of the terms of the sequence are defined. This can be proofed using the integral test (Theorem 14.11), which we will introduce in the next section.

Example 14.8

Determine the convergence of the following series.

1. $\sum_{n=1}^{+\infty} \frac{1}{n}$
2. $\sum_{n=1}^{+\infty} \frac{1}{n^2}$
3. $\sum_{n=1}^{+\infty} \frac{1}{\sqrt{n}}$
4. $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n}$

Solution

1. This is a p–series with $p = 1$. By Theorem 14.7, this series diverges.

   This series is a famous series, called the **harmonic series** (*harmonische reeks*), so named because of its relationship to harmonics in the study of music and sound.

2. This is a p–series with $p = 2$. By Theorem 14.7, it converges. Note that the theorem does not give a formula by which we can determine what the series converges to; we just know it converges. A famous, unexpected result is that this series converges to $\pi^2/6$.

3. This is a p–series with $p = 1/2$; the theorem states that it diverges.

4. This is not a p–series, it is a so-called **alternating harmonic series** (*alternerende harmonische reeks*); the definition does not allow for alternating signs. Therefore we cannot apply Theorem 14.7. Another famous result states that it converges to $\ln(2)$.

Later sections will provide tests by which we can determine whether or not a given series converges. This, in general, is much easier than determining what a given series converges to. There are many cases, though, where the sum can be determined.

Example 14.9

Evaluate the sum

$$\sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right).$$

It will help to write down some of the first few partial sums of this series.

$$\begin{aligned}
S_1 &= \frac{1}{1} - \frac{1}{2} && = 1 - \frac{1}{2} \\
S_2 &= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) && = 1 - \frac{1}{3} \\
S_3 &= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) && = 1 - \frac{1}{4} \\
S_4 &= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right) && = 1 - \frac{1}{5}
\end{aligned}$$

Note how most of the terms in each partial sum are cancelled out! In general, we see that $S_n = 1 - \frac{1}{n+1}$. The sequence $\{S_n\}$ converges, as $\lim_{n\to+\infty} S_n = \lim_{n\to+\infty} \left(1 - \frac{1}{n+1}\right) = 1$, and so we conclude that

$$\sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1.$$

Partial sums of the series are plotted in Figure 14.7.

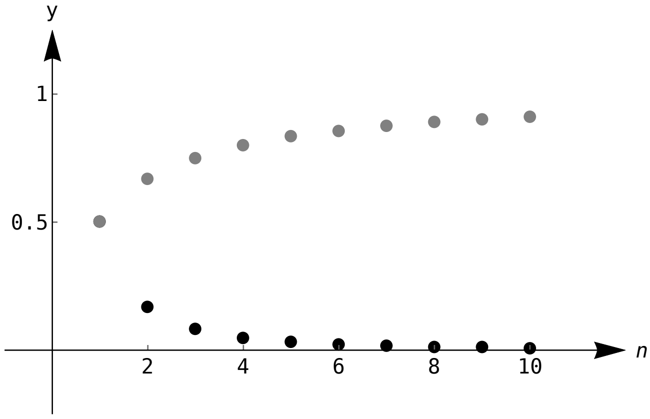


Figure 14.7: Scatter plots of a_n (black) and S_n (gray) relating to the series of Example 14.9.

The series in Example 14.9 is an example of a telescoping series. Informally, a telescoping series is one in which most terms cancel with preceding or following terms, reducing the number of terms in each partial sum. The partial sum S_n did not contain n terms, but rather just two: 1 and $1/(n+1)$.

When possible, seek a way to write an explicit formula for the n^{th} partial sum S_n. This makes evaluating the limit $\lim\limits_{n\to+\infty} S_n$ much more approachable. We do so in the next example.

Example 14.10

Evaluate each of the following infinite series.

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{2}{n^2+2n}$

2. $\displaystyle\sum_{n=1}^{+\infty} \ln\left(\frac{n+1}{n}\right)$

Solution

1. We can decompose the general term as

$$\frac{2}{n^2+2n} = \frac{1}{n} - \frac{1}{n+2}.$$

(See Section 12.4.5)

Expressing the terms of $\{S_n\}$ is now more instructive.

$$\begin{aligned}
S_1 &= 1-\frac{1}{3} && = 1-\frac{1}{3}\\
S_2 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right) && = 1+\frac{1}{2}-\frac{1}{3}-\frac{1}{4}\\
S_3 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{5}\right) && = 1+\frac{1}{2}-\frac{1}{4}-\frac{1}{5}\\
S_4 &= \left(1-\frac{1}{3}\right)+\left(\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{5}\right)+\left(\frac{1}{4}-\frac{1}{6}\right) && = 1+\frac{1}{2}-\frac{1}{5}-\frac{1}{6}
\end{aligned}$$

We again have a telescoping series. In each partial sum, most of the terms cancel and we obtain the formula

$$S_n = 1 + \frac{1}{2} - \frac{1}{n+1} - \frac{1}{n+2}.$$

Taking limits allows us to determine the convergence of the series:

$$\lim_{n\to+\infty} S_n = \lim_{n\to+\infty}\left(1 + \frac{1}{2} - \frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{3}{2},$$

so

$$\sum_{n=1}^{+\infty} \frac{1}{n^2+2n} = \frac{3}{2}.$$

This is illustrated in Figure 14.8(a).

2. We begin by writing the first few partial sums of the series.

$$\begin{aligned}
S_1 &= \ln(2) \\
S_2 &= \ln(2) + \ln\left(\frac{3}{2}\right) \\
S_3 &= \ln(2) + \ln\left(\frac{3}{2}\right) + \ln\left(\frac{4}{3}\right) \\
S_4 &= \ln(2) + \ln\left(\frac{3}{2}\right) + \ln\left(\frac{4}{3}\right) + \ln\left(\frac{5}{4}\right)
\end{aligned}$$

At first, this does not seem helpful, but recall the logarithmic identity: $\ln(x) + \ln(y) = \ln(xy)$. Applying this to S_4 gives:

$$S_4 = \ln(2) + \ln\left(\frac{3}{2}\right) + \ln\left(\frac{4}{3}\right) + \ln\left(\frac{5}{4}\right) = \ln\left(\frac{2}{1}\cdot\frac{3}{2}\cdot\frac{4}{3}\cdot\frac{5}{4}\right) = \ln(5).$$

We can conclude that $\{S_n\} = \{\ln(n+1)\}$. This sequence does not converge, as $\lim\limits_{n\to+\infty} S_n = +\infty$. Therefore, we have that

$$\sum_{n=1}^{+\infty} \ln\left(\frac{n+1}{n}\right) = +\infty,$$

which indicates that the series diverges. Note in Figure 14.8(b) how the sequence of partial sums grows slowly; after 100 terms, it is not yet over 5. Graphically we may be fooled into thinking the series converges, but our analysis above shows that it does not.

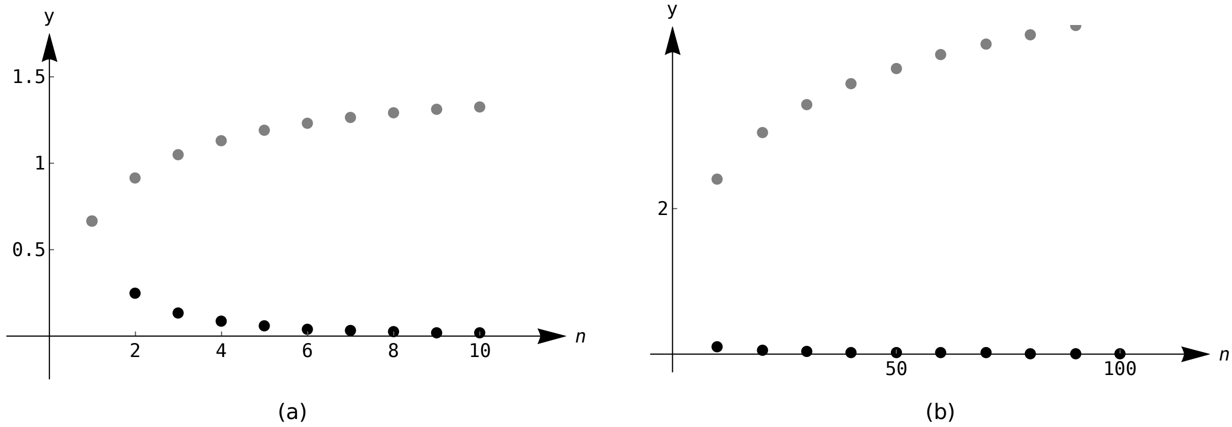


Figure 14.8: Scatter plots of a_n and S_n relating to Example 14.10.

In addition to the geometric, p-, harmonic and alternating harmonic series, there are a few more famous series that we will encounter from time to time later on. They are listed below for comprehenseveness.

$$\sum_{n=0}^{+\infty} \frac{1}{n!} = e \tag{14.1}$$

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{14.2}$$

$$\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12} \tag{14.3}$$

$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4} \tag{14.4}$$

14.2.3 Properties

Let $\sum_{n=1}^{+\infty} a_n = L$, $\sum_{n=1}^{+\infty} b_n = K$, and let c be a constant. Then, the following properties follow easily after some algebra.

1. Constant multiple rule:
$$\sum_{n=1}^{+\infty} c \cdot a_n = c \cdot \sum_{n=1}^{+\infty} a_n = c \cdot L.$$

2. Sum/difference rule:
$$\sum_{n=1}^{+\infty} \left(a_n \pm b_n\right) = \sum_{n=1}^{+\infty} a_n \pm \sum_{n=1}^{+\infty} b_n = L \pm K.$$

Example 14.11

Evaluate the given series.

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}\left(n^2-n\right)}{n^3}$

2. $\displaystyle\sum_{n=1}^{+\infty} \frac{1000}{n!}$

Solution

1. We start by using algebra to break the series apart:

$$\begin{aligned}\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}(n^2-n)}{n^3} &= \sum_{n=1}^{+\infty} \left(\frac{(-1)^{n+1}n^2}{n^3} - \frac{(-1)^{n+1}n}{n^3} \right) \\ &= \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n} - \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n^2} \\ &= \ln(2) - \frac{\pi^2}{12} \approx -0.1293.\end{aligned}$$

   This is illustrated in Figure 14.9(a).

2. This looks very similar to the series that involves e in the list above. Note, however, that the series given in this example starts with $n = 1$ and not $n = 0$. The first term of the series in the list above is $1/0! = 1$, so we will subtract this from our result below:

$$\begin{aligned}\sum_{n=1}^{+\infty} \frac{1000}{n!} &= 1000 \cdot \sum_{n=1}^{+\infty} \frac{1}{n!} \\ &= 1000 \cdot (e-1) \approx 1718.28.\end{aligned}$$

   This is illustrated in Figure 14.9(b). The graph shows how this particular series converges very rapidly.

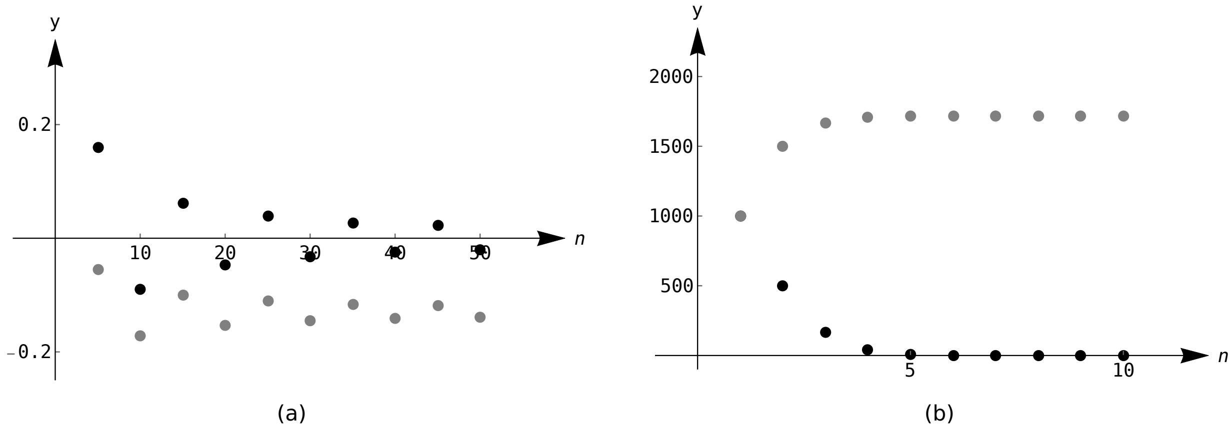


Figure 14.9: Scatter plots of a_n (black) and S_n (gray) relating to Example 14.11.

14.2.4 Divergence test

As one contemplates the behavior of series, a few facts become clear.

1. In order to add an infinite list of nonzero numbers and get a finite result, most of those numbers must be very near 0.
2. If a series diverges, it means that the sum of an infinite list of numbers is not finite (it may approach $\pm\infty$ or it may oscillate), and:

(a) The series will still diverge if the first term is removed.

(b) The series will still diverge if the first 10 terms are removed.

(c) The series will still diverge if the first 1000000 terms are removed.

(d) The series will still diverge if any finite number of terms from anywhere in the series are removed.

These concepts are very important and lie at the heart of the next two theorems.

Theorem 14.8 (n^{th}–term test for divergence)
Consider the series

$$\sum_{n=1}^{+\infty} a_n.$$

If $\lim_{n\to+\infty} a_n \neq 0$ *or if the limit does not exist, then this series diverges.*

Proof The theorem is typically proved in contrapositive form:

$$\sum_{n=1}^{+\infty} a_n$$

converges, then $\lim_{n\to+\infty} a_n = 0$.

If S_n are the partial sums of the series, then the assumption that the series converges means that

$$\lim_{n\to+\infty} S_n = L$$

for some number S. Then

$$\lim_{n\to+\infty} a_n = \lim_{n\to+\infty} (S_n - S_{n-1}) = \lim_{n\to+\infty} S_n - \lim_{n\to+\infty} S_{n-1} = L - L = 0.$$

It is important to underline that this theorem does not state that if $\lim_{n\to+\infty} a_n = 0$ then the corresponding series converges. The standard example of this is the harmonic series. The harmonic sequence, $\{1/n\}$, converges to 0, whereas the harmonic series, $\sum_{n=1}^{+\infty} \frac{1}{n}$, diverges.

Looking back, we can apply this theorem to the series in Example 14.6. In that example, the n^{th} terms of both sequences do not converge to 0, therefore we can quickly conclude that each series diverges.

The following theorem tells us something about what we may expect for what concerns the impact on a series' convergence or divergence when adding finite number of terms or subtracting a finite number of terms.

Theorem 14.9 (Infinite nature of series)
The convergence or divergence of an infinite series remains unchanged by the addition or subtraction of any finite number of terms. That is:

1. *A divergent series will remain divergent with the addition or subtraction of any finite number of terms.*

2. *A convergent series will remain convergent with the addition or subtraction of any finite number of terms.*

Consider once more the harmonic series that diverges; that is, the sequence of partial sums $\{S_n\}$ grows (very, very slowly) without bound. One might think that by removing the large terms of the sequence that perhaps the series will converge. This is simply not the case. For instance, the sum of the first 10 million terms of the harmonic series is about 16.7. Removing the first 10 million terms from the harmonic series changes the n^{th} partial sums, effectively subtracting 16.7 from the sum. However, a sequence that is growing without bound will still grow without bound when 16.7 is subtracted from it.

The equations below illustrate this. The first line shows the infinite sum of the harmonic series split into the sum of the first 10 million terms plus the sum of everything else. The next equation shows us subtracting these first 10 million terms from both sides. The final equation indicates that this still leaves one in the end with infinity.

$$\sum_{n=1}^{+\infty} \frac{1}{n} = \sum_{n=1}^{10000000} \frac{1}{n} + \sum_{n=10000001}^{+\infty} \frac{1}{n}$$

$$\sum_{n=1}^{+\infty} \frac{1}{n} - \sum_{n=1}^{10000000} \frac{1}{n} = \sum_{n=10000001}^{+\infty} \frac{1}{n}$$

$$+\infty - 16.7 = +\infty.$$

This section introduced us to series and defined a few special types of series whose convergence properties are well known: we know when a p-series or a geometric series converges or diverges. Most series that we encounter are not one of these types, but we are still interested in knowing whether or not they converge. The next sections introduce tests that help us determine whether or not a given series converges.

14.3 Convergence tests

14.3.1 Integral and comparison tests

Knowing whether or not a series converges is very important, especially when we discuss power series in Section 14.5. Theorems 14.6 and 14.7 give criteria for when geometric and p-series converge, and Theorem 14.8 gives a quick test to determine if a series diverges. There are many important series whose convergence cannot be determined by these theorems, though, so we introduce a set of tests that allow us to handle a broad range of series. We start with the direct comparison test.

14.3.1.1 Direct comparison test

First n that a sequence $\{a_n\}$ is a positive sequence if $a_n > 0$ for all n.

Theorem 14.10 (Direct comparison test)
Let $\{a_n\}$ and $\{b_n\}$ be positive sequences where $a_n \leq b_n$ for all $n \geq N$, for some $N \geq 1$.

1. *If $\sum_{n=1}^{+\infty} b_n$ converges, then $\sum_{n=1}^{+\infty} a_n$ converges.*

2. *If* $\sum_{n=1}^{+\infty} a_n$ *diverges, then* $\sum_{n=1}^{+\infty} b_n$ *diverges.*

Besides, because of Theorem 14.9, any theorem that relies on a positive sequence still holds true when $a_n > 0$ for all but a finite number of values of n.

Proof We will start off with the partial sums of each series, being

$$S_n = \sum_{i=1}^{n} a_i \qquad \text{and} \qquad T_n = \sum_{i=1}^{n} b_i\,.$$

Since $a_n, b_n \geq 0$ we know that $S_{n+1} \geq S_n$ and $T_{n+1} \geq T_n$. So, both partial sums are increasing sequences. Also, because $a_n \leq b_n$, we know that $S_n \leq T_n$ for all n.

Let us now assume that

$$\sum_{n=1}^{+\infty} b_n$$

converges. Since $b_n \geq 0$, we know that

$$T_n = \sum_{i=1}^{n} b_i \leq \sum_{i=1}^{+\infty} b_i\,.$$

Yet, we also established that $s_n \leq t_n$ for all n, so we have

$$S_n \leq \sum_{i=1}^{n} b_i\,.$$

Finally since $\sum_{n=1}^{+\infty} b_n$ is a convergent series it must have a finite value and so the partial sums S_n, are bounded above. We know that a monotonic and bounded sequence is also convergent (Theorem 14.5), so the sequence $\{S_n\}_{n=1}^{+\infty}$ is a convergent sequence and hence $\sum_{n=1}^{+\infty} a_n$ must converge.

The proof of the other statement in Theorem 14.10 is similar. □

Example 14.12

Determine the convergence of

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{3^n + n^2}$,
2. $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n - \ln(n)}$.

Solution

1. This series is neither a geometric or p-series, but seems related. We predict it will converge, so we look for a series with larger terms that converges.
   Since $3^n < 3^n + n^2$, it holds that
   $$\frac{1}{3^n} > \frac{1}{3^n + n^2},$$
   for all $n \geq 1$. The series $\sum_{n=1}^{+\infty} \frac{1}{3^n}$ is a convergent geometric series; by Theorem 14.10, the considered series converges.

2. We know the harmonic series diverges, and it seems that the given series is closely related to it, hence we predict it will diverge. Since $n \geq n - \ln(n)$ for all $n \geq 1$,
$$\frac{1}{n} \leq \frac{1}{n - \ln(n)},$$
for all $n \geq 1$.

The harmonic series diverges, so we conclude that the studied series diverges as well.

The concept of direct comparison is powerful and often relatively easy to apply. Practice helps one develop the necessary intuition to quickly pick a proper series with which to compare. However, it is easy to construct a series for which it is difficult to apply the direct comparison test.

For instance, consider
$$\sum_{n=1}^{+\infty} \frac{1}{n + \ln(n)}.$$
It is very similar to the divergent series given in Example 14.12. We suspect that it also diverges, as $\frac{1}{n} \approx \frac{1}{n + \ln(n)}$ for large n. However, the inequality that we naturally want to use goes the wrong way: since $n \leq n + \ln(n)$ for all $n \geq 1$, we have that
$$\frac{1}{n} \geq \frac{1}{n + \ln(n)},$$
for all $n \geq 1$. The given series has terms less than the terms of a divergent series, and we cannot conclude anything from this.

Fortunately, we can apply other tests to such problematic series to determine its convergence.

14.3.1.2 Integral test

We stated in Section 14.1 that a sequence $\{a_n\}$ is a function $a(n)$ whose domain is $\mathbb{N}_0$. If we can extend $a(n)$ to $\mathbb{R}$, the real numbers, and it is both positive and decreasing on $[1, +\infty[$, then the convergence of
$$\sum_{n=1}^{+\infty} a_n$$
is the same as
$$\int_1^{+\infty} a(x)\, dx.$$

Theorem 14.11 (Integral test)
Let a sequence $\{a_n\}$ be defined by $a_n = a(n)$, where $a(n)$ is continuous, positive and decreasing on $[1, +\infty[$. Then $\sum_{n=1}^{+\infty} a_n$ converges if and only if $\int_1^{+\infty} a(x)\, dx$ converges.

Note that Theorem 14.11 does not state that the integral and the summation have the same value. Moreover, Theorem 14.9 allows us to extend this theorem to series where $a(n)$ is positive and decreasing on $[b, +\infty[$ for some $b > 1$.

Proof Let's start off the proof of this theorem and estimate the area under the curve on the interval using the right-hand rule with rectangles of width 1 (Figure 14.10). So, we have $f(2) = a_2$, $f(3) = a_3$, and so on. Clearly, we will underestimate the area defined by

$$\int_1^{+\infty} a(x)\, dx.$$

Consequently, it holds for the approximate area that

$$\sum_{i=2}^{n} a_i < \int_1^{n} a(x)\, dx.$$

Now, let us suppose that

$$\int_1^{+\infty} a(x)\, dx$$

is convergent, so it must have a finite value. Also, because $a(x)$ is positive we know that

$$\int_1^{n} a(x)\, dx < \int_1^{+\infty} a(x)\, dx\,,$$

from which we infer that

$$\sum_{i=2}^{n} a_i < \int_1^{n} a(x)\, dx < \int_1^{+\infty} a(x)\, dx.$$

Our series of interest, however, starts at $n = 1$, so we manipulate the last result to get

$$\sum_{i=1}^{n} a_i = a_1 + \sum_{i=2}^{n} a_i < a_1 + \int_1^{+\infty} a(x)\, dx = M\,.$$

This tells us that the sequence of partial sums $\sum_{i=1}^{n} a_n$ is bounded above by M. Moreover, as the terms are positive we also know that,

$$S_n \leq S_{n+1},$$

i.e. the sequence $\{S_n\}_{n=1}^{+\infty}$ is also increasing and bounded above. Consequently, from Theorem 14.5

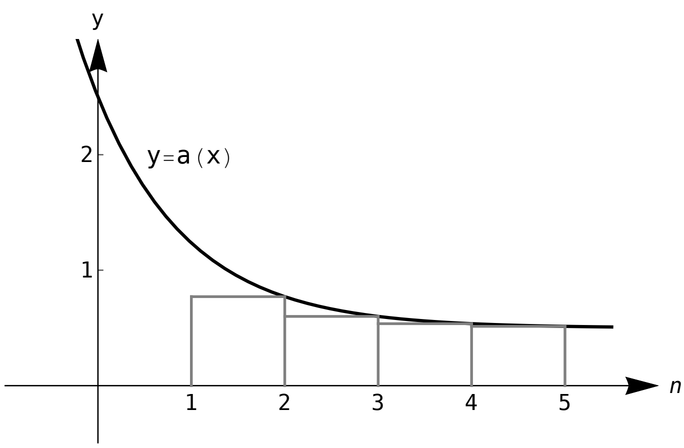


Figure 14.10: Proving the integral test.

this sequence of partial sums $\{S_n\}_{n=1}^{+\infty}$ is convergent, and hence is the considered series. □

Example 14.13

Determine the convergence of

$$\sum_{n=1}^{+\infty} \frac{\ln(n)}{n^2}.$$

The terms of the sequence $\{a_n\} = \{\ln(n)/n^2\}$ and the n[th] partial sums are given in Figure 14.11.

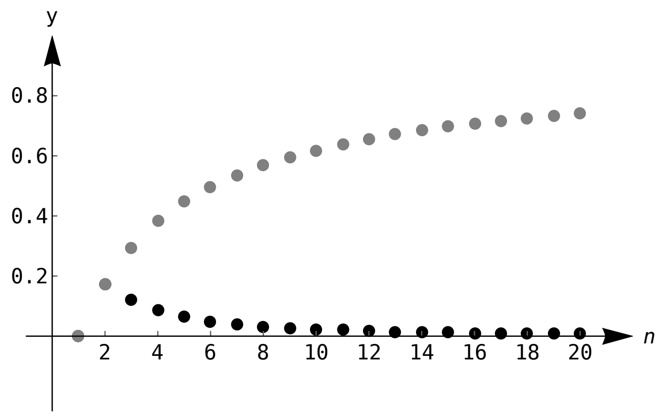


Figure 14.11: Plotting the terms (black) and partial series (gray) in Example 14.13.

Solution

Figure 14.11 implies that $a(n) = \ln(n)/n^2$ is positive and decreasing on $[2, +\infty[$. We can determine this analytically, too. We know $a(n)$ is positive as both $\ln(n)$ and n^2 are positive on $[2, +\infty[$. To determine that $a(n)$ is decreasing, consider the derivative of $a(n)$, i.e. $a'(n) = (1 - 2\ln(n))/n^3$, which is negative for $n \geq 2$. Since $a'(n)$ is negative, $a(n)$ is decreasing.

Applying the integral test, we test the convergence of

$$\int_1^{+\infty} \frac{\ln(x)}{x^2}\, dx.$$

Integrating this improper integral requires the use of integration by parts, with $u = \ln(x)$ and $dv = 1/x^2\, dx$.

$$\begin{aligned}
\int_1^{+\infty} \frac{\ln(x)}{x^2}\, dx &= \lim_{b\to+\infty} \int_1^{b} \frac{\ln(x)}{x^2}\, dx \\
&= \lim_{b\to+\infty} \left[-\frac{1}{x}\ln(x)\Big|_1^b + \int_1^b \frac{1}{x^2}\, dx \right] \\
&= \lim_{b\to+\infty} \left[-\frac{1}{x}\ln(x) - \frac{1}{x} \right]\Bigg|_1^b \\
&= \lim_{b\to+\infty} \left[1 - \frac{1}{b} - \frac{\ln(b)}{b} \right] \qquad \text{(Apply L'Hôpital's rule.)}
\end{aligned}$$

$$= 1.$$

Since this improper integral converges, so does the studied series.

Theorem 14.7 was given without justification, stating that the general p-series

$$\sum_{n=1}^{+\infty} \frac{1}{(an+b)^p}$$

converges if, and only if, $p > 1$. Using the integral test, we can now prove this.

Proof (of Theorem 14.7) For that purpose, consider the following integral; assuming $p \neq 1$,

$$\begin{aligned} \int_1^{+\infty} \frac{1}{(ax+b)^p}\, dx &= \lim_{c \to +\infty} \int_1^c \frac{1}{(ax+b)^p}\, dx \\ &= \lim_{c \to +\infty} \frac{1}{a(1-p)} (ax+b)^{1-p} \Big|_1^c \\ &= \lim_{c \to +\infty} \frac{1}{a(1-p)} \big((ac+b)^{1-p} - (a+b)^{1-p} \big). \end{aligned}$$

This limit converges if, and only if, $p > 1$. It is easy to show that the integral also diverges in the case of $p = 1$.

Therefore the general p-series converges if, and only if, $p > 1$. □

14.3.1.3 Limit comparison test

We study one more convergence test, the so called limit comparison test.

Theorem 14.12 (Limit comparison test)
Let $\{a_n\}$ and $\{b_n\}$ be positive sequences. Then, the following holds.

1. *If* $\lim_{n \to +\infty} \frac{a_n}{b_n} = L$, *where* $L \in \mathbb{R}_0^+$, *then* $\sum_{n=1}^{+\infty} a_n$ *and* $\sum_{n=1}^{+\infty} b_n$ *either both converge or both diverge.*
2. *If* $\lim_{n \to +\infty} \frac{a_n}{b_n} = 0$, *then if* $\sum_{n=1}^{+\infty} b_n$ *converges, so does* $\sum_{n=1}^{+\infty} a_n$.
3. *If* $\lim_{n \to +\infty} \frac{a_n}{b_n} = +\infty$, *then if* $\sum_{n=1}^{+\infty} b_n$ *diverges, so does* $\sum_{n=1}^{+\infty} a_n$.

Theorem 14.12 is most useful when the convergence of the series from $\{b_n\}$ is known and we are trying to determine the convergence of the series from $\{a_n\}$.

Proof We prove the first statement of Theorem 14.12.

Because $\lim_{n \to +\infty} \frac{a_n}{b_n} = L$ we know that for all $\varepsilon > 0$ there is a positive integer n_0 such that for all $n \geq n_0$ we have that $\left| \frac{a_n}{b_n} - L \right| < \varepsilon$, or equivalently

$$-\varepsilon < \frac{a_n}{b_n} - L < \varepsilon.$$

From this, we infer that

$$L - \varepsilon < \frac{a_n}{b_n} < L + \varepsilon,$$

or

$$(L-\varepsilon)b_n < a_n < (L+\varepsilon)b_n\,.$$

As $L > 0$ we can choose ε to be sufficiently small such that $L-\varepsilon$ is positive. So $b_n < \frac{1}{L-\varepsilon}a_n$ and by the direct comparison test, if $\sum_n a_n$ converges then so does $\sum_n b_n$. Similarly $a_n < (L+\varepsilon)b_n$, so if $\sum_n b_n$ converges, again by the direct comparison test, so does $\sum_n a_n$.

That is both series converge or both series diverge. ◻

Example 14.14

Determine the convergence of

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n+\ln(n)}$
2. $\displaystyle\sum_{n=1}^{+\infty} \frac{\sqrt{n}+3}{n^2-n+1}$

Solution

1. We compare the terms of the studied series to the terms of the harmonic sequence:

$$\begin{aligned}\lim_{n\to+\infty} \frac{1/(n+\ln(n))}{1/n} &= \lim_{n\to+\infty} \frac{n}{n+\ln(n)}\\ &= 1 \qquad \text{(after applying L'Hôpital's rule).}\end{aligned}$$

   Since the harmonic series diverges, we conclude that the investigated series diverges as well.

2. We note that the dominant term in the expression of the series is $1/n^2$. Knowing that $\displaystyle\sum_{n=1}^{+\infty}\frac{1}{n^2}$ converges, we attempt to apply the limit comparison test:

$$\begin{aligned}\lim_{n\to+\infty} \frac{(\sqrt{n}+3)/(n^2-n+1)}{1/n^2} &= \lim_{n\to+\infty} \frac{n^2(\sqrt{n}+3)}{n^2-n+1}\\ &= +\infty \qquad \text{(Apply L'Hôpital's rule).}\end{aligned}$$

   Theorem 14.12 part (3) only applies when $\displaystyle\sum_{n=1}^{+\infty} b_n$ diverges; in our case, it converges. Ultimately, our test has not revealed anything about the convergence of our series.

   The problem is that we chose a poor series with which to compare. Since the numerator and denominator of the terms of the series are both algebraic functions, we should have compared our series to the dominant term of the numerator divided by the dominant term of the denominator. The dominant term of the numerator is $n^{1/2}$ and the dominant term of the denominator is n^2. Thus we should compare the terms of the given series to $n^{1/2}/n^2 = 1/n^{3/2}$:

$$\begin{aligned}\lim_{n\to+\infty} \frac{(\sqrt{n}+3)/(n^2-n+1)}{1/n^{3/2}} &= \lim_{n\to+\infty} \frac{n^{3/2}(\sqrt{n}+3)}{n^2-n+1}\\ &= 1 \qquad \text{(Apply L'Hôpital's rule).}\end{aligned}$$

Since the p-series $\sum_{n=1}^{+\infty} \frac{1}{n^{3/2}}$ converges, we conclude that the investigated series converges as well.

The integral test does not work well with series containing factorial terms. The next section introduces the ratio test, which does handle such series well. We also introduce the root test, which is good for series where each term is raised to a power.

14.3.2 Ratio and root tests

This section introduces the ratio and root tests, which determine convergence by analysing the terms of a series to see if they approach 0 fast enough.

14.3.2.1 Ratio test

Theorem 14.13 (Ratio test)
Let $\{a_n\}$ be a positive sequence where $\lim_{n\to+\infty} \frac{a_{n+1}}{a_n} = L$.

1. *If $L < 1$, then $\sum_{n=1}^{+\infty} a_n$ converges.*
2. *If $L > 1$ or $L = +\infty$, then $\sum_{n=1}^{+\infty} a_n$ diverges.*
3. *If $L = 1$, the ratio test is inconclusive.*

Recall that Theorem 14.9 allows us to apply the ratio test to series where $\{a_n\}$ is positive for all but a finite number of terms. The principle of the ratio test is this: if $L < 1$, then for large n, each term of $\{a_n\}$ is significantly smaller than its previous term which is enough to ensure convergence.

Proof Suppose first that

$$L = \lim_{n\to+\infty} \frac{a_{n+1}}{a_n} < 1.$$

Consider r such that $L < r < 1$ and it consequently for sufficiently large n (say N) holds that

$$\frac{a_{n+1}}{a_n} \leq r\,.$$

Equivalently, for $n \geq N$, we have that

$$a_{n+1} \leq r\,a_n\,.$$

This implies that

$$\begin{array}{rrcl} & a_{N+1} & \leq & r\,a_N \\ \Leftrightarrow & a_{N+2} & \leq & r\,a_{N+1} \leq r^2\,a_N \\ \Leftrightarrow & a_{N+3} & \leq & r\,a_{N+2} \leq r^3\,a_N \\ & & \vdots & \\ \Leftrightarrow & a_{N+k} & \leq & r^k\,a_N \end{array}$$

for $k = 0, 1, 2, \ldots$. Now, we can consider the sum of the terms in the left-hand sides of the above inequalities and hence compare

$$\sum_{n=N}^{+\infty} a_n$$

with the geometric series

$$\sum_{k=0}^{+\infty} r^k$$

having $r < 1$, it is clear that the former series must be convergent as the latter is and it constitutes an upper bound on the former. Finally, this leads us to conclude that the series under study

$$\sum_{n=1}^{+\infty} a_n = \sum_{n=1}^{N-1} a_n + \sum_{n=N}^{+\infty} a_n$$

is convergent as well.

Suppose now that

$$L = \lim_{n\to+\infty} \frac{a_{n+1}}{a_n} > 1.$$

Consider r such that $1 < r < L$ and it consequently for sufficiently large n (say N) holds that

$$\frac{a_{n+1}}{a_n} \geq r\,.$$

Is N chosen such that $a_N > 0$, then

$$a_{N+k} \geq r^k a_N$$

for $k = 0, 1, 2, \ldots$. Since $r > 1$, it is immediately clear that $\lim_{n\to+\infty} a_n = +\infty$, from which we may conclude that the series under study is divergent.

Suppose finally that $L = 1$, then we can easily find an example that demonstrates the inconclusiveness of the ratio test for those cases. For instance, the harmonic series is known to be divergent, whereas a p-series with $p = 2$ is convergent (Theorem 14.7). □

Example 14.15

Determine the convergence of the following series:

1. 1. $\displaystyle\sum_{n=1}^{+\infty} \frac{2^n}{n!}$
2. $\displaystyle\sum_{n=1}^{+\infty} \frac{3^n}{n^3}$
3. $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2+1}$.

Solution

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{2^n}{n!}$

$$\begin{aligned}\lim_{n\to+\infty} \frac{2^{n+1}/(n+1)!}{2^n/n!} &= \lim_{n\to+\infty} \frac{2^{n+1}n!}{2^n(n+1)!} \\ &= \lim_{n\to+\infty} \frac{2}{n+1} \\ &= 0.\end{aligned}$$

Since the limit is $0 < 1$, by the ratio test the considered series converges.

2. $\displaystyle\sum_{n=1}^{+\infty} \frac{3^n}{n^3}$

$$\begin{aligned}\lim_{n\to+\infty} \frac{3^{n+1}/(n+1)^3}{3^n/n^3} &= \lim_{n\to+\infty} \frac{3^{n+1}n^3}{3^n(n+1)^3} \\ &= \lim_{n\to+\infty} \frac{3n^3}{(n+1)^3} \\ &= 3.\end{aligned}$$

Since the limit is $3 > 1$, by the ratio test the considered series diverges.

3. $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2+1}$

$$\begin{aligned}\lim_{n\to+\infty} \frac{1/\big((n+1)^2+1\big)}{1/(n^2+1)} &= \lim_{n\to+\infty} \frac{n^2+1}{(n+1)^2+1} \\ &= 1.\end{aligned}$$

Since the limit is 1, the ratio test is inconclusive. We can easily show this series converges using the direct or limit comparison tests, with each comparing to the series

$$\sum_{n=1}^{+\infty} \frac{1}{n^2}.$$

The ratio test is not effective when the terms of a series only contain algebraic functions (e.g. polynomials). It is most effective when the terms contain some factorials or exponentials. The previous example also reinforces our developing intuition: factorials dominate exponentials, which dominate algebraic functions, which dominate logarithmic functions. In Part 1 of the example, the factorial in the denominator dominated the exponential in the numerator, causing the series to converge. In Part 2, the exponential in the numerator dominated the algebraic function in the denominator, causing the series to diverge.

14.3.2.2 Root test

The final test we introduce is the root test, which works particularly well on series where each term is raised to a power, and does not work well with terms containing factorials.

Theorem 14.14 (Root test)
Let $\{a_n\}$ be a positive sequence, and let $\lim_{n\to+\infty} a_n^{1/n} = \lim_{n\to+\infty} \sqrt[n]{a_n} = L$.

1. *If $L < 1$, then $\displaystyle\sum_{n=1}^{+\infty} a_n$ converges.*
2. *If $L > 1$ or $L = +\infty$, then $\displaystyle\sum_{n=1}^{+\infty} a_n$ diverges.*
3. *If $L = 1$, the root test is inconclusive.*

Proof This proof of the convergence is an application of the comparison test. If for all $n \geq N$ (N some fixed natural number) we have $\sqrt[n]{a_n} \leq k < 1$, then $a_n \leq k^n < 1$. Since the geometric series

$$\sum_{n=N}^{+\infty} k^n$$

converges so does

$$\sum_{n=N}^{+\infty} a_n$$

by the comparison test. □

Example 14.16

Determine the convergence of the following series:

1. $\displaystyle\sum_{n=1}^{+\infty} \left(\frac{3n+1}{5n-2}\right)^n$
2. $\displaystyle\sum_{n=1}^{+\infty} \frac{2^n}{n^2}$.

Solution

1. We have that
$$\lim_{n\to+\infty} \left(\left(\frac{3n+1}{5n-2}\right)^n\right)^{1/n} = \lim_{n\to+\infty} \frac{3n+1}{5n-2} = \frac{3}{5}.$$
Since the limit is less than 1, we conclude the series converges.
2. We see that
$$\lim_{n\to+\infty} \left(\frac{2^n}{n^2}\right)^{1/n} = \lim_{n\to+\infty} \frac{2}{\left(n^{1/n}\right)^2} = 2.$$
Since this is greater than 1, we conclude the series diverges.

The next section considers alternating series, where the underlying sequence has terms that alternate between being positive and negative.

14.4 Alternating series

All of the series convergence tests we have used require that the underlying sequence $\{a_n\}$ is a positive sequence. We can relax this with Theorem 14.9 and state that there must be an $N > 0$ such that $a_n > 0$ for all $n > N$; that is, $\{a_n\}$ is positive for all but a finite number of values of n. Here, we explore series whose summation includes negative terms. We start with a very specific form of series, where the terms of the summation alternate between being positive and negative.

Definitie 14.8 (Alternating series)

Let $\{a_n\}$ be a positive sequence. An **alternating series** (*alternerende reeks*) is a series of either the form

$$\sum_{n=1}^{+\infty}(-1)^n a_n \qquad \text{or} \qquad \sum_{n=1}^{+\infty}(-1)^{n+1} a_n.$$

Recall the terms of harmonic series come from the harmonic sequence $\{a_n\} = \{1/n\}$. An important alternating series is the alternating harmonic series:

$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots.$$

Geometric series can also be alternating series when $r < 0$. For instance, if $r = -1/2$, the geometric series is

$$\sum_{n=0}^{+\infty}\left(\frac{-1}{2}\right)^n = 1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \frac{1}{16} - \frac{1}{32} + \cdots.$$

Here, we have $r = -1/2$, so from Theorem 14.6 we get

$$\sum_{n=0}^{+\infty}\left(\frac{-1}{2}\right)^n = \frac{1}{1-(-1/2)} = \frac{1}{3/2} = \frac{2}{3}.$$

A powerful convergence theorem exists for other alternating series that meet a few conditions.

Theorem 14.15 (Alternating series test)
Let $\{a_n\}$ be a monotonically decreasing, positive sequence with a_n where $\lim_{n\to+\infty} a_n = 0$. Then

$$\sum_{n=1}^{+\infty}(-1)^n a_n \qquad \textit{and} \qquad \sum_{n=1}^{+\infty}(-1)^{n+1} a_n$$

converge.

Proof Suppose we are given a series of the form $\sum_{n=1}^{+\infty}(-1)^{n+1}a_n$, where $\lim_{n\to+\infty} a_n = 0$ and $a_n \geq a_{n+1}$ for all natural numbers n.

We will prove that both the partial sums

$$S_{2m+1} = \sum_{n=1}^{2m+1}(-1)^{n-1}a_n$$

with odd number of terms, and

$$S_{2m} = \sum_{n=1}^{2m}(-1)^{n-1}a_n$$

with even number of terms, converge to the same number L. Thus the usual partial sum

$$S_k = \sum_{n=1}^{k}(-1)^{n-1}a_n$$

also converges to L.

We observe that the odd partial sums decrease monotonically:

$$S_{2(m+1)+1} = S_{2m+1} - a_{2m+2} + a_{2m+3} \leq S_{2m+1},$$

while the even partial sums increase monotonically:

$$S_{2(m+1)} = S_{2m} + a_{2m+1} - a_{2m+2} \geq S_{2m} \,,$$

because $a_n \geq a_{n+1}$. Moreover, since a_n are positive, $S_{2m+1} - S_{2m} = a_{2m+1} \geq 0$. Thus we can collect these facts to form the following inequality:

$$a_1 - a_2 = S_2 \leq S_{2m} < S_{2m+1} \leq S_1 = a_1.$$

Now, note that $a_1 - a_2$ is a lower bound of the monotonically decreasing sequence $\left\{S_{2m+1}\right\}_{m=1}^{+\infty}$, the monotone convergence theorem (Theorem 14.5) then implies that this sequence converges as m approaches infinity. Similarly, the sequence of even partial sum converges too.

Finally, they must converge to the same number because

$$\lim_{m\to+\infty}(S_{2m+1} - S_{2m}) = \lim_{m\to+\infty} a_{2m+1} = 0.$$

This is illustrated in Figure 14.12. Since $\left\{S_{2m+1}\right\}_{m=1}^{+\infty}$ and $\{S_{2m}\}_{m=1}^{+\infty}$ converge, also $\{S_m\}_{m=1}^{+\infty}$ converges, and hence the studied alternating sequence also must converge. □

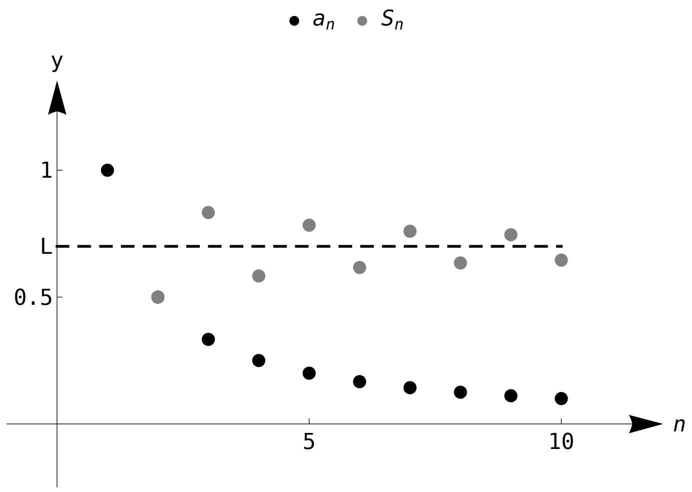


Figure 14.12: Illustrating convergence with the alternating series test.

Example 14.17

Determine if the alternating series test applies to each of the following series.

1. $\displaystyle\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{1}{n}$
2. $\displaystyle\sum_{n=1}^{+\infty}(-1)^{n}\frac{\ln(n)}{n}$
3. $\displaystyle\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{|\sin(n)|}{n^2}$

Solution

1. This is the alternating harmonic series. The underlying sequence is $\{a_n\} = \{1/n\}$, which is positive, decreasing, and approaches 0 as $n \to +\infty$. Therefore we can apply the alternating series test and conclude this series converges.
2. The underlying sequence is $\{a_n\} = \{\ln(n)/n\}$. This is positive and approaches 0 as $n \to +\infty$ (use L'Hôpital's rule). However, the sequence is not decreasing for all n. It is straightforward

to compute $a_1 = 0$, $a_2 \approx 0.347$, $a_3 \approx 0.366$, and $a_4 \approx 0.347$: the sequence is increasing for at least the first 3 terms.

We do not immediately conclude that we cannot apply the alternating series test. Rather, consider the long–term behaviour of $\{a_n\}$. Treating $a_n = a(n)$ as a continuous function of n defined on $[1, +\infty[$, we can take its derivative:

$$a'(n) = \frac{1 - \ln(n)}{n^2}.$$

The derivative is negative for all $n \geq 3$ (actually, for all $n \geq e$), meaning $a(n) = a_n$ is decreasing on $[3, +\infty[$. We can apply the alternating series test to the series when we start with $n = 3$ and conclude that the investigated series converges; adding the terms with $n = 1$ and $n = 2$ do not change the convergence (i.e., we apply Theorem 14.9). The important lesson here is that as before, if a series fails to meet the criteria of the alternating series test on only a finite number of terms, we can still apply the test.

3. The underlying sequence is $\{a_n\} = |\sin(n)|/n^2$. This sequence is positive and approaches 0 as $n \to +\infty$. However, it is not a decreasing sequence; the value of $|\sin(n)|$ oscillates between 0 and 1 as $n \to +\infty$. We cannot remove a finite number of terms to make $\{a_n\}$ decreasing, therefore we cannot apply the alternating series test. Keep in mind that this does not mean we conclude the series diverges; in fact, it does converge. We are just unable to conclude this based on Theorem 14.15.

While there are many factors involved when studying rates of convergence, the alternating structure of an alternating series gives us a powerful tool when approximating the sum of a convergent series.

Theorem 14.16 (The alternating series approximation theorem)
Let $\{a_n\}$ be a sequence that satisfies the hypotheses of the alternating series test, and let S_n and L be the n^{th} partial sums and sum, respectively, of either $\sum_{n=1}^{+\infty}(-1)^n a_n$ or $\sum_{n=1}^{+\infty}(-1)^{n+1} a_n$. Then

1. *$|S_n - L| < a_{n+1}$, and*
2. *L is between S_n and S_{n+1}.*

Proof Essentially, the limit of the converging sequences $\{S_{2m+1}\}_{m=1}^{+\infty}$ and $\{S_{2m}\}_{m=1}^{+\infty}$ in the proof of Theorem 14.15 is the sum L. So, the monotone convergence theorem also tells us that

$$S_{2m} \leq L \leq S_{2m+1} \tag{14.5}$$

for any m. This means the partial sums of an alternating series also alternates above and below the final limit. More precisely, when there are odd (even) number of terms, i.e. the last term is a plus (minus) term, then the partial sum is above (below) the final limit.

Now, we would like to show $|S_n - L| \leq a_{n+1}$ by splitting into two cases.

When $n = 2m + 1$, i.e. odd, then

$$\left|S_{2m+1} - L\right| = S_{2m+1} - L \leq S_{2m+1} - S_{2m+2} = a_{(2m+1)+1}$$

When $n = 2m$, i.e. even, then

$$|S_{2m} - L| = L - S_{2m} \leq S_{2m+1} - S_{2m} = a_{2m+1}$$

as desired. Both statements from Inequality (14.5). □

Part 1 of Theorem 14.16 states that the n^{th} partial sum of a convergent alternating series will be within a_{n+1} of its total sum. Consider the alternating series

$$\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n^2}.$$

Since $a_{14} = 1/14^2 \approx 0.0051$, we know that S_{13} is within 0.0051 of the total sum.

Moreover, Part 2 of the theorem states that since $S_{13} \approx 0.8252$ and $S_{14} \approx 0.8201$, we know the sum L lies between 0.8201 and 0.8252. One use of this is the knowledge that S_{14} is accurate to two places after the decimal.

Some alternating series converge slowly. In Example 14.17 we determined that series

$$\sum_{n=1}^{+\infty} (-1)^{n+1} \frac{\ln(n)}{n}$$

converged. With $n = 1001$, we find $\ln(n)/n \approx 0.0069$, meaning that $S_{1000} \approx 0.1633$ is accurate to one, maybe two, places after the decimal. Since $S_{1001} \approx 0.1564$, we know the sum L is $0.1564 \leq L \leq 0.1633$.

Sometimes, the signs of the terms in a series can have a significant impact on the convergence of a series. This leads us to the following definitions.

Definitie 14.9 (Absolute and conditional convergence)

1. A series $\sum_{n=1}^{+\infty} a_n$ converges **absolutely** (*absoluut*) if $\sum_{n=1}^{+\infty} |a_n|$ converges.
2. A series $\sum_{n=1}^{+\infty} a_n$ converges **conditionally** (*voorwaardelijk*) if $\sum_{n=1}^{+\infty} a_n$ converges but $\sum_{n=1}^{+\infty} |a_n|$ diverges.

Example 14.18

Determine if the following series converge absolutely, conditionally, or diverge.

1. $\sum_{n=1}^{+\infty} (-1)^n \frac{n+3}{n^2+2n+5}$
2. $\sum_{n=1}^{+\infty} (-1)^n \frac{n^2+2n+5}{2^n}$

Solution

1. We can show the series

$$\sum_{n=1}^{+\infty} \left| (-1)^n \frac{n+3}{n^2+2n+5} \right| = \sum_{n=1}^{+\infty} \frac{n+3}{n^2+2n+5}$$

   diverges using the limit comparison test, comparing with $1/n$.

   The investigated series converges using the alternating series test; we conclude it converges conditionally.

2. We can show the series

$$\sum_{n=1}^{+\infty} \left|(-1)^n \frac{n^2+2n+5}{2^n}\right| = \sum_{n=1}^{+\infty} \frac{n^2+2n+5}{2^n}$$

converges using the ratio test.

Therefore we conclude that the studied series converges absolutely.

Knowing that a series converges absolutely allows us to make two important statements, given in the following theorem.

Theorem 14.17 (Absolute convergence theorem)

Let $\sum_{n=1}^{+\infty} a_n$ be a series that converges absolutely.

1. *$\sum_{n=1}^{+\infty} a_n$ converges.*
2. *Let $\{b_n\}$ be any rearrangement of the sequence $\{a_n\}$. Then*

$$\sum_{n=1}^{+\infty} b_n = \sum_{n=1}^{+\infty} a_n.$$

Proof To prove the first statement of this theorem, first notice that $|a_n|$ is either a_n or it is $-a_n$ depending on its sign. This means that we can then say,

$$0 \le a_n + |a_n| \le 2\,|a_n|\ .$$

Now, since we are assuming that $\sum_{n=1}^{+\infty} |a_n|$ is convergent also $\sum_{n=1}^{+\infty} 2\,|a_n|$ is convergent since we can just factor the 2 out of the series and 2 times a finite value will still be finite. This however allows us to use the comparison test to say that

$$\sum_{n=1}^{+\infty} (a_n + |a_n|)$$

is also a convergent series. Finally, we can write

$$\sum_{n=1}^{+\infty} a_n = \sum_{n=1}^{+\infty} (a_n + |a_n|) - \sum_{n=1}^{+\infty} |a_n|\ ,$$

so $\sum_{n=1}^{+\infty} a_n$ is the difference of two convergent series and hence is also convergent. □

The first statement of Theorem 14.17 tells us that absolute convergence is stronger than regular convergence. That is, just because $\sum_{n=1}^{+\infty} a_n$ converges, we cannot conclude that $\sum_{n=1}^{+\infty} |a_n|$ will converge, but knowing a series converges absolutely tells us that $\sum_{n=1}^{+\infty} a_n$ will converge. For instance, in Example 14.18, we determined the series in Part 2 converges absolutely. Theorem 14.17 tells us the series converges

One reason this is important is that our convergence tests all require that the underlying sequence of terms be positive. By taking the absolute value of the terms of a series where not all terms are positive, we are often able to apply an appropriate test and determine absolute convergence. This, in turn, determines that the series we are given also converges.

The second statement relates to **rearrangements** of series. When dealing with a finite set of numbers, the sum of the numbers does not depend on the order which they are added. (So $1+2+3=3+1+2$.) One may be surprised to find out that when dealing with an infinite set of numbers, the same statement does not always hold true: some infinite lists of numbers may be rearranged in different orders to achieve different sums. The theorem states that the terms of an absolutely convergent series can be rearranged in any way without affecting the sum.

This implies that perhaps the sum of a conditionally convergent series can change based on the arrangement of terms. Indeed, it can. The so-called Riemann rearrangement theorem states that any conditionally convergent series can have its terms rearranged so that the sum is any desired value, including $+\infty$.

As an example, consider the alternating harmonic series once more. We have stated that

$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{1}{n}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}\cdots=\ln(2).$$

Consider the rearrangement where every positive term is followed by two negative terms:

$$1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\frac{1}{5}-\frac{1}{10}-\frac{1}{12}\cdots.$$

Now group some terms and simplify:

$$\begin{aligned}\left(1-\frac{1}{2}\right)-\frac{1}{4}+\left(\frac{1}{3}-\frac{1}{6}\right)-\frac{1}{8}+\left(\frac{1}{5}-\frac{1}{10}\right)-\frac{1}{12}+\cdots\\ =\frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{10}-\frac{1}{12}+\cdots\\ =\frac{1}{2}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots\right)=\frac{1}{2}\ln(2).\end{aligned}$$

By rearranging the terms of the series, we have arrived at a different sum! One could try to argue that the Alternating Harmonic Series does not actually converge to $\ln(2)$, because rearranging the terms of the series should not change the sum. However, the Alternating Series Test proves this series converges to L, for some number L, and if the rearrangement does not change the sum, then $L=L/2$, implying $L=0$. But the alternating series approximation theorem quickly shows that $L>0$. The only conclusion is that the rearrangement did change the sum. This is an incredible result.

While series are worthy of study in and of themselves, our ultimate goal within calculus is the study of power series, which we will consider in the next section. We will use power series to create functions where the output is the result of an infinite summation.

14.5 Power series

So far, our study of series has examined the question of “Is the sum of these infinite terms finite?,” i.e., “Does the series converge?” We now approach series from a different perspective: as a function. Given a value of x, we evaluate $f(x)$ by finding the sum of a particular series that depends on x (assuming the series converges). We start this new approach to series with a definition.

Definitie 14.10 (Power series)
Let $\{a_n\}$ be a sequence, let x be a variable, and let x_0 be a real number.

1. The **power series** (*machtreeks*) in x is the series

$$\sum_{n=0}^{+\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots$$

2. The power series in x centred at x_0 is the series

$$\sum_{n=0}^{+\infty} a_n (x - x_0)^n = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \ldots$$

For instance,

$$\sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x+1)^n}{n}$$

is a power series centred at $x_0 = -1$. Note how this series starts with $n = 1$. We could, however, rewrite it starting at $n = 0$ with the understanding that $a_0 = 0$, and hence the first term is 0. Anyhow, its first five terms are given by:

$$\sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x+1)^n}{n} = (x+1) - \frac{(x+1)^2}{2} + \frac{(x+1)^3}{3} - \frac{(x+1)^4}{4} + \frac{(x+1)^5}{5} \ldots$$

Of course, not every series converges. This makes us wonder for what values of x will a given power series converge? Which leads us to a theorem and definition.

Theorem 14.18 (Convergence of power series)
Let a power series $\sum_{n=0}^{+\infty} a_n (x - x_0)^n$ *be given. Then one of the following is true:*

1. *The series converges only at* $x = x_0$.
2. *There is an* $R > 0$ *such that the series converges for all* x *in* $]x_0 - R, x_0 + R[$ *and diverges for all* $x < x_0 - R$ *and* $x > x_0 + R$.
3. *The series converges for all* x.

Note that part 2 of this theorem makes a statement about the interval $]x_0 - R, x_0 + R[$, but not the endpoints of that interval. A series may/may not converge at these endpoints. The value of R is important when understanding a power series, hence it is given a name in the following definition.

Definitie 14.11 (Radius and interval of convergence)

1. The number R given in Theorem 14.18 is the **radius of convergence** (*convergentiestraal*) of a given series. When a series converges for only $x = x_0$, we say the radius of convergence is 0, i.e., $R = 0$. When a series converges for all x, we say the series has an infinite radius of convergence, i.e., $R = +\infty$.
2. The **interval of convergence** (*convergentie-interval*) is the set of all values of x for which the series converges.

The radius of convergence can be found using the following theorem.

Theorem 14.19 (Cauchy-Hadamard theorem)
The radius of convergence R of the power series

$$\sum_{n=0}^{+\infty} a_n(x-x_0)^n$$

is given by

$$R=\frac{1}{\limsup\limits_{n\to+\infty}\sqrt[n]{|a_n|}},$$

where $R=0$ if the lim sup diverges to $+\infty$, and $R=+\infty$ if the lim sup is 0.

Proof This theorem can be proved by considering the number used in the root test:

$$C=\limsup_{n\to+\infty}\sqrt[n]{|a_n(x-x_0)^n|}=\limsup_{n\to+\infty}\sqrt[n]{|a_n|}|x-x_0|.$$

The root test states that the series converges if $C<1$ and diverges if $C>1$. It follows immediately that the power series converges if the distance from x to the centre x_0 is less than

$$R=\frac{1}{\limsup\limits_{n\to+\infty}\sqrt[n]{|a_n|}}$$

and diverges if the distance exceeds that number. □

In practice, however, to find the values of x for which a given series converges, we will use the convergence tests we studied previously. However, the tests all required that the terms of a series be positive. The following theorem gives us a work–around to this problem.

Theorem 14.20 (The radius of convergence of a series and absolute convergence)
The series $\sum\limits_{n=0}^{+\infty} a_n(x-x_0)^n$ *and* $\sum\limits_{n=0}^{+\infty}\left|a_n(x-x_0)^n\right|$ *have the same radius of convergence R.*

Proof This theorem is an immediate consequence of the proof of the Cauchy-Hadamard theorem. □

Theorem 14.20 allows us to find the radius of convergence R of a series by applying the ratio test (or any applicable test) to the absolute value of the terms of the series. We practice this in the following example.

Example 14.19

Find the radius and interval of convergence for each of the following series:

1. $\sum\limits_{n=1}^{+\infty}(-1)^{n+1}\frac{x^n}{n}$
2. $\sum\limits_{n=0}^{+\infty}2^n(x-3)^n$
3. $\sum\limits_{n=0}^{+\infty}n!x^n$

Solution

1. We apply the ratio test to the series
$$\sum_{n=1}^{+\infty}\left|(-1)^{n+1}\frac{x^n}{n}\right| = \sum_{n=1}^{+\infty}\left|\frac{x^n}{n}\right|.$$
We have
$$\begin{aligned}\lim_{n\to+\infty}\frac{\left|x^{n+1}/(n+1)\right|}{\left|x^n/n\right|} &= \lim_{n\to+\infty}\left|\frac{x^{n+1}}{x^n}\cdot\frac{n}{n+1}\right| \\ &= \lim_{n\to+\infty}|x|\frac{n}{n+1} \\ &= |x|.\end{aligned}$$
The ratio test states a series converges if the limit of $|a_{n+1}/a_n| = L < 1$. We found the limit above to be $|x|$; therefore, the studied power series converges when $|x| < 1$, or when x is in $]-1,1[$. Thus the radius of convergence is $R = 1$.

To determine the interval of convergence, we need to check the endpoints of $]-1,1[$. When $x = -1$, we have the opposite of the harmonic series:
$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(-1)^n}{n} = \sum_{n=1}^{+\infty}\frac{-1}{n} = -\infty.$$
This series diverges when $x = -1$.

When $x = 1$, we have the series
$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(1)^n}{n},$$
which is the alternating harmonic series, which converges conditionally. Therefore the interval of convergence is $]-1,1]$.

2. We apply the ratio test to the series
$$\sum_{n=0}^{+\infty}\left|2^n(x-3)^n\right|.$$
We have
$$\begin{aligned}\lim_{n\to+\infty}\frac{\left|2^{n+1}(x-3)^{n+1}\right|}{\left|2^n(x-3)^n\right|} &= \lim_{n\to+\infty}\left|\frac{2^{n+1}}{2^n}\cdot\frac{(x-3)^{n+1}}{(x-3)^n}\right| \\ &= \lim_{n\to+\infty}\left|2(x-3)\right|.\end{aligned}$$
According to the ratio test, the series converges when $\left|2(x-3)\right| < 1$ so when $\left|x-3\right| < 1/2$. The series is centred at 3, and x must be within $1/2$ of 3 in order for the series to converge. Therefore the radius of convergence is $R = 1/2$, and we know that the series converges absolutely for all x in $]3-1/2, 3+1/2[=]2.5, 3.5[$.

We check for convergence at the endpoints to find the interval of convergence. When $x =$

2.5, we have:

$$\sum_{n=0}^{+\infty} 2^n(2.5-3)^n = \sum_{n=0}^{+\infty} 2^n\left(-\frac{1}{2}\right)^n$$
$$= \sum_{n=0}^{+\infty}(-1)^n,$$

which diverges. A similar process shows that the series also diverges at $x = 3.5$. Therefore the interval of convergence is $]2.5, 3.5[$.

3. We apply the ratio test to $\sum_{n=0}^{+\infty} |n!x^n|$.

   We have

$$\lim_{n\to+\infty} \frac{|(n+1)!x^{n+1}|}{|n!x^n|} = \lim_{n\to+\infty} |(n+1)x|$$
$$= +\infty$$

   for all x, except $x = 0$.

   The ratio test shows that the series diverges for all x except $x = 0$. Therefore the radius of convergence is $R = 0$.

We can use a power series to define a function:

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n,$$

where the domain of f is a subset of the interval of convergence of the power series. One can apply calculus techniques to such functions; in particular, we can find derivatives and antiderivatives.

Theorem 14.21 (Derivatives and indefinite integrals of power series)

Let $f(x) = \sum_{n=0}^{+\infty} a_n(x-x_0)^n$ be a function defined by a power series, with radius of convergence R. Then the following hold:

1. *$f(x)$ is continuous and differentiable on $[x_0 - R, x_0 + R]$.*
2. *$f'(x) = \sum_{n=1}^{+\infty} a_n \cdot n \cdot (x-x_0)^{n-1}$, with radius of convergence R.*
3. *$\int f(x)\, dx = x_0 + \sum_{n=0}^{+\infty} a_n \frac{(x-x_0)^{n+1}}{n+1}$, with radius of convergence R.*

Note that this theorem states that differentiation and integration do not change the radius of convergence. It does not state anything about the interval of convergence. They are not always the same. Moreover, differentiation and integration are simply calculated term–by–term using the power rules.

Proof We omit the proof of this theorem due to its technicality, but still its statements are relatively easy to understand by applying differentiation and integration termwise. □

We can learn a great deal from taking derivatives and indefinite integrals of power series functions.

For instance, consider

$$f(x) = \sum_{n=0}^{+\infty} x^n,$$

which is a geometric series. According to Theorem 14.6, this series converges to $1/(1-x)$ when $|x| < 1$. Thus we can say

$$f(x) = \sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}, \quad \text{on} \quad]-1,1[\,.$$

Integrating the power series, we find

$$F(x) = C_1 + \sum_{n=0}^{+\infty} \frac{x^{n+1}}{n+1}, \tag{14.6}$$

while integrating the function $f(x) = 1/(1-x)$ gives

$$F(x) = -\ln|1-x| + C_2. \tag{14.7}$$

Equating Equations (14.6) and (14.7), we have

$$F(x) = C_1 + \sum_{n=0}^{+\infty} \frac{x^{n+1}}{n+1} = -\ln|1-x| + C_2.$$

Letting $x = 0$, we have $F(0) = C_1 = C_2$. This implies that we can drop the constants and conclude

$$\sum_{n=0}^{+\infty} \frac{x^{n+1}}{n+1} = -\ln|1-x|. \tag{14.8}$$

At $x = -1$, we have

$$F(-1) = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots$$

Notice that this series is an alternating series whose terms converge to 0. By the alternating series test, this series converges. In fact, we can recognize that its terms are the opposite of the alternating harmonic series. We can thus say that $F(-1) = -\ln(2)$.

Since, the series on the left of Equation (14.8) converges at $x = -1$; substituting $x = -1$ on both sides of the above equality gives

$$-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots = -\ln(2).$$

We conclude that

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln(2).$$

This shows that the alternating harmonic series converges to $\ln(2)$.

Example 14.20

Let $f(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$. Find $f'(x)$ and $\int f(x)\, dx$, and use these to analyze the behavior of $f(x)$.

Solution

It can be verified easily that the interval of convergence of this power series is $\mathbb{R}$. Besides we will find it useful later to have a few terms of the series written out:

$$\sum_{n=0}^{+\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots \tag{14.9}$$

We now find the derivative:

$$\begin{aligned} f'(x) &= \sum_{n=1}^{+\infty} n \frac{x^{n-1}}{n!} \\ &= \sum_{n=1}^{+\infty} \frac{x^{n-1}}{(n-1)!} = 1 + x + \frac{x^2}{2!} + \cdots . \end{aligned}$$

Since the series starts at $n = 1$ and each term refers to $(n-1)$, we can re-index the series starting with $n = 0$:

$$\begin{aligned} f'(x) &= \sum_{n=0}^{+\infty} \frac{x^n}{n!} \\ &= f(x). \end{aligned}$$

We found the derivative of $f(x)$ is $f(x)$. The only functions for which this is true are of the form $y = Ce^x$ for some constant C. As $f(0) = 1$ (see Equation (14.9)), C must be 1. Therefore we conclude that

$$f(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} = e^x$$

for all x.

We can also find $\int f(x)\ dx$:

$$\begin{aligned} \int f(x)\ dx &= C + \sum_{n=0}^{+\infty} \frac{x^{n+1}}{n!(n+1)} \\ &= C + \sum_{n=0}^{+\infty} \frac{x^{n+1}}{(n+1)!}. \end{aligned}$$

We write out a few terms of this last series:

$$C + \sum_{n=0}^{+\infty} \frac{x^{n+1}}{(n+1)!} = C + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots .$$

The integral of $f(x)$ differs from $f(x)$ only by a constant, again indicating that $f(x) = e^x$.

This example established relationships between a power series function and regular functions that we have dealt with in the past. In general, given a power series function, it is difficult (if not impossible) to express the function in terms of elementary functions. We chose examples where things worked out nicely.

In general , it is far easier to start with a known function, expressed in terms of elementary functions, and represent it as a power series function. One may wonder why we would bother doing so, as the latter function probably seems more complicated. In the next two sections, we show both how to do this and why such a process can be beneficial.

14.6 Taylor polynomials

14.6.1 Definition

Consider a function $y = f(x)$ and a point $\big(c, f(c)\big)$. The derivative, $f'(c)$, gives the instantaneous rate of change of f at $x = c$. Of all lines that pass through the point $\big(c, f(c)\big)$, the line that best approximates f at this point is the tangent line; that is, the line whose slope (rate of change) is $f'(c)$.

In Figure 14.13(a), we see a function $y = f(x)$ graphed, while its derivatives at $x = 0$ are given in table 14.1:

Table 14.1: Derivatives of a function $f(x)$ evaluated at $x = 0$.

$f(0) = 2$	$f'''(0) = -1$
$f'(0) = 1$	$f^{(4)}(0) = -12$
$f''(0) = 2$	$f^{(5)}(0) = -19$

This table shows that $f(0) = 2$ and $f'(0) = 1$; therefore, the tangent line to f at $x = 0$ is $p_1(x) = 1(x-0)+2 = x+2$. The tangent line is also given in the figure. Note that near $x = 0$, $p_1(x) \approx f(x)$; that is, the tangent line approximates f well. One shortcoming of this approximation is that the tangent line only matches the slope of f; it does not, for instance, match the concavity of f. We can find a polynomial, $p_2(x)$, that does match the concavity without much difficulty, though. The table of derivatives gives the following information:

$$f(0) = 2 \qquad f'(0) = 1 \qquad f''(0) = 2.$$

Therefore, we want our polynomial $p_2(x)$ to have these same properties. That is, we need

$$p_2(0) = 2 \qquad p_2'(0) = 1 \qquad p_2''(0) = 2.$$

We can solve this as follows. To keep $p_2(x)$ as simple as possible, we will assume that not only $p_2''(0) = 2$, but that $p_2''(x) = 2$. That is, the second derivative of p_2 is constant. If $p_2''(x) = 2$, then $p_2'(x) = 2x + C$ for some constant C. Since we have determined that $p_2'(0) = 1$, we find that $C = 1$ and so $p_2'(x) = 2x+1$. Finally, we can compute $p_2(x) = x^2 + x + C$. Using our initial values, we know $p_2(0) = 2$ so $C = 2$. We conclude that $p_2(x) = x^2 + x + 2$. This function is plotted with f in Figure 14.13(b).

We can repeat this approximation process by creating polynomials of higher degree that match more of the derivatives of f at $x = 0$. In general, a polynomial of degree n can be created to match the first n derivatives of f. Figure 14.13(b) also shows $p_4(x) = -x^4/2 - x^3/6 + x^2 + x + 2$, whose first four derivatives at 0 match those of f.

As we use more and more derivatives, our polynomial approximation to f gets better and better. In this example, the interval on which the approximation is good gets bigger and bigger. Figure 14.13(c) shows $p_{13}(x)$; we can visually affirm that this polynomial approximates f very well on $[-2, 3]$. Note, however, that the polynomial $p_{13}(x)$ is not particularly nice. It is

$$p_{13}(x) = \frac{16901x^{13}}{6227020800} + \frac{13x^{12}}{1209600} - \frac{1321x^{11}}{39916800} - \frac{779x^{10}}{1814400} - \frac{359x^9}{362880} + \frac{x^8}{240} + \frac{139x^7}{5040} + \frac{11x^6}{360} - \frac{19x^5}{120} - \frac{x^4}{2} - \frac{x^3}{6} + x^2 + x + 2.$$

The polynomials we have created are examples of **Taylor polynomials** (*Taylor-veelterm*), named after the British mathematician Brook Taylor who made important discoveries about such functions. It can

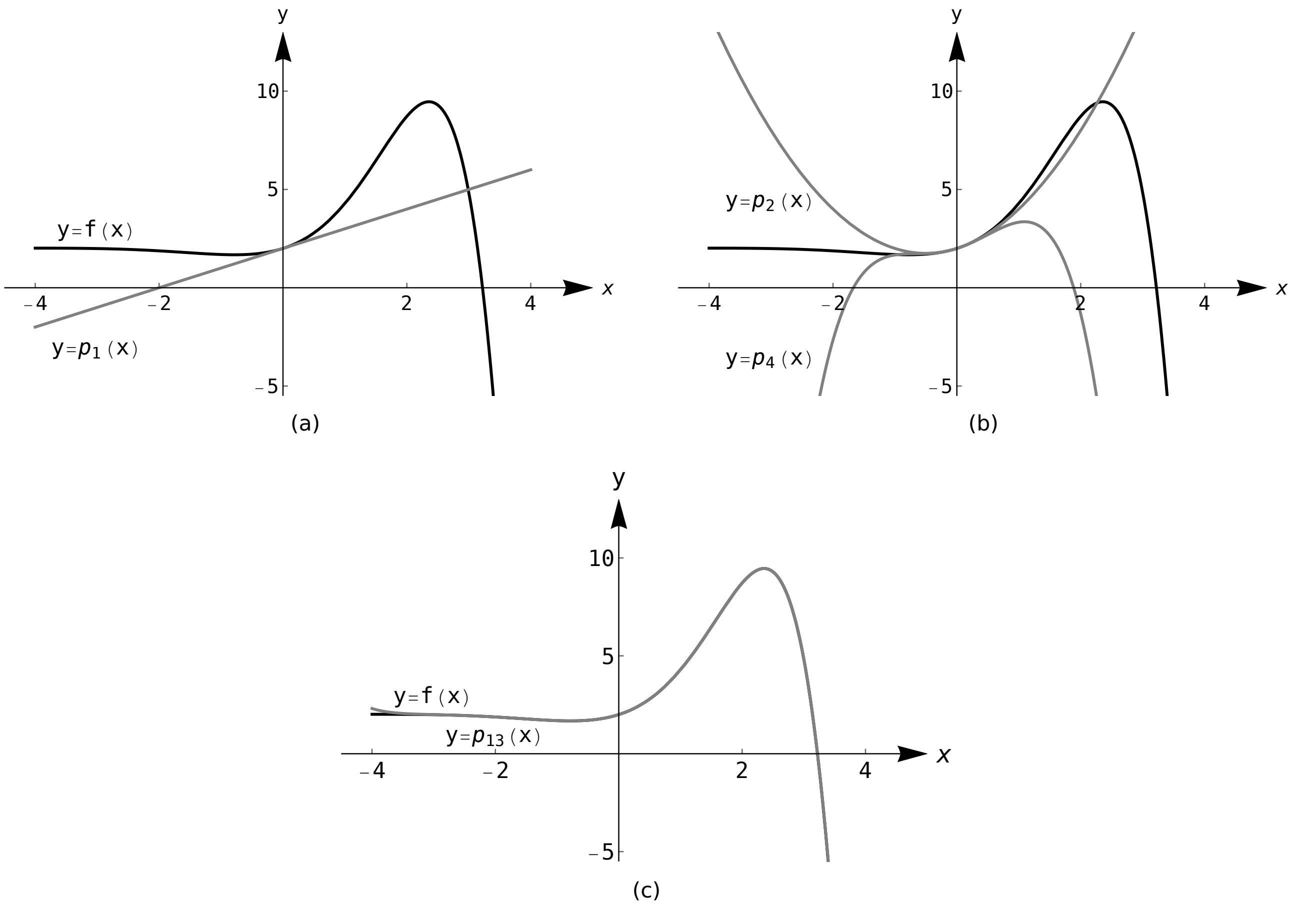


Figure 14.13: Plotting f and the tangent line at $x = 0$ (a) , f, p_2 and p_4 (b), and f and p_{13} (c).

be shown that Taylor polynomials follow a general pattern that make their formation much more direct. This is described in the following definition.

Definitie 14.12 (Taylor and MacLaurin polynomials)
Let f be a function whose first n derivatives exist at $x = x_0$.

1. The **Taylor polynomial of degree n of f at $x = x_0$** is

$$p_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \frac{f'''(x_0)}{3!}(x - x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

2. A special case of the Taylor polynomial is the Maclaurin polynomial, where $x_0 = 0$. That is, the **Maclaurin polynomial of degree n of f** is

$$p_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n.$$

We will practice creating such polynomials in the following example.

Example 14.21

1. Find the n^{th} Taylor polynomial of $y = \ln(x)$ at $x = 1$.
2. Use $p_6(x)$ to approximate the value of $\ln(1.5)$.

3. Use $p_6(x)$ to approximate the value of $\ln(2)$.

Solution

1. We begin by creating a table of derivatives of $\ln(x)$ evaluated at $x = 1$. While this is not as straightforward as it was in the previous example, a pattern does emerge, as shown in Table 14.2.

   Using Definition 14.12, we have

$$\begin{aligned} p_n(x) &= f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n \\ &= 0 + (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \cdots + \frac{(-1)^{n+1}}{n}(x-1)^n. \end{aligned}$$

   Note how the coefficients of the $(x-1)$ terms turn out to be nice.

Table 14.2: Derivatives of $\ln(x)$ evaluated at $x = 1$.

Derivative function	derivative at $x = 1$
$f(x) = \ln(x)$	$f(1) = 0$
$f'(x) = 1/x$	$f'(1) = 1$
$f''(x) = -1/x^2$	$f''(1) = -1$
$f'''(x) = 2/x^3$	$f'''(1) = 2$
$f^{(4)}(x) = -6/x^4$	$f^{(4)}(1) = -6$
$\vdots$	$\vdots$
$f^{(n)}(x) = \dfrac{(-1)^{n+1}(n-1)!}{x^n}$	$f^{(n)}(1) = (-1)^{n+1}(n-1)!$

2. We can compute $p_6(x)$ using our work above:

$$p_6(x) = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \frac{1}{5}(x-1)^5 - \frac{1}{6}(x-1)^6.$$

   Since $p_6(x)$ approximates $\ln(x)$ well near $x = 1$, we approximate $\ln(1.5) \approx p_6(1.5)$:

$$\begin{aligned} p_6(1.5) &= (1.5-1) - \frac{1}{2}(1.5-1)^2 + \frac{1}{3}(1.5-1)^3 - \frac{1}{4}(1.5-1)^4 + \cdots \\ &\quad \cdots + \frac{1}{5}(1.5-1)^5 - \frac{1}{6}(1.5-1)^6 \\ &= \frac{259}{640} \\ &\approx 0.404688. \end{aligned}$$

   This is a good approximation as a calculator shows that $\ln(1.5) \approx 0.4055$. Figure 14.14(a) plots $y = \ln(x)$ with $y = p_6(x)$. We can see that $\ln(1.5) \approx p_6(1.5)$.

3. We approximate $\ln(2)$ with $p_6(2)$:

$$\begin{aligned} p_6(2) &= (2-1) - \frac{1}{2}(2-1)^2 + \frac{1}{3}(2-1)^3 - \frac{1}{4}(2-1)^4 + \cdots \\ &\cdots + \frac{1}{5}(2-1)^5 - \frac{1}{6}(2-1)^6 \\ &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \\ &= \frac{37}{60} \\ &\approx 0.616667. \end{aligned}$$

This approximation is not terribly impressive: a hand held calculator shows that $\ln(2) \approx 0.693147$. The graph in Figure 14.14(a) shows that $p_6(x)$ provides less accurate approximations of $\ln(x)$ as x gets close to 0 or 2.

Surprisingly enough, even the 20$^{\text{th}}$ degree Taylor polynomial fails to approximate $\ln(x)$ for $x > 2$, as shown in Figure 14.14(b). We will soon discuss why this is.

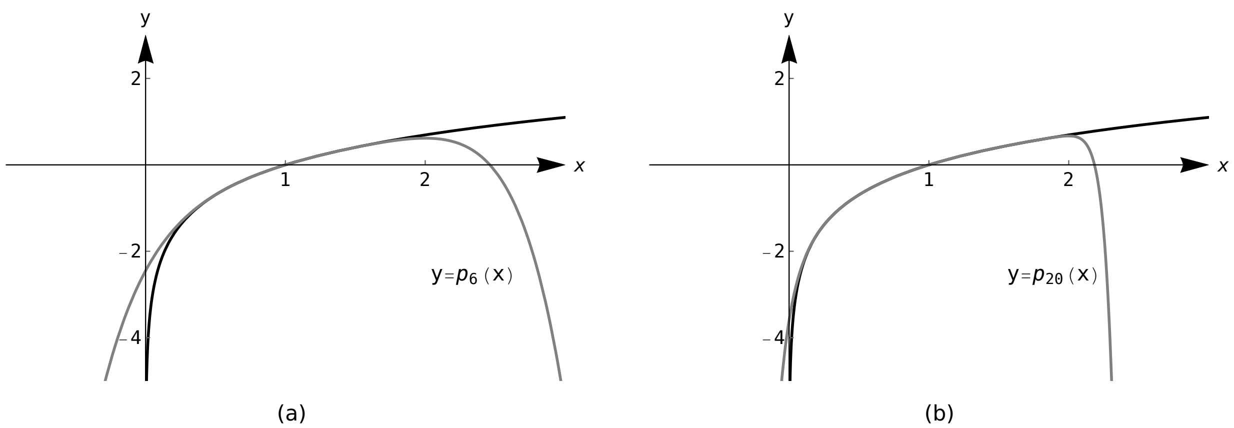


Figure 14.14: A plot of $y = \ln(x)$ and its 6$^{\text{th}}$ (a) and 20$^{\text{th}}$ (b) degree Taylor polynomial at $x = 1$.

14.6.2 Taylor's theorem

Taylor polynomials are used to approximate functions $f(x)$ in mainly two situations:

1. When $f(x)$ is known, but perhaps hard to compute directly. For instance, we can define $y = \cos(x)$ as either the ratio of sides of a right triangle or with the unit circle. However, neither of these provides a convenient way of computing $\cos(2)$. A Taylor polynomial of sufficiently high degree can provide a reasonable method of computing such values using only operations usually hard–wired into a computer ($+$, $-$, $\times$ and $\div$).

2. When $f(x)$ is not known, but information about its derivatives is known. This occurs more often than one might think, especially in the study of differential equations.

In both situations, it is crucial to know how good the approximation is. If we use a Taylor polynomial to compute $\cos(2)$, how do we know how accurate the approximation is?

The following theorem provides this kind of information for Taylor (and hence Maclaurin) polynomials.

Theorem 14.22 (Taylor's theorem)

1. *Let f be a function whose $(n+1)^{th}$ derivative exists on an interval I and let x_0 be in I. Then, for each x in I, there exists θ_x between x and x_0 such that*

$$\begin{aligned} f(x) &= \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i + \frac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-x_0)^{n+1} \\ &= f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_n(x), \end{aligned}$$

where $R_n(x) = \dfrac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-x_0)^{(n+1)}$ *is the remainder term.*

2. $\left|R_n(x)\right| \le \dfrac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(x-x_0)^{(n+1)}\right|$, *where θ is in I.*

Proof The proof of the first part of theorem requires some cleverness to set up, but then the details are quite elementary. We want to define a function $F(t)$. Start with the equation

$$F(t) = \sum_{i=0}^{n} \frac{f^{(i)}(t)}{i!}(x-t)^i + B(x-t)^{n+1}.$$

Here we have replaced x_0 by t in the first $n+1$ terms of the Taylor series, and added a carefully chosen term on the end, with B to be determined. Note that we are temporarily keeping x fixed, so the only variable in this equation is t, and we will be interested only in t between x_0 and x. Now substitute $t = x_0$:

$$F(x_0) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i + B(x-x_0)^{n+1}.$$

Set this equal to $f(x)$:

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i + B(x-x_0)^{n+1}.$$

Since $x \neq x_0$, we can solve this for B, which is a "constant"—it depends on x and x_0 but those are temporarily fixed. Now we have defined a function $F(t)$ with the property that $F(x_0) = f(x)$. Consider also $F(x)$: all terms with a positive power of $(x-t)$ become zero when we substitute x for t, so we are left with $F(x) = f^{(0)}(x)/0! = f(x)$. So $F(t)$ is a function with the same value on the endpoints of the interval $[x_0, x]$. By Rolle's theorem (10.5), we know that there is a value $z \in (x_0, x)$ such that $F'(z) = 0$. Let's look at $F'(t)$. Each term in $F(t)$, except the first term and the extra term involving B, is a product, so to take the derivative we use the product rule on each of these terms. It will help to write out the first few terms of the definition:

$$F(t) = f(t) + \frac{f^{(1)}(t)}{1!}(x-t)^1 + \frac{f^{(2)}(t)}{2!}(x-t)^2 + \frac{f^{(3)}(t)}{3!}(x-t)^3 + \cdots + \frac{f^{(n)}(t)}{n!}(x-t)^n + B(x-t)^{n+1}.$$

Now take the derivative:

$$\begin{aligned} F'(t) = f'(t) &+ \left(\frac{f^{(1)}(t)}{1!}(x-t)^0(-1) + \frac{f^{(2)}(t)}{1!}(x-t)^1\right) \\ &+ \left(\frac{f^{(2)}(t)}{1!}(x-t)^1(-1) + \frac{f^{(3)}(t)}{2!}(x-t)^2\right) \end{aligned}$$

$$+\left(\frac{f^{(3)}(t)}{2!}(x-t)^2(-1)+\frac{f^{(4)}(t)}{3!}(x-t)^3\right)+\cdots+$$
$$+\left(\frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1}(-1)+\frac{f^{(n+1)}(t)}{n!}(x-t)^n\right)+B(n+1)(x-t)^n(-1). \quad \square$$

Now most of the terms in this expression cancel out, leaving just

$$F'(t)=\frac{f^{(n+1)}(t)}{n!}(x-t)^n+B(n+1)(x-t)^n(-1).$$

At some θ_x, $F'(z_x)=0$ so

$$0=\frac{f^{(n+1)}(\theta_x)}{n!}(x-\theta_x)^n+B(n+1)(x-\theta_x)^n(-1)$$

from which

$$B(n+1)(x-\theta_x)^n=\frac{f^{(n+1)}(\theta_x)}{n!}(x-\theta_x)^n$$

and finally

$$B=\frac{f^{(n+1)}(\theta_x)}{(n+1)!}.$$

Now we can write

$$F(t)=\sum_{i=0}^{n}\frac{f^{(i)}(t)}{i!}(x-t)^i+\frac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-t)^{n+1}.$$

Recalling that $F(x_0)=f(x)$ we get

$$f(x)=\sum_{i=0}^{n}\frac{f^{(i)}(x_0)}{i!}(x-x_0)^i+\frac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-x_0)^{n+1},$$

which is what we wanted to show.

For what concerns the second part, it is clear that if you find an upper bound M on the absolute value of the $(n+1)$-st derivative of f between x_0 and x, then the error can be at most

$$\left|R_n(x)\right|\leq\frac{M}{(n+1)!}\left|(x-x_0)^{(n+1)}\right|$$

Basically, the first part of Taylor's theorem states that $f(x)=p_n(x)+R_n(x)$, where $p_n(x)$ is the n^{th} order Taylor polynomial and $R_n(x)$ is the remainder, or error, in the Taylor approximation. The second part gives bounds on how big that error can be. If the $(n+1)^{\text{th}}$ derivative is large on I, the error may be large; if x is far from x_0, the error may also be large. However, the $(n+1)!$ term in the denominator tends to ensure that the error gets smaller as n increases.

The following example computes error estimates for the approximations of $\ln(1.5)$ and $\ln(2)$ made in Example 14.21.

Example 14.22

Find error bounds when approximating 1) $\ln(1.5)$ and 2) $\ln(2)$ with $p_6(x)$, the Taylor polynomial of degree 6 of $f(x)=\ln(x)$ at $x=1$, as calculated in Example 14.21.

Solution

1. We start with the approximation of $\ln(1.5)$ with $p_6(1.5)$. The theorem references an open interval I that contains both x and x_0. The smaller the interval we use the better; it will give us a more accurate (and smaller!) approximation of the error. We let $I = \,]\,0.9, 1.6\,[$, as this interval contains both $c = 1$ and $x = 1.5$.

   The theorem references $\max\left|f^{(n+1)}(\theta)\right|$. In our situation, this is asking the question how big can the 7th derivative of $y = \ln(x)$ be on the interval $]0.9, 1.6[$. The seventh derivative is $y = -6!/x^7$. The largest value it attains on I is about 1506. Thus we can bound the error as:

$$\begin{aligned}\left|R_6(1.5)\right| &\le \frac{\max\left|f^{(7)}(\theta)\right|}{7!}\left|(1.5-1)^7\right|\\ &\le \frac{1506}{5040}\cdot\frac{1}{2^7}\\ &\approx 0.0023.\end{aligned}$$

   We computed $p_6(1.5) = 0.404688$; using a calculator, we find $\ln(1.5) \approx 0.405465$, so the actual error is about 0.000778, which is less than our bound of 0.0023. This affirms Taylor's theorem.

2. We again find an interval I that contains both $c = 1$ and $x = 2$; we choose $I = \,]\,0.9, 2.1\,[$. The maximum value of the seventh derivative of f on this interval is again about 1506 (as the largest values come near $x = 0.9$). Thus

$$\begin{aligned}\left|R_6(2)\right| &\le \frac{\max\left|f^{(7)}(\theta)\right|}{7!}\left|(2-1)^7\right|\\ &\le \frac{1506}{5040}\cdot 1^7\\ &\approx 0.30.\end{aligned}$$

   This bound is not as nearly as good as before. Using the degree 6 Taylor polynomial at $x = 1$ will bring us within 0.3 of the correct answer. As $p_6(2) \approx 0.61667$, our error estimate guarantees that the actual value of $\ln(2)$ is somewhere between 0.31667 and 0.91667. These bounds are not particularly useful.

   In reality, our approximation was only off by about 0.07. However, we are approximating ostensibly because we do not know the real answer. In order to be assured that we have a good approximation, we would have to resort to using a polynomial of higher degree.

We may also use Taylor's theorem to find n that guarantees our approximation is within a certain amount.

Example 14.23

Find n such that the n^{th} Taylor polynomial of $f(x) = \cos(x)$ at $x = 0$ approximates $\cos(2)$ to within 0.001 of the actual answer. What is $p_n(2)$?

Solution

Following Taylor's theorem, we need bounds on the size of the derivatives of $f(x) = \cos(x)$. In the case of this trigonometric function, this is easy. All derivatives of cosine are $\pm\sin(x)$ or $\pm\cos(x)$. In all cases, these functions are never greater than 1 in absolute value. We want the error to be less than 0.001. To find the appropriate n, consider the following inequalities:

$$\frac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(2-0)^{(n+1)}\right| \leq 0.001$$
$$\Leftrightarrow \quad \frac{1}{(n+1)!}\cdot 2^{(n+1)} \leq 0.001$$

We find an n that satisfies this last inequality with trial–and–error. When $n=8$, we have $\frac{2^{8+1}}{(8+1)!} \approx 0.0014$; when $n=9$, we have $\frac{2^{9+1}}{(9+1)!} \approx 0.000282 < 0.001$. Thus we want to approximate $\cos(2)$ with $p_9(2)$.

We now set out to compute $p_9(x)$. We again need a table of the derivatives of $f(x)=\cos(x)$ evaluated at $x=0$ (Table 14.3).

Table 14.3: The derivatives of $f(x)=\cos(x)$ evaluated at $x=0$.

Derivative function	derivative at $x=0$
$f(x)=\cos(x)$	$f(0)=1$
$f'(x)=-\sin(x)$	$f'(0)=0$
$f''(x)=-\cos(x)$	$f''(0)=-1$
$f'''(x)=\sin(x)$	$f'''(0)=0$
$f^{(4)}(x)=\cos(x)$	$f^{(4)}(0)=1$
$f^{(5)}(x)=-\sin(x)$	$f^{(5)}(0)=0$
$f^{(6)}(x)=-\cos(x)$	$f^{(6)}(0)=-1$
$f^{(7)}(x)=\sin(x)$	$f^{(7)}(0)=0$
$f^{(8)}(x)=\cos(x)$	$f^{(8)}(0)=1$
$f^{(9)}(x)=-\sin(x)$	$f^{(9)}(0)=0$

Notice how the derivatives, evaluated at $x=0$, follow a certain pattern. All the odd powers of x in the Taylor polynomial will disappear as their coefficient is 0. While our error bounds state that we need $p_9(x)$, our work shows that this will be the same as $p_8(x)$.

Since we are forming our polynomial at $x=0$, we are creating a Maclaurin polynomial, and:

$$\begin{aligned} p_8(x) &= f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots+\frac{f^{(8)}(0)}{8!}x^8 \\ &= 1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\frac{1}{6!}x^6+\frac{1}{8!}x^8 \end{aligned}$$

We finally approximate $\cos(2)$:

$$\cos(2)\approx p_8(2) = -\frac{131}{315} \approx -0.41587.$$

Our error bound guarantee that this approximation is within 0.001 of the correct answer. Technology shows us that our approximation is actually within about 0.0003 of the correct answer. Figure 14.15 shows a graph of $y = p_8(x)$ and $y = \cos(x)$. Note how well the two functions agree on about $]-\pi, \pi[$.

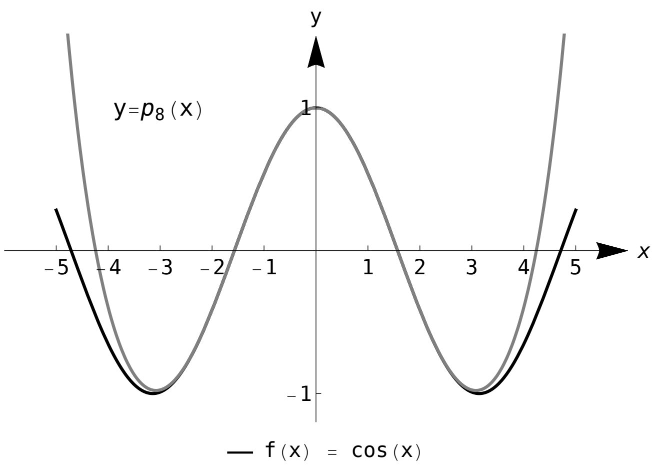


Figure 14.15: A graph of $f(x) = \cos(x)$ (black) and its degree 8 Maclaurin polynomial (gray).

Most of this chapter has been devoted to the study of infinite series. This section has taken a step back from this study, focusing instead on finite summation of terms. In the next section, we explore Taylor series, where we represent a function with an infinite series.

14.7 Taylor series

In Section 14.5, we showed how certain functions can be represented by a power series function. In Section 14.6, we showed how we can approximate functions with polynomials, given that enough derivative information is available. In this section we combine these concepts: if a function $f(x)$ is infinitely differentiable, we show how to represent it with a power series function.

Definitie 14.13 (Taylor and Maclaurin series)
Let $f(x)$ have derivatives of all orders at $x = x_0$.

1. The **Taylor series of** $f(x)$ (*Taylor-reeks*), centred at x_0 is

$$\sum_{n=0}^{+\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

2. Setting $x_0 = 0$ gives the **Maclaurin series of** $f(x)$ (*Maclaurin-reeks*):

$$\sum_{n=0}^{+\infty} \frac{f^{(n)}(0)}{n!}x^n.$$

Note that the order of a Taylor series is determined by the highest-order derivative that appears in it. So the third-order Taylor series expansion of a function f contains terms up to those containing x^3. If $p_n(x)$ is the n^{th} degree Taylor polynomial for $f(x)$ centred at $x = x_0$, we saw how $f(x)$ is approximately equal to $p_n(x)$ near $x = x_0$. We also saw how increasing the degree of the polynomial generally reduced the error. We are now considering series, where we sum an infinite set of terms. Our ultimate hope is to see the error vanish and claim a function is equal to its Taylor series.

When creating the Taylor polynomial of degree n for a function $f(x)$ at $x = x_0$, we needed to evaluate f, and the first n derivatives of f, at $x = x_0$. When creating the Taylor series of f, it helps to find a pattern that describes the n^{th} derivative of f at $x = x_0$. We demonstrate this in the next example.

Example 14.24

Find the Taylor series of $f(x) = \ln(x)$ centred at $x = 1$.

Solution

Table 14.4 shows the n^{th} derivative of $\ln(x)$ evaluated at $x = 1$ for $n = 0, \ldots, 5$, along with an expression for the n^{th} term:

$$f^{(n)}(1) = (-1)^{n+1}(n-1)! \quad \text{for } n \geq 1.$$

Remember that this is what distinguishes Taylor series from Taylor polynomials; we are very interested in finding a pattern for the n^{th} term, not just finding a finite set of coefficients for a polynomial.

Table 14.4: The derivatives of $\ln(x)$ evaluated at $x = 1$.

Derivative function	derivative at $x = 1$
$f(x) = \ln(x)$	$f(1) = 0$
$f'(x) = 1/x$	$f'(1) = 1$
$f''(x) = -1/x^2$	$f''(1) = -1$
$f'''(x) = 2/x^3$	$f'''(1) = 2$
$f^{(4)}(x) = -6/x^4$	$f^{(4)}(1) = -6$
$f^{(5)}(x) = 24/x^5$	$f^{(5)}(1) = 24$
$\vdots$	$\vdots$
$f^{(n)}(x) = \frac{(-1)^{n+1}(n-1)!}{x^n}$	$f^{(n)}(1) = (-1)^{n+1}(n-1)!$

Since $f(1) = \ln(1) = 0$, we skip the first term and start the summation with $n = 1$, giving the Taylor series for $\ln(x)$, centred at $x = 1$, as

$$\sum_{n=1}^{+\infty}(-1)^{n+1}(n-1)!\frac{1}{n!}(x-1)^n = \sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(x-1)^n}{n}.$$

We now determine the interval of convergence, using the ratio test.

$$\begin{aligned}\lim_{n\to+\infty}\left|(-1)^{n+2}\frac{(x-1)^{n+1}}{n+1}\right| \Bigg/ \left|(-1)^{n+1}\frac{(x-1)^n}{n}\right| &= \lim_{n\to+\infty}\left|\frac{(x-1)^{n+1}}{(x-1)^n}\right|\frac{n}{n+1} \\ &= \left|(x-1)\right|.\end{aligned}$$

By the ratio test, we have convergence when $\left|(x-1)\right| < 1$: the radius of convergence is 1, and we have convergence on $]\,0, 2\,[$. We now check the endpoints.

At $x = 0$, the series is

$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(-1)^n}{n} = -\sum_{n=1}^{+\infty}\frac{1}{n},$$

which diverges as it is the harmonic series times (-1).

At $x = 2$, the series is

$$\sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(1)^n}{n} = \sum_{n=1}^{+\infty}(-1)^{n+1}\frac{1}{n},$$

the alternating harmonic series, which converges conditionally.

We have found the Taylor series of $\ln(x)$ centred at $x = 1$, and have determined the series converges on $]0,2]$. We cannot (yet) say that $\ln(x)$ is equal to this Taylor series on $]0,2]$.

Also in Mathematica it is possible to determine the **series expansion** (*reeks-ontwikkeling*) of a function. For instance, to get the Taylor series of $f(x) = \ln(x)$ centred at $x = 1$, we can proceed as follows with the command **Series**.

```
In[23]:= Series[Log[x],{x,1,5}]

Out[23]= (x-1)-1/2 (x-1)^2+1/3 (x-1)^3-1/4 (x-1)^4+1/5 (x-1)^5+O[x-1]^6
```

The general syntax of this command is

```
In[24]:= Series[f[x],{x,x0,n},]
```

where `f[x]` is the function at stake, `x` the variable, `x0` the point at which the series is centred and `n` the order.

It is important to note that Definition 14.13 defines a Taylor series given a function $f(x)$, but makes no claim about their equality. We will find that most of the time they are equal, but we need to consider the conditions that allow us to conclude this.

Theorem 14.22 states that the error between a function $f(x)$ and its n^{th}–degree Taylor polynomial $p_n(x)$ is $R_n(x)$, where

$$|R_n(x)| \leq \frac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(x-x_0)^{(n+1)}\right|.$$

If $R_n(x)$ goes to 0 for each x in an interval I as n approaches infinity, we conclude that the function is equal to its Taylor series expansion. This formalized in the following theorem.

Theorem 14.23 (Function and Taylor series equality)
Let $f(x)$ have derivatives of all orders at $x = x_0$ i.e. $f(x)$ is a smooth function, let $R_n(x)$ be as stated in Theorem 14.22, and let I be an interval on which the Taylor series of $f(x)$ converges. If $\lim_{n\to+\infty} R_n(x) = 0$ for all x in I, then

$$f(x) = \sum_{n=0}^{+\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n \text{ on } I.$$

Proof This theorem is an immediate consequence of Theorem 14.22. □

We demonstrate the use of this theorem in an example.

Example 14.25

Show that $f(x) = \cos(x)$ is equal to its Maclaurin series, given by

$$\sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!},$$

for all x.

Solution

Given a value x, the magnitude of the error term $R_n(x)$ is bounded by

$$\left|R_n(x)\right| \le \frac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|x^{n+1}\right|.$$

Since all derivatives of $\cos x$ are $\pm\sin x$ or $\pm\cos x$, whose magnitudes are bounded by 1, we can state

$$\left|R_n(x)\right| \le \frac{1}{(n+1)!}\left|x^{n+1}\right|$$

which implies

$$-\frac{|x^{n+1}|}{(n+1)!} \le R_n(x) \le \frac{|x^{n+1}|}{(n+1)!}. \tag{14.10}$$

For any x, $\lim\limits_{n\to+\infty} \dfrac{x^{n+1}}{(n+1)!} = 0$. Applying the squeeze theorem to Equation (14.10), we conclude that $\lim\limits_{n\to+\infty} R_n(x) = 0$ for all x, and hence

$$\cos(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!} \quad \text{for all } x.$$

It is natural to assume that a function is equal to its Taylor series on the series' interval of convergence, but this is not always the case. In order to properly establish equality, one must use Theorem 14.23. This is a bit disappointing, as we developed beautiful techniques for determining the interval of convergence of a power series, and proving that $R_n(x) \to 0$ can be difficult. For instance, it is not a simple task to show that $\ln(x)$ equals its Taylor series on $]0,2]$ as found in Example 14.24.

Fourier series

A Fourier series is another kind of series that is used to represent a function as the sum of simple sine waves. Essentially, it decomposes any periodic function into the sum of a (possibly infinite) set of simple oscillating functions, namely sines and cosines. The Fourier series has many applications in electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing, quantum mechanics, econometrics, and so on. Figure 14.16 shows the Fourier series approximation of a square wave using 5 and 15 terms.

Fourier series

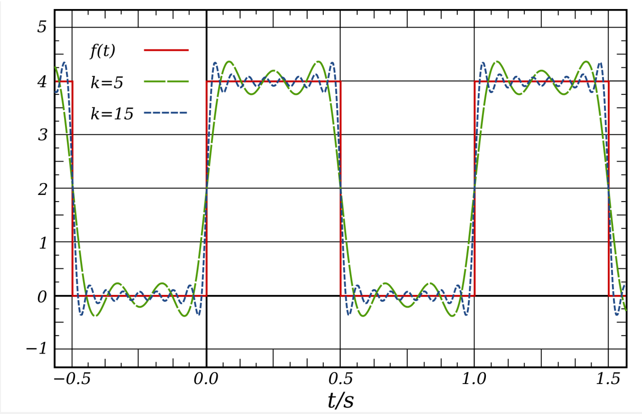


Figure 14.16: Fourier series approximation of a square wave using 5 and 15 terms.

A function $f(x)$ that is equal to its Taylor series, centered at any point of the domain of $f(x)$, is said to be an **analytic function** (*analytische functie*), and most functions that we will encounter are analytic functions. Generally speaking, any function that one creates with elementary functions (polynomials, exponentials, trigonometric functions, etc.) that is not piecewise defined is probably analytic. For most functions, we may assume the function is equal to its Taylor series on the series' interval of convergence and only use Theorem 14.23 when we suspect something may not work as expected.

We develop the Taylor series for one more important function, then give a table of the Taylor series for a number of common functions.

Example 14.26

Find the Maclaurin series of $f(x) = (1+x)^k$ with $k \neq 0$.

Solution

When k is a positive integer, the Maclaurin series is finite. For instance, when $k = 4$, we have

$$f(x) = (1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4.$$

The coefficients of x when k is a positive integer are known as the binomial coefficients, giving the series we are developing its name. When $k = 1/2$, we have $f(x) = \sqrt{1+x}$. Knowing a series representation of this function would give a useful way of approximating $\sqrt{1.3}$, for instance. To develop the Maclaurin series for $f(x) = (1+x)^k$ for any value of $k \neq 0$, we consider the derivatives of f evaluated at $x = 0$ as in Table 14.5

Table 14.5: The derivatives of $f(x) = (1+x)^k$ evaluated at $x = 0$.

Derivative function	derivative at $x = 0$
$f(x) = (1+x)^k$	$f(0) = 1$
$f'(x) = k(1+x)^{k-1}$	$f'(0) = k$
$f''(x) = k(k-1)(1+x)^{k-2}$	$f''(0) = k(k-1)$
$f'''(x) = k(k-1)(k-2)(1+x)^{k-3}$	$f'''(0) = k(k-1)(k-2)$
$\vdots$	$\vdots$
$f^{(n)}(x) = k(k-1)\cdots\big(k-(n-1)\big)(1+x)^{k-n}$	$f^{(n)}(0) = k(k-1)\cdots\big(k-(n-1)\big)$

Thus the Maclaurin series for $f(x) = (1+x)^k$ is

$$1 + kx + \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \ldots + \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!}x^n + \ldots.$$

It is important to determine the interval of convergence of this series. With

$$a_n = \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!}x^n,$$

we apply the ratio test:

$$\begin{aligned}
\lim_{n\to+\infty} \frac{|a_{n+1}|}{|a_n|} &= \lim_{n\to+\infty} \left| \frac{k(k-1)\cdots(k-n)}{(n+1)!}x^{n+1} \right| \Bigg/ \left| \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!}x^n \right| \\
&= \lim_{n\to+\infty} \left| \frac{k-n}{n+1} x \right| \\
&= |x|.
\end{aligned}$$

The series converges absolutely when the limit of the ratio test is less than 1; therefore, we have absolute convergence when $|x| < 1$.

It can be verified that the interval of convergence depends on the value of k. When $k > 0$, the interval of convergence is $[-1,1]$. When $-1 < k < 0$, the interval of convergence is $]-1,1]$. If $k \leq -1$, the interval of convergence is $]-1,1[$.

We learned that Taylor polynomials offer a way of approximating a difficult to compute function with a polynomial. Taylor series offer a way of exactly representing a function with a series. One probably can see the use of a good approximation; is there any use of representing a function exactly as a series? Yes, amongst other things they provide a valuable tool for solving a variety of problems, including problems relating to integration and differential equations.

In Table 14.6 we give the Taylor series of a number of common functions.

We also give a theorem about the algebra of power series, that is, how we can combine power series to create power series of new functions. This allows us to find the Taylor series of functions like $f(x) = e^x \cos(x)$ by knowing the Taylor series of e^x and $\cos(x)$.

Theorem 14.24 (Algebra of power series)

Let $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ and $g(x) = \sum_{n=0}^{+\infty} b_n x^n$ converge absolutely for $|x| < R$, and let $h(x)$ be continuous.

1. $f(x) \pm g(x) = \sum_{n=0}^{+\infty} (a_n \pm b_n)x^n$ *for* $|x| < R$.
2. $f(x)g(x) = \left(\sum_{n=0}^{+\infty} a_n x^n\right)\left(\sum_{n=0}^{+\infty} b_n x^n\right) = \sum_{n=0}^{+\infty} (a_0 b_n + a_1 b_{n-1} + \ldots a_n b_0)x^n$ *for* $|x| < R$.
3. $f\big(h(x)\big) = \sum_{n=0}^{+\infty} a_n \big(h(x)\big)^n$ *for* $|h(x)| < R$.

Proof We again omit the proof of this theorem because it is rather technical and tedious. Still it is

Table 14.6: Important Taylor series expansions.

Function and Series	First few terms	Interval of convergence
$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$	$\mathbb{R}$
$\sin(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$	$\mathbb{R}$
$\cos(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!}$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$	$\mathbb{R}$
$\ln(x) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x-1)^n}{n}$	$(x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$	$]0,2]$
$\frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n$	$1 + x + x^2 + x^3 + \cdots$	$[-1,1[$
$(1+x)^k = \sum_{n=0}^{+\infty} \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!} x^n$	$1 + kx + \frac{k(k-1)}{2!} x^2 + \cdots$	$]-1,1[$[a]
$\arctan(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$	$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$	$[-1,1]$

[a] Convergence at $x = \pm 1$ depends on the value of k.

clear that the first two statements follow from the termwise summation and multiplication, while the third statement is logic consequence upon introducing an auxiliary variable $u = h(x)$. □

Example 14.27

Create a series expansion for $y = \sin(x^2)$ and $y = \ln(\sqrt{x})$.

Solution

Given that

$$\sin(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots,$$

we simply substitute x^2 for x in the series, giving

$$\sin(x^2) = \sum_{n=0}^{+\infty} (-1)^n \frac{(x^2)^{2n+1}}{(2n+1)!} = x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \cdots.$$

Since the Taylor series for $\sin(x)$ has an infinite radius of convergence, so does the Taylor series for $\sin(x^2)$.

The Taylor expansion for $\ln(x)$ given in Table 14.6 is centred at $x = 1$, so we will centre the series for $\ln(\sqrt{x})$ at $x = 1$ as well. With

$$\ln(x) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x-1)^n}{n} = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots,$$

we substitute $\sqrt{x}$ for x to obtain

$$\ln(\sqrt{x}) = \sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(\sqrt{x}-1)^n}{n} = (\sqrt{x}-1) - \frac{(\sqrt{x}-1)^2}{2} + \frac{(\sqrt{x}-1)^3}{3} - \cdots.$$

While this is not strictly a power series, it is a series that allows us to study the function $\ln(\sqrt{x})$. Since the interval of convergence of $\ln(x)$ is $]0,2]$, and the range of $\sqrt{x}$ on $]0,4]$ is $]0,2]$, the interval of convergence of this series expansion of $\ln(\sqrt{x})$ is $]0,4]$.

Taylor series can as well be used to evaluate limits and definite integrals, as illustrated in the following example

Example 14.28

Use the Taylor series of e^{-x^2} to evaluate $\displaystyle\int_0^1 e^{-x^2}\,dx$.

Solution

It can be verified that e^{-x^2} does not have an antiderivative expressible in terms of elementary functions. This means any definite integral of this function must have its value approximated, and not computed exactly.

We can quickly write out the Taylor series for e^{-x^2} using the Taylor series of e^x:

$$e^x = \sum_{n=0}^{+\infty}\frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

and so

$$\begin{aligned} e^{-x^2} &= \sum_{n=0}^{+\infty}\frac{(-x^2)^n}{n!} \\ &= \sum_{n=0}^{+\infty}(-1)^n\frac{x^{2n}}{n!} \\ &= 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \cdots. \end{aligned}$$

We use Theorem 14.21 to integrate:

$$\int e^{-x^2}\,dx = C + x - \frac{x^3}{3} + \frac{x^5}{5\cdot 2!} - \frac{x^7}{7\cdot 3!} + \cdots + (-1)^n\frac{x^{2n+1}}{(2n+1)n!} + \cdots$$

This is the antiderivative of e^{-x^2}; while we can write it out as a series, we cannot write it out in terms of elementary functions. We can evaluate the definite integral at stake using this antiderivative; substituting 1 and 0 for x and subtracting gives

$$\int_0^1 e^{-x^2}\,dx = 1 - \frac{1}{3} + \frac{1}{5\cdot 2!} - \frac{1}{7\cdot 3!} + \frac{1}{9\cdot 4!}\cdots.$$

Summing the 5 terms shown above give the approximation of 0.74749. Since this is an alternating series, we can use the alternating series approximation theorem, (Theorem 14.16), to determine how accurate this approximation is. The next term of the series is $1/(11\cdot 5!) \approx 0.00075758$. Thus we know our approximation is within 0.00075758 of the actual value of the integral. This is arguably much less work than using Simpson's Rule to approximate the value of the integral.

This chapter introduced sequences, which are ordered lists of numbers, followed by series, wherein we add up the terms of a sequence. We quickly saw that such sums do not always add up to infinity, but rather converge. We studied tests for convergence, then ended the chapter with a formal way of defining functions based on series. Such series–defined functions are a valuable tool in solving a number of different problems throughout science and engineering.

14.8 Exercices

Sequences

Assignment 14.1 — Examine whether the given sequences are (a) bounded (above or below), (b) positive or negative, (c) increasing, decreasing or alternating, (d) convergent or divergent.

(a) $\left\{\dfrac{2n^2}{n^2+1}\right\}$

(b) $\left\{\dfrac{(-1)^n\, n}{e^n}\right\}$

(c) $\left\{\dfrac{e^n}{\pi^{n/2}}\right\}$

(d) $\left\{n\cos\left(\dfrac{n\pi}{2}\right)\right\}$

(e) $\left\{\dfrac{(n!)^2}{(2n)!}\right\}$

Assignment 14.2 — Find the limit of the sequence $\{a_n\}$ and investigate its convergence.

(a) $a_n = \dfrac{e^n - e^{-n}}{e^n + e^{-n}}$

(b) $a_n = \left(\dfrac{n-3}{n}\right)^n$

(c) $a_n = \left(\dfrac{n-1}{n+1}\right)^n$

(d) $a_n = \dfrac{n}{\ln(n+1)}$

(e) $a_n = n - \sqrt{n^2 - 4n}$

(f) $a_n = \dfrac{\pi^n}{1+2^{2n}}$

Assignment 14.3 — Examine the convergence of the sequences below.

(a) $\arctan(1), \arctan\left(\dfrac{4}{3}\right), \ldots, \arctan\left(\dfrac{2n}{n+1}\right), \ldots$

(b) $\sin\left(\dfrac{\pi}{3}\right), \sin\left(\dfrac{2\pi}{3}\right), \ldots, \sin\left(\dfrac{n\pi}{3}\right), \ldots$

(c) $2, 2, \dfrac{4}{3}, \ldots, \dfrac{2^n}{n!}, \ldots$

Assignment 14.4 — Consider the following recursively defined sequences. Show that the sequences are increasing and bounded from above. Examine their convergence and find their limit (in case of convergence).

(a) $a_1 = 1$ and $a_{n+1} = \sqrt{1+2a_n}, \quad n = 1, 2, 3, \ldots$

Hint: Prove that 3 is an upper bound.

(b) $a_1 = 3$ and $a_{n+1} = \sqrt{15+2a_n}, \quad n = 1, 2, 3, \ldots$

Hint: Prove that 5 is an upper bound.

Infinite series and Convergence tests

Assignment 14.5 — Examine the convergence of the series below.

(a) $\displaystyle\sum_{n=1}^{+\infty} n\sin\left(\frac{\alpha}{n}\right)$

(b) $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{\sqrt{n+1}-\sqrt{n}}$

(c) $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2}\sin\left(\frac{\pi}{n}\right)$

(d) $\displaystyle\sum_{n=1}^{+\infty} \frac{n^4}{4^n}$

(e) $\displaystyle\sum_{n=2}^{+\infty} \frac{\ln(n)}{n}$

(f) $\displaystyle\sum_{n=1}^{+\infty} \frac{3^n}{n^2 2^{n+1}}$

(g) $\displaystyle\sum_{n=2}^{+\infty} \frac{1}{n\ln^2(n)}$

(h) $\displaystyle\sum_{n=0}^{+\infty} \frac{1}{(2n+1)2^{2n+1}}$

(i) $\displaystyle\sum_{n=1}^{+\infty} \left|\sin\left(\frac{1}{n^2}\right)\right|$

(j) $\displaystyle\sum_{n=2}^{+\infty} \frac{1}{\ln^3(n)}$

(k) $\displaystyle\sum_{n=1}^{+\infty} \frac{n!}{n^2 e^n}$

(l) $\displaystyle\sum_{n=1}^{+\infty} \frac{(2n)!}{(n!)^3}$

(m) $\displaystyle\sum_{n=1}^{+\infty} \left(\frac{n^2-n}{n^2+n}\right)^n$

(n) $\displaystyle\sum_{n=1}^{+\infty} \frac{\sqrt{n}}{n^2+n+1}$

(o) $\displaystyle\sum_{n=3}^{+\infty} \frac{1}{n\ln(n)\sqrt{\ln(\ln(n))}}$

(p) $\displaystyle\sum_{n=2}^{+\infty} \frac{\sqrt{n}}{3^n\ln(n)}$

(q) $\displaystyle\sum_{n=1}^{+\infty} \frac{\ln(10+n)}{n}$

(r) $\displaystyle\sum_{n=1}^{+\infty} \left(\frac{n}{n+1}\right)^{n^2}$

(s) $\displaystyle\sum_{n=1}^{+\infty} \frac{n}{\sqrt{n^2+4n+1}}$

Alternating series

Assignment 14.6 — Examine the convergence of the alternating series below.

(a) $\displaystyle\sum_{n=1}^{+\infty} \frac{(-1)^n}{n^2+\ln(n)}$

(b) $\displaystyle\sum_{n=1}^{+\infty} \frac{\cos(n\pi)}{(n+1)\ln(n+1)}$

(c) $\displaystyle\sum_{n=1}^{+\infty} \frac{(-1)^{2n}}{2^n}$

(d) $\displaystyle\sum_{n=1}^{+\infty} \frac{(-2)^n}{n!}$

(e) $\displaystyle\sum_{n=0}^{+\infty} \frac{-n}{n^2+1}$

(f) $\displaystyle\sum_{n=10}^{+\infty} \frac{\sin\big((n+1/2)\pi\big)}{\ln\big(\ln(n)\big)}$

Power series

Assignment 14.7 — Determine the region of convergence of the following power series.

(a) $\displaystyle\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{5^n}$

(b) $\displaystyle\sum_{n=1}^{+\infty} \frac{n!}{x^n}$

(c) $\displaystyle\sum_{n=1}^{+\infty} \frac{e^n x^n}{n!}$

(d) $\displaystyle\sum_{n=1}^{+\infty} \frac{(x-1)^n}{n^2 2^n}$

(e) $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n}\left(\frac{x+2}{2}\right)^n$

(f) $\displaystyle\sum_{n=1}^{+\infty} \frac{(x-2)^n}{2n^2\, 2^n}$

(g) $\displaystyle\sum_{n=2}^{+\infty} \frac{\ln(n+1)}{3n}(x+1)^n$

(h) $\displaystyle\sum_{n=1}^{+\infty} \frac{\arctan(n)}{(n+1)^2}(x-3)^n$

(i) $\displaystyle\sum_{n=2}^{+\infty} \frac{(x-2)^n}{\sqrt[3]{n^2-1}\; 5^{n-2}}$

Taylor polynomials and Taylor series

Assignment 14.8 — Determine an appropriate power series for the functions below and give the corresponding interval of convergence.

(a) $f(x) = \dfrac{1}{(2-x)^2}$ in powers of x

(b) $f(x) = \ln(2-x)$ in powers of x

(c) $f(x) = \dfrac{x^3}{1-2x^2}$ in powers of x

(d) $f(x) = \ln(x)$ in powers of $x-4$

(e) $f(x) = \dfrac{1-x}{1+x}$ in powers of x

Assignment 14.9 — Find the MacLaurin series of the functions below. Also, give the interval of convergence.

(a) $f(x) = \cos^2\left(\dfrac{x}{2}\right)$

(b) $f(x) = \dfrac{e^{2x^2}-1}{x^2}$

(c) $f(x) = \sinh(x) - \sin(x)$

(d) $f(x) = \cosh(x) - \cos(x)$

(e) $f(x) = x^2 \sin\left(\dfrac{x}{3}\right)$

(f) $f(x) = (1+x)^{\frac{1}{2}} \cos(x)$

Assignment 14.10 — Find the Taylor series of the functions below. Also, give the interval of convergence.

(a) $f(x) = \sin(x) - \cos(x)$ around $\dfrac{\pi}{4}$

(b) $f(x) = x\ln(x)$ in powers of $x-1$

(c) $f(x) = xe^x$ in powers of $x+2$

(d) $f(x) = \ln(2+x)$ in powers of $x-2$

(e) $f(x) = \cos^2(x)$ at $\dfrac{\pi}{8}$

(f) $f(x) = \dfrac{1}{x^2}$ in powers of $x+2$

(g) $f(x) = \dfrac{1}{x}$ at 1

Assignment 14.11 —

(a) Find a MacLaurin series in powers of x for the function $\ln(1+x)$.

(b) Find $\ln(2)$ using the established series from (a) and observe that there is slow convergence.

(c) Find a MacLaurin series for $\ln\left(\dfrac{1+x}{1-x}\right)$.

(d) Find $\ln(2)$ using the established series for $x = \dfrac{1}{3}$ and observe that this one converges much faster.

Assignment 14.12 — Find the MacLaurin series of the functions below.

(a) $\displaystyle\int_0^{\sqrt{\pi}} \sin\left(x^2\right)\, dx$

(b) $\displaystyle\int_0^{\pi^2/4} \cos\left(\sqrt{x}\right)\, dx$

Assignment 14.13 — The so-called error function is used, amongst other things, to describe groundwater flow and is given by

$$f(x) = \int_0^x e^{-u^2}\, du\,.$$

Yet, the integral cannot be evaluated analytically and numerical integration is not evident here either. The function can be approximated using a MacLaurin series. Determine the MacLaurin series of this function up to fourth order terms.

Mathematics is the queen of science

— Carl Friedrich Gauss —

15

Vector–valued functions

In Chapter 6, we learned about vectors and we were introduced to the power of vectors within mathematics. In this chapter, we will build on this foundation to define functions whose input is a real number and whose output is a vector.

15.1 Vector–valued functions

15.1.1 Definition

We are very familiar with real–valued functions, that is, functions whose output is a real number. This section introduces vector–valued functions – functions whose output is a vector.

Definitie 15.1 (Vector–valued functions)
A n-dimensional **vector–valued function** (*vectorfunctie*) is a function of the form

$$\vec{\boldsymbol{r}}(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big),$$

where the f_i are real-valued functions, and are called the **component functions**.

The domain of $\vec{\boldsymbol{r}}$ is the set of all values of t for which $\vec{\boldsymbol{r}}(t)$ is defined. The range of $\vec{\boldsymbol{r}}$ is the set of all possible output vectors $\vec{\boldsymbol{r}}(t)$.

Evaluating a vector–valued function at a specific value of t is straightforward; simply evaluate each component function at that value of t. For instance, if $\vec{\boldsymbol{r}}(t) = \left(t^2, t^2 + t - 1\right)$, then $\vec{\boldsymbol{r}}(-2) = (4, 1)$. We can sketch this vector, as is done in Figure 15.1(a). Plotting lots of vectors is cumbersome, though, so generally we do not sketch the whole vector but just the terminal point. The graph of a vector–valued function is the set of all terminal points of $\vec{r}(t)$, where the initial point of each vector is always the origin. In Figure 15.1(b) we sketch the graph of $\vec{r}$; we can indicate individual points on the graph with their respective vector, as shown.

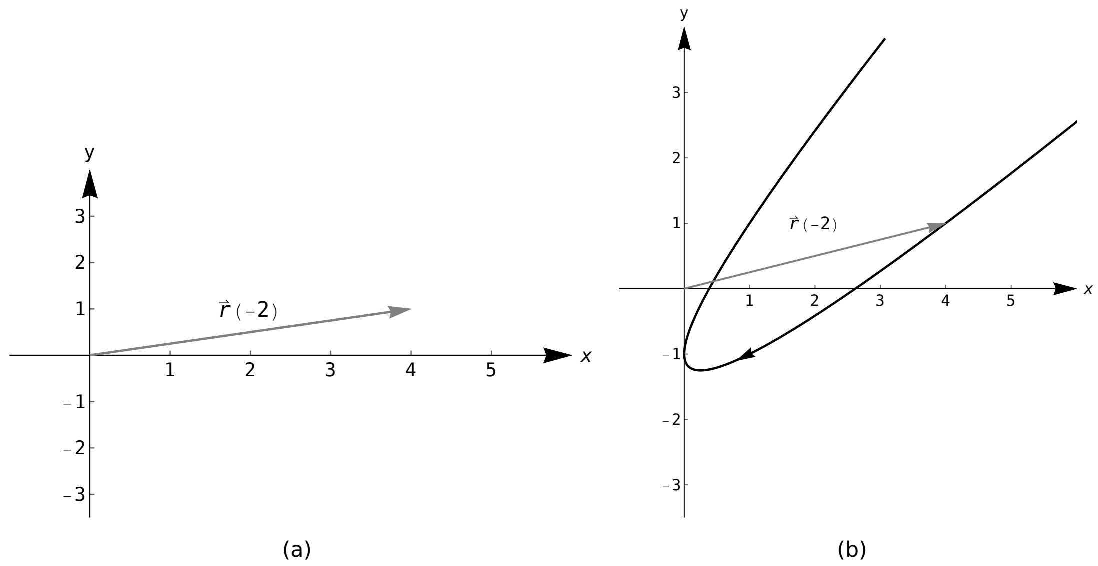


Figure 15.1: Sketching the graph of a vector–valued function.

Vector–valued functions are closely related to parametric equations of graphs. While in both methods we plot points $\big(x(t), y(t)\big)$ or $\big(x(t), y(t), z(t)\big)$ to produce a graph, in the context of vector–valued functions each such point represents a vector. The implications of this will be more fully realized in the next section as we apply calculus ideas to these functions.

Example 15.1

Graph $\vec{r}(t) = \big(\cos(t), \sin(t), t\big)$ for $0 \le t \le 4\pi$.

Solution

We can plot points, but careful consideration of this function is very revealing. Momentarily ignoring the third component, we see that the x- and y- components trace out a circle of radius 1 centred at the origin. Noticing that the z component is t, we see that as the graph winds around the z-axis, it is also increasing at a constant rate in the positive z direction, forming a spiral. This is graphed in Figure 15.2. In the graph, $\vec{r}(7\pi/4) \approx (0.707, -0.707, 5.498)$ is highlighted to help us understand the graph.

15.1.2 Algebra of vector–valued functions

Let $\vec{r}_1(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big)$ and $\vec{r}_2(t) = \big(g_1(t), g_2(t), \ldots, g_n(t)\big)$ be vector–valued functions in $\mathbb{R}^n$ and let c be a scalar. Then:

1. $\vec{r}_1(t) \pm \vec{r}_2(t) = \big(f_1(t) \pm g_1(t), f_2(t) \pm g_2(t), \ldots, f_n(t) \pm g_n(t)\big)$,
2. $c\vec{r}_1(t) = \big(cf_1(t), cf_2(t), \ldots, cf_n(t)\big)$.

This shows that we add, subtract and scale vector-valued functions component–wise. Combining vector–valued functions in this way can be very useful.

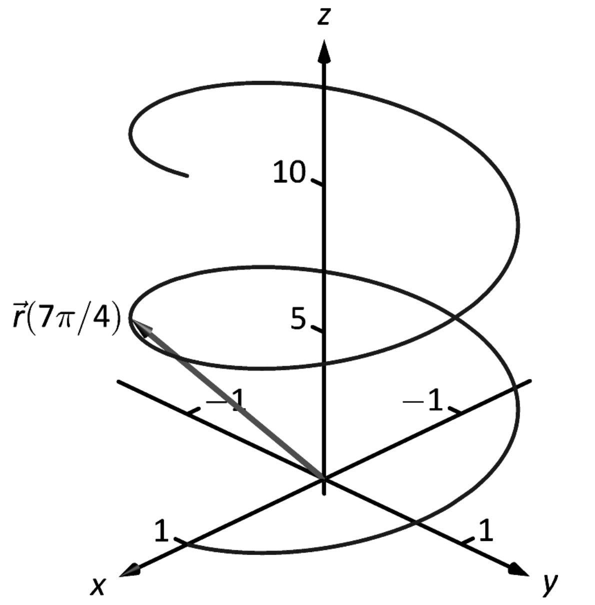


Figure 15.2: The graph of $\vec{r}(t)$ in Example 15.1.

Example 15.2

Let $\vec{r}_1(t) = (0.2t, 0.3t)$, $\vec{r}_2(t) = \big(\cos(t), \sin(t)\big)$ and $\vec{r}(t) = \vec{r}_1(t) + \vec{r}_2(t)$. Graph $\vec{r}_1(t)$, $\vec{r}_2(t)$, $\vec{r}(t)$ and $5\vec{r}(t)$ for $-10 \le t \le 10$.

Solution

We can graph $\vec{r}_1$ and $\vec{r}_2$ easily by plotting points. Let us think about each for a moment to better understand how vector–valued functions work.

We can rewrite $\vec{r}_1(t) = (0.2t, 0.3t)$ as $\vec{r}_1(t) = t\,(0.2, 0.3)$. That is, the function $\vec{r}_1$ scales the vector $(0.2, 0.3)$ by t. This scaling of a vector produces a line in the direction of $(0.2, 0.3)$. We are familiar with $\vec{r}_2(t) = \big(\cos(t), \sin(t)\big)$; it traces out a circle, centered at the origin, of radius 1. Figure 15.3(a) graphs $\vec{r}_1(t)$ and $\vec{r}_2(t)$.

Adding $\vec{r}_1(t)$ to $\vec{r}_2(t)$ produces $\vec{r}(t) = \big(\cos(t) + 0.2t, \sin(t) + 0.3t\big)$, graphed in Figure 15.3(b). The linear movement of the line combines with the circle to create loops that move in the direction of $(0.2, 0.3)$.

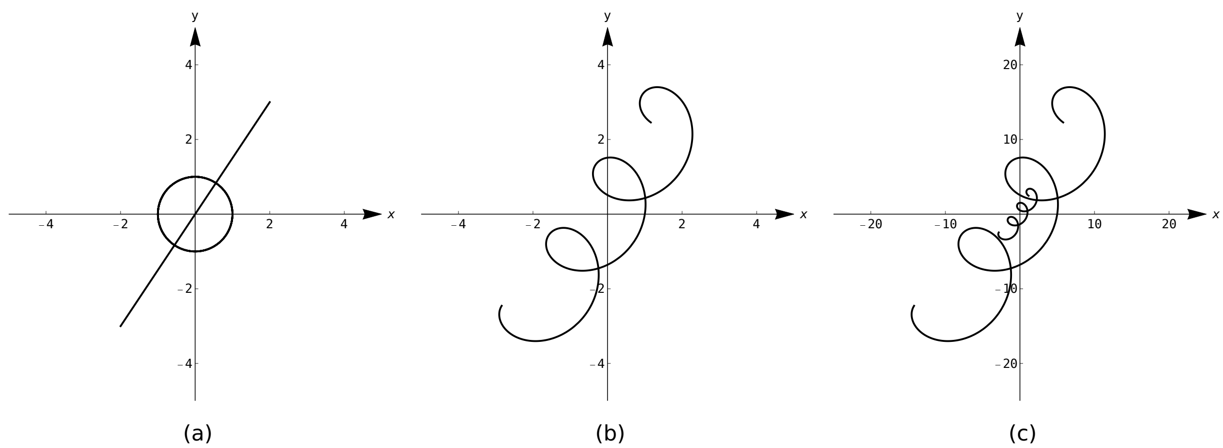


Figure 15.3: Graphing the functions in Example 15.2.

Multiplying $\vec{r}(t)$ by 5 scales the function by 5, producing $5\vec{r}(t) = \big(5\cos(t)+1, 5\sin(t)+1.5\big)$, which is graphed in Figure 15.3(c) along with $\vec{r}(t)$. The new function is 5 times bigger than $\vec{r}(t)$. Note how the graph of $5\vec{r}(t)$ in (c) looks identical to the graph of $\vec{r}(t)$ in (b). This is due to the fact that the x- and y- bounds of the plot in (c) are exactly 5 times larger than the bounds in (b).

A vector–valued function $\vec{r}(t)$ is often used to describe the position of a moving object at time t. At $t = t_0$, the object is at $\vec{r}(t_0)$; at $t = t_1$, the object is at $\vec{r}(t_1)$. Knowing the locations $\vec{r}(t_0)$ and $\vec{r}(t_1)$ gives no indication of the path taken between them, but often we only care about the difference of the locations, $\vec{r}(t_1)-\vec{r}(t_0)$, the **displacement** (*verplaatsing*).

Definitie 15.2 (Displacement)
Let $\vec{r}(t)$ be a vector–valued function and let $t_0 < t_1$ be values in the domain. The **displacement** $\vec{d}$ of $\vec{r}$, from $t = t_0$ to $t = t_1$, is

$$\vec{d} = \vec{r}(t_1)-\vec{r}(t_0).$$

When the displacement vector is drawn with initial point at $\vec{r}(t_0)$, its terminal point is $\vec{r}(t_1)$. We think of it as the vector which points from a starting position to an ending position.

Example 15.3

Let $\vec{r}(t) = \Big(\cos\big(\frac{\pi}{2}t\big), \sin\big(\frac{\pi}{2}t\big)\Big)$. Graph $\vec{r}(t)$ on $-1 \le t \le 1$, and find the displacement of $\vec{r}(t)$ on this interval.

Solution

The function $\vec{r}(t)$ traces out the unit circle, though at a different rate than the usual $\big(\cos(t), \sin(t)\big)$ parametrization. At $t_0 = -1$, we have $\vec{r}(t_0) = (0,-1)$; at $t_1 = 1$, we have $\vec{r}(t_1) = (0,1)$. The displacement of $\vec{r}(t)$ on $[-1,1]$ is thus $\vec{d} = (0,1)-(0,-1) = (0,2)$. A graph of $\vec{r}(t)$ on $[-1,1]$ is given in Figure 15.4, along with the displacement vector $\vec{d}$ on this interval.

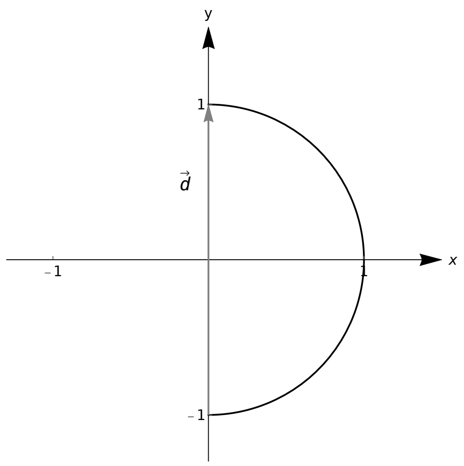


Figure 15.4: Graphing the displacement of a position function in Example 15.3.

Measuring displacement makes us contemplate related, yet very different, concepts. Considering the semi–circular path the object in Example 15.3 took, we can quickly verify that the object ended up a distance of 2 units from its initial location. That is, we can compute $\left\|\vec{d}\right\| = 2$. However, measuring distance from the starting point is different from measuring distance travelled. Being a semi–circle, we can measure the distance traveled by this object as $\pi \approx 3.14$ units. Knowing distance from the starting point allows us to compute average rate of change.

Definitie 15.3 (Average rate of change)
Let $\vec{r}(t)$ be a vector–valued function, where each of its component functions is continuous on its domain, and let $t_0 < t_1$. The **average rate of change of $\vec{r}(t)$** on $[t_0, t_1]$ is

$$\text{average rate of change} = \frac{\vec{r}(t_1) - \vec{r}(t_0)}{t_1 - t_0}.$$

Example 15.4

Let $\vec{r}(t) = \left(\cos\left(\frac{\pi}{2}t\right), \sin\left(\frac{\pi}{2}t\right)\right)$ as in Example 15.3. Find the average rate of change of $\vec{r}(t)$ on $[-1, 1]$ and on $[-1, 5]$.

Solution

We computed in Example 15.3 that the displacement of $\vec{r}(t)$ on $[-1, 1]$ was $\vec{d} = (0, 2)$. Thus the average rate of change of $\vec{r}(t)$ on $[-1, 1]$ is:

$$\frac{\vec{r}(1) - \vec{r}(-1)}{1 - (-1)} = \frac{(0, 2)}{2} = (0, 1).$$

We interpret this as follows: the object followed a semi–circular path, meaning it moved towards the right then moved back to the left, while climbing slowly, then quickly, then slowly again. On average, however, it progressed straight up at a constant rate of $(0, 1)$ per unit of time.

We considered average rates of change in Sections 8.1 and 9.1 as we studied limits and derivatives. The same is true here; in the following section we apply calculus concepts to vector–valued functions as we find limits, derivatives, and integrals. Understanding the average rate of change will give us an understanding of the derivative; displacement gives us one application of integration.

15.2 Calculus and vector–valued functions

The previous section introduced us to a new mathematical object, the vector–valued function. We now apply calculus concepts to these functions. We start with the limit, then work our way through derivatives to integrals.

15.2.1 Limits of vector-valued functions

The initial definition of the limit of a vector–valued function is a bit intimidating, as was the definition of the limit in Definition 8.1. The theorem following the definition shows that in practice, taking limits of vector–valued functions is no more difficult than taking limits of real–valued functions. Of course, we can define one-sided limits in a manner very similar to Definition 15.4.

Definitie 15.4 (Limits of vector–valued functions)
Let I be an open interval containing c, and let $\vec{r}(t)$ be a vector–valued function defined on I, except possibly at c. The **limit of $\vec{r}(t)$** (*limiet*), as t approaches c, is $\vec{L}$, expressed as

$$\lim_{t \to c} \vec{r}(t) = \vec{L},$$

means that given any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $t \neq c$, if $|t-c| < \delta$, we have $\left\|\vec{\boldsymbol{r}}(t)-\vec{\boldsymbol{L}}\right\| < \varepsilon$.

Note how the measurement of distance between real numbers is the absolute value of their difference; the measure of distance between vectors is the vector norm, or magnitude, of their difference.

The following theorem affirms that we can compute limits of vector–valued functions component–wise.

Theorem 15.1 (Limits of vector-valued functions)
Let $\vec{\boldsymbol{r}}(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big)$ be a n-dimensional vector–valued function in $\mathbb{R}^n$ defined on an open interval I containing c, except possibly at c. Then

$$\lim_{t\to c}\vec{\boldsymbol{r}}(t) = \left(\lim_{t\to c} f_1(t)\,, \lim_{t\to c} f_2(t), \ldots,\, \lim_{t\to c} f_n(t)\right).$$

So, for instance, if

$$\vec{\boldsymbol{r}}(t) = \left(\frac{\sin(t)}{t},\, t^2-3t+3,\, \cos(t)\right). \tag{15.1}$$

Then, $\lim\limits_{t\to 0}\vec{\boldsymbol{r}}(t)$ is given by:

$$\lim_{t\to 0}\vec{\boldsymbol{r}}(t) = \left(\lim_{t\to 0}\frac{\sin(t)}{t}\,,\, \lim_{t\to 0}(t^2-3t+3)\,,\, \lim_{t\to 0}\cos(t)\right) = (1,3,1)\,.$$

15.2.2 Continuity

Having defined limits of vector–valued functions, it makes sense to explore the continuity of such functions.

Definitie 15.5 (Continuity of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function defined on an open interval I containing c.

1. $\vec{\boldsymbol{r}}(t)$ is **continuous** at c if $\lim\limits_{t\to c}\vec{\boldsymbol{r}}(t) = \vec{\boldsymbol{r}}(c)$.
2. If $\vec{\boldsymbol{r}}(t)$ is continuous at all c in I, then $\vec{\boldsymbol{r}}(t)$ is continuous on I.

Using one-sided limits, we can of course also define continuity on closed intervals as done before. Moreover, we again have a theorem that lets us evaluate continuity component–wise.

Theorem 15.2 (Continuity of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function defined on an open interval I containing c. Then $\vec{\boldsymbol{r}}(t)$ is continuous at c if, and only if, each of its component functions is continuous at c.

In the case of the vector–valued function defined by Equation (15.1), for instance, $\vec{\boldsymbol{r}}(t)$ is continuous at $t=1$ because each of the component functions is continuous at $t=1$. On the other hand, at $t=0$, the second and third components of $\vec{\boldsymbol{r}}(t)$ are defined, but the first component, $(\sin(t))/t$, is not. Since the first component is not even defined at $t=0$, $\vec{\boldsymbol{r}}(t)$ is not defined at $t=0$, and hence it is not continuous at $t=0$.

15.2.3 Derivatives

15.2.3.1 Definition and properties

Consider a vector–valued function $\vec{r}$ defined on an open interval I containing t_0 and t_1. We can compute the displacement of $\vec{r}$ on $[t_0, t_1]$, as shown in Figure 15.5(a). Recall that dividing the displacement vector by $t_1 - t_0$ gives the average rate of change on $[t_0, t_1]$, as shown in Figure 15.5(b).

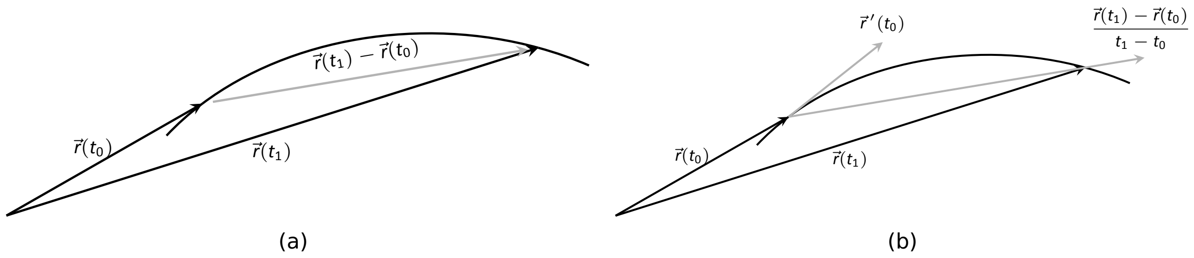


Figure 15.5: Illustrating displacement, leading to an understanding of the derivative of vector–valued functions.

The **derivative** (*afgeleide*) of a vector–valued function is a measure of the instantaneous rate of change, measured by taking the limit as the length of $[t_0, t_1]$ goes to 0. Instead of thinking of an interval as $[t_0, t_1]$, we think of it as $[c, c+h]$ for some value of h (hence the interval has length h). The average rate of change is

$$\frac{\vec{r}(c+h) - \vec{r}(c)}{h}$$

for any value of $h \neq 0$. We take the limit as $h \to 0$ to measure the instantaneous rate of change; this is the derivative of $\vec{r}$.

Definitie 15.6 (Derivative of a vector–valued function)
Let $\vec{r}(t)$ be continuous on an open interval I containing c.

1. The **derivative of $\vec{r}$ at** $t = c$ is
$$\vec{r}\,'(c) = \lim_{h\to 0} \frac{\vec{r}(c+h) - \vec{r}(c)}{h}.$$

2. The **derivative of $\vec{r}$** is
$$\vec{r}\,'(t) = \lim_{h\to 0} \frac{\vec{r}(t+h) - \vec{r}(t)}{h}.$$

Alternate notations for the derivative of $\vec{r}$ include:

$$\vec{r}\,'(t) = \frac{d}{dt}\big(\vec{r}(t)\big) = \frac{d\vec{r}}{dt}.$$

If a vector–valued function has a derivative for all c in an open interval I, we say that $\vec{r}(t)$ is **differentiable** (*afleidbaar*) on I. Again, of course, using one-sided limits, we can define differentiability on closed intervals. We might view Definition 15.6 as intimidating, but recall that we can evaluate limits component–wise. The following theorem verifies that this means we can compute derivatives component–wise as well, making the task not too difficult.

Theorem 15.3 (Derivative of a vector–valued function)
Let $\vec{r}(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big)$*. Then*

$$\vec{r}\,'(t) = \Big(f_1'(t), f_2'(t), \ldots, f_n'(t)\Big).$$

Proof This theorem follows easily by combining Definition 15.3 and vector arithmetic (Section 6.2). For instance, for a three-dimensional vector-valued function $\vec{r}(t) = \big(f(t), g(t), h(t)\big)$, we get:

$$\begin{aligned}
\vec{r}\,'(t) &= \lim_{\Delta t \to 0} \frac{\vec{r}(t+\Delta t) - \vec{r}(t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\big(f(t+\Delta t) - f(t), g(t+\Delta t) - g(t), h(t+\Delta t) - h(t)\big)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \left(\frac{f(t+\Delta t) - f(t)}{\Delta t}, \frac{g(t+\Delta t) - g(t)}{\Delta t}, \frac{h(t+\Delta t) - h(t)}{\Delta t}\right) \\
&= \big(f'(t), g'(t), h'(t)\big)
\end{aligned}$$

Example 15.5

Let $\vec{r}(t) = \Big(t^2, t\Big)$.

1. Sketch $\vec{r}(t)$ and $\vec{r}\,'(t)$ on the same axes.
2. Compute $\vec{r}\,'(1)$ and sketch this vector with its initial point at the origin and at $\vec{r}(1)$.

Solution

1. Theorem 15.3 allows us to compute derivatives component–wise, so

$$\vec{r}\,'(t) = (2t, 1).$$

   $\vec{r}(t)$ and $\vec{r}\,'(t)$ are graphed together in Figure 15.6(a). Note how plotting the two of these together, in this way, is not very illuminating. When dealing with real–valued functions, plotting $f(x)$ with $f\,'(x)$ gave us useful information as we were able to compare f and $f\,'$ at the same x-values. When dealing with vector–valued functions, it is hard to tell which points on the graph of $\vec{r}\,'$ correspond to which points on the graph of $\vec{r}$.

2. We easily compute $\vec{r}\,'(1) = (2, 1)$, which is drawn in Figure 15.6(b) with its initial point at the origin, as well as at $\vec{r}(1) = (1, 1)$. These are sketched in Figure 15.6(b).

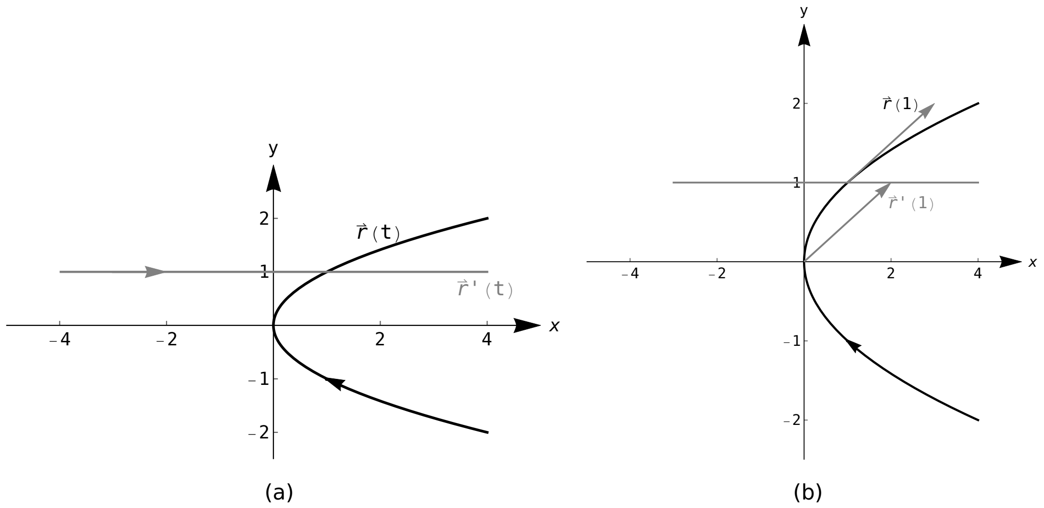


Figure 15.6: Graphing the derivative of a vector–valued function in Example 15.5.

Having established derivatives of vector–valued functions, we now explore the relationships between the derivative and other vector operations. The following properties show how the derivative interacts with vector addition and the various vector products. For that purpose, let $\vec{r}$ and $\vec{s}$ be differentiable vector–valued functions, let f be a differentiable real–valued function, and let c be a real number. Then the following properties hold.

- $\dfrac{d}{dt}\Big(\vec{r}(t) \pm \vec{s}(t)\Big) = \vec{r}\,'(t) \pm \vec{s}\,'(t)$
- $\dfrac{d}{dt}\Big(c\vec{r}(t)\Big) = c\vec{r}\,'(t)$
- $\dfrac{d}{dt}\Big(f(t)\vec{r}(t)\Big) = f\,'(t)\vec{r}(t) + f(t)\vec{r}\,'(t)$
- $\dfrac{d}{dt}\Big(\vec{r}(t) \cdot \vec{s}(t)\Big) = \vec{r}\,'(t) \cdot \vec{s}(t) + \vec{r}(t) \cdot \vec{s}\,'(t)$
- $\dfrac{d}{dt}\Big(\vec{r}(t) \times \vec{s}(t)\Big) = \vec{r}\,'(t) \times \vec{s}(t) + \vec{r}(t) \times \vec{s}\,'(t)$
- $\dfrac{d}{dt}\Big(\vec{r}\big(f(t)\big)\Big) = \vec{r}\,'\big(f(t)\big)f\,'(t)$

Example 15.6

Let $\vec{r}(t) = \big\langle t, t^2-1\big\rangle$ and let $\vec{u}(t)$ be the unit vector that points in the direction of $\vec{r}(t)$.

1. Graph $\vec{r}(t)$ and $\vec{u}(t)$ on the same axes, on $[-2,2]$.
2. Find $\vec{u}\,'(t)$ and sketch $\vec{u}\,'(-2)$, $\vec{u}\,'(-1)$ and $\vec{u}\,'(0)$. Sketch each with initial point the corresponding point on the graph of $\vec{u}$.

Solution

1. To form the unit vector that points in the direction of $\vec{r}$, we need to divide $\vec{r}(t)$ by its magnitude.

$$\big\|\vec{r}(t)\big\| = \sqrt{t^2+(t^2-1)^2} \quad \Rightarrow \quad \vec{u}(t) = \frac{1}{\sqrt{t^2+(t^2-1)^2}}\vec{r}(t)$$

$\vec{r}(t)$ and $\vec{u}(t)$ are graphed in Figure 15.7(a). Note how the graph of $\vec{u}(t)$ forms part of a circle; this must be the case, as the length of $\vec{u}(t)$ is 1 for all t.

2. To compute $\vec{u}\,'(t)$, we rely on the above properties and write

$$\vec{u}(t) = f(t)\vec{r}(t), \quad \text{where} \quad f(t) = \frac{1}{\sqrt{t^2 + (t^2-1)^2}} = \left(t^2 + (t^2-1)^2\right)^{-1/2}.$$

We find $f'(t)$ using the chain rule:

$$\begin{aligned} f'(t) &= -\frac{1}{2}\left(t^2 + (t^2-1)^2\right)^{-3/2}\left(2t + 2(t^2-1)(2t)\right) \\ &= -\frac{2t(2t^2-1)}{2\left(\sqrt{t^2+(t^2-1)^2}\right)^3}. \end{aligned}$$

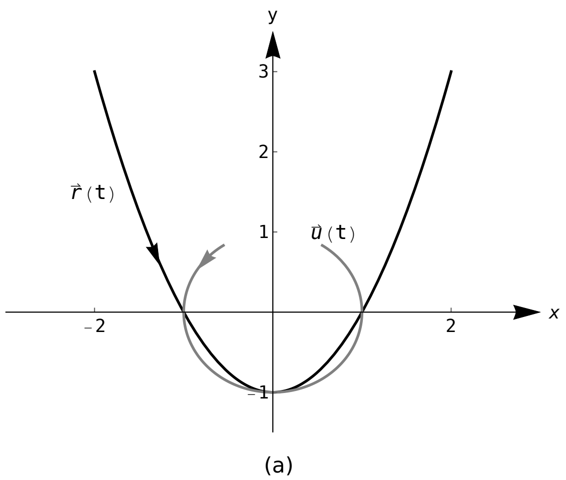
We now find $\vec{u}\,'(t)$:

$$\begin{aligned} \vec{u}\,'(t) &= f'(t)\vec{u}(t) + f(t)\vec{u}\,'(t) \\ &= -\frac{2t(2t^2-1)}{2\left(\sqrt{t^2+(t^2-1)^2}\right)^3}\left(t, t^2-1\right) + \frac{1}{\sqrt{t^2+(t^2-1)^2}}\left(1, 2t\right). \end{aligned}$$

We can use this formula to compute $\vec{u}\,'(-2)$, $\vec{u}\,'(-1)$ and $\vec{u}\,'(0)$:

$$\begin{aligned} \vec{u}\,'(-2) &= \left(-\frac{15}{13\sqrt{13}}, -\frac{10}{13\sqrt{13}}\right) \approx (-0.320, -0.213), \\ \vec{u}\,'(-1) &= (0, -2), \\ \vec{u}\,'(0) &= (1, 0). \end{aligned}$$

Each of these is sketched in Figure 15.7(b). Note how the length of the vector gives an indication of how quickly the circle is being traced at that point. When $t = -2$, the circle is being drawn relatively slow; when $t = -1$, the circle is being traced much more quickly.



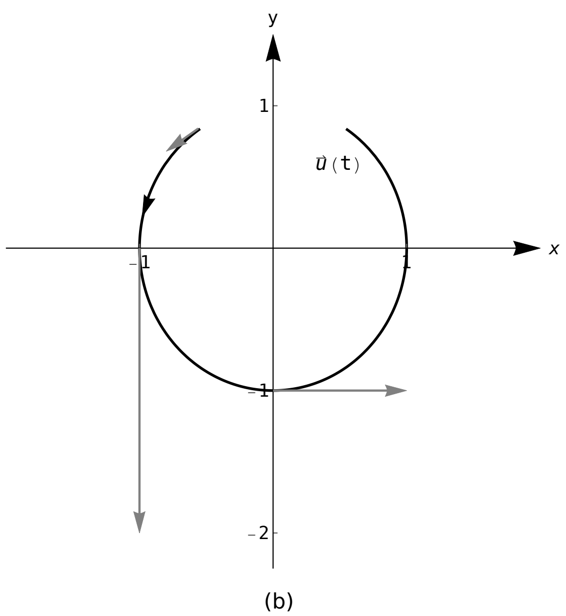


Figure 15.7: Graphing $\vec{r}(t)$ and $\vec{u}(t)$ (a) and some of the derivatives of $\vec{u}(t)$ (b) in Example 15.6.

15.2.3.2 Tangent vector and lines

In Example 15.5, sketching a particular derivative with its initial point at the origin did not seem to reveal anything significant. However, when we sketched the vector with its initial point on the corresponding point on the graph, we did see something significant: the vector appeared to be tangent to the graph. We have not yet defined what tangent means in terms of curves in space; in fact, we use the derivative to define this term.

Definitie 15.7 (Tangent vector and line)
Let $\vec{r}(t)$ be a differentiable vector–valued function on an open interval I containing c, where $\vec{r}\,'(c) \neq \vec{0}$.

1. A vector $\vec{v}$ is **tangent** (*rakend*) to the graph of $\vec{r}(t)$ at $t = c$ if $\vec{v}$ is parallel to $\vec{r}\,'(c)$.
2. The tangent line to the graph of $\vec{r}(t)$ at $t = c$ is the line through $\vec{r}(c)$ with direction parallel to $\vec{r}\,'(c)$. An equation of the **tangent line** (*raaklijn*) is

$$\vec{y} = \vec{\ell}(t) = \vec{r}(c) + t\,\vec{r}\,'(c).$$

Example 15.7

Find the equations of the lines tangent to $\vec{r}(t) = \left(t^3, t^2\right)$ at $t = -1$ and $t = 0$.

Solution

We find that $\vec{r}\,'(t) = \left(3t^2, 2t\right)$. At $t = -1$, we have

$$\vec{r}(-1) = (-1, 1) \quad \text{and} \quad \vec{r}\,'(-1) = (3, -2),$$

so the equation of the line tangent to the graph of $\vec{r}(t)$ at $t = -1$ is

$$\vec{\ell}(t) = (-1, 1) + t\,(3, -2).$$

This line is graphed with $\vec{r}(t)$ in Figure 15.8.

At $t = 0$, we have $\vec{r}\,'(0) = (0, 0) = \vec{0}$! This implies that the tangent line has no direction. We cannot apply Definition 15.7, hence cannot find the equation of the tangent line.

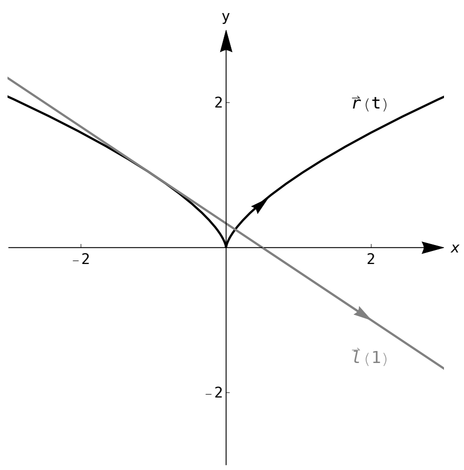


Figure 15.8: Graphing $\vec{r}(t)$ and its tangent line in Example 15.7.

15.2.3.3 Smoothness

We were unable to compute the equation of the tangent line to $\vec{r}(t) = \left(t^3, t^2\right)$ at $t = 0$ because $\vec{r}\,'(0) = \vec{0}$. The graph in Figure 15.8 shows that there is a cusp at this point. This leads us to another definition of **smooth** (*glad*), previously defined by Definition 9.7.

Definitie 15.8 (Smooth vector–valued function)
Let $\vec{r}(t)$ be a differentiable vector–valued function on an open interval I where $\vec{r}\,'(t)$ is continuous on I. $\vec{r}(t)$ is **smooth** on I if $\vec{r}\,'(t) \neq \vec{0}$ on I.

It is a basic geometric fact that a line tangent to a circle at a point P is perpendicular to the line passing through the center of the circle and P. This is illustrated in Figure 15.7(b); each tangent vector is perpendicular to the line that passes through its initial point and the centre of the circle. Since the centre of the circle is the origin, we can state this another way: $\vec{u}\,'(t)$ is orthogonal to $\vec{u}(t)$.

Recall that the dot product serves as a test for orthogonality: if $\vec{u} \cdot \vec{v} = 0$, then $\vec{u}$ is orthogonal to $\vec{v}$. Thus in the above example, $\vec{u}(t) \cdot \vec{u}\,'(t) = 0$. This is true for any vector–valued function that has a constant length, that is, that traces out part of a circle. It has important implications later on, so we state it as a theorem.

Theorem 15.4 (Vector–valued functions of constant length)
Let $\vec{r}(t)$ be a vector–valued function of constant length that is differentiable on an open interval I. That is, $\left\|\vec{r}(t)\right\| = c$ for all t in I. Then $\vec{r}(t) \cdot \vec{r}\,'(t) = 0$ for all t in I.

15.2.4 Integration

Before formally defining integrals of vector–valued functions, consider the following equation that our calculus experience tells us should be true:

$$\int_a^b \vec{r}\,'(t)\, dt = \vec{r}(b) - \vec{r}(a).$$

That is, the integral of a rate of change function should give total change. In the context of vector–valued functions, this total change is displacement. The above equation is true; we now develop the theory to show why.

We can define antiderivatives and the indefinite integral of vector–valued functions in the same manner we defined indefinite integrals in Definition 12.1. However, we cannot define the definite integral of a vector–valued function as we did in Definition 12.2. That definition was based on the signed area between a function $y = f(x)$ and the x-axis. An area–based definition will not be useful in the context of vector–valued functions. Instead, we define the definite integral of a vector–valued function in a manner similar to that of Theorem 12.4, utilizing Riemann sums.

Definitie 15.9 (Antiderivatives, integrals of vector–valued functions)
Let $\vec{r}(t)$ be a continuous vector–valued function on $[a, b]$. An **antiderivative of $\vec{r}(t)$** (*primitieve functie*) is a function $\vec{R}(t)$ such that $\vec{R}\,'(t) = \vec{r}(t)$.

The set of all antiderivatives of $\vec{r}(t)$ is the **indefinite integral of $\vec{r}(t)$** (*onbepaalde integraal*), denoted by

$$\int \vec{\boldsymbol{r}}(t)\ dt.$$

The **definite integral of $\vec{\boldsymbol{r}}(t)$** (*bepaalde integraal*) on $[a,b]$ is

$$\int_a^b \vec{\boldsymbol{r}}(t)\ dt = \lim_{\mathcal{T}\to 0}\sum_{i=1}^{n}\vec{\boldsymbol{r}}(c_i)\Delta t_i,$$

where Δt_i is the length of the i^{th} subinterval of a partition of $[a,b]$, $\mathcal{T}$ is the length of the largest subinterval in the partition, and c_i is any value in the i^{th} subinterval of the partition.

It is probably difficult to infer meaning from the definition of the definite integral. The important thing to realize from the definition is that it is built upon limits, which we can evaluate component–wise. The following theorem simplifies the computation of definite integrals.

Theorem 15.5 (Indefinite and definite integrals of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big)$ be a vector–valued function in $\mathbb{R}^n$ that is continuous on $[a,b]$.

1. $\displaystyle\int \vec{\boldsymbol{r}}(t)\ dt = \left(\int f_1(t)\ dt, \int f_2(t)\ dt, \ldots, \int f_n(t)\ dt\right)$

2. $\displaystyle\int_a^b \vec{\boldsymbol{r}}(t)\ dt = \left(\int_a^b f_1(t)\ dt, \int_a^b f_2(t)\ dt, \ldots, \int_a^b f_n(t)\ dt\right)$

Proof This theorem is an immediate consequence of our understanding of the derivative of a vector function through Theorem 15.3. □

Example 15.8

Let $\vec{\boldsymbol{r}}(t) = \big(e^{2t}, \sin(t)\big)$. Evaluate

$$\int_0^1 \vec{\boldsymbol{r}}(t)\ dt.$$

Solution

We follow Theorem 15.5.

$$\begin{aligned}
\int_0^1 \vec{\boldsymbol{r}}(t)\ dt &= \int_0^1 \big(e^{2t}, \sin(t)\big)\ dt \\
&= \left(\int_0^1 e^{2t}\ dt\ , \int_0^1 \sin(t)\ dt\right) \\
&= \left(\frac{1}{2}e^{2t}\Big|_0^1\ , (-\cos(t))\Big|_0^1\right)
\end{aligned}$$

$$= \left(\frac{1}{2}(e^2-1)\ , -\cos(1)+1\right)$$
$$\approx (3.19, 0.460)$$

What does the integration of a vector–valued function mean? There are many applications, but none as direct as the area under the curve that we used in understanding the integral of a real–valued function. A key understanding for us comes from considering the integral of a derivative:

$$\int_a^b \vec{\boldsymbol{r}}\,'(t)\ dt = \vec{\boldsymbol{r}}(t)\Big|_a^b = \vec{\boldsymbol{r}}(b) - \vec{\boldsymbol{r}}(a).$$

This indicates integrating a rate of change function gives displacement.

Noting that vector–valued functions are closely related to parametric equations, we can describe the arc length of the graph of a vector–valued function as an integral. Given parametric equations $x = f(t)$, $y = g(t)$, the arc length on $[a, b]$ of the graph is

$$L = \int_a^b \sqrt{f\,'(t)^2 + g'(t)^2}\ dt,$$

as stated in Theorem 13.8. If $\vec{\boldsymbol{r}}(t) = \big(f(t), g(t)\big)$, note that $\sqrt{f\,'(t)^2 + g'(t)^2} = \left\|\vec{\boldsymbol{r}}\,'(t)\right\|$. Therefore we can express the arc length of the graph of a vector–valued function as an integral of the magnitude of its derivative.

Theorem 15.6 (Arc length of a vector–valued function)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function where $\vec{\boldsymbol{r}}\,'(t)$ is continuous on $[a, b]$. The ***arc length*** *(booglengte) L of the graph of $\vec{\boldsymbol{r}}(t)$ is*

$$L = \int_a^b \left\|\vec{\boldsymbol{r}}\,'(t)\right\|\ dt.$$

Note that we are actually integrating a scalar function here, not a vector–valued function.

The remainder of this section takes what we have established thus far and applies it to objects in motion.

15.2.5 The calculus of motion

A common use of vector–valued functions is to describe the motion of an object in the plane or in space. A position function $\vec{\boldsymbol{r}}(t)$ gives the position of an object at time t. This section explores how derivatives and integrals are used to study the motion described by such a function.

Definitie 15.10 (Velocity, speed and acceleration)
Let $\vec{\boldsymbol{r}}(t)$ be a position function in $\mathbb{R}^2$ or $\mathbb{R}^3$.

1. **Velocity**, denoted $\vec{\boldsymbol{v}}(t)$, is the instantaneous rate of position change; that is, $\vec{\boldsymbol{v}}(t) = \vec{\boldsymbol{r}}\,'(t)$.
2. **Speed** is the magnitude of velocity, $\left\|\vec{\boldsymbol{v}}(t)\right\|$.
3. **Acceleration**, denoted $\vec{\boldsymbol{a}}(t)$, is the instantaneous rate of velocity change; that is,

$\vec{a}(t) = \vec{v}\,'(t) = \vec{r}\,''(t)$.

Example 15.9

An object is moving with position function $\vec{r}(t) = \left(t^2 - t, t^2 + t\right)$, $-3 \le t \le 3$, where distances are measured in metres and time is measured in seconds.

1. Find $\vec{v}(t)$ and $\vec{a}(t)$.
2. Sketch $\vec{r}(t)$; plot $\vec{v}(-1)$, $\vec{a}(-1)$, $\vec{v}(1)$ and $\vec{a}(1)$, each with their initial point at their corresponding point on the graph of $\vec{r}(t)$.
3. When is the object's speed minimized?

Solution

1. Taking derivatives, we find

$$\vec{v}(t) = \vec{r}\,'(t) = (2t-1, 2t+1) \quad \text{and} \quad \vec{a}(t) = \vec{r}\,''(t) = (2,2).$$

   Note that the acceleration is constant.

2. $\vec{v}(-1) = (-3,-1)$, $\vec{a}(-1) = (2,2)$; $\quad \vec{v}(1) = (1,3)$, $\vec{a}(1) = (2,2)$. These are plotted with $\vec{r}(t)$ in Figure 15.9(a).

   We can think of acceleration as pulling the velocity vector in a certain direction. At $t = -1$, the velocity vector points down and to the left; at $t = 1$, the velocity vector has been pulled in the $(2,2)$ direction and is now pointing up and to the right. In Figure 15.9(b) we plot more velocity/acceleration vectors, making more clear the effect acceleration has on velocity.

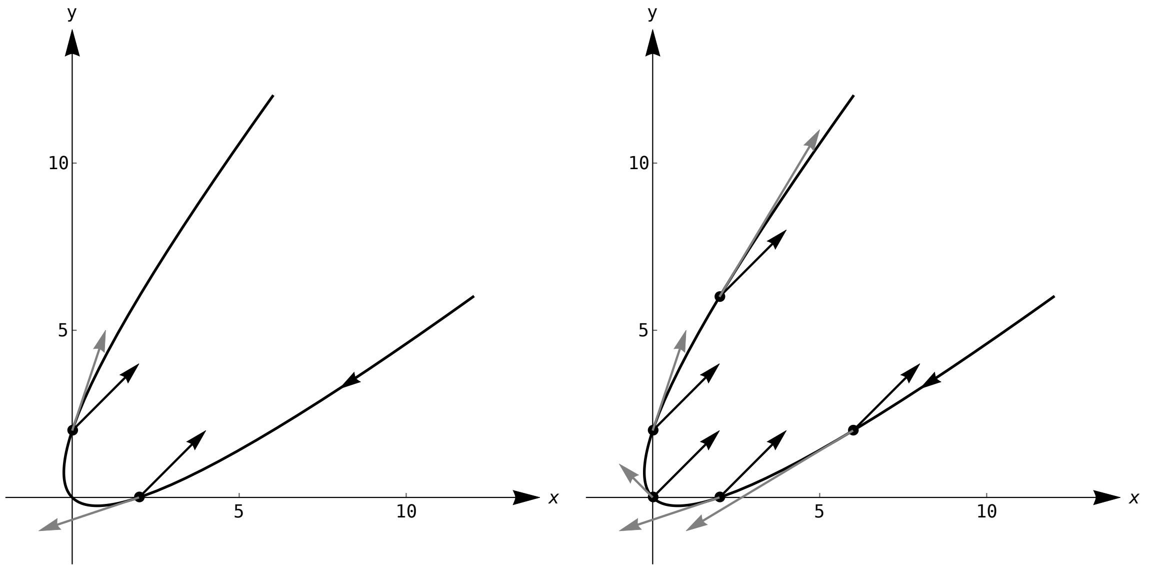
   Since $\vec{a}(t)$ is constant in this example, as t grows large $\vec{v}(t)$ becomes almost parallel to $\vec{a}(t)$. For instance, when $t = 10$, $\vec{v}(10) = (19,21)$, which is nearly parallel to $(2,2)$.

3. The object's speed is given by

$$\left\|\vec{v}(t)\right\| = \sqrt{(2t-1)^2 + (2t+1)^2} = \sqrt{8t^2+2}.$$

   To find the minimal speed, we could apply calculus techniques (such as set the derivative equal to 0 and solve for t, etc.) but we can find it by inspection. Inside the square root we have a quadratic which is minimized when $t = 0$. Thus the speed is minimized at $t = 0$, with a speed of $\sqrt{2}$ m/s.

   The graph in Figure 15.9(b) also implies speed is minimized here. The filled dots on the graph are located at integer values of t between -3 and 3. Dots that are far apart imply the object travelled a far distance in 1 second, indicating high speed; dots that are close together imply the object did not travel far in 1 second, indicating a low speed. The dots are closest together near $t = 0$, implying the speed is minimized near that value.



(a) (b)

Figure 15.9: Graphing the position (curve), velocity (gray) and acceleration (black) of an object in Example 15.9.

If an object travels at a constant speed, we have $\left\|\vec{\boldsymbol{v}}(t)\right\| = c$ for some constant c. Recall Theorem 15.4, which states that if a vector–valued function $\vec{\boldsymbol{r}}(t)$ has constant length, then $\vec{\boldsymbol{r}}(t)$ is perpendicular to its derivative: $\vec{\boldsymbol{r}}(t)\cdot\vec{\boldsymbol{r}}\,'(t) = 0$. So, the corresponding velocity function has constant length, therefore we can conclude that the velocity is perpendicular to the acceleration: $\vec{\boldsymbol{v}}(t)\cdot\vec{\boldsymbol{a}}(t) = 0$.

An important application of vector–valued position functions is projectile motion: the motion of objects under only the influence of gravity. We will measure time in seconds, and distances will either be in meters or feet. We will show that we can completely describe the path of such an object knowing its initial position and initial velocity.

Suppose an object has initial position $\vec{\boldsymbol{r}}(0) = (x_0, y_0)$ and initial velocity $\vec{\boldsymbol{v}}(0) = \big(v_x, v_y\big)$. It is customary to rewrite $\vec{\boldsymbol{v}}(0)$ in terms of its speed v_0 and direction $\vec{\boldsymbol{u}}$, where $\vec{\boldsymbol{u}}$ is a unit vector. Recall all unit vectors in $\mathbb{R}^2$ can be written as $\big(\cos(\theta), \sin(\theta)\big)$, where θ is an angle measure counter–clockwise from the x-axis. We refer to θ as the angle of elevation. Thus $\vec{\boldsymbol{v}}(0) = v_0\big(\cos(\theta), \sin(\theta)\big)$.

Since the acceleration of the object is known, namely $\vec{\boldsymbol{a}}(t) = (0, -g)$, where g is the gravitational constant, we can find $\vec{\boldsymbol{r}}(t)$ knowing our two initial conditions. We first find $\vec{\boldsymbol{v}}(t)$:

$$\begin{aligned}
\vec{\boldsymbol{v}}(t) &= \int \vec{\boldsymbol{a}}(t)\ dt\\
\Rightarrow\ \vec{\boldsymbol{v}}(t) &= \int (0, -g)\ dt\\
\Leftrightarrow\ \vec{\boldsymbol{v}}(t) &= (0, -gt) + \vec{\boldsymbol{C}}.
\end{aligned}$$

Knowing $\vec{\boldsymbol{v}}(0) = v_0\big(\cos(\theta), \sin(\theta)\big)$, we have $\vec{\boldsymbol{C}} = v_0\big(\cos(\theta), \sin(\theta)\big)$ and so

$$\vec{\boldsymbol{v}}(t) = \big(v_0\cos(\theta), -gt + v_0\sin(\theta)\big).$$

We integrate once more to find $\vec{\boldsymbol{r}}(t)$:

$$\begin{aligned}
\vec{\boldsymbol{r}}(t) &= \int \vec{\boldsymbol{v}}(t)\ dt\\
\vec{\boldsymbol{r}}(t) &= \int \big(v_0\cos(\theta), -gt + v_0\sin(\theta)\big)\ dt
\end{aligned}$$

$$\vec{\boldsymbol{r}}(t) = \left(\big(v_0\cos(\theta)\big)t, -\frac{1}{2}gt^2 + \big(v_0\sin(\theta)\big)t\right) + \vec{\boldsymbol{C}}.$$

Knowing $\vec{\boldsymbol{r}}(0) = (x_0, y_0)$, we conclude $\vec{\boldsymbol{C}} = (x_0, y_0)$ and

$$\vec{\boldsymbol{r}}(t) = \left(\big(v_0\cos(\theta)\big)t + x_0\ , -\frac{1}{2}gt^2 + \big(v_0\sin(\theta)\big)t + y_0\ \right).$$

This is the position function of a projectile propelled from an initial position of $\vec{\boldsymbol{r}}_0 = (x_0, y_0)$, with initial speed v_0, with angle of elevation θ and neglecting all accelerations but gravity.

We can also rely on vector–valued functions to compute the distance travelled. For instance, consider a driver who sets her cruise–control to 60 km/h, and travels at this speed for an hour. We can ask:

1. How far did the driver travel?
2. How far from her starting position is the driver?

The first is easy to answer: she travelled 60 kilometres. The second is impossible to answer with the given information. We do not know if she travelled in a straight line, on an oval racetrack, or along a slowly–winding highway.

This highlights an important fact: to compute distance travelled, we need only to know the speed, given by $\left\|\vec{\boldsymbol{v}}(t)\right\|$.

Definitie 15.11 (Distance travelled)
Let $\vec{\boldsymbol{v}}(t)$ be a velocity function for a moving object. The **distance travelled** (*afgelegde afstand*) by the object on $[a, b]$ is:

$$\text{distance travelled} = \int_a^b \left\|\vec{\boldsymbol{v}}(t)\right\|\ dt.$$

Note that this is just a restatement of Theorem 15.6: arc length is the same as distance travelled, just viewed in a different context.

Example 15.10

A particle moves in space with position function $\vec{\boldsymbol{r}}(t) = \big(t, t^2, \sin(\pi t)\big)$ on $[-2, 2]$, where t is measured in seconds and distances are in meters. Find:

1. The distance travelled by the particle on $[-2, 2]$.
2. The displacement of the particle on $[-2, 2]$.
3. The particle's average speed.

Solution

1. We use Definition 15.11 to establish the integral:

$$\text{distance travelled} = \int_{-2}^{2} \left\|\vec{\boldsymbol{v}}(t)\right\|\ dt$$

$$= \int_{-2}^{2} \sqrt{1+(2t)^2+\pi^2\cos^2(\pi t)}\, dt.$$

This cannot be solved in terms of elementary functions so we turn to numerical integration, finding the distance to be 12.88m.

2. The displacement is the vector

$$\vec{r}(2)-\vec{r}(-2) = (2,4,0)-(-2,4,0) = (4,0,0).$$

That is, the particle ends with an x-value increased by 4 and with y- and z-values the same.

3. We found above that the particle travelled 12.88m over 4 seconds. We can compute average speed by dividing: 12.88/4 = 3.22 m/s. We should also consider Definition 12.5 to compute the average value of the speed as

$$\text{average speed} = \frac{1}{2-(-2)}\int_{-2}^{2} \left\|\vec{v}(t)\right\|\, dt \approx \frac{1}{4}12.88 = 3.22\text{m/s}.$$

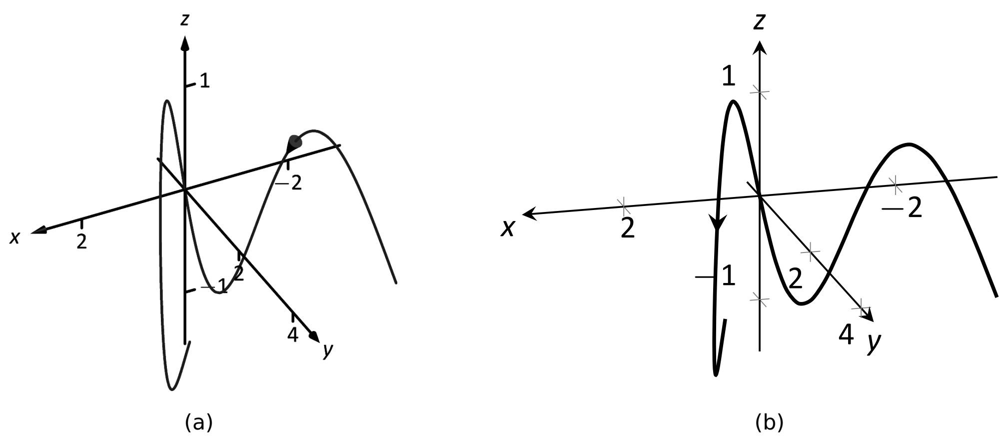


Figure 15.10: The path of the particle, from two perspectives, in Example 15.10.

Note how in Example 15.10 we computed the average speed as the average value of $\left\|\vec{v}(t)\right\|$ on $[-2,2]$. Likewise, given the position function $\vec{r}(t)$, the average velocity on $[a,b]$ is

$$\frac{\text{displacement}}{\text{travel time}} = \frac{1}{b-a}\int_{a}^{b} \vec{r}\,'(t)\, dt = \frac{\vec{r}(b)-\vec{r}(a)}{b-a},$$

that is, it is the average value of $\vec{r}\,'(t)$, or $\vec{v}(t)$, on $[a,b]$.

15.3 Unit tangent and normal vectors

Given a smooth vector–valued function $\vec{r}(t)$, we defined in Definition 15.7 that any vector parallel to $\vec{r}\,'(t_0)$ is tangent to the graph of $\vec{r}(t)$ at $t=t_0$. It is often useful to consider just the direction of $\vec{r}\,'(t)$ and not its magnitude. Therefore we are interested in the unit vector in the direction of $\vec{r}\,'(t)$. This leads to a definition.

Definitie 15.12 (Unit tangent vector)
Let $\vec{\boldsymbol{r}}(t)$ be a smooth function on an open interval I. The **unit tangent vector** $\widehat{\boldsymbol{T}}(t)$ (*eenheids-raakvector*) is

$$\widehat{\boldsymbol{T}}(t) = \frac{1}{\left\|\vec{\boldsymbol{r}}\,'(t)\right\|}\vec{\boldsymbol{r}}\,'(t).$$

Just as knowing the direction tangent to a path is important, knowing a direction orthogonal to a path is important. When dealing with real-valued functions, we defined the normal line at a point to be the line through the point that was perpendicular to the tangent line at that point. We can do a similar thing with vector–valued functions. Given $\vec{\boldsymbol{r}}(t)$ in $\mathbb{R}^2$, we have 2 directions perpendicular to the tangent vector, as shown in Figure 15.11.

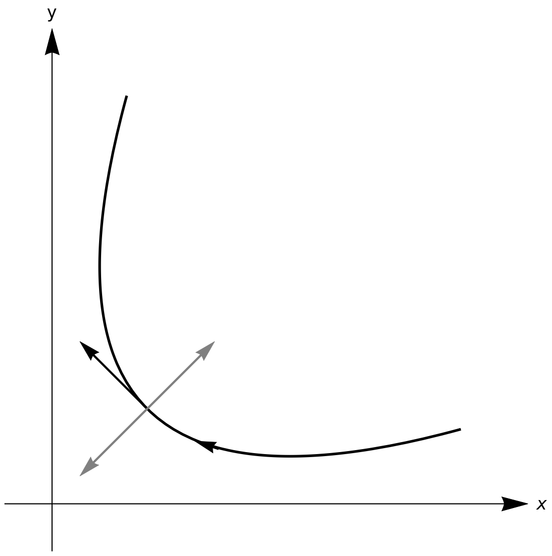


Figure 15.11: Given a direction in the plane, there are always two directions orthogonal to it.

Given $\vec{\boldsymbol{r}}(t)$ in $\mathbb{R}^3$, there are infinitely many vectors orthogonal to the tangent vector at a given point. We might wonder whether one of these infinite choices preferable over the others.

The answer in both $\mathbb{R}^2$ and $\mathbb{R}^3$ is “Yes, there is one vector that is not only preferable, it is the right one to choose. Recall that if $\vec{\boldsymbol{r}}(t)$ has constant length, then $\vec{\boldsymbol{r}}(t)$ is orthogonal to $\vec{\boldsymbol{r}}\,'(t)$ for all t. We know $\widehat{\boldsymbol{T}}(t)$, the unit tangent vector, has constant length. Therefore $\widehat{\boldsymbol{T}}(t)$ is orthogonal to $\widehat{\boldsymbol{T}}\,'(t)$.

We will see that $\widehat{\boldsymbol{T}}\,'(t)$ is more than just a convenient choice of vector that is orthogonal to $\vec{\boldsymbol{r}}\,'(t)$; rather, it is the right choice. Since all we care about is the direction, we define this newly found vector to be a unit vector. Note that if $\widehat{\boldsymbol{T}}(t)$ is a unit vector, this does not imply that $\widehat{\boldsymbol{T}}\,'(t)$ is also a unit vector.

Definitie 15.13 (Unit normal vector)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function where the unit tangent vector, $\widehat{\boldsymbol{T}}(t)$, is smooth on an open interval I. The **unit normal vector** $\widehat{\boldsymbol{N}}(t)$ (*eenheidsnormaalvector*) is

$$\widehat{\boldsymbol{N}}(t) = \frac{1}{\left\|\widehat{\boldsymbol{T}}\,'(t)\right\|}\widehat{\boldsymbol{T}}\,'(t).$$

Example 15.11

Let

$$\vec{r}(t) = \left(t^2 - t, t^2 + t\right).$$

1. Find $\widehat{T}(t)$ and compute $\widehat{T}(0)$ and $\widehat{T}(1)$.
2. Find $\widehat{N}(t)$ and sketch $\vec{r}(t)$ with the unit tangent and normal vectors at $t = -1, 0$ and 1.

Solution

1. We find $\vec{r}\,'(t) = (2t - 1, 2t + 1)$, and

$$\left\|\vec{r}\,'(t)\right\| = \sqrt{(2t-1)^2 + (2t+1)^2} = \sqrt{8t^2 + 2}.$$

   Therefore

$$\widehat{T}(t) = \frac{1}{\left\|\vec{r}\,'(t)\right\|}\vec{r}\,'(t) = \frac{1}{\sqrt{8t^2+2}}(2t-1, 2t+1) = \left(\frac{2t-1}{\sqrt{8t^2+2}}, \frac{2t+1}{\sqrt{8t^2+2}}\right).$$

   When $t = 0$, we have $\widehat{T}(0) = \left(-1/\sqrt{2}, 1/\sqrt{2}\right)$; when $t = 1$, we have $\widehat{T}(1) = \left(1/\sqrt{10}, 3/\sqrt{10}\right)$. We leave it to the reader to verify each of these is a unit vector.

2. Given $\widehat{T}(t)$, finding $\widehat{T}\,'(t)$ requires two applications of the quotient rule:

$$T\,'(t) = \left(\frac{\sqrt{8t^2+2}(2) - (2t-1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2},\right.$$
$$\left.\frac{\sqrt{8t^2+2}(2) - (2t+1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2}\right)$$
$$= \left(\frac{4(2t+1)}{\left(8t^2+2\right)^{3/2}}, \frac{4(1-2t)}{\left(8t^2+2\right)^{3/2}}\right).$$

   This is not a unit vector; to find $\widehat{N}(t)$, we need to divide $\widehat{T}\,'(t)$ by its magnitude:

$$\begin{aligned}\left\|\widehat{T}\,'(t)\right\| &= \sqrt{\frac{16(2t+1)^2}{(8t^2+2)^3} + \frac{16(1-2t)^2}{(8t^2+2)^3}}\\ &= \sqrt{\frac{16(8t^2+2)}{(8t^2+2)^3}}\\ &= \frac{4}{8t^2+2}.\end{aligned}$$

   Finally,

$$\widehat{N}(t) = \frac{1}{\left\|\widehat{T}'(t)\right\|}\widehat{T}'(t) = \frac{1}{4/(8t^2+2)}\left(\frac{4(2t+1)}{\left(8t^2+2\right)^{3/2}}, \frac{4(1-2t)}{\left(8t^2+2\right)^{3/2}}\right)$$

$$= \left(\frac{2t+1}{\sqrt{8t^2+2}}, -\frac{2t-1}{\sqrt{8t^2+2}} \right).$$

Using this formula for $\widehat{\boldsymbol{N}}(t)$, we compute the unit tangent and normal vectors for $t = -1, 0$ and 1 and sketch them in Figure 15.12.

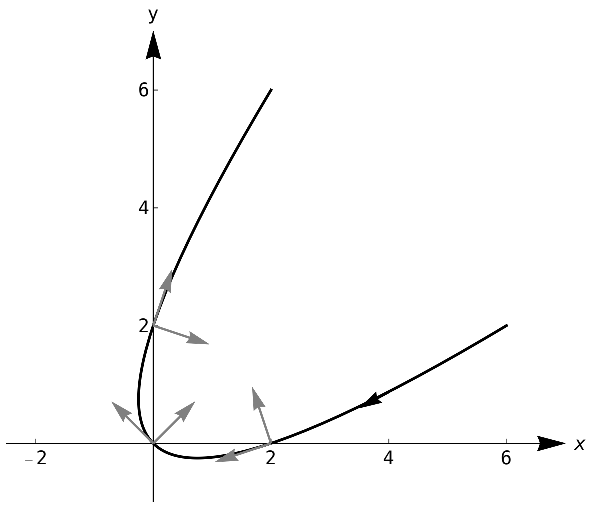


Figure 15.12: Unit tangent and normal vectors from Example 15.11.

The final result for $\widehat{\boldsymbol{N}}(t)$ in Example 15.11 is suspiciously similar to $\widehat{\boldsymbol{T}}(t)$. There is a clear reason for this. If $\vec{\boldsymbol{u}} = (u_1, u_2)$ is a unit vector in $\mathbb{R}^2$, then the only unit vectors orthogonal to $\vec{\boldsymbol{u}}$ are $(-u_2, u_1)$ and $(u_2, -u_1)$. Given $\widehat{\boldsymbol{T}}(t)$, we can quickly determine $\widehat{\boldsymbol{N}}(t)$ if we know which term to multiply by (-1). Consider again Figure 15.12, where we have plotted some unit tangent and normal vectors. Note how $\widehat{\boldsymbol{N}}(t)$ always points inside the curve, or to the concave side of the curve. This is not a coincidence; this is true in general. Knowing the direction that $\vec{\boldsymbol{r}}(t)$ turns allows us to quickly find $\widehat{\boldsymbol{N}}(t)$.

Theorem 15.7 (Unit normal vectors in $\mathbb{R}^2$)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function in $\mathbb{R}^2$ where $\widehat{\boldsymbol{T}}\,'(t)$ is smooth on an open interval I. Let t_0 be in I and $\widehat{\boldsymbol{T}}(t_0) = (t_1, t_2)$ Then $\widehat{\boldsymbol{N}}(t_0)$ is either

$$\widehat{\boldsymbol{N}}(t_0) = (-t_2, t_1) \quad \textit{or} \quad \widehat{\boldsymbol{N}}(t_0) = (t_2, -t_1),$$

whichever is the vector that points to the concave side of the graph of $\vec{\boldsymbol{r}}$.

15.4 Arc length and curvature

15.4.1 Arc length

Currently, our vector–valued functions have defined points with a parameter t, which we often take to represent time. Consider Figure 15.13(a), where $\vec{\boldsymbol{r}}(t) = \left(t^2 - t, t^2 + t\right)$ is graphed and the points corresponding to $t = 0,\ 1$ and 2 are shown. Note how the arc length between $t = 0$ and $t = 1$ is smaller than the arc length between $t = 1$ and $t = 2$; if the parameter t is time and $\vec{\boldsymbol{r}}$ is position, we can say that the particle travelled faster on $[1, 2]$ than on $[0, 1]$.

Now consider Figure 15.13(b), where the same graph is parametrized by a different variable s. Points corresponding to $s = 0$ through $s = 6$ are plotted. The arc length of the graph between each adjacent pair of points is 1. We can view this parameter s as distance; that is, the arc length of the graph from $s = 0$ to $s = 3$ is 3, the arc length from $s = 2$ to $s = 6$ is 4, etc. If one wants to find the point 2.5 units from an initial location (i.e., $s = 0$), one would compute $\vec{r}(2.5)$. This parameter s is very useful, and is called the **arc length parameter**.

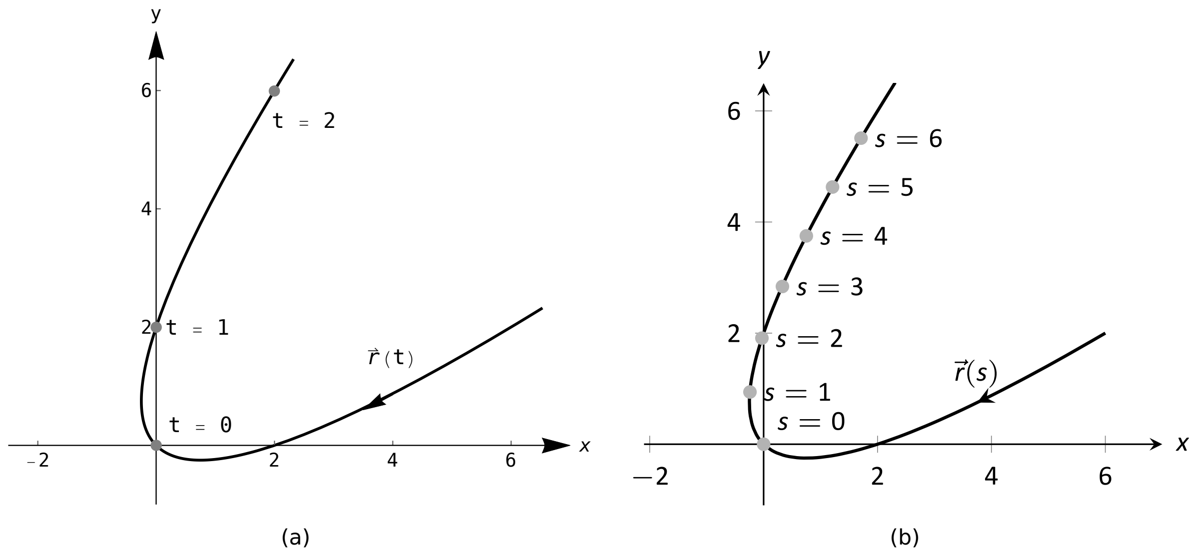


Figure 15.13: Introducing the arc length parameter.

How do we find the arc length parameter?

Start with any parametrization of $\vec{r}$. We can compute the arc length of the graph of $\vec{r}$ on the interval $[0, t]$ with

$$\text{arc length } = \int_0^t \left\|\vec{r}\,'(u)\right\|\, du.$$

We can turn this into a function: as t varies, we find the arc length s from 0 to t. This function is

$$s(t) = \int_0^t \left\|\vec{r}\,'(u)\right\|\, du. \tag{15.2}$$

This establishes a relationship between s and t. Knowing this relationship explicitly, we can rewrite $\vec{r}(t)$ as a function of s: $\vec{r}(s)$. We demonstrate this in an example.

Example 15.12

Let $\vec{r}(t) = \langle 3t-1, 4t+2\rangle$. Parametrize $\vec{r}$ with the arc length parameter s.

Solution

Using Equation (15.2), we write

$$s(t) = \int_0^t \left\|\vec{r}\,'(u)\right\|\, du.$$

We can integrate this, explicitly finding a relationship between s and t:

$$\begin{aligned} s(t) &= \int_0^t \left\| \vec{r}\,'(u) \right\| \, du \\ &= \int_0^t \sqrt{3^2+4^2}\, du \\ &= \int_0^t 5\, du \\ &= 5t. \end{aligned}$$

Since $s = 5t$, we can write $t = s/5$ and replace t in $\vec{r}(t)$ with $s/5$:

$$\vec{r}(s) = \left(3\left(\frac{s}{5}\right)-1, 4\left(\frac{s}{5}\right)+2\right) = \left(\frac{3}{5}s-1, \frac{4}{5}s+2\right).$$

Clearly, as shown in Figure 15.14, the graph of $\vec{r}$ is a line, where $t = 0$ corresponds to the point $(-1, 2)$. What point on the line is 2 units away from this initial point? We find it with $\vec{r}(2) = (1/5, 18/5)$.

Is the point $(1/5, 18/5)$ really 2 units away from $(-1, 2)$? We use the distance formula to check:

$$d = \sqrt{\left(\frac{1}{5}-(-1)\right)^2 + \left(\frac{18}{5}-2\right)^2} = \sqrt{\frac{36}{25}+\frac{64}{25}} = \sqrt{4} = 2.$$

Yes, $\vec{r}(2)$ is indeed 2 units away, in the direction of travel, from the initial point.

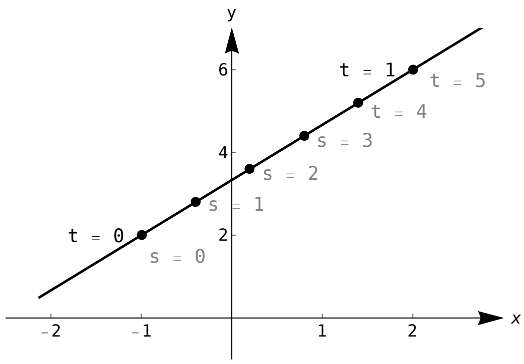


Figure 15.14: Graphing $\vec{r}$ in Example 15.12 with parameters t and s.

Things worked out very nicely in Example 15.12; we were able to establish directly that $s = 5t$. Usually, the arc length parameter is much more difficult to describe in terms of t, a result of integrating a square–root. There are a number of things that we can learn about the arc length parameter from Equation (15.2), though, that are useful.

First, take the derivative of s with respect to t. The fundamental theorem of calculus (see Theorem 12.7) states that

$$\frac{ds}{dt} = s'(t) = \left\| \vec{r}\,'(t) \right\|. \tag{15.3}$$

Letting t represent time and $\vec{r}(t)$ represent position, we see that the rate of change of s with respect to t is speed; that is, the rate of change of distance travelled is speed, which should match our intuition.

The chain rule states that

$$\frac{d\vec{r}}{dt} = \frac{d\vec{r}}{ds}\frac{ds}{dt}$$
$$\vec{r}\,'(t) = \vec{r}\,'(s)\|\vec{r}\,'(t)\|.$$

Solving for $\vec{r}\,'(s)$, we have

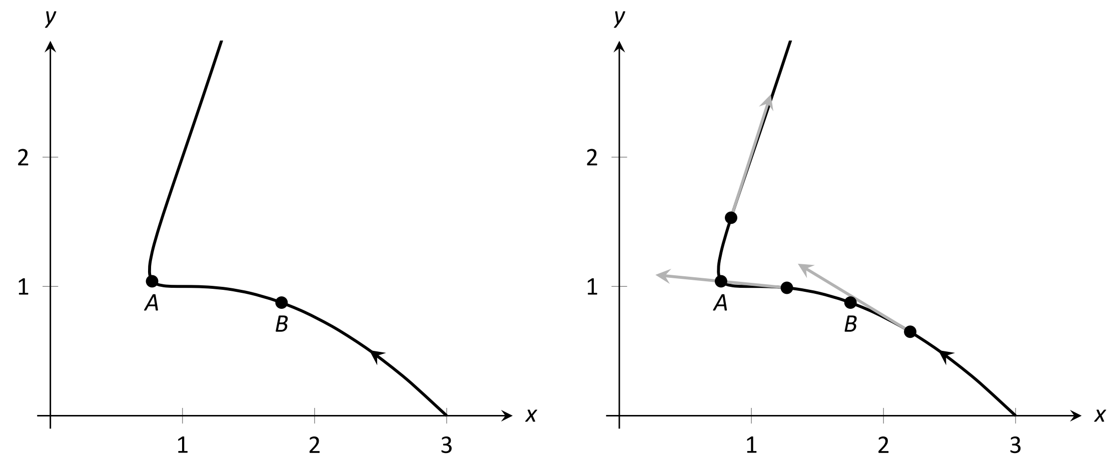
$$\vec{r}\,'(s) = \frac{\vec{r}\,'(t)}{\|\vec{r}\,'(t)\|} = \widehat{T}(t), \tag{15.4}$$

where $\widehat{T}(t)$ is the unit tangent vector. Equation (15.4) is often misinterpreted, as one is tempted to think it states $\vec{r}\,'(t) = \widehat{T}(t)$, but there is a big difference between $\vec{r}\,'(s)$ and $\vec{r}\,'(t)$. The key to take from it is that $\vec{r}\,'(s)$ is a unit vector. In fact, the following definition states that this characterizes the arc length parameter.

Definitie 15.14 (Arc length parameter)
Let $\vec{r}(s)$ be a vector–valued function. The parameter s is the **arc length parameter** if, and only if, $\|\vec{r}\,'(s)\| = 1$.

15.4.2 Curvature

Consider points A and B on the curve graphed in Figure 15.15(a). One can readily argue that the curve curves more sharply at A than at B. It is useful to use a number to describe how sharply the curve bends; that number is the **curvature** (*kromming*) of the curve.



(a) (b)

Figure 15.15: Establishing the concept of curvature.

We derive this number in the following way. Consider Figure 15.15(b), where unit tangent vectors are graphed around points A and B. Notice how the direction of the unit tangent vector changes quite a bit near A, whereas it does not change as much around B. This leads to an important concept: measuring the rate of change of the unit tangent vector with respect to arc length gives us a measurement of curvature.

Definitie 15.15 (Curvature)
Let $\vec{r}(s)$ be a vector–valued function where s is the arc length parameter. The **curvature** κ of the graph of $\vec{r}(s)$ is

$$\kappa(t) = \left\| \frac{d\widehat{\boldsymbol{T}}}{ds} \right\| = \left\| \widehat{\boldsymbol{T}}'(s) \right\|.$$

If $\vec{r}(s)$ is parametrized by the arc length parameter, then

$$\widehat{\boldsymbol{T}}(s) = \frac{\vec{r}\,'(s)}{\|\vec{r}\,'(s)\|} \quad \text{and} \quad \widehat{\boldsymbol{N}}(s) = \frac{\widehat{\boldsymbol{T}}'(s)}{\left\|\widehat{\boldsymbol{T}}'(s)\right\|}.$$

Having defined $\left\|\widehat{\boldsymbol{T}}'(s)\right\| = \kappa$, we can rewrite the second equation as

$$\widehat{\boldsymbol{T}}'(s) = \kappa \widehat{\boldsymbol{N}}(s). \tag{15.5}$$

We already knew that $\widehat{\boldsymbol{T}}'(s)$ is in the same direction as $\widehat{\boldsymbol{N}}(s)$; that is, we can think of $\widehat{\boldsymbol{T}}(s)$ as being pulled in the direction of $\widehat{\boldsymbol{N}}(s)$. How hard is it being pulled? By a factor of κ. When the curvature is large, $\widehat{\boldsymbol{T}}(s)$ is being pulled hard and the direction of $\widehat{\boldsymbol{T}}(s)$ changes rapidly. When κ is small, $T(s)$ is not being pulled hard and hence its direction is not changing rapidly.

Example 15.13

Find the curvature of $\vec{r}(t) = (3t-1, 4t+2)$.

Solution

In Example 15.12, we found that the arc length parameter was defined by $s = 5t$, so $\vec{r}(s) = (3s/5-1, 4s/5+2)$ parametrized $\vec{r}$ with the arc length parameter. To find κ, we need to find $\widehat{\boldsymbol{T}}'(s)$.

$$\begin{aligned} \widehat{\boldsymbol{T}}(s) &= \vec{r}\,'(s) \qquad \text{(recall this is a unit vector)} \\ &= \left(\frac{3}{5}, \frac{4}{5}\right). \end{aligned}$$

Therefore

$$\widehat{\boldsymbol{T}}'(s) = (0, 0)$$

and

$$\kappa = \left\| \widehat{\boldsymbol{T}}'(s) \right\| = 0.$$

It probably comes as no surprise that the curvature of a line is 0.

While the definition of curvature is a beautiful mathematical concept, it is nearly impossible to use most of the time; writing $\vec{r}$ in terms of the arc length parameter is generally very hard. Fortunately, there are other methods of calculating this value that are much easier.

Theorem 15.8 (Formulas for curvature)
Let C be a smooth curve in the plane or in space.

1. *If C is defined in space by a vector–valued function* $\vec{\boldsymbol{r}}(t)$*, then*

$$\kappa(t) = \frac{\left\|\widehat{\boldsymbol{T}}'(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|} = \frac{\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|^3} = \frac{\vec{\boldsymbol{a}}(t)\cdot\widehat{\boldsymbol{N}}(t)}{\left\|\vec{\boldsymbol{v}}(t)\right\|^2}.$$

2. *If C is defined by* $y = f(x)$*, then*

$$\kappa = \frac{|f''(x)|}{\left(1+\left(f'(x)\right)^2\right)^{3/2}}.$$

3. *If C is defined as a vector–valued function in the plane,* $\vec{\boldsymbol{r}}(t) = \left(x(t), y(t)\right)$*, then*

$$\kappa = \frac{|x'y''-x''y'|}{\left((x')^2+(y')^2\right)^{3/2}}.$$

Proof To show the first statement in this theorem, we resort to the chain rule, i.e.

$$\frac{d\widehat{\boldsymbol{T}}}{dt} = \frac{d\widehat{\boldsymbol{T}}}{ds}\frac{ds}{dt} = \frac{d\widehat{\boldsymbol{T}}}{ds}\left\|\vec{\boldsymbol{r}}'(t)\right\|,$$

or after rearranging

$$\frac{d\widehat{\boldsymbol{T}}}{ds} = \frac{\frac{d\widehat{\boldsymbol{T}}}{dt}}{\left\|\vec{\boldsymbol{r}}'(t)\right\|} = \frac{\widehat{\boldsymbol{T}}'(t)}{\left\|\vec{\boldsymbol{r}}'(t)\right\|}.$$

Consequently, we find that

$$\kappa(t) = \left\|\frac{d\widehat{\boldsymbol{T}}}{ds}\right\| = \frac{\left\|\widehat{\boldsymbol{T}}'(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|}.$$

Subsequently, we can show that

$$\kappa(t) = \frac{\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|^3}$$

by expressing $\vec{\boldsymbol{r}}'(t)$ and $\vec{\boldsymbol{r}}''(t)$ in terms of $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{T}}'$ and subsequently compute their cross product. For what concerns $\vec{\boldsymbol{r}}'(t)$, we recall that $\frac{ds}{dt} = \left\|\vec{\boldsymbol{r}}'(t)\right\|$ and $\widehat{\boldsymbol{T}}(t)\frac{\vec{\boldsymbol{r}}'(t)}{\left\|\vec{\boldsymbol{r}}'(t)\right\|}$, so we infer

$$\vec{\boldsymbol{r}}'(t) = \frac{ds}{dt}\widehat{\boldsymbol{T}}.$$

Computing the derivative of this expression with respect to t yields

$$\vec{\boldsymbol{r}}''(t) = \frac{d^2s}{dt^2}\widehat{\boldsymbol{T}} + \frac{ds}{dt}\widehat{\boldsymbol{T}}'$$

so that the cross product $\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)$ becomes

$$\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t) = \frac{ds}{dt}\frac{d^2s}{dt^2}\left(\widehat{\boldsymbol{T}}\times\widehat{\boldsymbol{T}}\right) + \left(\frac{ds}{dt}\right)^2\left(\widehat{\boldsymbol{T}}\times\widehat{\boldsymbol{T}}'\right).$$

Since the cross product of a vector by itself is always the zero vector,we see that

$$\begin{aligned}\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\| &= \left(\frac{ds}{dt}\right)^2\left\|\widehat{\boldsymbol{T}}\times\widehat{\boldsymbol{T}}'\right\| \\ &= \left(\frac{ds}{dt}\right)^2\left\|\widehat{\boldsymbol{T}}\right\|\left\|\widehat{\boldsymbol{T}}'\right\|\sin(\theta),\end{aligned}$$

where θ is the angle between $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{T}}'$. Since $\left\|\widehat{\boldsymbol{T}}\right\| = 1$ as a consequence of Equation (15.4) and Definition 15.14, and $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{T}}'$ are perpendicular, we get that

$$\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\| = \left(\frac{ds}{dt}\right)^2\left\|\widehat{\boldsymbol{T}}'\right\| = \left\|\vec{\boldsymbol{r}}'(t)\right\|^2\left\|\widehat{\boldsymbol{T}}'\right\|.$$

Therefore,

$$\left\|\widehat{\boldsymbol{T}}'\right\| = \frac{\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|^2}$$

and

$$\kappa(t) = \frac{\left\|\widehat{\boldsymbol{T}}'\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|} = \frac{\left\|\vec{\boldsymbol{r}}'(t)\times\vec{\boldsymbol{r}}''(t)\right\|}{\left\|\vec{\boldsymbol{r}}'(t)\right\|^3}. \tag{15.6}$$

The proof of the second statement in the theorem follows easily from Equation (15.6) after acknowledging that it is easy to parameterise the curve given by $y = f(x)$ as a 3D parametric curve. Indeed, we can simply use $x = x$, $y = f(x)$ and $z = 0$, where x is considered a parameter.

Starting from Equation (15.6) and letting $\vec{\boldsymbol{r}}(t) = \big(x(t), y(t)\big)$ we easily get the third statement in the theorem. □

Example 15.14

Find the curvature of a circle with radius r, defined by $\vec{\boldsymbol{c}}(t) = \big(r\cos(t), r\sin(t)\big)$.

Solution

Before we start, we should expect the curvature of a circle to be constant, and not dependent on t.

We compute κ using the second part of Theorem 15.8:

$$\begin{aligned}\kappa &= \frac{\Big|\big(-r\sin(t)\big)\big(-r\sin(t)\big) - \big(-r\cos(t)\big)\big(r\cos(t)\big)\Big|}{\Big(\big(-r\sin(t)\big)^2 + \big(r\cos(t)\big)^2\Big)^{3/2}} \\ &= \frac{r^2\big(\sin^2(t)+\cos^2(t)\big)}{\Big(r^2\big(\sin^2(t)+\cos^2(t)\big)\Big)^{3/2}} \\ &= \frac{r^2}{r^3} = \frac{1}{r}.\end{aligned}$$

We have found that a circle with radius r has curvature $\kappa = 1/r$.

Example 15.14 gives a great result. Before this example, if we were told "The curve has a curvature of 5 at point A," we would have no idea what this really meant. Is 5 big – does is correspond to a really sharp turn, or a not-so-sharp turn? Now we can think of 5 in terms of a circle with radius 1/5. Knowing the units allows us to determine how sharply the curve is curving.

Let a point P on a smooth curve C be given, and let κ be the curvature of the curve at P. A circle that:

- passes through P,
- lies on the concave side of C,
- has a common tangent line as C at P, and
- has radius $r = 1/\kappa$ (hence has curvature κ)

is the **osculating circle** (*osculatiecirkel*), or **circle of curvature**, to C at P, and r is the **radius of curvature** (*kromtestraal*). Figure 15.16 shows the graph of the curve seen earlier in Figure 15.15(a) and its osculating circles at A and B. A sharp turn corresponds to a circle with a small radius; a gradual turn corresponds to a circle with a large radius. Being able to think of curvature in terms of the radius of a circle is very useful. The word osculating comes from a Latin word related to kissing; an osculating circle kisses the graph at a particular point. Many beautiful ideas in mathematics have come from studying the osculating circles to a curve.

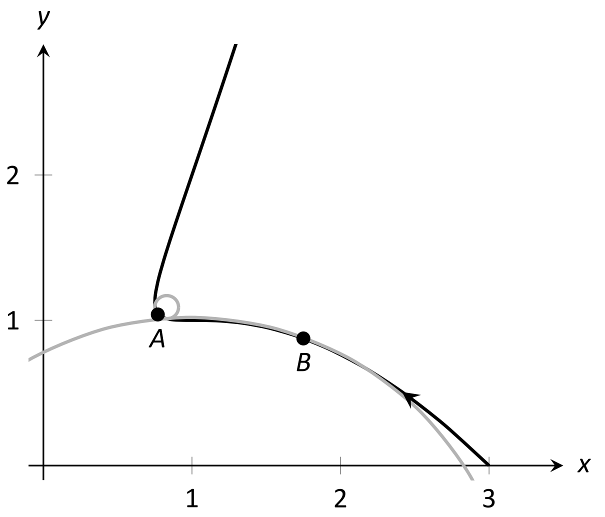


Figure 15.16: Illustrating the osculating circles for the curve seen in Figure 15.15(a).

Example 15.15

Find where the curvature of $\vec{r}(t) = \left(t, t^2, 2t^3\right)$ is maximized.

Solution

We use the third formula in Theorem 15.8 as $\vec{r}(t)$ is defined in space. We leave it to the reader to verify that

$$\vec{r}\,'(t) = \left(1, 2t, 6t^2\right), \quad \vec{r}\,''(t) = (0, 2, 12t), \quad \text{and} \quad \vec{r}\,'(t) \times \vec{r}\,''(t) = \left(12t^2, -12t, 2\right).$$

Thus

$$\kappa(t) = \frac{\left\|\vec{r}\,'(t) \times \vec{r}\,''(t)\right\|}{\left\|\vec{r}\,'(t)\right\|^3}$$

$$= \frac{\left\| \left(12t^2, -12t, 2 \right) \right\|}{\left\| \left(1, 2t, 6t^2 \right) \right\|^3}$$

$$= \frac{\sqrt{144t^4 + 144t^2 + 4}}{\left(\sqrt{1 + 4t^2 + 36t^4} \right)^3}.$$

While this is not a particularly nice formula, it does explicitly tell us what the curvature is at a given t value. To maximize $\kappa(t)$, we should solve $\kappa'(t) = 0$ for t. This is doable, but time consuming. Instead, consider the graph of $\kappa(t)$ as given in Figure 15.17(a). We see that κ is maximized at two t values; using a numerical solver, we find these values are $t \approx \pm 0.189$. In Figure 15.17(b) we graph $\vec{r}(t)$ and indicate the points where curvature is maximized.

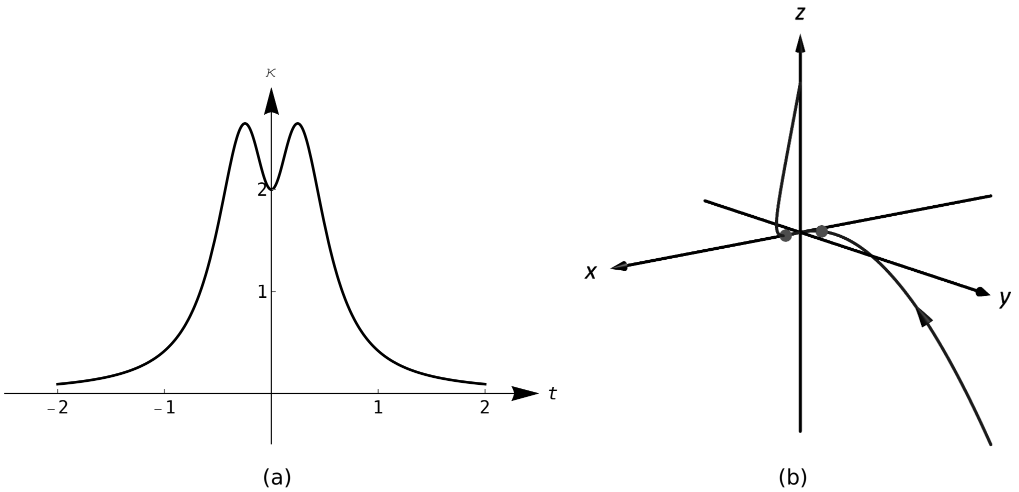


Figure 15.17: Understanding the curvature of a curve in space.

We started this chapter with vector–valued functions, which may have seemed at the time to be just another way of writing parametric equations. However, we have seen that the vector perspective has given us great insight into the behaviour of functions and the study of motion. Vector–valued position functions convey displacement, distance travelled, speed, velocity, acceleration and curvature information, each of which has great importance in science and engineering.

15.5 Exercices

Algebra of vector–valued functions

Assignment 15.1 — Show that the vector functions below all represent the same curve. Which curve is being described here?

$$\vec{\boldsymbol{r}}_1(t) = \big(\sin(t), \cos(t)\big) \qquad \text{with} \quad -\frac{\pi}{2} \leq t \leq \frac{\pi}{2}$$

$$\vec{\boldsymbol{r}}_2(t) = \left(t-1, \sqrt{2t-t^2}\right) \qquad \text{with} \quad 0 \leq t \leq 2$$

$$\vec{\boldsymbol{r}}_3(t) = \left(t\sqrt{2-t^2}, 1-t^2\right) \qquad \text{with} \quad -1 \leq t \leq 1$$

Calculus and vector–valued functions

Assignment 15.2 — Find the limits below.

(a) $\lim\limits_{t\to 5}\big(2t+1, 3t^2-1, \sin(t)\big)$

(b) $\lim\limits_{t\to 3}\left(e^t, \dfrac{t^2-9}{t+3}\right)$

(c) $\lim\limits_{t\to 0}\left(\dfrac{t}{\sin(t)}, (1+t)^{\frac{1}{t}}\right)$

(d) $\lim\limits_{h\to 0}\dfrac{\vec{\boldsymbol{r}}\,(t+h)-\vec{\boldsymbol{r}}\,(t)}{h} \qquad \text{with} \quad \vec{\boldsymbol{r}}(t) = \left(t^2, t, 1\right)$

Assignment 15.3 — Find the derivative of the vector functions below.

(a) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{1}{t}, \dfrac{2t-1}{3t+1}, \tan(t)\right)$

(b) $\vec{\boldsymbol{r}}(t) = \left(t^2\right)\big(\sin(t), 2t+5\big)$

(c) $\vec{\boldsymbol{r}}(t) = \left(t^2+1, t-1\right)\cdot\big(\sin(t), 2t+5\big)$

(d) $\vec{\boldsymbol{r}}(t) = \left(t^2+1, t-1, 1\right)\times\big(\sin(t), 2t+5, 1\big)$

Assignment 15.4 — Determine the values of t at which $\vec{\boldsymbol{r}}(t)$ is not smooth.

(a) $\vec{\boldsymbol{r}}(t) = \big(\cos(t), \sin(t)-t\big)$

(b) $\vec{\boldsymbol{r}}(t) = \left(t^2-2t+1, t^3+t^2-5t+3\right)$

(c) $\vec{\boldsymbol{r}}(t) = \big(\cos(t)-\sin(t), \sin(t)-\cos(t), \cos(4t)\big)$

(d) $\vec{\boldsymbol{r}}(t)=\left(t^3-3t+2,-\cos(\pi t),\sin^2(\pi t)\right)$

Assignment 15.5 — Evaluate the integrals below.

(a) $\displaystyle\int \left(t^3,\cos(t),te^t\right)\,dt$

(b) $\displaystyle\int \left(\frac{1}{1+t^2},\sec^2(t)\right)\,dt$

(c) $\displaystyle\int \left(\cos(t)e^{\sin(t)},t\sin^2(t),-1\right)\,dt$

(d) $\displaystyle\int_0^{\pi} \left(\sin^2(t)\cos(t),\cos^2(t)\sin(t)\right)\,dt$

(e) $\displaystyle\int_0^{1} \left(\frac{1}{2}e^{-\frac{t}{2}},\frac{1}{2}e^{\frac{t}{2}},e^t\right)\,dt$

The calculus of motion

Assignment 15.6 — Determine the location vector of an object if its acceleration, initial velocity and position are given.

(a) $\vec{\boldsymbol{a}}(t)=(2,3),\quad \vec{\boldsymbol{v}}(1)=(1,2),\quad \vec{\boldsymbol{r}}(1)=(5,-2)$

(b) $\vec{\boldsymbol{a}}(t)=\left(\cos(t),-\sin(t)\right)\quad \vec{\boldsymbol{v}}(0)=(0,1),\quad \vec{\boldsymbol{r}}(0)=(0,0)$

(c) $\vec{\boldsymbol{a}}(t)=(0,-32),\quad \vec{\boldsymbol{v}}(0)=(10,50),\quad \vec{\boldsymbol{r}}(0)=(0,0)$

Assignment 15.7 — Determine the velocity vector $\vec{\boldsymbol{v}}(t)$, the velocity $\left\|\vec{\boldsymbol{v}}(t)\right\|$ and the acceleration $\vec{\boldsymbol{a}}(t)$ at time t of the object with position function $\vec{\boldsymbol{r}}(t)$. Also describe the path followed by the object.

(a) $\vec{\boldsymbol{r}}(t)=(1,t)$

(b) $\vec{\boldsymbol{r}}(t)=\left(0,t^2,t\right)$

(c) $\vec{\boldsymbol{r}}(t)=(1,t,t)$

(d) $\vec{\boldsymbol{r}}(t)=\left(t^2,-t^2,1\right)$

(e) $\vec{\boldsymbol{r}}(t)=\left(3\cos(t),4\cos(t),5\sin(t)\right)$

(f) $\vec{\boldsymbol{r}}(t)=\left(3\cos(t),4\sin(t),t\right)$

Assignment 15.8 — An object moves at a constant speed of 5 to the right along the parabola $y=x^2$. Determine the velocity vector $\vec{\boldsymbol{v}}(t)$ and the acceleration $\vec{\boldsymbol{a}}(t)$ of the object at $(1,1)$.

Assignment 15.9 — Show that the velocity of a moving object remains constant over a time interval if and only if its acceleration is perpendicular to the velocity.

Unit tangent and normal vectors

Assignment 15.10 — Find the unit tangent vector $\widehat{\boldsymbol{T}}(t)$ of the curves below.

(a) $\vec{\boldsymbol{r}}(t) = \left(2t^2, t^2 - t\right)$

(b) $\vec{\boldsymbol{r}}(t) = \left(t, -2t^2, 3t^3\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(t, \dfrac{t^2}{2}, \dfrac{t^3}{3}\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(\cos(t)\sin(t), \sin^2(t), \cos(t)\right)$

(e) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{\cos^3(t)}{3}, \dfrac{\sin^3(t)}{3}\right)$ in $t = \dfrac{\pi}{6}$

Assignment 15.11 — Find the unite normal vector $\widehat{\boldsymbol{N}}(t)$ of the curves below.

(a) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{t^3}{3} - t, t^2\right)$ in $t = 3$

(b) $\vec{\boldsymbol{r}}(t) = \left(3\cos(t), 3\sin(t)\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(e^t, e^{-t}\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(4t, 2\sin(t), 2\cos(t)\right)$

(e) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{\cos^3(t)}{3}, \dfrac{\sin^3(t)}{3}\right)$ in $t = \dfrac{\pi}{6}$

(f) $\vec{\boldsymbol{r}}(t) = \left(a\cos(t), a\sin(t), bt\right)$, $a > 0$

Arc length and curvature

Assignment 15.12 — Determine the arc length of the curves below between the indicated points.

(a) $\vec{\boldsymbol{r}}(t) = \left(t^2, t^2, t^3\right)$ between $t = 0$ and $t = 1$

(b) $\vec{\boldsymbol{r}}(t) = \left(e^t\cos(t), e^t\sin(t), t\right)$ between $t = 0$ and $t = 2\pi$

(c) $\vec{\boldsymbol{r}}(t) = \left(t^3, t^2\right)$ between $t = -1$ and $t = 2$

Assignment 15.13 — Parameterize the vector functions below with the arc length parameter s starting from the point where $t = 0$.

(a) $\vec{\boldsymbol{r}}(t) = (2t, t, -2t)$

(b) $\vec{\boldsymbol{r}}(t) = \left(7\cos(t), 7\sin(t)\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(3\cos(t), 3\sin(t), 2t\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(e^t, \sqrt{2}t, -e^{-t}\right)$

Assignment 15.14 — Parameterize the first arc of the cycloid $\vec{\boldsymbol{r}}(\theta) = \left(a(\theta - \sin(\theta)), a(1 - \cos(\theta))\right)$ $(0 \le \theta \le 2\pi)$ with the arc length parameter s.

Assignment 15.15 — Determine the radius of curvature r of the curves below at the indicated points.

(a) $y = x^2$ at $x = 0$ and $x = \sqrt{2}$

(b) $y = \cos(x)$ at $x = 0$ en $x = \pi/2$

(c) $y = \tan(x)$ at $x = \pi/4$

(d) $\vec{r}(t) = \left(2t, \frac{1}{t}, -2t\right)$ at $(2, 1, -2)$

(e) $\vec{r}(t) = \left(t^3, t^2, t\right)$ at $t = 1$

(f) $16y^2 = 4x^4 - x^6$ at $x = 2$

(g) $\vec{r}(t) = \left(3t^2, 3t - t^3\right)$ at $t = 1$

(h) $\vec{r}(t) = \left(\cos(t), \sin(3t)\right)$ at $t = 0$

Assignment 15.16 — Determine thecurvature κ and the radius of curvature r at a generic point on the given curve.

(a) $y = \dfrac{1}{x^2 + 1}$

(b) $y = \sqrt{1 - x^2}$

(c) $x(t) = 2 + \sqrt{2}\cos(t), \quad y(t) = 1 - \sin(t), \quad z(t) = 3 + \sin(t)$

(d) $y = e^x$

(e) $r(\theta) = a(1 - \cos(\theta))$

(f) $\vec{r}(t) = \left(2\cos(t), \sin(t)\right)$

(g) $\vec{r}(t) = \left(t, \ln(\sin(t))\right)$ with $0 < t < \pi$

Assignment 15.17 — Show that for the catenary $y = a\cosh\left(\frac{x}{a}\right)$ the radius of curvature r is given by y^2/a.

Assignment 15.18 — Determine the radius of curvature of the cycloid.

$$\begin{cases} x(\theta) = r(\theta - \sin(\theta)) \\ y(\theta) = r(1 - \cos(\theta)) \end{cases}$$

at $\theta = \dfrac{\pi}{2}$.

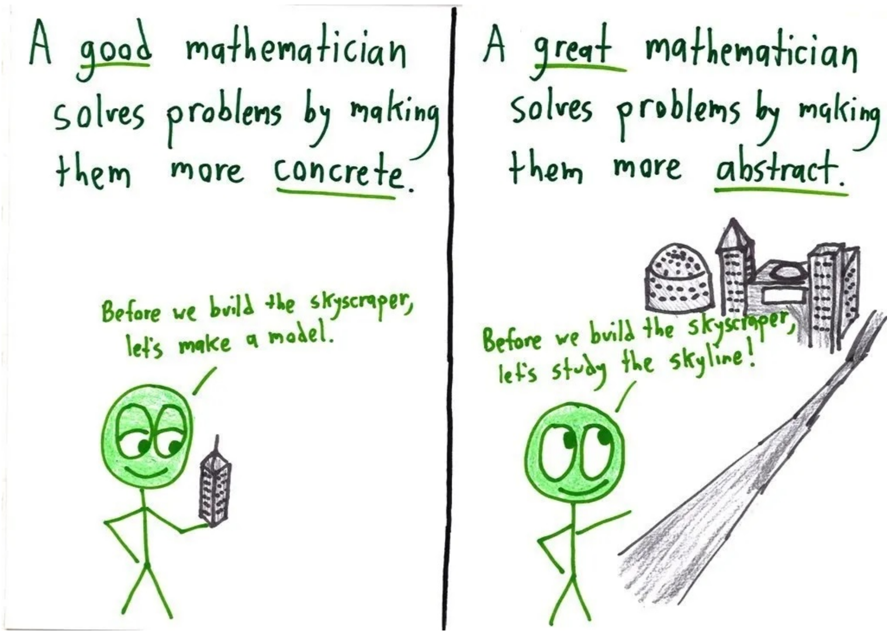
Review exercises

Assignment 15.19 — Determine the requested parametrization of the circle $x^2 + y^2 = a^2$ in the first quadrant.

(a) in terms of the y-coordinate, counterclockwise orientation

(b) in terms of the angle between the tangent at a point (x, y) and the positive x-axis, counterclockwise orientation

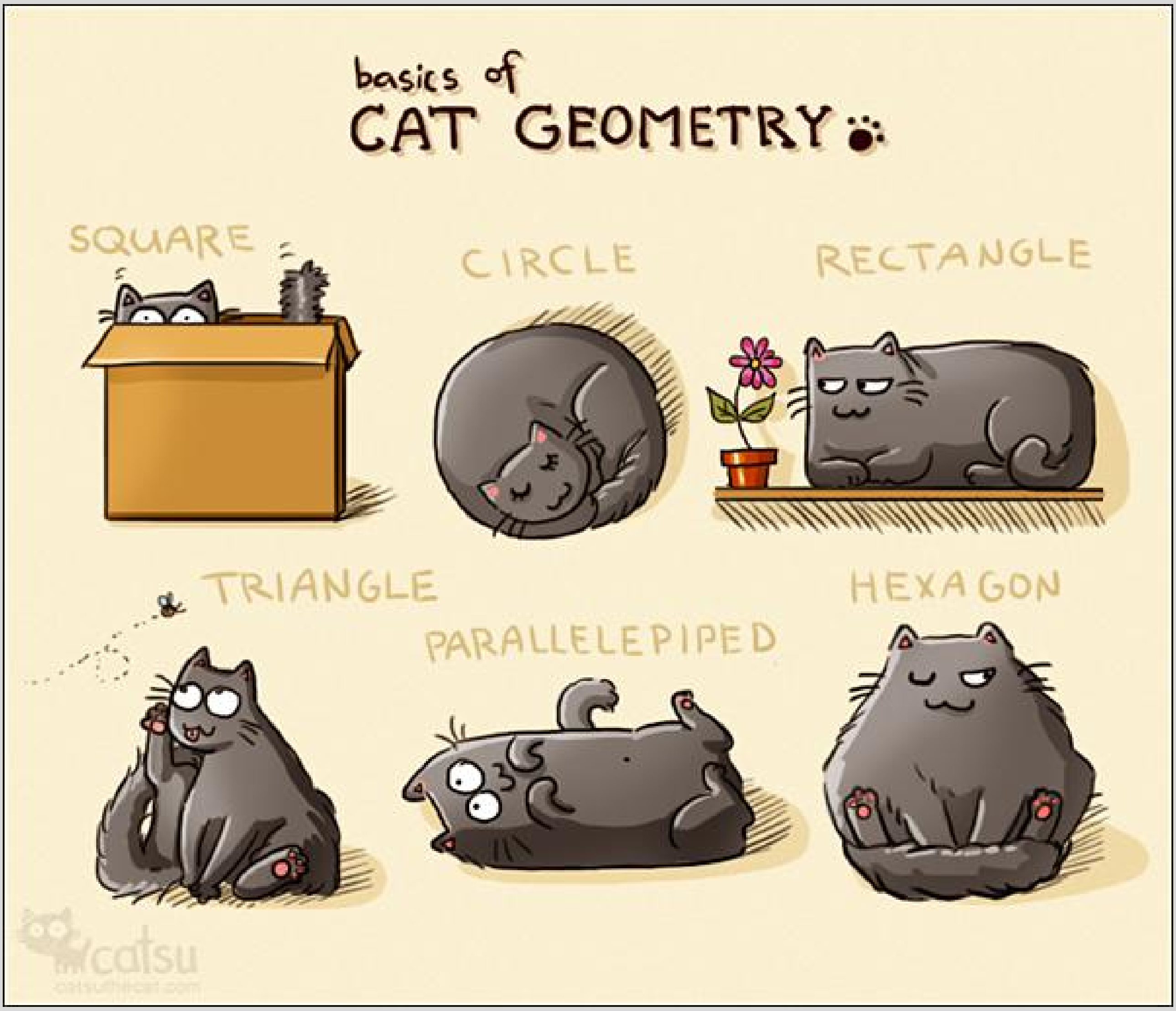
(c) in terms of the arc length measured from $(0, a)$, clockwise orientation



From *Math with Bad Drawings*, used by permission of Ben Orlin.

PART III

MULTIVARIABLE CALCULUS



As you will find in multivariable calculus, there is often a number of solutions for any given problem.

— John Nash —

16

Functions of several variables

A function of the form $y = f(x)$ is a function of a single variable; given a value of x, we can find a value y. Even the vector–valued functions of Chapter 15 are single–variable functions; the input is a single variable though the output is a vector.

There are many situations where a desired quantity is a function of two or more variables. For instance, wind chill is measured by knowing the temperature and wind speed; the volume of a gas can be computed knowing the pressure and temperature of the gas; to compute a baseball player's batting average, one needs to know the number of hits and the number of at–bats.

This chapter studies multivariable functions, that is, functions with more than one input.

16.1 Introduction to multivariable functions

16.1.1 Functions of two variables

We start with a definition of a function of two variables.

Definitie 16.1 (Function of two variables)
Let D be a subset of $\mathbb{R}^2$. A **function f of two variables** (*functie van twee veranderlijken*) is a rule that assigns each pair (x,y) in D a value $z = f(x,y)$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

Example 16.1

Let

$$f(x,y) = \sqrt{1 - \frac{x^2}{9} - \frac{y^2}{4}}.$$

Find the domain and range of f.

Solution

The domain is all pairs (x,y) allowable as input in f. Because of the square root, we need (x,y) such that:

$$\begin{aligned} 1-\frac{x^2}{9}-\frac{y^2}{4} &\geq 0 \\ \Leftrightarrow \qquad \frac{x^2}{9}+\frac{y^2}{4} &\leq 1 \end{aligned}$$

The above equation describes an ellipse and its interior. We can represent the domain D in set notation as

$$D=\left\{(x,y)\;\middle|\;\frac{x^2}{9}+\frac{y^2}{4}\leq 1\right\}.$$

The range is the set of all possible output values. The square root ensures that all output is positive. Since the x and y terms are squared, then subtracted, inside the square root, the largest output value comes at $x=0$, $y=0$: $f(0,0)=1$. Thus the range R is the interval $[0,1]$.

Definitie 16.2 (Graph of a function of two variables)
The **graph** of a function f of two variables is the set of all points $\big(x,y,f(x,y)\big)$ where (x,y) is in the domain of f. This creates a **surface** (*oppervlak*) in space.

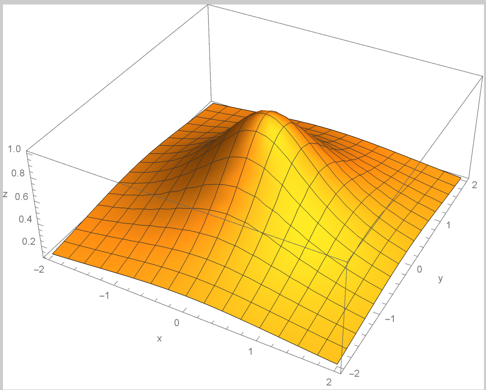
One can begin sketching a graph by plotting points, but this has limitations. Consider Figure 16.1(a) where 25 points have been plotted of

$$f(x,y)=\frac{1}{x^2+y^2+1}.$$

More points have been plotted than one would reasonably want to do by hand, yet it is not clear at all what the graph of the function looks like. Technology allows us to plot lots of points, connect adjacent points with lines and add shading to create a graph like Figure 16.1(b) which does a far better job of illustrating the behaviour of f. More specifically, in Mathematica, a function of two variables can be plotted using the command **Plot3D** as follows

```
In[25]:=  Plot3D[{1/(x^2 + y^2 + 1)}, {x, -2, 2}, {y, -2, 2},  AxesLabel → {"x", "y", "z"}]
```

Out[25]=



Of course, many options are available to format such graphs according to one's preference. These can be checked in the Documentation Center.

While technology is readily available to help us graph functions of two variables, there is still a paper–and–pencil approach that is useful to understand and master as it, combined with high–quality graph-

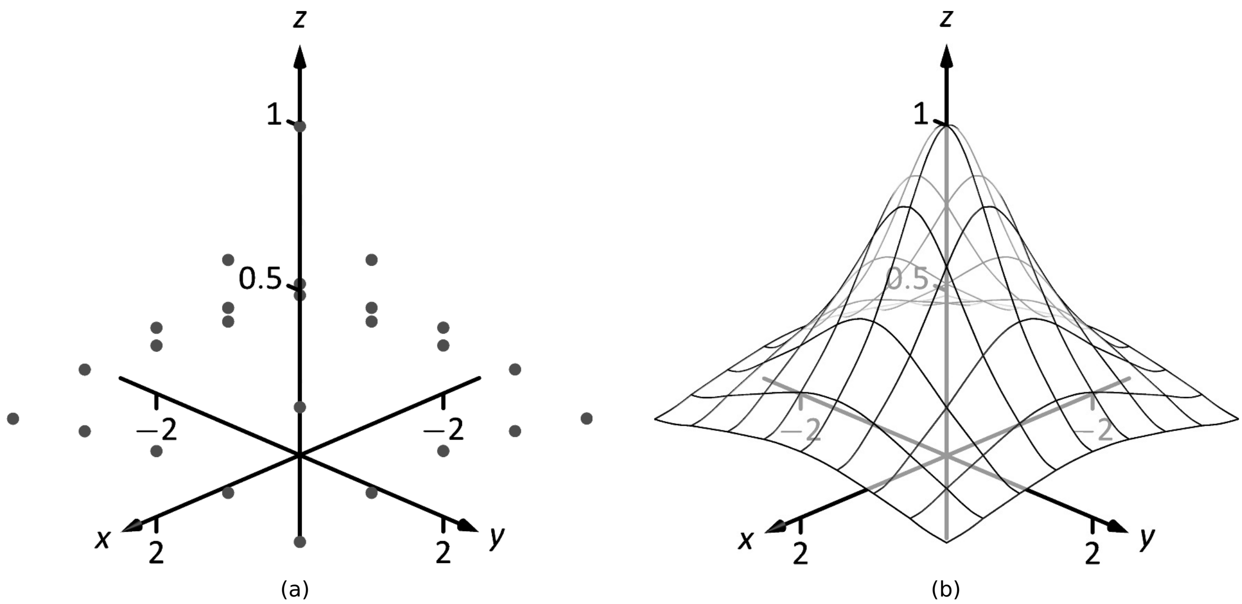


Figure 16.1: Graphing a function of two variables.

ics, gives one great insight into the behaviour of a function. This technique is known as sketching **level curves** (*niveaukromme*).

It may be surprising to find that the problem of representing a three dimensional surface on paper is familiar to most people. Topographical maps, like the one of Dinant shown in Figure 16.2, represent the surface of Earth by indicating points with the same elevation with **contour lines** (*countourlijn*). The elevations marked are equally spaced; in this example, each thin line indicates an elevation change in 10m increments and each thick line indicates a change of 50m. When lines are drawn close together, elevation changes rapidly. When lines are far apart, elevation changes more gradually as one has to walk farther to rise 10m.

Given a function $z = f(x, y)$, we can draw a topographical map of f by drawing **level curves** (or, contour lines). A level curve at $z = c$ is a curve in the xy-plane such that for all points (x, y) on the curve, $f(x, y) = c$. When drawing level curves, it is important that the c-values are spaced equally apart as that gives the best insight to how quickly the elevation is changing.

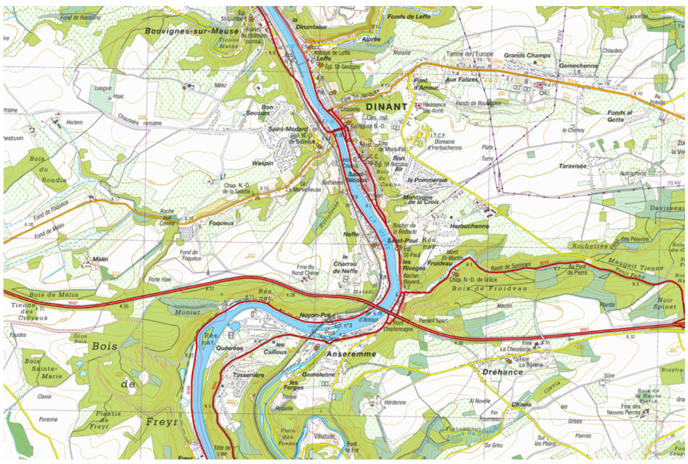


Figure 16.2: The topographical map of Dinant displays elevation by drawing contour lines, along which the elevation is constant.

Example 16.2

Let

$$f(x,y) = \frac{x+y}{x^2+y^2+1}.$$

Find level curves.

Solution

We begin by setting $f(x,y) = c$ for an arbitrary c and seeing if algebraic manipulation of the equation reveals anything significant.

$$\frac{x+y}{x^2+y^2+1} = c \qquad \Leftrightarrow \qquad x^2 - \frac{1}{c}x + y^2 - \frac{1}{c}y = -1.$$

We recognize this as a circle, though the centre and radius are not yet clear. By completing the square, we obtain:

$$\left(x - \frac{1}{2c}\right)^2 + \left(y - \frac{1}{2c}\right)^2 = \frac{1}{2c^2} - 1$$

a circle centred at $\big(1/(2c), 1/(2c)\big)$ with radius $\sqrt{1/(2c^2) - 1}$, where $|c| < 1/\sqrt{2}$. The level curves for $c = \pm 0.2$, ± 0.4 and ± 0.6 are sketched in Figure 16.3(a). To help illustrate elevation, we use dashed lines where $c < 0$. There is one special level curve, when $c = 0$. The level curve in this situation is $x + y = 0$, the line $y = -x$.

In Figure 16.3(b) we see a graph of the surface. Note how the y-axis is pointing away from the viewer to more closely resemble the orientation of the level curves in Figure 16.3(a). Seeing the level curves helps us understand the graph. For instance, the graph does not make it clear that one can walk along the line $y = -x$ without elevation change, though the level curve does.

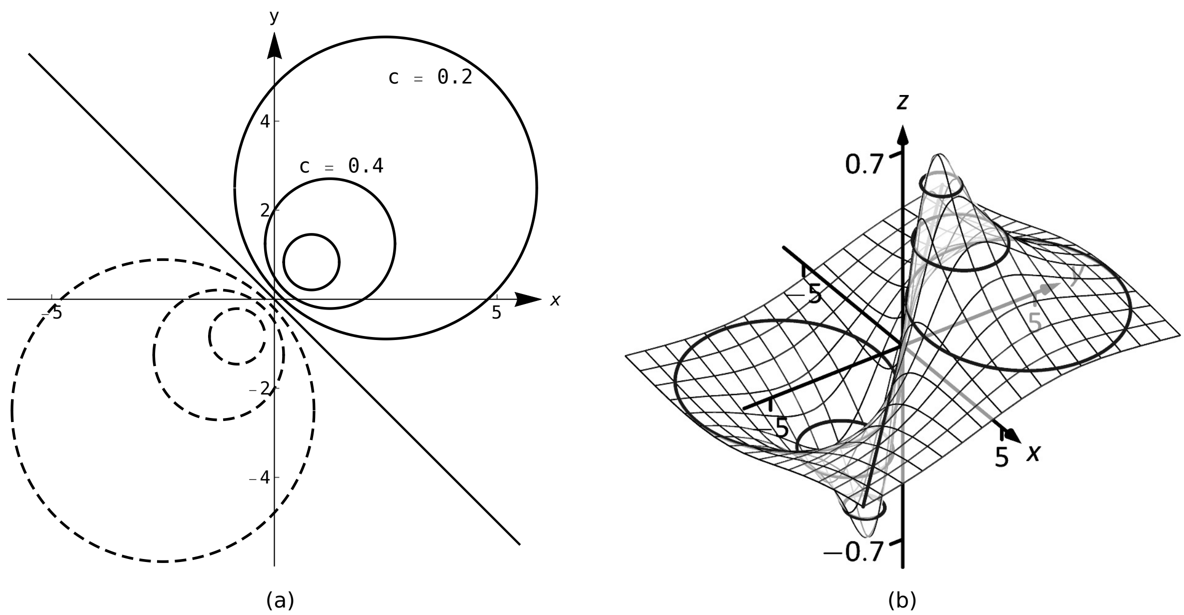


Figure 16.3: Graphing the level curves in Example 16.2.

The contour lines are established as intersection between the surface defined by $z = f(x,y)$ with the horizontal plane $z = c$. However, there are two more special types of planes with which we can intersect the surface and which improve our understanding of functions of two variables, namely $x = x_0$ and $y = y_0$. These are planes perpendicular to the x- or y-axis. The curves of intersection that we thus obtain are nothing more than the graphs of the so-called partial functions to y or x.

Definitie 16.3 (Partial functions)
Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables, then

- the partial function of f with regard to x, or the first partial function is given by:

$$f_1 : \mathbb{R} \to \mathbb{R} : x \to z = f_1(x) = f(x, y_0)\,,$$

  with y_0 constant;

- the partial function of f with regard to y, or the second partial function is given by:

$$f_2 : \mathbb{R} \to \mathbb{R} : x \to z = f_2(y) = f(x_0, y)\,,$$

  with x_0 constant.

16.1.2 Functions of n variables

We extend our study of multivariable functions to functions of n variables.

Definitie 16.4 (Function of *n* variables)
Let D be a subset of $\mathbb{R}^n$. A **function** f **of** n **variables** is a rule that assigns each $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in D a value $w = f(x_1, x_2, \ldots, x_n) = f(\mathbf{x})$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

Note that in this definition, we are using the notation $\mathbf{x}$ to abbreviate the n-tuple $(x_1, x_2, \ldots, x_n)$. It is very difficult to produce a meaningful graph of a function of three variables. A function of one variable is a curve drawn in 2 dimensions; a function of two variables is a surface drawn in 3 dimensions; a function of n variables is a **hypersurface** (*hyperoppervlak*) drawn in $n+1$ dimensions.

There are a few techniques one can employ to try to picture a graph of three variables. One is an analogue of level curves: **level surfaces** (*niveau-oppervlak*). Given $w = f(x, y, z)$, the level surface at $w = c$ is the surface in space formed by all points (x, y, z) where $f(x, y, z) = c$.

Example 16.3

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0, 0, 0)$,

$$I(x, y, z) = \frac{k}{x^2 + y^2 + z^2}$$

for some constant k.

Let $k = 1$; find the level surfaces of I.

Solution

We can answer this question using common sense. If energy is emanating from the origin, its intensity will be the same at all points equidistant from the origin. That is, at any point on the surface of a sphere centred at the origin, the intensity should be the same. Therefore, the level surfaces are spheres. We now find this mathematically. The level surface at $I = c$ for $c > 0$ is defined by

$$c = \frac{1}{x^2 + y^2 + z^2}.$$

A small amount of algebra reveals

$$x^2 + y^2 + z^2 = \frac{1}{c}.$$

Given an intensity c, the level surface $I = c$ is a sphere of radius $1/\sqrt{c}$, centred at the origin. Table 16.1 gives the radii of the spheres for given c-values. Normally one would use equally spaced c-values, but these values have been chosen purposefully. At a distance of 0.25 from the point source, the intensity is 16; to move to a point of half that intensity, one just moves out 0.1 to 0.35 – not much at all. To again halve the intensity, one moves 0.15, a little more than before.

Table 16.1: A table of c-values and corresponding radius r of the spheres of constant value in Example 16.3.

c	16.	8.	4.	2.	1.	0.5	0.25	0.125	0.0625
r	0.25	0.35	0.5	0.71	1	1.41	2	2.83	4

Note how each time the intensity is halved, the distance required to move away grows. We conclude that the closer one is to the source, the more rapidly the intensity changes.

Finally, we note that the notion of partial functions can be extended to functions of n variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ a function of n variables, than we call:

$$f_i : \mathbb{R} \to \mathbb{R} : x_i \to z = f_i(x_i) = f(c_1, c_2, \ldots, c_{i-1}, x_i, c_{i+1}, \ldots, c_n),$$

with c_j constant for $i = 1, \ldots, n\,(j \neq i)$, the i-th partial function of f.

16.2 Limits and continuity of multivariable functions

This section investigates what it means for multivariable functions to be continuous.

16.2.1 Introductory concepts and definitions

We begin with a series of definitions. We are used to open and closed intervals. We need analogous definitions for open and closed sets in the n-dimensional space.

Definitie 16.5 (Points and sets)
An **open ball** (*open bal*) B in $\mathbb{R}^n$ centred at $\mathbf{x}_0$ with radius r is the set of all points $\mathbf{x}$ such that $d(\mathbf{x}, \mathbf{x}_0) < r$.

Let S be a set of points in $\mathbb{R}^n$. A point P in $\mathbb{R}^n$ is a **boundary point** (*randpunt*) of S if all open disks centred at P contain both points in S and points not in S.

A point P in S is an **interior point** (*inwendig punt*) of S if there is an open disk centred at P that contains only points in S.

A set S is **open** (*open*) if every point in S is an interior point.

A set S is **closed** (*gesloten*) if it contains all of its boundary points.

A set S is **bounded** (*begrensd*) if there is an $M > 0$ such that the open disk, centred at the origin with radius M, contains S. A set that is not bounded is unbounded.
A set S is **convex** (*convex*) if, given any two points, it contains the whole line segment that joins. A set that is not convex is called non-convex.

Figure 16.4 shows several sets in the xy-plane. In each set, point P_1 lies on the boundary of the set as all open disks centred there contain both points in, and not in, the set. In contrast, point P_2 is an interior point for there is an open disk centred there that lies entirely within the set. The set depicted in Figure 16.4(a) is a closed set as it contains all of its boundary points. The set in Figure 16.4(b) is open, for all of its points are interior points (or, equivalently, it does not contain any of its boundary points). The set in Figure 16.4(c) is neither open nor closed as it contains some of its boundary points. Finally, it should be clear that all sets shown in Figure 16.4 are non-convex because we can easily find pairs of points that can only be connected by a straight line that is not completely contained in the considered sets.

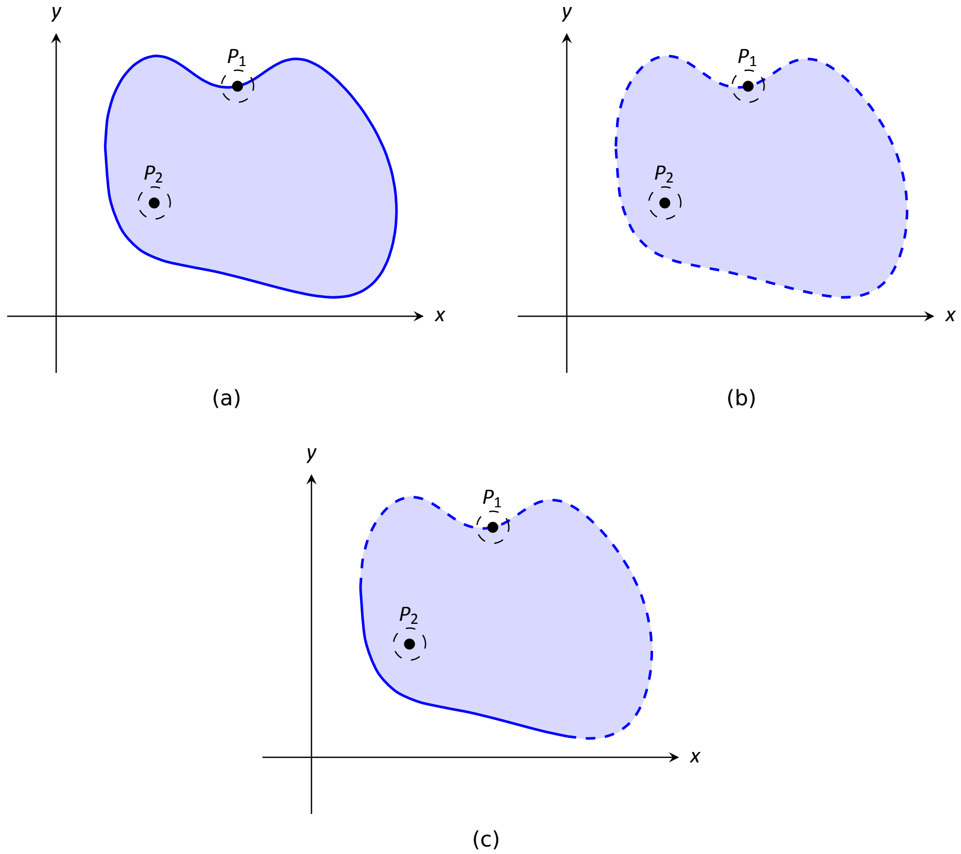


Figure 16.4: Illustrating open and closed sets in the xy-plane.

Example 16.4

Determine if the domain of the functions

1. $f(x,y) = \sqrt{1 - \dfrac{x^2}{9} - \dfrac{y^2}{4}}$
2. $g(x,y) = \dfrac{1}{x-y}$

is open, closed, or neither, and if it is bounded.

Solution

1. The domain of this function was found in Example 16.1 to be

$$D = \left\{(x,y) \;\middle|\; \frac{x^2}{9} + \frac{y^2}{4} \le 1\right\},$$

the region bounded by the ellipse $\frac{x^2}{9} + \frac{y^2}{4} = 1$. Since the region includes the boundary, the set contains all of its boundary points and hence is closed. The region is bounded as a disk of radius 4, centred at the origin, contains D.

2. As we cannot divide by 0, we find the domain to be $D = \{(x,y) \mid x - y \neq 0\}$. In other words, the domain is the set of all points (x,y) not on the line $y = x$. The domain is sketched in Figure 16.5. Note how we can draw an open disk around any point in the domain that lies

entirely inside the domain, and also note how the only boundary points of the domain are the points on the line $y = x$. We conclude the domain is an open set. The set is unbounded.

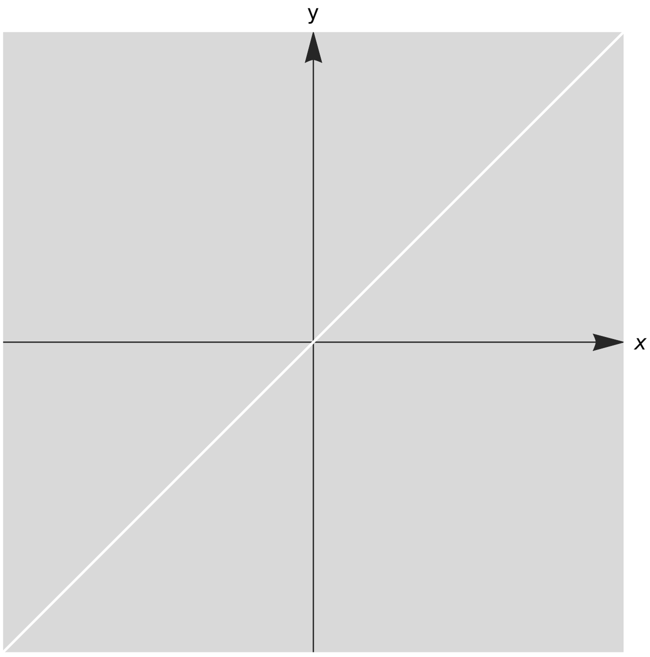


Figure 16.5: Sketching the domain of the function in Example 16.4.2.

16.2.2 Limits

We will say that

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$$

means if the point (x,y) is really close to the point (x_0,y_0), then $f(x,y)$ is really close to L. The formal definition for a function of n variables is given below.

Definitie 16.6 (Limit of a function of *n* variables)
Let S be a set containing $P = \mathbf{x}_0$ where every open disk centred at P contains points in S other than P, i.e. P is a limit point, let f be a function of two variables defined on S, except possibly at P, and let L be a real number. The **limit of $f(\mathbf{x})$ as $\mathbf{x}$ approaches $\mathbf{x}_0$** is L, denoted

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = L,$$

means that given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{x}$ in S, where $\mathbf{x} \neq \mathbf{x}_0$, if $\mathbf{x}$ is in the open ball centred at $\mathbf{x}_0$ with radius δ, then $|f(\mathbf{x}) - L| < \varepsilon$.

Note that we now define limits over a set S in the plane (where S does not have to be open). As planar sets can be far more complicated than intervals, our definition adds the restriction "*. . .* where every open disk centred at P contains points in S other than P." This means that P should be a so-called **limit point** (*ophopingspunt*) of the set S. This in contrast to a so-called **isolated point** (*gëisoleerd punt*) Q of S for which there exists a neighbourhood of Q which does not contain any other points of S.

The concept behind Definition 16.6 is sketched in Figure 16.6. Given $\varepsilon > 0$, find $\delta > 0$ such that if (x, y) is any point in the open disk centred at (x_0, y_0) in the xy-plane with radius δ, then $f(x, y)$ should be within ε of L.

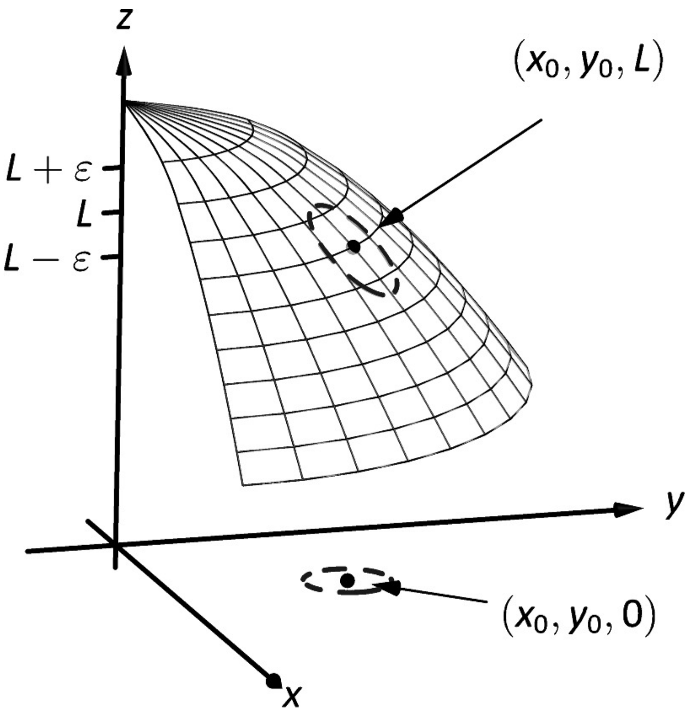


Figure 16.6: Illustrating the definition of a limit of a function of two variables.

Computing limits using this definition is rather cumbersome. The following properties allow us to evaluate limits much more easily. For that purpose, let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, let $\mathbf{x}_0$ have real components $(x_0)_i$ and let b, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = L \quad \text{and} \quad \lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) = K.$$

The following limits hold.

- **Constants:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} b = b$$

- **Identity**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} x_i = (x_0)_i$$

- **Sums/Differences:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \big(f(\mathbf{x}) \pm g(\mathbf{x})\big) = L \pm K$$

- **Scalar Multiples:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} b \cdot f(\mathbf{x}) = bL$$

- **Products:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) \cdot g(\mathbf{x}) = LK$$

- **Quotients:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{f(\mathbf{x})}{g(\mathbf{x}} = \frac{L}{K}, \qquad (K \neq 0)$$

- **Powers:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} [f(\mathbf{x})]^n = L^n$$

These properties, combined with the ones we introduced in Chapter 8, allow us to evaluate many limits. For instance, we can easily evaluate

$$\lim_{(x,y)\to(1,\pi)} \left(\frac{y}{x} + \cos(xy)\right) = \frac{\pi}{1} + \cos(\pi) = \pi - 1.$$

This limit may as well be evaluated in Mathematica with a nested application of the command **`Limit`**.

```
In[26]:= Limit[Limit[y/x + Cos[x*y], x → 1], y → Pi]

Out[26]= -1+π
```

When dealing with functions of a single variable we also considered one–sided limits and stated

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

if, and only if

$$\lim_{\substack{x\to c\\ >}} f(x) = L \quad \text{and} \quad \lim_{\substack{x\to c\\ <}} f(x) = L.$$

That is, the limit is L if and only if $f(x)$ approaches L when x approaches c from either direction, the left or the right.

In the plane, there are infinitely many directions from which (x,y) might approach (x_0,y_0). In fact, we do not have to restrict ourselves to approaching (x_0,y_0) from a particular direction, but rather we can approach that point along a path that is not a straight line. It is possible to arrive at different limiting values by approaching (x_0,y_0) along different paths. If this happens, we say that $\lim_{(x,y)\to(x_0,y_0)} f(x,y)$ does not exist. This is analogous to the left and right hand limits of single variable functions not being equal.

Our theorems tell us that we can evaluate most limits quite simply, without worrying about paths. When indeterminate forms arise, the limit may or may not exist. If it does exist, it can be difficult to prove this as we need to show the same limiting value is obtained regardless of the path chosen. The case where the limit does not exist is often easier to deal with, for we can often pick two paths along which the limit is different.

Example 16.5

1. Show

$$\lim_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2}$$

does not exist by finding the limits along the lines $y = mx$.

2. Show

$$\lim_{(x,y)\to(0,0)} \frac{\sin(xy)}{x+y}$$

does not exist by finding the limit along the path $y = -\sin(x)$.

Solution

1. Evaluating this limit along the lines $y = mx$ means replace all y's with mx and evaluating the resulting limit:

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{3x(mx)}{x^2+(mx)^2} &= \lim_{x\to 0} \frac{3mx^2}{x^2(m^2+1)}\\ &= \lim_{x\to 0} \frac{3m}{m^2+1}\\ &= \frac{3m}{m^2+1}.\end{aligned}$$

While the limit exists for each choice of m, we get a different limit for each choice of m. That is, along different lines we get differing limiting values, meaning the limit does not exist.

2. We are to show that $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist by finding the limit along the path $y = -\sin(x)$. First, however, consider the limits found along the lines $y = mx$ as done above.

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{\sin\big(x(mx)\big)}{x+mx} &= \lim_{x\to 0} \frac{\sin(mx^2)}{x(m+1)}\\ &= \lim_{x\to 0} \frac{\sin(mx^2)}{x}\cdot\frac{1}{m+1}.\end{aligned}$$

By applying L'Hôpital's rule, we can show this limit is 0 except when $m = -1$, that is, along the line $y = -x$. This line is not in the domain of f, so we have found the following fact: along every line $y = mx$ in the domain of f,

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

Now consider the limit along the path $y = -\sin(x)$:

$$\lim_{(x,-\sin(x))\to(0,0)} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)} = \lim_{x\to 0} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)}.$$

Now apply L'Hôpital's rule twice:

$$\begin{aligned}&= \lim_{x\to 0} \frac{\cos\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))}{1-\cos(x)} \qquad \left(= \frac{0}{0}\right)\\ &= \lim_{x\to 0} \frac{-\sin\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))^2+\cos\big(-x\sin(x)\big)(-2\cos(x)+x\sin(x))}{\sin(x)}\\ &= \frac{-2}{0}.\end{aligned}$$

It follows that the limit does not exist. Step back and consider what we have just discovered. Along any line $y = mx$ in the domain of the $f(x,y)$, the limit is 0. However, along the path $y = -\sin(x)$, which lies in the domain of $f(x,y)$ for all $x \neq 0$, the limit does not exist. Since the limit is not the same along every path to $(0,0)$, we say that the studied limit does not exist.

Example 16.6

Let

$$f(x,y)=\frac{5x^2y^2}{x^2+y^2}.$$

Find $\lim_{(x,y)\to(0,0)} f(x,y)$.

Solution

It is relatively easy to show that along any line $y=mx$, the limit is 0. This is not enough to prove that the limit exists, as demonstrated in the previous example, but it tells us that if the limit does exist then it must be 0.

To prove the limit is 0, we apply Definition 16.6. Let $\varepsilon>0$ be given. We want to find $\delta>0$ such that if $\sqrt{(x-0)^2+(y-0)^2}<\delta$, then $|f(x,y)-0|<\varepsilon$.

Set $\delta<\sqrt{\varepsilon/5}$. Note that $\left|\frac{5y^2}{x^2+y^2}\right|<5$ for all $(x,y)\neq(0,0)$, and that if $\sqrt{x^2+y^2}<\delta$, then $x^2<\delta^2$.

Let $\sqrt{(x-0)^2+(y-0)^2}=\sqrt{x^2+y^2}<\delta$. Consider $|f(x,y)-0|$:

$$\begin{aligned}|f(x,y)-0| &= \left|\frac{5x^2y^2}{x^2+y^2}-0\right| \\ &= \left|x^2\cdot\frac{5y^2}{x^2+y^2}\right| \\ &< \delta^2\cdot 5 \\ &< \frac{\varepsilon}{5}\cdot 5 \\ &= \varepsilon.\end{aligned}$$

Thus if $\sqrt{(x-0)^2+(y-0)^2}<\delta$ then $|f(x,y)-0|<\varepsilon$, which is what we wanted to show. Thus

$$\lim_{(x,y)\to(0,0)}\frac{5x^2y^2}{x^2+y^2}=0\,.$$

16.2.3 Continuity

Definition 8.3 defines what it means for a function of one variable to be continuous. In brief, it meant that the graph of the function did not have breaks, holes, jumps, etc. We define continuity for functions of two variables in a similar way as we did for functions of one variable.

Definitie 16.7 (Continuity)
Let a function $f(\mathbf{x})$ be defined on a set S containing the point $\mathbf{x}_0$.

1. f is continuous at $\mathbf{x}_0$ if $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})=f(\mathbf{x}_0)$.
2. f is **continuous on** S (*continu over*) if f is continuous at all points in S. If f is continuous at all points in $\mathbb{R}^n$, we say that f is continuous everywhere.

Example 16.7

Let

$$f(x,y) = \begin{cases} \dfrac{\cos(y)\sin(x)}{x}, & x \neq 0 \\ \cos(y), & x = 0. \end{cases}$$

Is f continuous at $(0,0)$? Is f continuous everywhere?

Solution

To determine if f is continuous at $(0,0)$, we need to compare $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ to $f(0,0)$. Applying the definition of f, we see that $f(0,0) = \cos(0) = 1$.

We now consider the limit $\lim\limits_{(x,y)\to(0,0)} f(x,y)$. Substituting 0 for x and y in $f(x,y)$ returns the indeterminate form "0/0", so we need to do more work to evaluate this limit.

Consider two related limits:

$$\lim_{(x,y)\to(0,0)} \cos(y) \qquad \text{and} \qquad \lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x}.$$

The first limit does not contain x, and since $\cos(y)$ is continuous,

$$\lim_{(x,y)\to(0,0)} \cos(y) = \lim_{y\to 0} \cos(y) = \cos(0) = 1.$$

The second limit does not contain y. By Theorem 8.6 we can say

$$\lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x} = \lim_{x\to 0} \frac{\sin(x)}{x} = 1.$$

Finally, following the properties of limits we can combine these two limits as follows:

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x} &= \lim_{(x,y)\to(0,0)} \left(\cos(y)\left(\frac{\sin(x)}{x}\right)\right) \\ &= \left(\lim_{(x,y)\to(0,0)} \cos(y)\right)\left(\lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x}\right) \\ &= (1)(1) = 1. \end{aligned}$$

We have found that

$$\lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x} = f(0,0),$$

so f is continuous at $(0,0)$.

A similar analysis shows that f is continuous at all points in $\mathbb{R}^2$. As long as $x \neq 0$, we can evaluate the limit directly; when $x = 0$, a similar analysis shows that the limit is $\cos(y)$. Thus we can say that f is continuous everywhere. A graph of f is given in Figure 16.7. Notice how it has no breaks, jumps, etc.

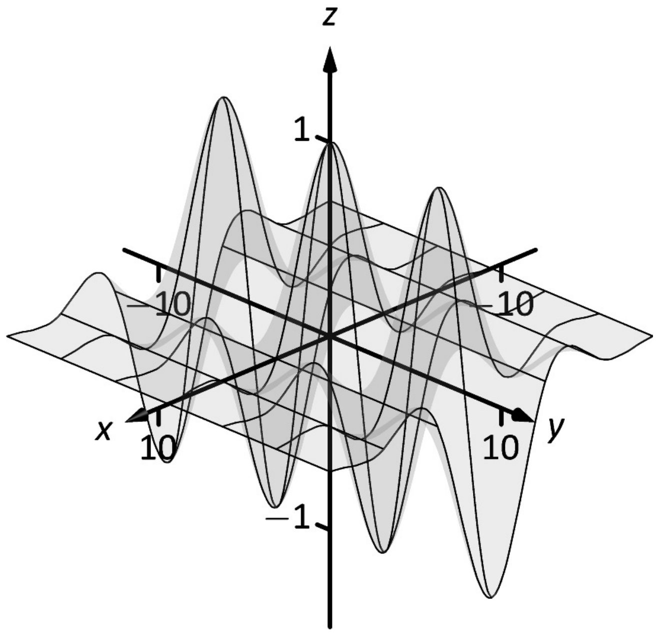


Figure 16.7: A graph of $f(x,y)$ in Example 16.7.

Of course, as with functions of one variable, we may combine continuous functions to create other continuous functions. More specifically, let f and g be continuous on a set S, let c be a real number, and let n be a positive integer. Then, the following functions are continuous on S.

- **Sums/Differences:** $f \pm g$
- **Constant multiples:** $c \cdot f$
- **Products:** $f \cdot g$
- **Quotients:** f/g (if $g \neq 0$ on S)
- **Powers:** f^n

For roots of a continuous function, we have that $\sqrt[n]{f}$ is continuous provided that $f \geq 0$ on S if n is even, whereas, if n is odd, this is true for all values of f on S. For what concerns function compositions, we let f be continuous on S, where the range of f on S is J, and let g be a single variable function that is continuous on J. Then $g \circ f$, i.e., $g(f(x,y))$, is continuous on S.

Having introduced the notion of continuity for functions of n variables, the multivariable counterpart of the indermediate value theorem follows intuitively.

Theorem 16.1 (Intermediate value theorem for functions of *n* variables)
Let f be a continuous function on D and, without loss of generality, let $f(\boldsymbol{\alpha}) < f(\mathbf{b})$. Then for every value u, where $f(\boldsymbol{\alpha}) < u < f(\mathbf{b})$, there is at least one interior point $\mathbf{c}$ in D such that $f(\mathbf{c}) = u$.

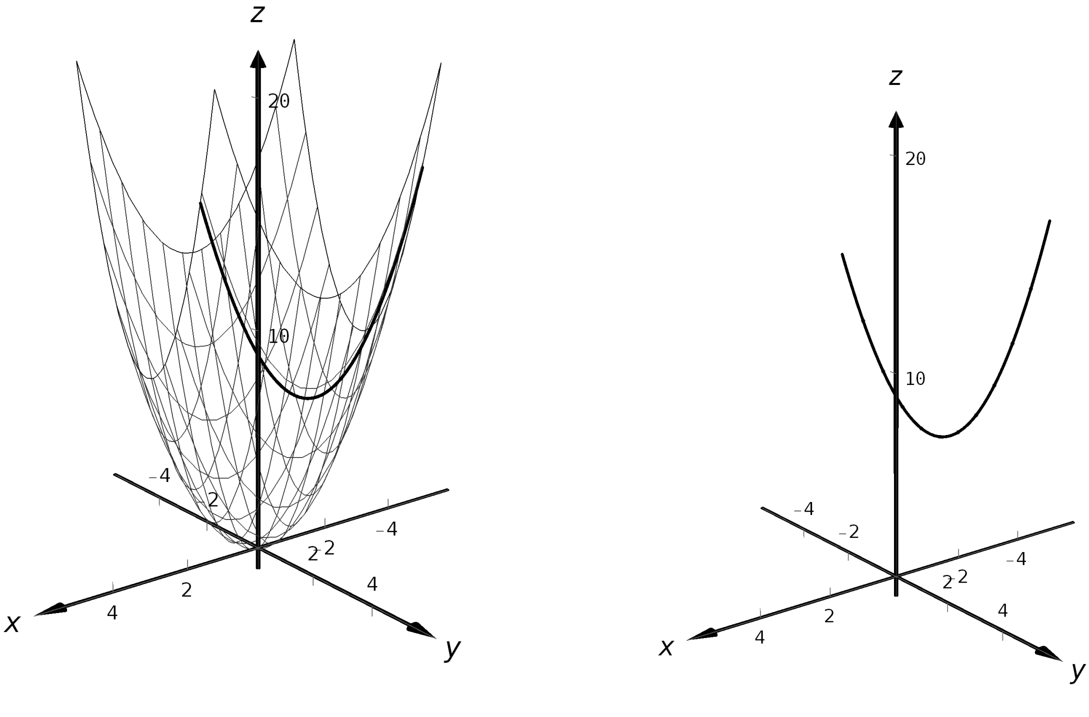
When considering single variable functions, we studied limits, then continuity, then the derivative. In our current study of multivariable functions, we have studied limits and continuity. In the next section we study derivation, which takes on a slight twist as we are in a multivarible context.

16.3 Partial derivatives

16.3.1 First partial derivatives

Let y be a function of x. We have studied in great detail the derivative of y with respect to x, that is, which measures the rate at which y changes with respect to x. Consider now $z = f(x,y)$. It makes sense to want to know how z changes with respect to x and/or y. This section begins our investigation into these rates of change.

Consider the function $z = f(x,y) = x^2 + 2y^2$, as graphed in Figure 16.8(a). By fixing $y = 2$, we focus our attention to all points on the surface where the y-value is 2, shown in Figures 16.8(a) and 16.8(b). These points form a curve in space: $z = f(x,2) = x^2 + 8$ which is a function of just one variable. We can take the derivative of z with respect to x along this curve and find equations of tangent lines, etc.



(a) (b)

Figure 16.8: By fixing $y = 2$, the surface $f(x,y) = x^2 + 2y^2$ is a curve in space.

The key notion to extract from this example is: by treating y as constant (it does not vary) we can consider how z changes with respect to x. In a similar fashion, we can hold x constant and consider how z changes with respect to y. This is the underlying principle of **partial derivatives** (*partiële afgeleide*). We state the formal, limit–based definition first, then show how to compute these partial derivatives without directly taking limits.

Definitie 16.8 (Partial derivative)
Let $z = f(x,y)$ be a continuous function on a set S in $\mathbb{R}^2$.

1. The **partial derivative of** f **with respect to** x is:

$$f_x(x,y) = \lim_{h \to 0} \frac{f(x+h,y) - f(x,y)}{h}.$$

2. The **partial derivative of f with respect to y** is:

$$f_y(x,y) = \lim_{h\to 0} \frac{f(x,y+h)-f(x,y)}{h}.$$

Alternate notations for $f_x(x,y)$ include:

$$\frac{\partial}{\partial x} f(x,y), \quad \frac{\partial f}{\partial x}, \quad \frac{\partial z}{\partial x}, \quad \text{and } z_x,$$

with similar notations for $f_y(x,y)$. For ease of notation, $f_x(x,y)$ is often abbreviated as f_x.

Example 16.8

Let $f(x,y) = x^2y+2x+y^3$. Find $f_x(x,y)$ using the limit definition.

Solution

Using Definition 16.8, we have:

$$\begin{aligned}
f_x(x,y) &= \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h} \\
&= \lim_{h\to 0} \frac{(x+h)^2y+2(x+h)+y^3-(x^2y+2x+y^3)}{h} \\
&= \lim_{h\to 0} \frac{x^2y+2xhy+h^2y+2x+2h+y^3-(x^2y+2x+y^3)}{h} \\
&= \lim_{h\to 0} \frac{2xhy+h^2y+2h}{h} \\
&= \lim_{h\to 0} (2xy+hy+2) \\
&= 2xy+2.
\end{aligned}$$

We have found $f_x(x,y) = 2xy+2$.

Using limits to compute partial derivatives is not necessary, as we can rely on our previous knowledge of derivatives to compute partial derivatives easily. When computing $f_x(x,y)$, we hold y fixed – it does not vary. Therefore we can compute the derivative with respect to x by treating y as a constant or coefficient.

Example 16.9

Find $f_x(x,y)$ and $f_y(x,y)$ for each of the following functions.

1. $f(x,y) = \cos(xy^2)+\sin(x)$
2. $f(x,y) = e^{x^2y^3}\sqrt{x^2+1}$

Solution

1. Begin with $f_x(x,y)$. We need to apply the chain rule with the cosine term; y^2 is the coefficient of the x-term inside the cosine function.

$$f_x(x,y) = -\sin(xy^2)(y^2)+\cos(x) = -y^2\sin(xy^2)+\cos(x).$$

To find $f_y(x,y)$, note that x is the coefficient of the y^2-term inside of the cosine term; also note that since x is fixed, $\sin(x)$ is also fixed, and we treat it as a constant.

$$f_y(x,y) = -\sin(xy^2)(2xy) = -2xy\sin(xy^2).$$

We may check our answer for what concerns f_x in Mathematica as follows:

```
In[27]:= D[Cos[x*y^2] + Sin[x], x]

Out[27]= Cos[x]-y^2 Sin[x y^2]
```

And likewise for what concerns f_y:

```
In[28]:= D[Cos[x*y^2] + Sin[x], y]

Out[28]= -2 x y Sin[x y^2]
```

2. Beginning with $f_x(x,y)$, note how we need to apply the product rule.

$$\begin{aligned} f_x(x,y) &= e^{x^2y^3}(2xy^3)\sqrt{x^2+1} + e^{x^2y^3}\frac{1}{2}(x^2+1)^{-1/2}(2x) \\ &= 2xy^3e^{x^2y^3}\sqrt{x^2+1} + \frac{xe^{x^2y^3}}{\sqrt{x^2+1}}. \end{aligned}$$

Note that when finding $f_y(x,y)$ we do not have to apply the product rule; since $\sqrt{x^2+1}$ does not contain y, we treat it as fixed and hence becomes a coefficient of the $e^{x^2y^3}$-term.

$$f_y(x,y) = e^{x^2y^3}(3x^2y^2)\sqrt{x^2+1} = 3x^2y^2e^{x^2y^3}\sqrt{x^2+1}.$$

We have shown how to compute a partial derivative, but it may still not be clear what a partial derivative means. Given $z = f(x,y)$, $f_x(x,y)$ measures the rate at which z changes as only x varies: y is held constant.

Imagine standing in a rolling meadow, then beginning to walk due east. Depending on your location, you might walk up, sharply down, or perhaps not change elevation at all. This is similar to measuring z_x: you are moving only east (in the x-direction) and not north/south at all. Going back to your original location, imagine now walking due north (in the y-direction). Perhaps walking due north does not change your elevation at all. This is analogous to $z_y = 0$: z does not change with respect to y. We can see that z_x and z_y do not have to be the same, or even similar, as it is easy to imagine circumstances where walking east means you walk downhill, though walking north makes you walk uphill.

16.3.2 Second partial derivatives

Let $z = f(x,y)$. We have learned to find the partial derivatives $f_x(x,y)$ and $f_y(x,y)$, which are each functions of x and y. Therefore we can take partial derivatives of them, each with respect to x and y. We define these second partials along with the notation, give examples, then discuss their meaning.

Definitie 16.9 (Second and mixed partial derivative)
Let $z = f(x,y)$ be continuous on a set S.

1. The **second partial derivative of** f **with respect to** x **then** x (*tweede partiële afgeleide van f naar x*) is
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \left(f_x\right)_x = f_{xx}.$$
2. The **second partial derivative of** f **with respect to** x **then** y **is** (*tweede partiële afgeleide van f naar x en y*) is
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = \left(f_x\right)_y = f_{xy}.$$

Similar definitions hold for $\dfrac{\partial^2 f}{\partial y^2} = f_{yy}$ and $\dfrac{\partial^2 f}{\partial x \partial y} = f_{yx}$. The second partial derivatives f_{xy} and f_{yx} are **mixed partial derivatives** (*gemengde partiële afgeleide*).

The terms in Definition 16.9 all depend on limits, so each definition comes with the caveat where the limit exists.

Example 16.10

For each of the following functions, find all 6 first and second partial derivatives. That is, find

$$f_x, \quad f_y, \quad f_{xx}, \quad f_{yy}, \quad f_{xy} \quad \text{and} \quad f_{yx}\,.$$

1. $f(x,y) = \dfrac{x^3}{y^2}$
2. $f(x,y) = e^x \sin(x^2 y)$

Solution

In each, we give f_x and f_y immediately and then spend time deriving the second partial derivatives.

1. $f(x,y) = \dfrac{x^3}{y^2} = x^3 y^{-2}$

$$f_x(x,y) = \frac{3x^2}{y^2}$$
$$f_y(x,y) = -\frac{2x^3}{y^3}$$
$$f_{xx}(x,y) = \frac{\partial}{\partial x}(f_x) = \frac{\partial}{\partial x}\left(\frac{3x^2}{y^2}\right) = \frac{6x}{y^2}$$
$$f_{yy}(x,y) = \frac{\partial}{\partial y}(f_y) = \frac{\partial}{\partial y}\left(-\frac{2x^3}{y^3}\right) = \frac{6x^3}{y^4}$$
$$f_{xy}(x,y) = \frac{\partial}{\partial y}(f_x) = \frac{\partial}{\partial y}\left(\frac{3x^2}{y^2}\right) = -\frac{6x^2}{y^3}$$
$$f_{yx}(x,y) = \frac{\partial}{\partial x}(f_y) = \frac{\partial}{\partial x}\left(-\frac{2x^3}{y^3}\right) = -\frac{6x^2}{y^3}$$

2. $f(x,y) = e^x \sin(x^2 y)$

   Because the following partial derivatives get rather long, we omit the extra notation and just give the results. In several cases, multiple applications of the product and chain rules will be necessary, followed by some basic combination of like terms.

   $f_x(x,y) = e^x \sin(x^2 y) + 2xye^x \cos(x^2 y)$

$$f_y(x,y) = x^2e^x\cos(x^2y)$$
$$f_{xx}(x,y) = e^x\sin(x^2y)+4xye^x\cos(x^2y)+2ye^x\cos(x^2y)-4x^2y^2e^x\sin(x^2y)$$
$$f_{yy}(x,y) = -x^4e^x\sin(x^2y)$$
$$f_{xy}(x,y) = x^2e^x\cos(x^2y)+2xe^x\cos(x^2y)-2x^3ye^x\sin(x^2y)$$
$$f_{yx}(x,y) = x^2e^x\cos(x^2y)+2xe^x\cos(x^2y)-2x^3ye^x\sin(x^2y)$$

Higher-order partial derivatives can also be computed in Mathematica. For instance, given $f(x,y) = x^3/y^2$ from Example 16.10, we can find f_{xy} as follows:

```
In[29]:= D[x^3/y^2, x, y]

Out[29]= - 6 x^2
           -----
            y^3
```

Notice how for each of the two functions in Example 16.10, $f_{xy} = f_{yx}$. Due to the complexity of the examples, this likely is not a coincidence. The following theorem states that it is not. It is also known as **Schwarz's theorem**, **Clairaut's theorem**, or **Young's theorem**.

Theorem 16.2 (Symmetry of second derivatives)
Let f be defined such that f_{xy} and f_{yx} are continuous on a set S. Then for each point (x,y) in S, $f_{xy}(x,y) = f_{yx}(x,y)$.

Now that we know how to find second partials, we investigate what they tell us.

Again we refer back to a function $y = f(x)$ of a single variable. The second derivative of f is "the derivative of the derivative," or "the rate of change of the rate of change." The second derivative measures how much the derivative is changing. If $f''(x) < 0$, then the derivative is getting smaller (so the graph of f is concave down); if $f''(x) > 0$, then the derivative is growing, making the graph of f concave up.

Now consider $z = f(x,y)$. Similar statements can be made about f_{xx} and f_{yy} as could be made about $f''(x)$ above. When taking derivatives with respect to x twice, we measure how much f_x changes with respect to x. If $f_{xx}(x,y) < 0$, it means that as x increases, f_x decreases, and the graph of f will be concave down in the x-direction. Using the analogy of standing in the rolling meadow used earlier in this section, f_{xx} measures whether one's path is concave up/down when walking due east. Similarly, f_{yy} measures the concavity in the y-direction. If $f_{yy}(x,y) > 0$, then f_y is increasing with respect to y and the graph of f will be concave up in the y-direction. Appealing to the rolling meadow analogy again, f_{yy} measures whether one's path is concave up/down when walking due north.

We now consider the mixed partials f_{xy} and f_{yx}. The mixed partial f_{xy} measures how much f_x changes with respect to y. Once again using the rolling meadow analogy, f_x measures the slope if one walks due east. Looking east, begin walking north (side–stepping). Is the path towards the east getting steeper? If so, $f_{xy} > 0$. Is the path towards the east not changing in steepness? If so, then $f_{xy} = 0$. A similar thing can be said about f_{yx}: consider the steepness of paths heading north while side–stepping to the east.

The following example examines these ideas with concrete numbers and graphs.

Example 16.11

Let $z = x^2 - y^2 + xy$. Evaluate the 6 first and second partial derivatives at $(-1/2, 1/2)$ and interpret what each of these numbers mean.

Solution

We find that:

$f_x(x,y) = 2x + y$, $f_y(x,y) = -2y + x$, $f_{xx}(x,y) = 2$, $f_{yy}(x,y) = -2$ and $f_{xy}(x,y) = f_{yx}(x,y) = 1$. Thus at $(-1/2, 1/2)$ we have

$$f_x\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}, \qquad f_y\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{3}{2}.$$

The slope of the tangent line at $(-1/2, 1/2, -1/4)$ in the direction of x is $-1/2$: if one moves from that point parallel to the x-axis, the instantaneous rate of change will be $-1/2$. The slope of the tangent line at this point in the direction of y is $-3/2$: if one moves from this point parallel to the y-axis, the instantaneous rate of change will be $-3/2$. These tangents lines are graphed in Figure 16.9(a) and 16.9(b), together with the curve where $x = -1/2$ and $y = 1/2$, respectively, where the tangent lines are drawn in a solid line.

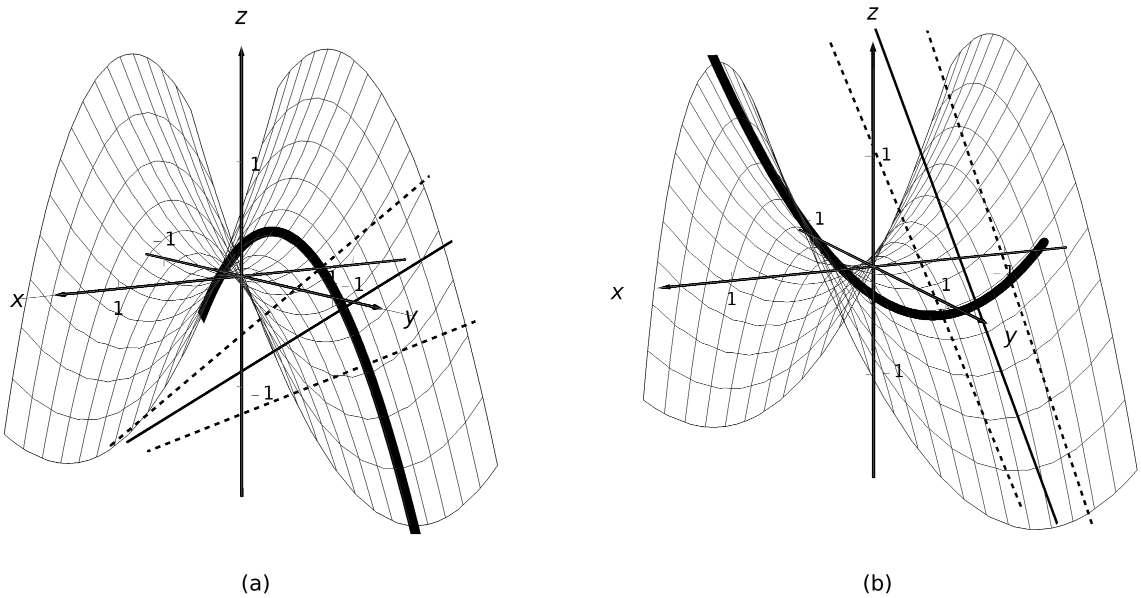


Figure 16.9: Understanding the second partial derivatives in Example 16.11.

Now consider only Figure 16.9(a). Three directed tangent lines are drawn (two are dashed), each in the direction of x; that is, each has a slope determined by f_x. Note how as y increases, the slope of these lines get closer to 0. Since the slopes are all negative, getting closer to 0 means the slopes are increasing. The slopes given by f_x are increasing as y increases, meaning f_{xy} must be positive.

Since $f_{xy} = f_{yx}$, we also expect f_y to increase as x increases. Consider Figure 16.9(b) where again three directed tangent lines are drawn, this time each in the direction of y with slopes determined by f_y. As x increases, the slopes become less steep (closer to 0). Since these are negative slopes, this means the slopes are increasing.

Thus far we have a visual understanding of f_x, f_y, and $f_{xy} = f_{yx}$. We now interpret f_{xx} and f_{yy}. In Figure 16.9(a), we see a curve drawn where x is held constant at $x = -1/2$: only y varies. This curve is clearly concave down, corresponding to the fact that $f_{yy} < 0$. In Figure 16.9(b), we see a similar curve where y is constant and only x varies. This curve is concave up, corresponding to the fact that $f_{xx} > 0$.

16.3.3 Higher-order partial derivatives

Essentially, we can continue taking partial derivatives of partial derivatives of partial derivatives of . . . ; we do not have to stop with second partial derivatives. These higher order partial derivatives do not have a tidy graphical interpretation; nevertheless they are not hard to compute and worthy of some practice.

We do not formally define each higher order derivative, but rather give just a few examples of the notation. For instance,

$$f_{xyx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right) \quad \text{and} \quad f_{xxz}(x,y) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\right).$$

Example 16.12

Let

$$f(x,y) = x^2y^2 + \sin(xy).$$

Find f_{xxy} and f_{yxx}.

Solution

To find f_{xxy}, we first find f_x, then f_{xx}, then f_{xxy}:

$$f_x = 2xy^2 + y\cos(xy)$$
$$f_{xx} = 2y^2 - y^2\sin(xy)$$
$$f_{xxy} = 4y - 2y\sin(xy) - xy^2\cos(xy).$$

To find f_{yxx}, we first find f_y, then f_{yx}, then f_{yxx}:

$$\begin{aligned} f_y &= 2x^2y + x\cos(xy) \\ f_{yx} &= 4xy + \cos(xy) - xy\sin(xy) \\ f_{yxx} &= 4y - y\sin(xy) - \big(y\sin(xy) + xy^2\cos(xy)\big) \\ &= 4y - 2y\sin(xy) - xy^2\cos(xy). \end{aligned}$$

Note how $f_{xxy} = f_{yxx}$.

In the previous example we saw that $f_{xxy} = f_{yxx}$; this is not a coincidence. While we do not state this as a formal theorem, as long as each partial derivative is continuous, it does not matter the order in which the partial derivatives are taken. For instance, $f_{xxy} = f_{xyx} = f_{yxx}$.

With $z = f(x,y)$, the partial derivatives f_x and f_y measure the instantaneous rate of change of z when moving parallel to the x- and y-axes, respectively. How do we measure the rate of change at a point when we do not move parallel to one of these axes? What if we move in the direction given by the vector $\langle 2,1\rangle$? Can we measure that rate of change? The answer is, of course, yes, we can. This is the topic of Section 16.6. First, we need to define what it means for a function of two variables to be differentiable.

16.3.4 Functions of n variables

The concepts underlying partial derivatives can be easily extend to n variables.

Definitie 16.10 (Partial derivative with n variables)
Let $w = f(\mathbf{x})$ be a continuous function on a set D in $\mathbb{R}^n$.

The **partial derivative of f with respect to x_i** is:

$$f_{x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_i + h, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n)}{h}.$$

By taking partial derivatives of partial derivatives, we can find second partial derivatives of f with respect to z then y, for instance, just as before.

Example 16.13

For each of the following functions, find f_x, f_y, f_z, f_{xz}, f_{yz}, and f_{zz}.

1. $f(x,y,z) = x^2y^3z^4 + x^2y^2 + x^3z^3 + y^4z^4$
2. $f(x,y,z) = x\sin(yz)$

Solution

1. $f_x = 2xy^3z^4 + 2xy^2 + 3x^2z^3$
   $f_y = 3x^2y^2z^4 + 2x^2y + 4y^3z^4$
   $f_z = 4x^2y^3z^3 + 3x^3z^2 + 4y^4z^3$
   $f_{xz} = 8xy^3z^3 + 9x^2z^2$
   $f_{yz} = 12x^2y^2z^3 + 16y^3z^3$
   $f_{zz} = 12x^2y^3z^2 + 6x^3z + 12y^4z^2$

2. $f_x = \sin(yz)$
   $f_y = xz\cos(yz)$
   $f_z = xy\cos(yz)$
   $f_{xz} = y\cos(yz)$
   $f_{yz} = x\cos(yz) - xyz\sin(yz)$
   $f_{zz} = -xy^2\sin(xy)$

16.4 Total differential and differentiability

16.4.1 Total differential

We studied differentials in Section 9.7.3, where Definition 9.8 states that if $y = f(x)$ and f is differentiable, then $dy = f'(x)dx$. One important use of this differential is in integration by substitution. Another important application is approximation. Let $\Delta x = dx$ represent a change in x. When dx is small, $dy \approx \Delta y$, the change in y resulting from the change in x. So, as dx goes to 0, the error in approximating Δy with dy goes to 0.

We extend this idea to functions of two variables. Let $z = f(x,y)$, and let $\Delta x = dx$ and $\Delta y = dy$ represent changes in x and y, respectively (Figure 16.10). Let $\Delta z = f(x+dx, y+dy) - f(x,y)$ be the change in z over the change in x and y. Recalling that f_x and f_y give the instantaneous rates of z-change in the x- and y-directions, respectively, we can approximate Δz with $dz = f_x dx + f_y dy$; in words, the total change in z is approximately the change caused by changing x plus the change caused by changing y. In a moment we give an indication of whether or not this approximation is any good. First we give a name to dz.

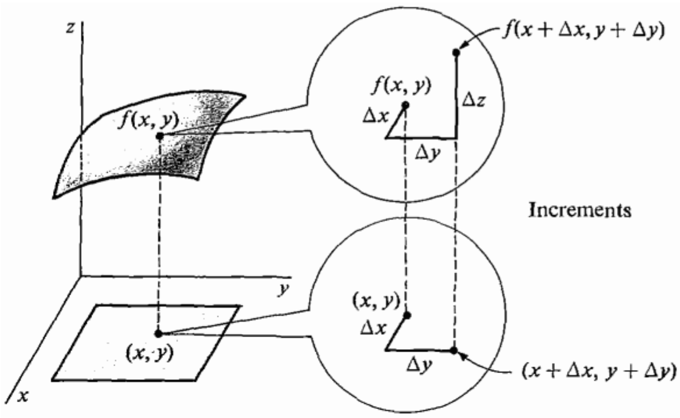


Figure 16.10: Understanding the total differential of a function of two variables.

Definitie 16.11 (Total differential)
Let $z = f(x,y)$ be continuous on a set S. Let dx and dy represent changes in x and y, respectively. Where the partial derivatives f_x and f_y exist, the **total differential of** z (*totale differentiaal van z*) is

$$dz = f_x(x,y)dx + f_y(x,y)dy.$$

Note that from Definition 16.11, we can as well use vector notation:

$$dz = \left(f_x, f_y\right) \cdot (dx, dy).$$

16.4.2 Differentiability

16.4.2.1 Definition

We can approximate Δz with dz, but as with all approximations, there is error involved. A good approximation is one in which the error is small. At a given point (x_0, y_0), let E_x and E_y be functions of dx and dy such that $E_x dx + E_y dy$ describes this error. Then

$$\begin{aligned}\Delta z &= dz + E_x dx + E_y dy \\ &= f_x(x_0,y_0)dx + f_y(x_0,y_0)dy + E_x dx + E_y dy.\end{aligned}$$

If the approximation of Δz by dz is good, then as dx and dy get small, so does $E_x dx + E_y dy$. The approximation of Δz by dz is even better if, as dx and dy go to 0, so do E_x and E_y. This leads us to our definition of differentiability.

Definitie 16.12 (Multivariable differentiability)
Let $z = f(x,y)$ be defined on a set S containing (x_0, y_0) where $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist. Let dz be the total differential of z at (x_0,y_0), let $\Delta z = f(x_0 + dx, y_0 + dy) - f(x_0, y_0)$, and let E_x and E_y be functions of dx and dy such that

$$\Delta z = dz + E_x dx + E_y dy.$$

1. We say f **is differentiable at** (x_0, y_0) (*afleidbaar*) if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||(dx,dy)|| < \delta$, then $||\left(E_x, E_y\right)|| < \varepsilon$. That is, as dx and dy go to 0, so do E_x and E_y.

2. We say f **is differentiable on** S (*afleidbaar over S*) if f is differentiable at every point in S. If f is differentiable on $\mathbb{R}^2$, we say that f is differentiable everywhere.

Example 16.14

Show $f(x,y) = xy + 3y^2$ is differentiable.

Solution

We begin by finding $f(x+dx, y+dy)$, Δz, f_x and f_y.

$$\begin{aligned} f(x+dx,y+dy) &= (x+dx)(y+dy)+3(y+dy)^2 \\ &= xy+xdy+ydx+dxdy+3y^2+6ydy+3dy^2. \end{aligned}$$

$\Delta z = f(x+dx, y+dy) - f(x,y)$, so

$$\Delta z = xdy + ydx + dxdy + 6ydy + 3dy^2.$$

It is straightforward to compute $f_x = y$ and $f_y = x + 6y$. Consider once more Δz:

$$\begin{aligned} \Delta z &= xdy+ydx+dxdy+6ydy+3dy^2 \qquad \text{(Now reorder.)} \\ &= ydx+xdy+6ydy+dxdy+3dy^2 \\ &= \underbrace{(y)}_{f_x}dx+\underbrace{(x+6y)}_{f_y}dy+\underbrace{(dy)}_{E_x}dx+\underbrace{(3dy)}_{E_y}dy \\ &= f_xdx+f_ydy+E_xdx+E_ydy. \end{aligned}$$

With $E_x = dy$ and $E_y = 3dy$, it is clear that as dx and dy go to 0, E_x and E_y also go to 0. Since this did not depend on a specific point (x_0, y_0), we can say that $f(x,y)$ is differentiable for all pairs (x,y) in $\mathbb{R}^2$, or, equivalently, that f is differentiable everywhere.

Our intuitive understanding of differentiability of functions $y = f(x)$ of one variable was that the graph of f was **smooth** (*glad*). A similar intuitive understanding of functions $z = f(x,y)$ of two variables is that the surface defined by f is also smooth, not containing cusps, edges, breaks, etc. The following theorem states that differentiable functions are continuous, followed by another theorem that provides a more tangible way of determining whether a great number of functions are differentiable or not.

Theorem 16.3 (Continuity and differentiability of multivariable functions)
Let $z = f(x,y)$ be defined on a set S containing (x_0, y_0). If f is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

Theorem 16.4 (Differentiability of multivariable functions)
Let $z = f(x,y)$ be defined on a set S. If f_x and f_y are both continuous on S, then f is differentiable on S.

These theorems assure us that essentially all functions that we see in the course of our studies here are differentiable (and hence continuous) on their natural domains. There is a difference between Definition 16.12 and Theorem 16.4, though: it is possible for a function f to be differentiable yet f_x and/or f_y is not continuous. So when f_x and f_y exist at a point but are not continuous at that point, we need to use other methods to determine whether or not f is differentiable at that point.

For instance, consider the function

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0), \end{cases}$$

We can find $f_x(0,0)$ and $f_y(0,0)$ using Definition 16.8:

$$\begin{aligned} f_x(0,0) &= \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0; \\ f_y(0,0) &= \lim_{h\to 0} \frac{f(0,0+h)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0. \end{aligned}$$

Both f_x and f_y exist at $(0,0)$, but they are not continuous at $(0,0)$, as

$$f_x(x,y) = \frac{y(y^2-x^2)}{(x^2+y^2)^2} \qquad \text{and} \qquad f_y(x,y) = \frac{x(x^2-y^2)}{(x^2+y^2)^2}$$

are not continuous at $(0,0)$. Take the limit of f_x as $(x,y) \to (0,0)$ along the x- and y-axes; they give different results. So even though f_x and f_y exist at every point in the xy-plane, they are not continuous. Therefore it is possible, by Theorem 16.4, for f to not be differentiable.

Indeed, it is not. One can show that f is not continuous at $(0,0)$ (see Example 16.5), and by Theorem 16.3, this means f is not differentiable at $(0,0)$.

16.4.2.2 Approximating with differentials

By the definition, when f is differentiable dz is a good approximation for Δz when dx and dy are small. We give a simple example of how this is used here. We can use this to approximate error propagation; that is, if the input is a little off from what it should be, how far from correct will the output be? We demonstrate this in an example.

Example 16.15

A cylindrical steel storage tank is to be built that is 10m tall and 4m across in diameter. It is known that the steel will expand/contract with temperature changes; is the overall volume of the tank more sensitive to changes in the diameter or in the height of the tank?

Solution

A cylindrical solid with height h and radius r has volume $V = \pi r^2 h$. We can view V as a function of two variables, r and h. We can compute partial derivatives of V:

$$\frac{\partial V}{\partial r} = V_r(r,h) = 2\pi rh \qquad \text{and} \qquad \frac{\partial V}{\partial h} = V_h(r,h) = \pi r^2.$$

The total differential is $dV = (2\pi rh)dr + (\pi r^2)dh$. When $h = 10$ and $r = 2$, we have $dV = 40\pi dr + 4\pi dh$. Note that the coefficient of dr is $40\pi \approx 125.7$; the coefficient of dh is a tenth of that, approximately 12.57. A small change in radius will be multiplied by 125.7, whereas a small change in height will be multiplied by 12.57. Thus the volume of the tank is more sensitive to changes in radius than in height.

The previous example showed that the volume of a particular tank was more sensitive to changes in radius than in height. Keep in mind that this analysis only applies to a tank of those dimensions. A tank with a height of 0.3 m and radius of 5 m would be more sensitive to changes in height than in radius. One could make a chart of small changes in radius and height and find exact changes in volume given specific changes.

16.4.3 Functions of n variables

The definition of differentiability for functions of n variables is very similar to that of functions of two variables. We again start with the total differential.

Definitie 16.13 (Total differential)
Let $w = f(\mathbf{x})$ be continuous on a set D. Let dx_i represent change in x_i. Where the partial derivatives f_i, $i = 1,...,n$, exist, **the total differential of** w is

$$dw = \sum_{i=1}^{n} f_{x_i}(\mathbf{x})dx_i.$$

Of course, assuming that we stick to the same increments dx_i, it is relatively straightforward to see that we can extend this definition to higher-order differentials.

Definitie 16.14 (The n-th order total differential)
Let $w = f(\mathbf{x})$ be continuous on a set D. Let dx_i represent change in x_i. Where the partial derivatives f_i, $i = 1,...,n$, exist, **the n-th order total differential of** w is

$$d^n w = \left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x})dx_i\right)^n.$$

To understand this definition correctly, it is important to realize that $\frac{\partial^n}{\partial x_i^n}$ represents the n-th power of $\frac{\partial}{\partial x_i}$ when expanding the power n according to the binomial theorem.

The first-order total differential given in Definition 16.13 can be a good approximation of the change in w when $w = f(\mathbf{x})$ is differentiable.

Definitie 16.15 (Multivariable differentiability)
Let $w = f(\mathbf{x})$ be defined on a set D containing $\mathbf{c}$ where $f_{x_i}(\mathbf{c})$, $i = 1,...,n$ exist. Let dw be the total differential of w at $\mathbf{c}$, let $\Delta w = f(c_1 + dx_1, c_2 + dx_2, ..., c_n + dx_n) - f(c_1, c_2, ..., c_n)$ $= f(\mathbf{c} + \mathbf{dx}) - f(\mathbf{c})$, and let E_{x_i} be functions of dx_i for $i = 1,...,n$ such that

$$\Delta w = dw + \sum_{i=1}^{n} E_{x_i} dx_i.$$

1. We say f **is differentiable at c** if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||\mathbf{dx}|| < \delta$, then $||\mathbf{E_x}|| < \varepsilon$.
2. We say f is differentiable on B if f is differentiable at every point in B. If f is differentiable on

$\mathbb{R}^n$, we say that f is differentiable everywhere.

Just as before, this definition gives a rigorous statement about what it means to be differentiable that is not very intuitive. We follow it with a theorem similar to Theorem 16.4.

Theorem 16.5 (Continuity and differentiability of functions of three variables)
Let $w = f(\mathbf{x})$ be defined on a set D containing $\mathbf{c}$.

1. *If f is differentiable at $\mathbf{c}$, then f is continuous at $\mathbf{c}$.*
2. *If the f_{x_i}, $i = 1, \ldots, n$ are continuous on B, then f is differentiable on B.*

16.5 The multivariable chain rule and implicit function theorem

16.5.1 Rationale

Consider driving an off-road vehicle along a dirt road. As you drive, your elevation likely changes. What factors determine how quickly your elevation rises and falls? After some thought, generally one recognizes that one's velocity (speed and direction) and the terrain influence your rise and fall.

One can represent the terrain as the surface defined by a multivariable function $z = f(x, y)$; one can represent the path of the off-road vehicle, as seen from above, with a vector–valued function $\vec{r}(t) = \big(x(t), y(t)\big)$; the velocity of the vehicle is thus $\vec{r}\,'(t) = \big(x'(t), y'(t)\big)$.

Consider Figure 16.11 in which a surface $z = f(x, y)$ is drawn, along with a dashed curve in the xy-plane. Restricting f to just the points on this circle gives the curve shown on the surface (i.e., the path of the vehicle.) The derivative $\frac{df}{dt}$ gives the instantaneous rate of change of f with respect to t. If we consider an object travelling along this path, $\frac{df}{dt} = \frac{dz}{dt}$ gives the rate at which the object rises/falls. Conceptually, the multivariable chain rule combines terrain and velocity information properly to compute this rate of elevation change.

Abstractly, let z be a function of x and y; that is, $z = f(x, y)$ for some function f, and let x and y each be functions of t. By choosing a t-value, x- and y-values are determined, which in turn determine z: this defines z as a function of t. The multivariable chain rule gives a method of computing $\frac{dz}{dt}$.

Theorem 16.6 (Multivariable chain rule, Part I)
Let $z = f(x, y)$, $x = g(t)$ and $y = h(t)$, where f, g and h are differentiable functions. Then $z = f(x, y) = f\big(g(t), h(t)\big)$ is a function of t, and

$$\begin{aligned}\frac{dz}{dt} = \frac{df}{dt} &= f_x(x, y)\frac{dx}{dt} + f_y(x, y)\frac{dy}{dt} \\ &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \qquad (16.1) \\ &= \big(f_x, f_y\big)\cdot\big(x', y'\big).\end{aligned}$$

Although this theorem should be clear if one has a good understanding of the chain rule, which forces you to work from outer function to the inner one, it can be shown more formally, for instance, by resorting to Definition 16.11, which allows us to write that

$$dz = f_x(x, y)\,dx + f_y(x, y)\,dy\,.$$

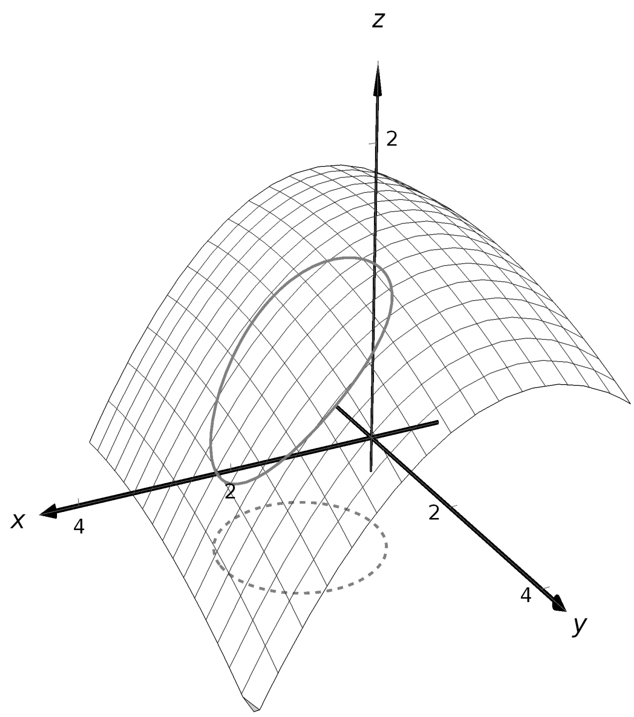


Figure 16.11: Understanding the application of the multivariable chain rule.

Dividing both sides by dt and recalling that $x = g(t)$ and $y = h(t)$ are differentiable functions, we get

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Notice, the third line of equations in Theorem 16.6. The vector $\left(f_x, f_y\right)$ contains information about the surface (terrain); the vector $\left(x', y'\right)$ can represent velocity. In the context measuring the rate of elevation change of the off-road vehicle, the multivariable chain rule states it can be found through a product of terrain and velocity information.

We now practice applying the multivariable chain rule.

Example 16.16

Let $z = f(x, y) = x^2y + x$, where $x = \sin(t)$ and $y = e^{5t}$. Find $\frac{dz}{dt}$ using the chain rule.

Solution

Following Theorem 16.6, we first find

$$f_x(x, y) = 2xy + 1, \qquad f_y(x, y) = x^2, \qquad \frac{dx}{dt} = \cos(t), \qquad \frac{dy}{dt} = 5e^{5t}.$$

Applying the theorem, we have

$$\frac{dz}{dt} = (2xy + 1)\cos(t) + 5x^2e^{5t}.$$

This may look odd, as it seems that $\frac{dz}{dt}$ is a function of x, y and t. Since x and y are functions of t,

$\frac{dz}{dt}$ is really just a function of t, and we can replace x with $\sin(t)$ and y with e^{5t} to arrive of:

$$\frac{dz}{dt} = (2\sin(t)e^{5t} + 1)\cos(t) + 5e^{5t}\sin^2(t).$$

The previous example can make us wonder: if we substituted for x and y at the end to show that $\frac{dz}{dt}$ is really just a function of t, why not substitute before differentiating, showing clearly that z is a function of t?

That is, $z = x^2y + x = \sin^2(t)e^{5t} + \sin(t)$. Applying the chain and product rules, we have

$$\frac{dz}{dt} = 2\sin(t)\cos(t)\,e^{5t} + 5\sin^2(t)\,e^{5t} + \cos(t),$$

which matches the result from the example.

This may now make one wonder "What's the point? If we could already find the derivative, why learn another way of finding it?" In some cases, applying this rule makes deriving simpler, but this is hardly the power of the chain rule. Rather, in the case where $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, the chain rule is extremely powerful when we do not know what f, g and/or h are. We demonstrate this in the next example.

Example 16.17

An object travels along a path on a surface. The exact path and surface are not known, but at time $t = t_0$ it is known that :

$$\frac{\partial z}{\partial x} = 5, \qquad \frac{\partial z}{\partial y} = -2, \qquad \frac{dx}{dt} = 3 \qquad \text{and} \qquad \frac{dy}{dt} = 7.$$

Find $\frac{dz}{dt}$ at time t_0.

Solution

The multivariable chain rule states that

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= 5(3) + (-2)(7)\\ &= 1.\end{aligned}$$

By knowing certain rates–of–change information about the surface and about the path of the particle in the xy-plane, we can determine how quickly the object is rising/falling.

Of course, we might as well be interested in the second derivative of $z = f(x,y)$. To compute that derivative, it is important to realize that one again gets a function of t upon substituting $x = g(t)$ and $y = h(t)$ in the right-hand side of Equation (16.1). So we may compute

$$\frac{d^2z}{dt^2} = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right).$$

Applying the product rule, we get

$$\frac{d^2z}{dt^2} = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\right)\frac{dx}{dt} + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{d}{dt}\left(\frac{\partial f}{\partial y}\right)\frac{dy}{dt} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2},$$

and after expanding the derivatives in the first and third term:

$$\frac{d^2z}{dt^2} = \left(\frac{\partial^2 f}{\partial x^2}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y \partial x}\frac{dy}{dt}\right)\frac{dx}{dt} + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \left(\frac{\partial^2 f}{\partial x \partial y}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y^2}\frac{dy}{dt}\right)\frac{dy}{dt} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2}.$$

After simplification, we arrive at

$$\frac{d^2z}{dt^2} = \frac{\partial^2 f}{\partial x^2}\left(\frac{dx}{dt}\right)^2 + 2\frac{\partial^2 f}{\partial x \partial y}\frac{dx}{dt}\frac{dy}{dt} + \frac{\partial^2 f}{\partial y^2}\left(\frac{dy}{dt}\right)^2 + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2},$$

which we can write as

$$\frac{d^2z}{dt^2} = \left(\frac{\partial}{\partial x}\frac{dx}{dt} + \frac{\partial}{\partial y}\frac{dy}{dt}\right)^2 f + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2}.$$

A similar reasoning is possible for higher-order derivatives. Note that, again, just as in Definition 16.14, $\left(\frac{\partial}{\partial x}\right)^2$ is to be understood as $\frac{\partial^2}{\partial x^2}$.

We can also extend the chain rule to include the situation where z is a function of more than one variable, and each of these variables is also a function of more than one variable. The basic case of this is where $z = f(x,y)$, and x and y are functions of two variables, say s and t.

Theorem 16.7 (Multivariable chain rule, Part II)

1. *Let $z = f(x,y)$, $x = g(s,t)$ and $y = h(s,t)$, where f, g and h are differentiable functions. Then z is a function of s and t, and*

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

2. *Let $z = f(\mathbf{x})$ be a differentiable function of n variables, where each of the x_i is a differentiable function of the variables $t_1, t_2, \ldots, t_n$. Then z is a function of the t_j, and*

$$\frac{\partial z}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j}.$$

Example 16.18

Let $z = f(x,y) = x^2y + x$, $x = s^2 + 3t$ and $y = 2s - t$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$, and evaluate each when $s = 1$ and $t = 2$.

Solution

Following Theorem 16.7, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = 2xy + 1 \qquad \frac{\partial f}{\partial y} = x^2,$$

$$\frac{\partial x}{\partial s} = 2s \qquad \frac{\partial x}{\partial t} = 3 \qquad \frac{\partial y}{\partial s} = 2 \qquad \frac{\partial y}{\partial t} = -1.$$

Thus

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} = (2xy+1)(2s) + (x^2)(2) = 4xys + 2s + 2x^2,$$

and

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = (2xy+1)(3) + (x^2)(-1) = 6xy - x^2 + 3.$$

When $s = 1$ and $t = 2$, $x = 7$ and $y = 0$, so

$$\frac{\partial z}{\partial s} = 100 \qquad \text{and} \qquad \frac{\partial z}{\partial t} = -46.$$

Example 16.19

Let $w = xy + z^2$, where $x = t^2 e^s$, $y = t\cos(s)$, and $z = s\sin(t)$. Find $\frac{\partial w}{\partial t}$ when $s = 0$ and $t = \pi$.

Solution

Following Theorem 16.7, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \qquad \frac{\partial f}{\partial z} = 2z,$$

$$\frac{\partial x}{\partial t} = 2te^s \qquad \frac{\partial y}{\partial t} = \cos(s) \qquad \frac{\partial z}{\partial t} = s\cos(t).$$

Thus

$$\frac{\partial w}{\partial t} = y(2te^s) + x(\cos(s)) + 2z(s\cos(t)).$$

When $s = 0$ and $t = \pi$, we have $x = \pi^2$, $y = \pi$ and $z = 0$. Thus

$$\frac{\partial w}{\partial t} = \pi(2\pi) + \pi^2 = 3\pi^2.$$

As indicated before, the real strength of the multivariable chain rule lies in the fact that one can compute derivatives and differentials of multivariable functions even not knowing the underlying function definitions. This is illustrated in the following example.

Example 16.20

In each of the following cases, determine dz and d^2z

1. $z = f(u)$ and $u = g(x,y)$
2. $z = f(u,v)$, $u = g(x,y)$ and $v = h(x,y)$

Solution

1. (a)
$$\begin{aligned} dz &= \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \\ &= \frac{df}{du}\frac{\partial u}{\partial x}\,dx + \frac{df}{du}\frac{\partial u}{\partial y}\,dy \\ &= \frac{df}{du}\left[\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right] \\ &= \frac{df}{du}\,du \end{aligned}$$

(b)
$$d^2z = \frac{\partial^2 z}{\partial x^2}\,dx^2 + 2\frac{\partial^2 z}{\partial x \partial y}\,dx\,dy + \frac{\partial^2 z}{\partial y^2}\,dy^2$$

$$
\begin{aligned}
&= \left[\frac{d^2f}{du^2}\left(\frac{\partial u}{\partial x}\right)^2 + \frac{df}{du}\frac{\partial^2 u}{\partial x^2}\right] dx^2 + 2\left[\frac{d^2f}{du^2}\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{df}{du}\frac{\partial^2 u}{\partial x \partial y}\right] dx\, dy \\
&\qquad + \left[\frac{d^2f}{du^2}\left(\frac{\partial u}{\partial y}\right)^2 + \frac{df}{du}\frac{\partial^2 u}{\partial y^2}\right] dy^2 \\
&= \frac{d^2f}{du^2}\left[\frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy\right]^2 + \frac{df}{du}\left[\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right]^2 u \\
&= \frac{d^2f}{du^2}du^2 + \frac{df}{du}d^2u
\end{aligned}
$$

2. (a)

$$
\begin{aligned}
dz &= \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \\
&= \left(\frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}\right)dx + \left(\frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y}\right)dy \\
&= \frac{\partial f}{\partial u}\left(\frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy\right) + \frac{\partial f}{\partial v}\left(\frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy\right) \\
&= \frac{\partial f}{\partial u}du + \frac{\partial f}{\partial v}dv
\end{aligned}
$$

(b)

$$
\begin{aligned}
d^2z &= \frac{\partial^2 z}{\partial x^2}dx^2 + 2\frac{\partial^2 z}{\partial x \partial y}dx\, dy + \frac{\partial^2 z}{\partial y^2}dy^2 \\
&= \left[\frac{\partial^2 f}{\partial u^2}\left(\frac{\partial u}{\partial x}\right)^2 + 2\frac{\partial^2 f}{\partial u \partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial^2 f}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2 + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial x^2} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial x^2}\right] dx^2 \\
&\quad + 2\left[\frac{\partial^2 f}{\partial u^2}\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial^2 f}{\partial u \partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} + \frac{\partial^2 f}{\partial u \partial v}\frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right. \\
&\qquad \left. + \frac{\partial^2 f}{\partial v^2}\frac{\partial v}{\partial x}\frac{\partial v}{\partial y} + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial x \partial y} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial x \partial y}\right] dxdy \\
&\quad + \left[\frac{\partial^2 f}{\partial u^2}\left(\frac{\partial u}{\partial y}\right)^2 + 2\frac{\partial^2 f}{\partial u \partial v}\frac{\partial u}{\partial y}\frac{\partial v}{\partial y} + \frac{\partial^2 f}{\partial v^2}\left(\frac{\partial v}{\partial y}\right)^2 + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial y^2} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial y^2}\right] dy^2 \\
&= \frac{\partial^2 f}{\partial u^2}\left[\left(\frac{\partial u}{\partial x}\right)^2 dx^2 + 2\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}dx\, dy + \left(\frac{\partial u}{\partial y}\right)^2 dy^2\right] \\
&\quad + 2\frac{\partial^2 f}{\partial u \partial v}\left[\frac{\partial u}{\partial x}\frac{\partial v}{\partial x}dx^2 + \frac{\partial u}{\partial x}\frac{\partial v}{\partial y}dx\, dy + \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}dx\, dy + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}dy^2\right] \\
&\quad + \frac{\partial^2 f}{\partial v^2}\left[\left(\frac{\partial v}{\partial x}\right)^2 dx^2 + 2\frac{\partial v}{\partial x}\frac{\partial v}{\partial y}dx\, dy + \left(\frac{\partial v}{\partial y}\right)^2 dy^2\right] \\
&\quad + \frac{\partial f}{\partial u}\left[\frac{\partial^2 u}{\partial x^2}dx^2 + 2\frac{\partial^2 u}{\partial x \partial y}dx\, dy + \frac{\partial^2 u}{\partial y^2}dy^2\right] \\
&\quad + \frac{\partial f}{\partial v}\left[\frac{\partial^2 v}{\partial x^2}dx^2 + 2\frac{\partial^2 v}{\partial x \partial y}dx\, dy + \frac{\partial^2 v}{\partial y^2}dy^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\partial f^2}{\partial u^2}\,du^2 + 2\frac{\partial^2 f}{\partial u \partial v}\,du\,dv + \frac{\partial^2 f}{\partial v^2}\,dv^2 + \frac{\partial f}{\partial u}\,d^2u + \frac{\partial f}{\partial v}\,d^2v \\
&= \left[\frac{\partial}{\partial u}\,du + \frac{\partial}{\partial v}\,dv\right]^2 f + \frac{\partial f}{\partial u}\,d^2u + \frac{\partial f}{\partial v}\,d^2v
\end{aligned}
$$

16.5.2 The implicit function theorem

We studied finding $\frac{dy}{dx}$ when y is given as an implicit function of x in detail in Section 9.4. We find here that the multivariable chain rule gives a simpler method of finding $\frac{dy}{dx}$.

For instance, consider the implicit function $x^2y - xy^3 = 3$. We learned to use the following steps to find $\frac{dy}{dx}$:

$$
\begin{aligned}
& \frac{d}{dx}\left(x^2y - xy^3\right) = \frac{d}{dx}\left(3\right) \\
\Rightarrow \quad & 2xy + x^2\frac{dy}{dx} - y^3 - 3xy^2\frac{dy}{dx} = 0 \\
\Leftrightarrow \quad & \frac{dy}{dx} = -\frac{2xy - y^3}{x^2 - 3xy^2}. \qquad (16.2)
\end{aligned}
$$

Instead of using this method, consider $z = x^2y - xy^3$. The implicit function above describes the level curve $z = 3$. Considering x and y as functions of x, the multivariable chain rule states that

$$
\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx}. \qquad (16.3)
$$

Since z is constant (in our example, $z = 3$) it holds that, $\frac{dz}{dx} = 0$. We also know $\frac{dx}{dx} = 1$. Consequently, equation (16.3) becomes

$$
0 = \frac{\partial z}{\partial x}(1) + \frac{\partial z}{\partial y}\frac{dy}{dx}.
$$

Consequently,

$$
\frac{dy}{dx} = -\frac{\partial z}{\partial x}\bigg/\frac{\partial z}{\partial y} = -\frac{f_x}{f_y} \qquad (16.4)
$$

Note how our solution for $\frac{dy}{dx}$ in Equation (16.2) is just the partial derivative of z with respect to x, divided by the partial derivative of z with respect to y, all multiplied by (-1).

Actually, Equation (16.4) is a consequence of the powerful implicit function theorem (*impliciete functiestelling*), which can be formulated as follows for implicit functions of two variables $F(x,y) = 0$.

Theorem 16.8 (The implicit function theorem)
Let F be a continuously differentiable implicit function of two variables x and y, i.e. it is of differentiability class C^1, let (x_0, y_0) be a point for which $F(x_0, y_0) = 0$ and let

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0\,.$$

Then there exists an open interval I containing x_0 and a unique C^1 function $f: I \mapsto \mathbb{R}$ such that $f(x_0) = y_0$ and $F(x, f(x)) = 0$ for all $x \in I$.

Essentially, the implicit function theorem allows relations

$$\left\{(x,y) \in \mathbb{R}^2 \mid F(x,y) = 0\right\}$$

to be converted to functions of a real variable. It does so by representing the relation as the graph of a function. There may, however, not be a single function whose graph can represent the entire relation, but there may be such a function on a restriction of the domain of the relation. The implicit function theorem gives a sufficient condition to ensure that there is such a function.

Example 16.21

Let us consider $F(x,y) = x^2 + y^2 - 1 = 0$. We know that the equation $x^2 + y^2 - 1 = 0$ cuts out the unit circle. There is no way to represent the unit circle as the graph of a function of one variable $y = f(x)$ because for each choice of $x \in [-1,1]$, there are two choices of y, namely $y = \pm\sqrt{1-x^2}$. It even works in situations where we do not have a formula for $F(x,y)$.

Still, as long as $y_0 \neq 0$, the conditions of Theorem 16.8 are fulfilled and we are guaranteed that there exists a unique function f of one variable x on $]-1,1[$ for which $f(x_0) = y_0$. More specifically, for $-1 < x_0 < 1$ and $y_0 > 0$, we let $f_1(x) = \sqrt{1-x^2}$ and the graph of $y = f_1(x)$ provides the upper half of the circle. Similarly, for $-1 < x_0 < 1$ and $y_0 < 0$, we let $f_2(x) = -\sqrt{1-x^2}$ and the graph of $y = f_2(x)$ gives the lower half of the circle. So, it is possible to represent part of the circle as the graph of a function of one variable.

At the intersections of the unit circle with x-axis, i.e. where $y = 0$, however, we observe that

$$\frac{\partial F}{\partial y} = 0\,,$$

which implies that the conditions in Theorem 16.8 are not met and we hence may not use it. We observe that these points of intersection lie on the graphs of both f_1 and f_2 when these are extended to the closed interval $[-1,1]$.

The purpose of the implicit function theorem is to tell us the existence of functions like $f_1(x)$ and $f_2(x)$, even in situations where we cannot write them down with explicit formulas. The theorem guarantees that $f_1(x)$ and $f_2(x)$ are differentiable. For instance, for what concerns the implicit function at stake, we may take advantage of the fact that $F(x, f_i(x)) = 0$ as long as $y_0 \neq 0$ to write

$$\begin{aligned} \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{df_i}{dx} &= 0\,, \qquad (16.5) \\ \Leftrightarrow \quad 2x + 2y\frac{df_i}{dx} &= 0\,, \end{aligned}$$

from which we conclude that

$$\frac{df_i}{dx} = -\frac{x}{y}\,.$$

Consequently, we find that

$$\frac{df_1}{dx} = -\frac{x}{\sqrt{1-x^2}}$$

or

$$\frac{df_2}{dx} = \frac{x}{\sqrt{1-x^2}}\,.$$

Of course, the second-order derivatives can computed by applying the chain rule once more to Equation (16.5) to arrive at

$$\frac{\partial^2 F}{\partial x^2} + 2\frac{\partial^2 F}{\partial x \partial y}\frac{df_i}{dx} + \frac{\partial F}{\partial y}\frac{d^2 f_i}{dx^2} + \frac{\partial^2 F}{\partial y^2}\left(\frac{df_i}{dx}\right)^2 = 0\,.$$

From this expression, it is easy to compute $\frac{d^2 f_i}{dx^2}$.

Having introduced the implicit function theorem for functions of two variables, we now extend it to implicit functions of $n+1$ variables $F(\mathbf{x}, y)$.

Theorem 16.9 (The general implicit function theorem)
Let F be a continuously differentiable implicit function of $n+1$ variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and y, i.e. it is of differentiability class C^1, let $(\mathbf{x}_0, y_0)$ be a point for which $F(\mathbf{x}_0, y_0) = 0$ and let

$$\frac{\partial F}{\partial y}(\mathbf{x}_0, y_0) \neq 0\,.$$

Then there exists an open n—dimensional ball B containing $\mathbf{x}_0$ and a unique C^1 function $f : B \mapsto \mathbb{R}$ such that $f(\mathbf{x}_0) = y_0$ and $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $\mathbf{x} \in B$.

To settle the mind, let us consider one more example.

Example 16.22

Consider the following implicit function of three variables

$$F(x, y, z) = ze^z - x - 3y = 0\,.$$

Clearly, near the point $(e, 0, 1)$ the conditions of Theorem 16.9 are fulfilled since $F(e, 0, 1) = 0$ and

$$\frac{\partial F}{\partial z}(e, 0, 1) = 2e\,.$$

So, there exists a function $z = f(x, y)$, and we may proceed by computing its first-order derivatives as follows

$$\begin{aligned} \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial f}{\partial x} = -1 + (z+1)e^z\frac{\partial f}{\partial x} &= 0, \\ \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial f}{\partial y} = -3 + (z+1)e^z\frac{\partial f}{\partial y} &= 0, \end{aligned} \tag{16.6}$$

from which we obtain

$$\frac{\partial f}{\partial x} = \frac{e^{-z}}{z+1} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{3e^{-z}}{z+1}.$$

Taking the derivatives of the expressions in Equation (16.6) yields

$$\begin{aligned}
(z+2)e^z\left(\frac{\partial f}{\partial x}\right)^2+(z+1)e^z\frac{\partial^2 f}{\partial x^2} &= 0,\\
(z+2)e^z\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}+(z+1)e^z\frac{\partial^2 f}{\partial x\partial y} &= 0,\\
(z+2)e^z\left(\frac{\partial f}{\partial y}\right)^2+(z+1)e^z\frac{\partial^2 f}{\partial y^2} &= 0,
\end{aligned}$$

such that

$$\frac{\partial^2 f}{\partial x^2}=-\frac{(z+2)e^z}{(z+1)e^z}\left(\frac{\partial f}{\partial x}\right)^2=-\frac{(z+2)e^{-2z}}{(z+1)^3}.$$

In Section 16.3 we learned how partial derivatives give certain instantaneous rate of change information about a function $z=f(x,y)$. In that section, we measured the rate of change of f by holding one variable constant and letting the other vary. We can visualize this change by considering the surface defined by f at a point and moving parallel to the x-axis.

What if we want to move in a direction that is not parallel to a coordinate axis? Can we still measure instantaneous rates of change? Yes; we find out how in the next section. In doing so, we will see how the multivariable chain rule informs our understanding of these directional derivatives.

16.6 Directional derivatives

16.6.1 Definition

Partial derivatives give us an understanding of how a surface changes when we move in the x- and y-directions. We made the comparison to standing in a rolling meadow and heading due east: the amount of rise/fall in doing so is comparable to f_x. Likewise, the rise/fall in moving due north is comparable to f_y. The steeper the slope, the greater in magnitude f_y.

But what if we did not move due north or east? What if we needed to move northeast and wanted to measure the amount of rise/fall? Partial derivatives alone cannot measure this. This section investigates **directional derivatives** (*richtingsafgeleide*), which do measure this rate of change.

We begin with a definition.

Definitie 16.16 (Directional derivative)
Let $z=f(x,y)$ be continuous on a set S and let $\hat{\boldsymbol{u}}=(u_1,u_2)$ be a unit vector. For all points (x,y), the **directional derivative of f at (x,y) in the direction of $\hat{\boldsymbol{u}}$** is

$$D_{\hat{\boldsymbol{u}}}f(x,y)=\lim_{h\to 0}\frac{f(x+hu_1,y+hu_2)-f(x,y)}{h}.$$

The partial derivatives f_x and f_y are defined with similar limits, but only x or y varies with h, not both. Here both x and y vary with a weighted h, determined by a particular unit vector $\hat{\boldsymbol{u}}$. In practice it can be a very difficult limit to evaluate, so we need an easier way of taking directional derivatives.

For that purpose, let us define a new function of a single variable,

$$g(z)=f(x_0+az,y_0+bz),$$

where x_0, y_0, a, and b are some fixed numbers. Then, by the definition of the derivative for functions of a single variable we have,

$$g'(z) = \lim_{h\to 0} \frac{g(z+h)-g(z)}{h},$$

while the derivative at $z=0$ is given by,

$$g'(0) = \lim_{h\to 0} \frac{g(h)-g(0)}{h}.$$

If we now substitute our expression for $g(z)$ we get,

$$g'(0) = \lim_{h\to 0} \frac{g(h)-g(0)}{h} = \lim_{h\to 0} \frac{f(x_0+ah, y_0+bh)-f(x_0,y_0)}{h} = D_{\hat{\mathbf{u}}} f(x_0,y_0). \tag{16.7}$$

Now, let us look at this from another perspective and rewrite $g(z)$ as follows,

$$g(z) = f(x,y),$$

where $x = x_0 + az$ and $y = y_0 + bz$. We can now use the chain rule to compute,

$$g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = f_x(x,y)a + f_y(x,y)b. \tag{16.8}$$

If we now take $z=0$ we will get that $x = x_0$ and $y = y_0$ and plug these into Equation (16.8) we get

$$g'(0) = f_x(x_0,y_0)a + f_y(x_0,y_0)b. \tag{16.9}$$

Now, simply equate Equations (16.7) and (16.9) to get that

$$D_{\hat{\mathbf{u}}} f(x_0,y_0) = f_x(x_0,y_0)a + f_y(x_0,y_0)b.$$

If we now go back to allowing x and y to be any number we get the following formula for computing directional derivatives:

$$D_{\hat{\mathbf{u}}} f(x,y) = f_x(x,y)a + f_y(x,y)b.$$

This leads to the following theorem.

Theorem 16.10 (Directional derivatives)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0), and let $\hat{\mathbf{u}} = (u_1,u_2)$ be a unit vector. The directional derivative of f at (x_0,y_0) in the direction of $\hat{\mathbf{u}}$ is

$$D_{\hat{\mathbf{u}}} f(x_0,y_0) = \left(f_x, f_y\right)\cdot(u_1,u_2) = f_x(x_0,y_0)u_1 + f_y(x_0,y_0)u_2.$$

Example 16.23

Let $z = 14 - x^2 - y^2$ and let $P = (1,2)$. Find the directional derivative of f, at P, in the following directions:

1. toward the point $Q = (3,4)$,
2. in the direction of $(2,-1)$, and
3. toward the origin.

The surface is plotted in Figure 16.12, where the point $P = (1,2)$ is indicated in the xy-plane as well as the point $(1,2,9)$ which lies on the surface of f.

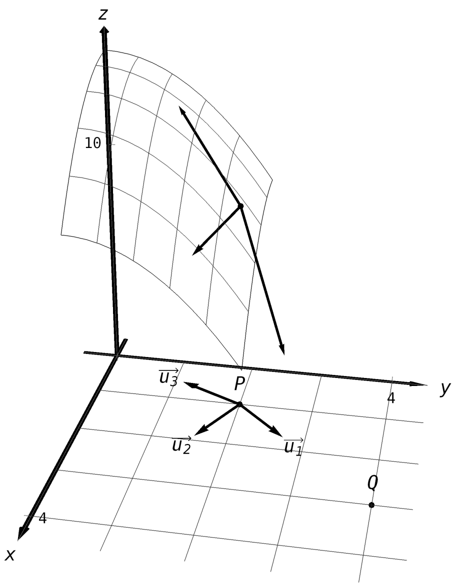


Figure 16.12: Understanding the directional derivative in Example 16.23.

Solution

We find that $f_x(x,y) = -2x$ and $f_x(1,2) = -2$; $f_y(x,y) = -2y$ and $f_y(1,2) = -4$.

1. Let $\vec{u}_1$ be the unit vector that points from the point $P = (1,2)$ to the point $Q = (3,4)$, as shown in the figure. The vector $\overrightarrow{PQ} = \langle 2,2 \rangle$; the unit vector in this direction is $\vec{u}_1 = \left\langle 1/\sqrt{2}, 1/\sqrt{2} \right\rangle$. Thus the directional derivative of f at $(1,2)$ in the direction of $\vec{u}_1$ is
$$D_{\vec{u}_1} f(1,2) = -2\left(\frac{1}{\sqrt{2}}\right) + (-4)\left(\frac{1}{\sqrt{2}}\right) = -\frac{6}{\sqrt{2}} \approx -4.24.$$
Thus the instantaneous rate of change in moving from the point $(1,2,9)$ on the surface in the direction of $\vec{u}_1$ (which points toward the point Q) is about -4.24. Moving in this direction moves one steeply downward.

2. We seek the directional derivative in the direction of $\langle 2,-1 \rangle$. The unit vector in this direction is $\vec{u}_2 = \left\langle 2/\sqrt{5}, -1/\sqrt{5} \right\rangle$. Thus the directional derivative of f at $(1,2)$ in the direction of $\vec{u}_2$ is
$$D_{\vec{u}_2} f(1,2) = -2\left(\frac{2}{\sqrt{5}}\right) + (-4)\left(-\frac{1}{\sqrt{5}}\right) = 0.$$
Starting on the surface of f at $(1,2)$ and moving in the direction of $\langle 2,-1 \rangle$ (or $\vec{u}_2$) results in no instantaneous change in z-value. This is analogous to standing on the side of a hill and choosing a direction to walk that does not change the elevation. One neither walks up nor down, rather just along the side of the hill.

3. At $P = (1,2)$, the direction towards the origin is given by the vector $\langle -1,-2 \rangle$; the unit vector in this direction is $\vec{u}_3 = \left\langle -1/\sqrt{5}, -2/\sqrt{5} \right\rangle$. The directional derivative of f at P in the direction of the origin is
$$D_{\vec{u}_3} f(1,2) = -2\left(-\frac{1}{\sqrt{5}}\right) + (-4)\left(-\frac{2}{\sqrt{5}}\right) = \frac{10}{\sqrt{5}} \approx 4.47.$$

Moving towards the origin means walking uphill quite steeply, with an initial slope of about 4.47.

16.6.2 The gradient

As we study directional derivatives, it will help to make an important connection between the unit vector $\hat{\boldsymbol{u}} = (u_1, u_2)$ that describes the direction and the partial derivatives f_x and f_y. We start with a definition.

Definitie 16.17 (Gradient)
Let $z = f(x, y)$ be differentiable on a set S that contains the point (x_0, y_0).

1. The **gradient** (*gradiënt*) of f is

$$\vec{\nabla} f(x, y) = \big(f_x(x, y), f_y(x, y)\big).$$

2. The **gradient** of f at (x_0, y_0) is

$$\vec{\nabla} f(x_0, y_0) = \big(f_x(x_0, y_0), f_y(x_0, y_0)\big).$$

The symbol $\vec{\nabla}$ is named **nabla**, derived from the Greek name of a Jewish harp. Oddly enough, in mathematics the expression $\vec{\nabla} f$ is pronounced del f.

To simplify notation, we often express the gradient as $\vec{\nabla} f = \big(f_x, f_y\big)$. The gradient allows us to compute directional derivatives in terms of a dot product:

$$D_{\hat{\boldsymbol{u}}} f = \vec{\nabla} f \cdot \hat{\boldsymbol{u}}. \tag{16.10}$$

The properties of the dot product studied in Chapter 6 allow us to investigate the properties of the directional derivative. Given that the directional derivative gives the instantaneous rate of change of z when moving in the direction of $\hat{\boldsymbol{u}}$, three questions naturally arise:

1. In what direction(s) is the change in z the greatest (i.e., the steepest uphill)?
2. In what direction(s) is the change in z the least (i.e., the steepest downhill)?
3. In what direction(s) is there no change in z?

Relying on the geometric interpretation of the dot product (Theorem 6.2), we have

$$\vec{\nabla} f \cdot \hat{\boldsymbol{u}} = \left\|\vec{\nabla} f\right\| \|\vec{\boldsymbol{u}}\| \cos(\theta) = \left\|\vec{\nabla} f\right\| \cos(\theta), \tag{16.11}$$

where θ is the angle between the gradient and $\hat{\boldsymbol{u}}$. (Since $\hat{\boldsymbol{u}}$ is a unit vector, $\|\vec{\boldsymbol{u}}\| = 1$.) This equation allows us to answer the three questions stated previously.

1. Equation (16.11) is maximized when $\cos(\theta) = 1$, i.e., when the gradient and $\hat{\boldsymbol{u}}$ have the same direction. We conclude the gradient points in the direction of greatest z change.
2. Equation (16.11) is minimized when $\cos(\theta) = -1$, i.e., when the gradient and $\hat{\boldsymbol{u}}$ have opposite directions. We conclude the gradient points in the opposite direction of the least z change.
3. Equation (16.11) is 0 when $\cos(\theta) = 0$, i.e., when the gradient and $\hat{\boldsymbol{u}}$ are orthogonal to each other. We conclude the gradient is orthogonal to directions of no z change.

This result is rather amazing. Once again imagine standing in a rolling meadow and face the direction that leads you steepest uphill. Then the direction that leads steepest downhill is directly behind you, and side–stepping either left or right (i.e., moving perpendicularly to the direction you face) does not change your elevation at all.

Recall that a level curve is defined as a curve in the xy-plane along which the z-values of a function do not change. Let a surface $z = f(x,y)$ be given, and let us represent one such level curve as a vector–valued function, $\vec{\boldsymbol{r}}(t) = \big(x(t), y(t)\big)$. As the output of f does not change along this curve, $f\big(x(t), y(t)\big) = c$ for all t, for some constant c.

Since f is constant for all t, $\frac{df}{dt} = 0$. By the multivariable chain rule, we also know

$$\begin{aligned}\frac{df}{dt} &= f_x(x,y)x'(t) + f_y(x,y)y'(t)\\ &= \big(f_x(x,y), f_y(x,y)\big)\cdot\big(x'(t), y'(t)\big)\\ &= \vec{\nabla} f\cdot\vec{\boldsymbol{r}}\,'(t)\\ &= 0.\end{aligned}$$

This last equality states $\vec{\nabla} f\cdot\vec{\boldsymbol{r}}\,'(t) = 0$: the gradient is orthogonal to the derivative of $\vec{\boldsymbol{r}}$, meaning the gradient is orthogonal to the graph of $\vec{\boldsymbol{r}}$. Our conclusion: at any point on a surface, the gradient at that point is orthogonal to the level curve that passes through that point.

We restate these ideas in a theorem, then use them in an example.

Theorem 16.11 (The gradient and directional derivatives)
Let $z = f(x,y)$ be differentiable on a set S with gradient $\vec{\nabla} f$, let $P = (x_0, y_0)$ be a point in S and let $\vec{\boldsymbol{u}}$ be a unit vector.

1. *The maximum value of $D_{\vec{\boldsymbol{u}}}\, f(x_0,y_0)$ is $\left\|\vec{\nabla} f(x_0,y_0)\right\|$; the direction of maximal z increase is $\vec{\nabla} f(x_0,y_0)$.*
2. *The minimum value of $D_{\vec{\boldsymbol{u}}}\, f(x_0,y_0)$ is $-\left\|\vec{\nabla} f(x_0,y_0)\right\|$; the direction of maximal z decrease is $-\vec{\nabla} f(x_0,y_0)$.*
3. *At P, $\vec{\nabla} f(x_0,y_0)$ is orthogonal to the level curve passing through $\big(x_0, y_0, f(x_0,y_0)\big)$.*

We now illustrate how to find the directions of maximal increase and decrease.

Example 16.24

Let $f(x,y) = \sin(x)\cos(y)$ and let $P = (\pi/3, \pi/3)$. Find the directions of maximal increase and decrease, and find a direction where the instantaneous rate of z change is 0.

Solution

We begin by finding the gradient. We have that $f_x = \cos(x)\cos(y)$ and $f_y = -\sin(x)\sin(y)$, thus

$$\vec{\nabla} f = \big(\cos(x)\cos(y), -\sin(x)\sin(y)\big)$$

and, at P

$$\vec{\nabla} f\left(\frac{\pi}{3}, \frac{\pi}{3}\right) = \left(\frac{1}{4}, -\frac{3}{4}\right).$$

Thus the direction of maximal increase is $(1/4, -3/4)$. In this direction, the instantaneous rate of z change is $||\,(1/4, -3/4)\,|| = \sqrt{10}/4 \approx 0.79$.

Figure 16.13 shows the surface. The gradient is drawn at P with a dashed line (because of the nature of this surface, the gradient points into the surface). Let $\vec{u} = \langle u_1, u_2 \rangle$ be the unit vector in the direction of $\vec{\nabla} f$ at P. The graph also contains the vector $\langle u_1, u_2, ||\vec{\nabla} f|| \rangle$. This vector has a run of 1 (because in the xy-plane it moves 1 unit) and a rise of $||\vec{\nabla} f||$, hence we can think of it as a vector with slope of $||\vec{\nabla} f||$ in the direction of $\vec{\nabla} f$, helping us visualize how steep the surface is in its steepest direction.

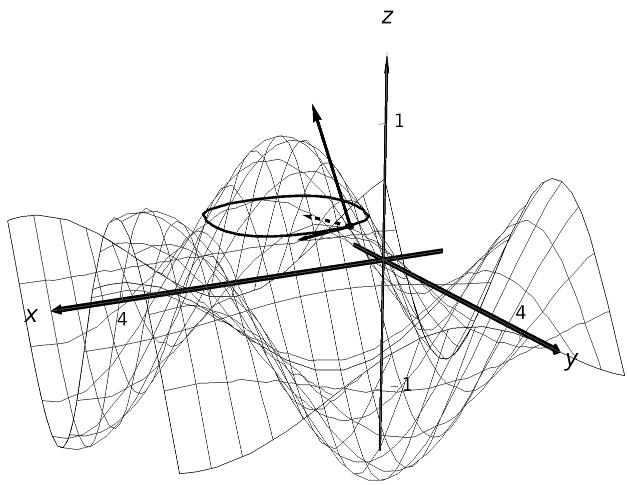


Figure 16.13: Graphing the surface and important directions in Example 16.24.

The direction of maximal decrease is $\langle -1/4, 3/4 \rangle$; in this direction the instantaneous rate of z change is $-\sqrt{10}/4 \approx -0.79$.

Any direction orthogonal to $\vec{\nabla} f$ is a direction of no z change. We have two choices: the direction of $\langle 3, 1 \rangle$ and the direction of $\langle -3, -1 \rangle$. The unit vector in the direction of $\langle 3, 1 \rangle$ is shown in the graph as well. The level curve at $z = \sqrt{3}/4$ is drawn: recall that along this curve the z-values do not change. Since $\langle 3, 1 \rangle$ is a direction of no z-change, this vector is tangent to the level curve at P.

It is as well important to figure out when $\vec{\nabla} f = 0$.

Example 16.25

Let $f(x, y) = -x^2 + 2x - y^2 + 2y + 1$. Find the directional derivative of f in any direction at $P = (1, 1)$.

Solution

We find $\vec{\nabla} f = \langle -2x + 2, -2y + 2 \rangle$. At P, we have $\vec{\nabla} f(1, 1) = \langle 0, 0 \rangle$. According to Theorem 16.11, this is the direction of maximal increase. However, $\langle 0, 0 \rangle$ is directionless; it has no displacement. And regardless of the unit vector $\vec{u}$ chosen, $D_{\vec{u}} f = 0$.

Figure 16.14 helps us understand what this means. We can see that P lies at the top of a paraboloid. In all directions, the instantaneous rate of change is 0. So what is the direction of maximal increase? It is fine to give an answer of $\vec{0} = \langle 0, 0 \rangle$, as this indicates that all directional derivatives are 0.

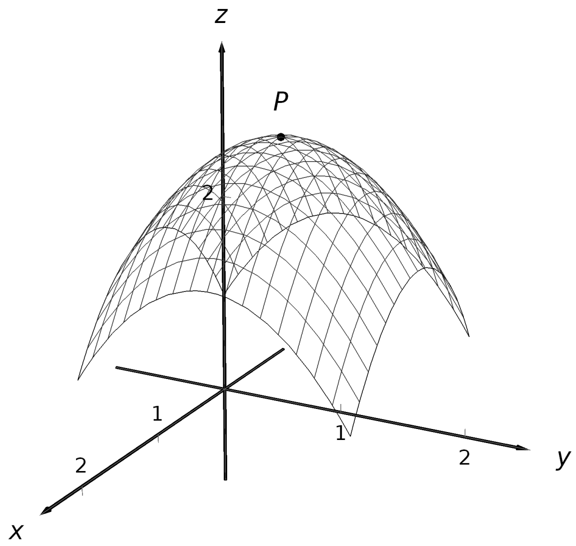


Figure 16.14: At the top of a paraboloid, all directional derivatives are 0.

In Mathematica, we could have computed the gradient in Example 16.25 using the command `Grad` as follows

```
In[30]:= Grad[-x^2 + 2*x - y^2 + 2*y + 1, {x, y}]

Out[30]= {2-2 x,2-2 y}
```

The fact that the gradient of a surface always points in the direction of steepest increase/decrease is very useful, as illustrated in the following example.

Example 16.26

Consider the surface given by $f(x,y) = 20 - x^2 - 2y^2$. Water is poured on the surface at $(1, 1/4)$. What path does it take as it flows downhill?

Solution

Let $\vec{r}(t) = \big(x(t), y(t)\big)$ be the vector–valued function describing the path of the water in the xy-plane; we seek $x(t)$ and $y(t)$. We know that water will always flow downhill in the steepest direction; therefore, at any point on its path, it will be moving in the direction of $-\vec{\nabla} f$. We ignore the physical effects of momentum on the water. Thus $\vec{r}\,'(t)$ will be parallel to $\vec{\nabla} f$, and there is some constant c such that $c\vec{\nabla} f = \vec{r}\,'(t) = \big(x'(t), y'(t)\big)$.

We find $\vec{\nabla} f = (-2x, -4y)$ and write $x'(t)$ as $\frac{dx}{dt}$ and $y'(t)$ as $\frac{dy}{dt}$. Then

$$\begin{aligned} c\vec{\nabla} f &= \big(x'(t), y'(t)\big) \\ \iff \qquad (-2cx, -4cy) &= \left(\frac{dx}{dt}, \frac{dy}{dt}\right). \end{aligned}$$

This implies

$$-2cx = \frac{dx}{dt} \qquad \text{and} \qquad -4cy = \frac{dy}{dt},$$

i.e.

$$c = -\frac{1}{2x}\frac{dx}{dt} \qquad \text{and} \qquad c = -\frac{1}{4y}\frac{dy}{dt}.$$

As c equals both expressions, we have

$$\frac{1}{2x}\frac{dx}{dt} = \frac{1}{4y}\frac{dy}{dt}.$$

To find an explicit relationship between x and y, we can integrate both sides with respect to t. Recall from our study of differentials that $\dfrac{dx}{dt}dt = dx$. Thus:

$$\begin{aligned} \int \frac{1}{2x}\frac{dx}{dt}\,dt &= \int \frac{1}{4y}\frac{dy}{dt}\,dt \\ \Leftrightarrow \quad \int \frac{1}{2x}\,dx &= \int \frac{1}{4y}\,dy \\ \Leftrightarrow \quad \frac{1}{2}\ln|x| &= \frac{1}{4}\ln|y| + C_1 \\ \Leftrightarrow \quad 2\ln|x| &= \ln|y| + 4C_1 \\ \Leftrightarrow \quad \ln(x^2) &= \ln|y| + 4C_1. \end{aligned}$$

Now raise both sides as a power of e:

$$\begin{aligned} x^2 &= e^{\ln|y|+4C_1} \\ \Leftrightarrow \quad x^2 &= e^{\ln|y|}e^{4C_1}. \end{aligned}$$

From which it follows that

$$x^2 = yC_2,$$

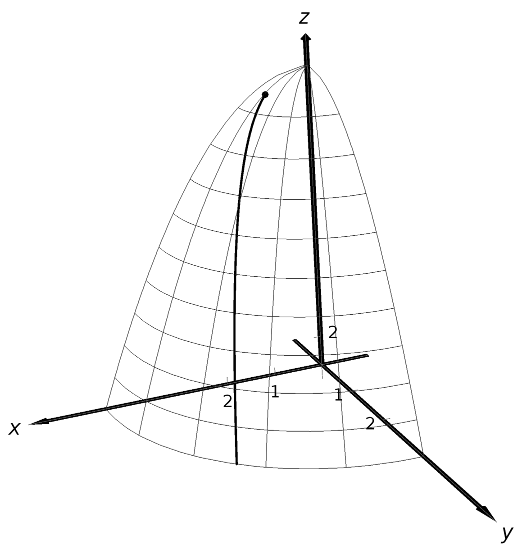
where $C_2 = \pm e^{4C_1}$, or alternatively

$$Cx^2 = y,$$

where $C = 1/C_2$. As the water started at the point $(1, 1/4)$, we can solve for C:

$$C(1)^2 = \frac{1}{4} \quad \Leftrightarrow \quad C = \frac{1}{4}.$$

Thus the water follows the curve $y = x^2/4$ in the xy-plane. The surface and the path of the water is graphed in Figure 16.15(a). In Figure 16.15(b), the level curves of the surface are plotted in the xy-plane, along with the curve $y = x^2/4$. Notice how the path intersects the level curves at right angles. As the path follows the gradient downhill, this reinforces the fact that the gradient is orthogonal to level curves.



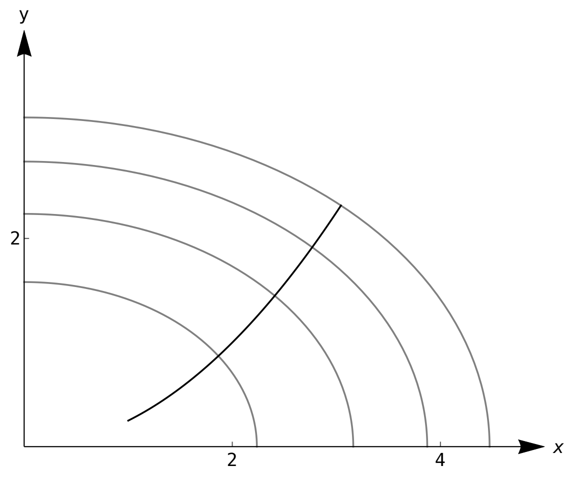


Figure 16.15: A graph of the surface described in Example 16.26 along with the path in the xy-plane with the level curves.

16.6.3 Functions of n variables

The concepts of directional derivatives and the gradient are easily extended to n variables. We combine the concepts behind Definitions 16.16 and 16.17 and Theorem 16.10 into one set of definitions.

Definitie 16.18 (Directional derivatives and gradient with n variables)
Let $w = f(\mathbf{x})$ be differentiable on a set D and let $\hat{\boldsymbol{u}}$ be a unit vector in $\mathbb{R}^n$.

1. The **gradient of** f is $\vec{\nabla} f = \mathbf{f_x} = \left(f_{x_1}, f_{x_2}, \ldots, f_{x_n}\right)$.
2. The **directional derivative of** f **in the direction of** $\hat{\boldsymbol{u}}$ is

$$D_{\hat{\boldsymbol{u}}} f = \vec{\nabla} f \cdot \hat{\boldsymbol{u}}.$$

The same properties of the gradient given in Theorem 16.11, hold for a function of three variables.

Theorem 16.12 (The gradient and directional derivatives with n variables)
Let $w = f(\mathbf{x})$ be differentiable on a set D, let $\vec{\nabla} f$ be the gradient of f, and let $\hat{\mathbf{u}}$ be a unit vector.

1. *The maximum value of $D_{\hat{\mathbf{u}}} f$ is $\left\|\vec{\nabla} f\right\|$, obtained when the angle between $\vec{\nabla} f$ and $\hat{\mathbf{u}}$ is 0, i.e., the direction of maximal increase is $\vec{\nabla} f$.*
2. *The minimum value of $D_{\hat{\mathbf{u}}} f$ is $-\left\|\vec{\nabla} f\right\|$, obtained when the angle between $\vec{\nabla} f$ and $\hat{\mathbf{u}}$ is π, i.e., the direction of maximal decrease is $-\vec{\nabla} f$.*
3. *$D_{\hat{\mathbf{u}}} f = 0$ when $\vec{\nabla} f$ and $\hat{\mathbf{u}}$ are orthogonal.*

We interpret the third statement of the theorem as the gradient is orthogonal to level surfaces, the three–variable analogue to level curves.

Example 16.27

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0,0,0)$, it holds that

$$I(x,y,z) = \frac{k}{x^2+y^2+z^2}$$

for some constant k.

Let $k = 1$, let $\hat{\boldsymbol{u}} = (2/3, 2/3, 1/3)$ be a unit vector, and let $P = (2,5,3)$. Measure distances in centimetres. Find the directional derivative of I at P in the direction of $\hat{\boldsymbol{u}}$, and find the direction of greatest intensity increase at P.

Solution

We need the gradient $\vec{\nabla} I$, so we compute I_x, I_y and I_z:

$$\begin{aligned}
\vec{\nabla} I &= \left(\frac{-2x}{(x^2+y^2+z^2)^2}, \frac{-2y}{(x^2+y^2+z^2)^2}, \frac{-2z}{(x^2+y^2+z^2)^2} \right) \\
\Rightarrow \quad \vec{\nabla} I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444} \right) \approx (-0.003, -0.007, -0.004) \\
\Rightarrow \quad D_{\hat{\boldsymbol{u}}} I &= \vec{\nabla} I(2,5,3) \cdot \hat{\boldsymbol{u}} \\
&= -\frac{17}{2166} \approx -0.0078.
\end{aligned}$$

The directional derivative tells us that moving in the direction of $\hat{\boldsymbol{u}}$ from P results in a decrease in intensity of about -0.008 units per centimetre. The intensity is decreasing as $\hat{\boldsymbol{u}}$ moves one farther from the origin than P.

The gradient gives the direction of greatest intensity increase. Notice that

$$\begin{aligned}
\vec{\nabla} I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444} \right) \\
&= \frac{2}{1444} (-2, -5, -3).
\end{aligned}$$

That is, the gradient at $(2,5,3)$ is pointing in the direction of $(-2,-5,-3)$, that is, towards the origin. That should make intuitive sense: the greatest increase in intensity is found by moving towards to source of the energy.

The directional derivative allows us to find the instantaneous rate of z change in any direction at a point. We can use these instantaneous rates of change to define lines and planes that are tangent to a surface at a point, which is the topic of the next section.

Using the gradient of functions of n variables introduced in Definition 16.18, we can easily state the multivariable counterpart of the mean value theorem of differentiation (Theorem 10.4).

Theorem 16.13 (Mean value theorem for functions of *n* variables)
Let f be differentiable on D, let $\vec{\boldsymbol{a}}$ and $\vec{\boldsymbol{b}}$ belong to D and let $\vec{\nabla} f$ be the gradient of f, then there exists a point $\vec{\boldsymbol{c}}$ on the line segment L connecting $\vec{\boldsymbol{a}}$ and $\vec{\boldsymbol{b}}$ for which it holds that

$$\vec{\nabla} f(\vec{\boldsymbol{c}}) \cdot \left(\vec{\boldsymbol{b}} - \vec{\boldsymbol{a}}\right) = f(\vec{\boldsymbol{b}}) - f(\vec{\boldsymbol{a}})\,.$$

Proof The key of the proof of this theorem is to design an appropriate function of one variable to which we can then apply the mean value theorem for functions of one variable (Theorem 10.4). For that purpose, we ought to realize that as we let t range from 0 to 1, $\vec{\boldsymbol{a}} + t(\vec{\boldsymbol{b}} - \vec{\boldsymbol{a}})$ traces out the line segment L. So, the idea of the proof is to apply the one-variable mean-value theorem to the function

$$\vec{\boldsymbol{\phi}}(t) = \vec{\boldsymbol{a}} + t\left(\vec{\boldsymbol{b}} - \vec{\boldsymbol{a}}\right) = (1-t)\vec{\boldsymbol{a}} + t\vec{\boldsymbol{b}}\,,$$

for $t \in [0,1]$. Clearly, $\vec{\boldsymbol{\phi}}(t)$ is differentiable with $\vec{\boldsymbol{\phi}}'(t) = \vec{\boldsymbol{b}} - \vec{\boldsymbol{a}}$. Consequently, the composition $g = f \circ \vec{\boldsymbol{\phi}}$ is differentiable by the chain rule. Since g provides a mapping from the unit interval to the real numbers, i.e. $g : [0,1] \mapsto \mathbb{R}$, the mean value theorem for functions of one variable applied to g gives us a point ξ in $[0,1]$, such that

$$g(1) - g(0) = g'(\xi)\,.$$

Now, let $\mathbf{c} = \vec{\boldsymbol{\phi}}(\xi) \in L$. We then have

$$\begin{aligned} f(\vec{\boldsymbol{b}}) - f(\vec{\boldsymbol{a}}) &= g(1) - g(0) \\ &= g'(\xi)\,, \\ &= \vec{\nabla} f\left(\vec{\boldsymbol{\phi}}(\xi)\right) \cdot \vec{\boldsymbol{\phi}}'(\xi) \\ &= \vec{\nabla} f(\vec{\boldsymbol{c}}) \cdot (\vec{\boldsymbol{b}} - \vec{\boldsymbol{a}})\,, \end{aligned}$$

which concludes the proof. □

For comprehensiveness, it should be noted that the transition from $g'(\xi)$ to $\vec{\nabla} f\left(\vec{\boldsymbol{\phi}}(\xi)\right) \cdot \vec{\boldsymbol{\phi}}'(\xi)$ in the right-hand side follows from clever use of the chain rule. More precisely, f is a function of n variables x_i applied to the output of the single-variable vector-valued function $\vec{\boldsymbol{\phi}}(t)$ with components $x_i = \phi_i(t)$, so the chain rule (Theorem 16.6) immediately gives

$$\frac{dg}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}\,,$$

which can be written using the dot product as

$$\frac{dg}{dt} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) \cdot \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}\right).$$

16.7 Tangent lines, normal lines, and tangent planes

16.7.1 Tangent and normal lines

Derivatives and tangent lines go hand–in–hand. Given $y = f(x)$, the line tangent to the graph of f at $x = x_0$ is the line through $\left(x_0, f(x_0)\right)$ with slope $f'(x_0)$; that is, the slope of the tangent line is the instantaneous rate of change of f at x_0. When dealing with functions of two variables, the graph is no longer a curve but a surface. At a given point on the surface, it seems there are many lines that fit our intuition of being tangent to the surface.

In Figure 16.16 we see lines that are tangent to curves in space. Since each curve lies on a surface, it makes sense to say that the lines are also tangent to the surface. The next definition formally defines what it means to be tangent to a surface.

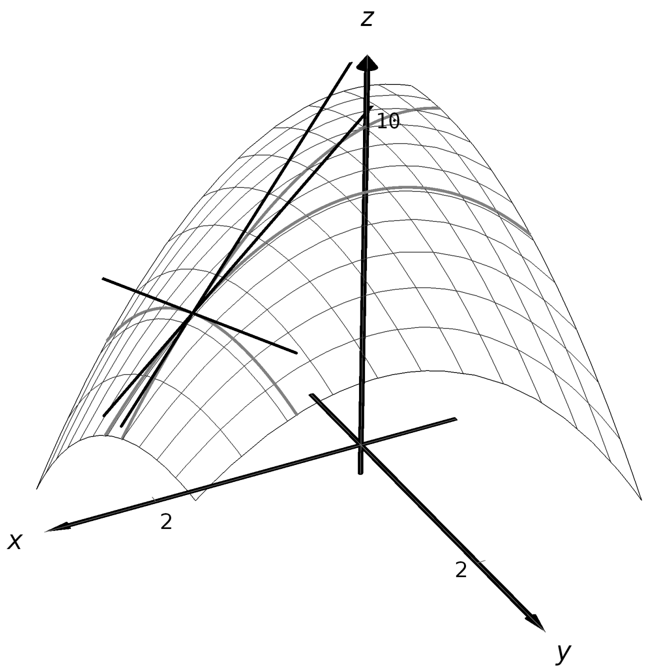


Figure 16.16: Showing various lines tangent to a surface.

Definitie 16.19 (Directional tangent line)
Let $z = f(x, y)$ be differentiable on a set S containing (x_0, y_0) and let $\hat{\boldsymbol{u}} = (u_1, u_2)$ be a unit vector.

1. The line ℓ_x through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\big(1, 0, f_x(x_0, y_0)\big)$ is **the tangent line to f in the direction of x at** (x_0, y_0).
2. The line ℓ_y through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\big(0, 1, f_y(x_0, y_0)\big)$ is **the tangent line to f in the direction of y at** (x_0, y_0).
3. The line $\ell_{\hat{\boldsymbol{u}}}$ through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\big(u_1, u_2, D_{\hat{\boldsymbol{u}}} f(x_0, y_0)\big)$ is **the tangent line to f in the direction of $\hat{\boldsymbol{u}}$ at** (x_0, y_0).

It is instructive to consider each of three directions given in the definition in terms of slope. The direction of ℓ_x is $\big(1, 0, f_x(x_0, y_0)\big)$; that is, the "run" is one unit in the x-direction and the rise is $f_x(x_0, y_0)$ units in the z-direction. Note how the slope is just the partial derivative with respect to x. A similar statement can be made for ℓ_y. The direction of $\ell_{\hat{\boldsymbol{u}}}$ is $\big(u_1, u_2, D_{\hat{\boldsymbol{u}}} f(x_0, y_0)\big)$; the run is one unit in the $\hat{\boldsymbol{u}}$ direction (where $\hat{\boldsymbol{u}}$ is a unit vector) and the rise is the directional derivative of z in that direction.

Definition 16.19 leads to the following parametric equations of directional tangent lines:

$$\ell_x(t) = \begin{cases} x &= x_0 + t \\ y &= y_0 \\ z &= z_0 + f_x(x_0, y_0)t, \end{cases} \quad \ell_y(t) = \begin{cases} x &= x_0 \\ y &= y_0 + t \\ z &= z_0 + f_y(x_0, y_0)t \end{cases} \quad \text{and}\ \ell_{\hat{\boldsymbol{u}}}(t) = \begin{cases} x &= x_0 + u_1 t \\ y &= y_0 + u_2 t \\ z &= z_0 + D_{\hat{\boldsymbol{u}}} f(x_0, y_0)t. \end{cases}$$

where $z_0 = f(x_0, y_0)$.

Example 16.28

Find the lines tangent to the surface $z = \sin(x)\cos(y)$ at $(\pi/2, \pi/2)$ in the x- and y- directions and also in the direction of $\vec{v} = \langle -1, 1 \rangle$.

Solution

The partial derivatives with respect to x and y are:

$$f_x(x,y) = \cos(x)\cos(y)$$
$$f_y(x,y) = -\sin(x)\sin(y).$$

from which it follows that $f_x(\pi/2, \pi/2) = 0$ and $f_y(\pi/2, \pi/2) = -1$. At $(\pi/2, \pi/2)$, the z-value is 0.

Thus the parametric equations of the line tangent to f at $(\pi/2, \pi/2)$ in the directions of x and y are:

$$\ell_x(t) = \begin{cases} x = \pi/2 + t \\ y = \pi/2 \\ z = 0 \end{cases} \quad \text{and} \quad \ell_y(t) = \begin{cases} x = \pi/2 \\ y = \pi/2 + t \\ z = -t. \end{cases}$$

The two lines are shown with the surface in Figure 16.17(a).

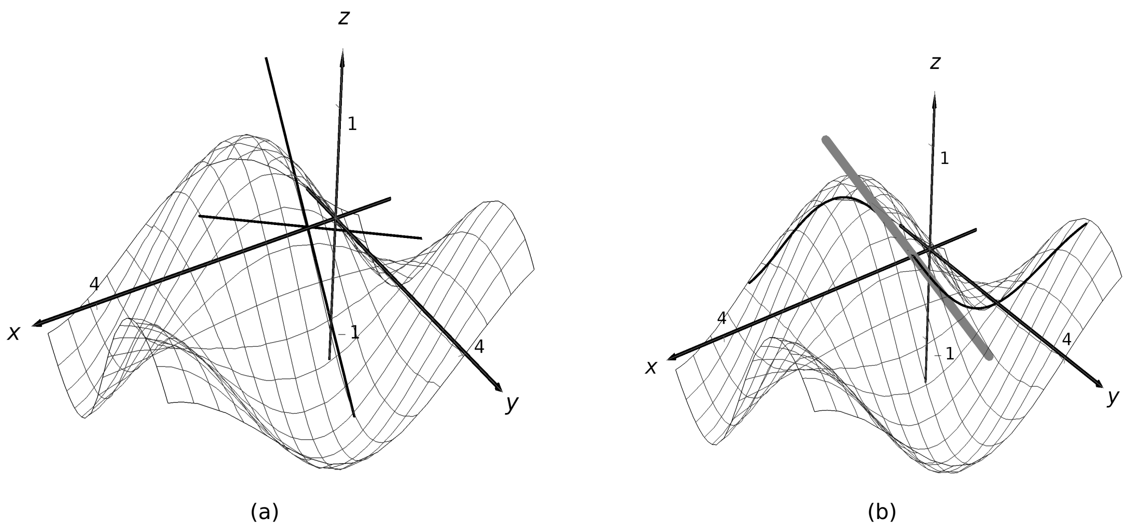


Figure 16.17: A surface and directional tangent lines in Example 16.28.

To find the equation of the tangent line in the direction of $\vec{v}$, we first find the unit vector in the direction of $\vec{v}$: $\hat{u} = \left\langle -1/\sqrt{2}, 1/\sqrt{2} \right\rangle$. The directional derivative at $(\pi/2, \pi/2)$ in the direction of $\hat{u}$ is

$$D_{\hat{u}} f\left(\frac{\pi}{2}, \frac{\pi}{2}\right) = \langle 0, -1 \rangle \cdot \left\langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = -\frac{1}{\sqrt{2}}.$$

Thus the directional tangent line is

$$\ell_{\hat{u}}(t) = \begin{cases} x = \dfrac{\pi}{2} - \dfrac{t}{\sqrt{2}} \\ y = \dfrac{\pi}{2} + \dfrac{t}{\sqrt{2}} \\ z = -\dfrac{t}{\sqrt{2}} \end{cases}.$$

The curve through $(\pi/2, \pi/2, 0)$ in the direction of $\vec{v}$ is shown in Figure 16.17(b) along with $\ell_{\hat{u}}(t)$.

The following example shows that the points on surfaces where all tangent lines have a slope of 0 can give us some information about the extrema of functions of several variables.

Example 16.29

Let $f(x,y) = 4xy - x^4 - y^4$. Find the equations of all directional tangent lines to f at $(1,1)$.

Solution

First note that $f(1,1) = 2$. We need to compute directional derivatives, so we need $\vec{\nabla} f$. We begin by computing partial derivatives.

$$f_x = 4y - 4x^3 \quad \Rightarrow \quad f_x(1,1) = 0; \qquad f_y = 4x - 4y^3 \quad \Rightarrow \quad f_y(1,1) = 0.$$

Thus $\vec{\nabla} f(1,1) = (0,0)$. Let $\hat{u} = (u_1, u_2)$ be any unit vector. The directional derivative of f at $(1,1)$ will be $D_{\hat{u}} f(1,1) = (0,0) \cdot (u_1, u_2) = 0$. It does not matter what direction we choose; the directional derivative is always 0. Therefore

$$\ell_{\hat{u}}(t) = \begin{cases} x = 1 + u_1 t \\ y = 1 + u_2 t \\ z = 2. \end{cases}$$

Figure 16.18 shows a graph of f and the point $(1,1,2)$. Note that this point comes at the top of a hill, and therefore every tangent line through this point will have a slope of 0.

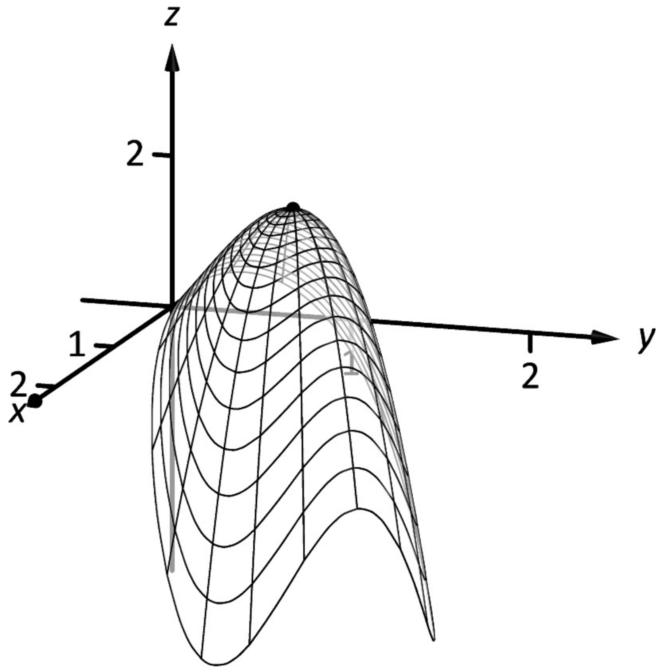


Figure 16.18: Graphing f in Example 16.29.

That is, consider any curve on the surface that goes through this point. Each curve will have a relative maximum at this point, hence its tangent line will have a slope of 0. The following section investigates the points on surfaces where all tangent lines have a slope of 0.

When dealing with a function $y = f(x)$ of one variable, we stated that a line through $(c, f(c))$ was tangent to f if the line had a slope of $f'(c)$ and was normal to f if it had a slope of $-1/f'(c)$. We extend the concept of normal, or orthogonal, to functions of two variables.

Let $z = f(x,y)$ be a differentiable function of two variables. By Definition 16.19, at (x_0, y_0), $\ell_x(t)$ is a line parallel to the vector $\vec{\boldsymbol{d}}_x = \big(1, 0, f_x(x_0, y_0)\big)$ and $\ell_y(t)$ is a line parallel to $\vec{\boldsymbol{d}}_y = \big(0, 1, f_y(x_0, y_0)\big)$. Since lines in these directions through $\big(x_0, y_0, f(x_0, y_0)\big)$ are tangent to the surface, a line through this point and orthogonal to these directions would be orthogonal, or normal, to the surface. We can use this direction to create a normal line.

The direction of the normal line is orthogonal to $\vec{\boldsymbol{d}}_x$ and $\vec{\boldsymbol{d}}_y$, hence the direction is parallel to $\vec{\boldsymbol{d}}_n = \vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y$. It turns out this cross product has a very simple form:

$$\vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y = (1, 0, f_x) \times \big(0, 1, f_y\big) = \big(-f_x, -f_y, 1\big).$$

It is often more convenient to refer to the opposite of this direction, namely $\big(f_x, f_y, -1\big)$. This leads to a definition.

Definitie 16.20 (Normal line)
Let $z = f(x, y)$ be differentiable on a set S containing (x_0, y_0) where

$$a = f_x(x_0, y_0) \quad \text{and} \quad b = f_y(x_0, y_0)$$

are defined.

1. A nonzero vector parallel to $\vec{\boldsymbol{n}} = (a, b, -1)$ is orthogonal to f at $P = \big(x_0, y_0, f(x_0, y_0)\big)$.
2. The line ℓ_n through P with direction parallel to $\vec{\boldsymbol{n}}$ is the **normal line** (*normaal*) to f at P.

Thus the parametric equations of the normal line to a surface f at $\big(x_0, y_0, f(x_0, y_0)\big)$ are:

$$\ell_n(t) = \begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = f(x_0, y_0) - t. \end{cases}$$

The direction of the normal line has many uses, one of which is the definition of the **tangent plane** (*raakvlak*) which we define shortly. Another use is in measuring distances from the surface to a point. Given a point Q in space, it is a general geometric concept to define the distance from Q to the surface as being the length of the shortest line segment $\overrightarrow{PQ}$ over all points P on the surface. This, in turn, implies that $\overrightarrow{PQ}$ will be orthogonal to the surface at P. Therefore we can measure the distance from Q to the surface f by finding a point P on the surface such that $\overrightarrow{PQ}$ is parallel to the normal line to f at P.

Example 16.30

Let $f(x, y) = 2 - x^2 - y^2$ and let $Q = (2, 2, 2)$. Find the distance from Q to the surface defined by f.

Solution

From Definition 16.20, we know that at (x, y) the direction of the normal line will be $\vec{\boldsymbol{d}}_n = (-2x, -2y, -1)$. A point P on the surface will have coordinates $(x, y, 2 - x^2 - y^2)$, so we have $\overrightarrow{PQ} = \big(2 - x, 2 - y, x^2 + y^2\big)$. To find where $\overrightarrow{PQ}$ is parallel to $\vec{\boldsymbol{d}}_n$, we need to find x, y and c such that $c\overrightarrow{PQ} = \vec{\boldsymbol{d}}_n$.

$$\begin{aligned} c\overrightarrow{PQ} &= \vec{\boldsymbol{d}}_n \\ \Rightarrow \quad c\big(2 - x, 2 - y, x^2 + y^2\big) &= (-2x, -2y, -1) \end{aligned}$$

This implies

$$c(2 - x) = -2x,$$

$$c(2-y) = -2y,$$
$$c(x^2+y^2) = -1.$$

In each equation, we can solve for c:

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

The first two fractions imply $x = y$, and so the last fraction can be rewritten as $c = -1/(2x^2)$. Then

$$\begin{aligned} & \frac{-2x}{2-x} &=& \frac{-1}{2x^2} \\ \Leftrightarrow & -2x(2x^2) &=& -1(2-x) \\ \Leftrightarrow & 4x^3 &=& 2-x \\ \Leftrightarrow & 4x^3+x-2 &=& 0. \end{aligned}$$

This last equation is a cubic, which is not difficult to solve with a numeric solver. We find that $x \approx 0.689$, hence $P = (0.689, 0.689, 1.051)$. We find the distance from Q to the surface of f is

$$\left\|\overrightarrow{PQ}\right\| = \sqrt{(2-0.689)^2+(2-0.689)^2+(2-1.051)^2} = 2.083.$$

We can of course take the concept of measuring the distance from a point to a surface to find a point Q a particular distance from a surface at a given point P on the surface.

16.7.2 Tangent plane

We can use the direction of the normal line to define a plane. With $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$ and $P = \big(x_0,y_0,f(x_0,y_0)\big)$, the vector $\vec{\boldsymbol{n}} = (a,b,-1)$ is orthogonal to f at P. The plane through P with normal vector $\vec{\boldsymbol{n}}$ is therefore tangent to f at P.

Definitie 16.21 (Tangent plane)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0), where $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$, $\vec{\boldsymbol{n}} = (a,b,-1)$ and $P = \big(x_0,y_0,f(x_0,y_0)\big)$.

The plane through P with normal vector $\vec{\boldsymbol{n}}$ is the **tangent plane to f at P** (*raakvlak aan f in P*). The standard form of this plane is

$$\begin{aligned} & \vec{\boldsymbol{n}}\cdot\big((x-x_0),(y-y_0),(z-f(x_0,y_0))\big) &=& 0 \\ \Leftrightarrow & a(x-x_0)+b(y-y_0)-\big(z-f(x_0,y_0)\big) &=& 0. \end{aligned}$$

Example 16.31

Find the equation of the normal line and tangent plane to $z = -x^2-y^2+2$ at $(0,1)$.

Solution

We find $z_x(x,y) = -2x$ and $z_y(x,y) = -2y$; at $(0,1)$, we have $z_x = 0$ and $z_y = -2$. We take the direction of the normal line, following Definition 16.20, to be $\vec{n} = (0,-2,-1)$. The line with this direction going through the point $(0,1,1)$ is

$$\ell_n(t) = \begin{cases} x = 0 \\ y = 1-2t \\ z = 1-t \end{cases} \quad \text{or} \quad \ell_n(t) = (0,1,1) + (0,-2,-1)\,t.$$

The surface $z = -x^2 - y^2 + 2$, along with the found normal line, is graphed in Figure 16.19(a).

Since we have that $\vec{n} = (0,-2,-1)$ and $P = (0,1,1)$, the equation of the tangent plane is

$$-2(y-1) - (z-1) = 0.$$

The surface $z = -x^2 - y^2 + 2$ and tangent plane are graphed in Figure 16.19(b).

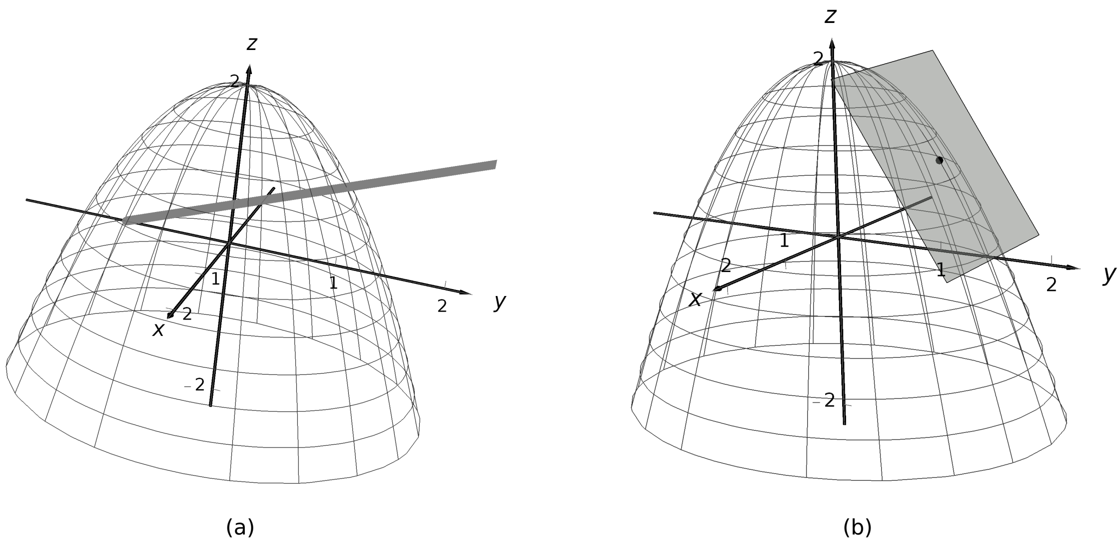


Figure 16.19: Graphing a surface with normal line and tangent plane from Example 16.31.

Just as tangent lines can be used to approximate function values of functions of one variable, tangent planes can be used to achieve this for functions of two variables.

Example 16.32

The point $(3,-1,4)$ lies on the surface of an unknown differentiable function f where $f_x(3,-1) = 2$ and $f_y(3,-1) = -1/2$. Find the equation of the tangent plane to f at P, and use this to approximate the value of $f(2.9,-0.8)$.

Solution

Knowing the partial derivatives at $(3,-1)$ allows us to form the normal vector to the tangent plane, $\vec{n} = (2,-1/2,-1)$. Thus the equation of the tangent line to f at P is:

$$2(x-3) - \frac{1}{2}(y+1) - (z-4) = 0 \quad \Longleftrightarrow \quad z = 2(x-3) - \frac{1}{2}(y+1) + 4. \tag{16.12}$$

Just as tangent lines provide excellent approximations of curves near their point of intersection, tangent planes provide excellent approximations of surfaces near their point of intersection. So

$f(2.9,-0.8) \approx z(2.9,-0.8) = 3.7.$

This is not a new method of approximation. Compare the right hand expression for z in Equation (16.12) to the total differential:

$$dz = f_x dx + f_y dy \quad \text{and} \quad z = \underbrace{\underbrace{2}_{f_x}\underbrace{(x-3)}_{dx} + \underbrace{-1/2}_{f_y}\underbrace{(y+1)}_{dy}}_{dz} + 4.$$

Thus the new z-value is the sum of the change in z (i.e., dz) and the old z-value (4). As mentioned when studying the total differential, it is not uncommon to know partial derivative information about an unknown function f, and tangent planes are used to give accurate approximations of f.

Recall that when $w = f(x,y,z)$, the gradient $\vec{\nabla} f = \left(f_x, f_y, f_z\right)$ is orthogonal to level surfaces of f. Given a point (x_0, y_0, z_0), let $c = f(x_0, y_0, z_0)$. Then $f(x,y,z) = c$ is a level surface that contains the point (x_0, y_0, z_0) and $\vec{\nabla} f(x_0, y_0, z_0)$ is orthogonal to this level surface. So, the gradient at a point gives a vector orthogonal to the surface at that point. This direction can be used to find tangent planes and normal lines.

Example 16.33

Find the equation of the plane tangent to the ellipsoid

$$\frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4} = 1$$

at $P = (1,2,1)$.

Solution

We consider the equation of the ellipsoid as a level surface of a function F of three variables, where $F(x,y,z) = \frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4}$. The gradient is:

$$\begin{aligned} \vec{\nabla} F(x,y,z) &= \left(F_x, F_y, F_z\right) \\ &= \left(\frac{x}{6}, \frac{y}{3}, \frac{z}{2}\right). \end{aligned}$$

At P, the gradient is $\vec{\nabla} F(1,2,1) = (1/6, 2/3, 1/2)$. Thus the equation of the plane tangent to the ellipsoid at P is

$$\frac{1}{6}(x-1) + \frac{2}{3}(y-2) + \frac{1}{2}(z-1) = 0.$$

The ellipsoid and tangent plane are graphed in Figure 16.20.

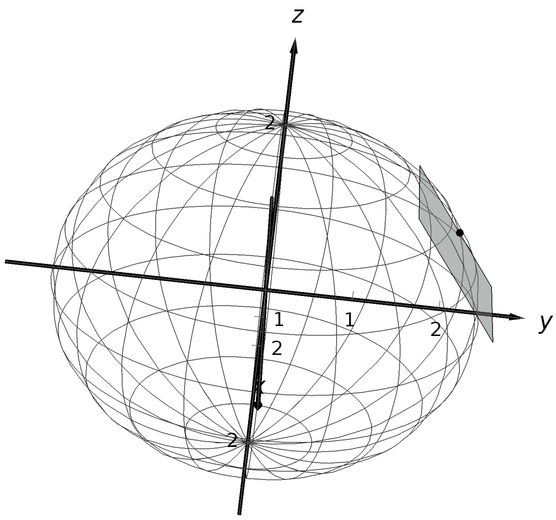


Figure 16.20: An ellipsoid and its tangent plane at a point.

Tangent lines and planes to surfaces have many uses, including the study of instantaneous rates of changes and making approximations. Normal lines also have many uses. In this section we focused on using them to measure distances from a surface. Another interesting application is in computer graphics, where the effects of light on a surface are determined using normal vectors.

The next section investigates another use of partial derivatives: Taylor series expansions of functions of several variables.

16.8 Taylor series expansions

Recall that we found in Section 14.7 a way of rewriting a continuous function of one variable as a series by relying on Taylor's theorem. Having introduced all mathematical tools that are needed to analyse functions of several variables, we are now ready to introduce Taylor's theorem for a function of two variables.

Theorem 16.14 (Taylor's theorem for a function of two variables)
Let f be a C^{n+1} function on a set D containing (x_0, y_0). Then, for each (x, y) in D, there exists (θ_x, θ_y) between (x, y) and (x_0, y_0) such that

$$f(x,y) = \sum_{i=0}^{n} \frac{1}{i!}\left(\frac{\partial}{\partial x}(x-x_0) + \frac{\partial}{\partial y}(y-y_0)\right)^i f(x_0,y_0) + R_n(x,y)\,,$$

where

$$f(x,y) = f(x_0,y_0) + \frac{1}{1!}f_x(x_0,y_0)(x-x_0) + \frac{1}{1!}f_y(x_0,y_0)(y-y_0) + \frac{1}{2!}f_{xx}(x_0,y_0)(x-x_0)^2 + \frac{2}{2!}f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \frac{1}{2!}f_{yy}(x_0,y_0)(y-y_0)^2 + \ldots$$

and the remainder term is given by

$$R_n(x,y) = \frac{1}{(n+1)!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^{n+1} f(\theta_x,\theta_y).$$

Proof For the sake of brevity, we will restrict the proof to functions of two variables, though the same reasoning may be followed to arrive at a full proof of the n variables case.

Essentially, the proof starts off in a similar way as the one of the mean value theorem for functions of n variables (Theorem 16.13) by devising an appropriate one-variable function to which we can apply Taylor's theorem of functions of one variable (Theorem 14.22).

We start by considering a point in the neighbourhood of (x_0,y_0). Let us say, the point $(x_0+\Delta x, y_0+\Delta y)$ and we then look for a formula for $f(x_0+\Delta x, y_0+\Delta y)$. As we will try to use Taylor's theorem for functions of one variable, we consider the vector-valued function $\vec{\boldsymbol{\phi}}:\mathbb{R}\mapsto\mathbb{R}^2$ of one variable:

$$\vec{\boldsymbol{\phi}}(t) = (x_0+t\Delta x, y_0+t\Delta y)$$

for $t\in[0,1]$, which traces out the line segment connecting (x_0,y_0) and $(x_0+\Delta x, y_0+\Delta y)$. Subsequently, we consider the composition

$$g = f\circ\vec{\boldsymbol{\phi}},$$

which is C^{n+1} in agreement with what we concluded in the proof of Theorem 16.13. Consequently, we may apply Taylor's theorem of functions of one variable to g, meaning that for every $t\in[0,1]$ there exists a $\theta_t\in]0,1[$ such that

$$g(t) = g(0)+g'(0)t+\frac{g''(0)}{2!}t^2+\cdots+\frac{g^{(n)}(0)}{n!}t^n+R_n(t), \tag{16.13}$$

where

$$R_n(t) = \frac{g^{(n+1)}(\theta_t)}{(n+1)!}t^{(n+1)}.$$

Now, note that

$$g(0) = f(x_0,y_0) \qquad \text{and} \qquad g(1) = f(x_0+\Delta x, y_0+\Delta y),$$

while the higher-order derivatives in Equation (16.13) can be obtained using the chain rule. In this way, we obtain

$$\begin{aligned}
g(t) &= f(x_0+t\Delta x, y_0+t\Delta y)\\
g'(t) &= \Delta x\frac{\partial f}{\partial x}(x_0+t\Delta x, y_0+t\Delta y)+\Delta y\frac{\partial f}{\partial y}(x_0+t\Delta x, y_0+t\Delta y)\\
&= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right)f(x_0+t\Delta x, y_0+t\Delta y)\\
g''(t) &= \Delta x^2\frac{\partial^2 f}{\partial x^2}(x_0+t\Delta x, y_0+t\Delta y)+2\Delta x\Delta y\frac{\partial^2 f}{\partial x\partial y}(x_0+t\Delta x, y_0+t\Delta y)+\Delta y^2\frac{\partial^2 f}{\partial y^2}(x_0+t\Delta x, y_0+t\Delta y)\\
&= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right)^2 f(x_0+t\Delta x, y_0+t\Delta y)\\
&\cdots\\
g^{(n)}(t) &= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right)^n f(x_0+t\Delta x, y_0+t\Delta y)
\end{aligned}$$

More specifically, we immediately get

$$g^{(n)}(0) = \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y}\right)^n f(x_0, y_0)\,.$$

Consequently, choosing $t = 1$ in Equation (16.13) we arrive at

$$g(1) = f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y}\right) f(x_0, y_0) + \cdots$$
$$+ \frac{1}{2!}\left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y}\right)^2 f(x_0, y_0) + \cdots + \frac{1}{n!}\left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y}\right)^n f(x_0, y_0) + R_n(x, y)\,,$$

where

$$R_n(x, y) = \frac{1}{(n+1)!}\left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y}\right)^{n+1} f(x_0 + \theta_t \Delta x, y_0 + \theta_t \Delta y)\,.$$

Finally, acknowledging that $\Delta x = x - x_0$ and $\Delta y = y - y_0$, we can restate the previous expression as

$$g(1) = f(x, y) = f(x_0, y_0) + \left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right) f(x_0, y_0) + \cdots$$
$$+ \frac{1}{2}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^2 f(x_0, y_0) + \cdots + \frac{1}{n!}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^n f(x_0, y_0) + R_n(x, y)\,,$$

$$R_n(x, y) = \frac{1}{(n+1)!}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^{n+1} f(\theta_x, \theta_y)\,,$$

where $\theta_x = x_0 + \theta_t(x - x_0)$ and $\theta_y = y_0 + \theta_t(y - y_0)$. □

As for functions of one variable, the Talor polynomial of degree n provides the best n-th degree polynomial approximation of $f(x, y)$ near a point (x_0, y_0). For instance, letting $n = 1$, we obtain

$$f(x, y) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Example 16.34

Find a second-degree polynomial approximation to the function

$$f(x, y) = \sqrt{x^2 + y^3}$$

near the point $(1, 2)$ and use it to estimate the value of $\sqrt{1.02^2 + 1.97^3}$.

Solution

For a second-degree approximation we need the values of the partial derivatives of f up to the second order at the point $(1, 2)$. We have

Derivative function	Derivative at $(1,2)$
$f(x,y) = \sqrt{x^2+y^3}$	$f(1,2) = 3$
$f_x(x,y) = \dfrac{x}{\sqrt{x^2+y^3}}$	$f_x(1,2) = \dfrac{1}{3}$
$f_y(x,y) = \dfrac{3y^2}{2\sqrt{x^2+y^3}}$	$f_x(1,2) = 2$
$f_{xx}(x,y) = \dfrac{y^3}{\left(x^2+y^3\right)^{3/2}}$	$f_{xx}(1,2) = \dfrac{8}{27}$
$f_{xy}(x,y) = \dfrac{-3xy^2}{2\left(x^2+y^3\right)^{3/2}}$	$f_{xy}(1,2) = -\dfrac{2}{9}$
$f_{yy}(x,y) = \dfrac{12x^2y+3y^4}{4\left(x^2+y^3\right)^{3/2}}$	$f_{yy}(1,2) = \dfrac{2}{3}.$

Thus, we get after evaluating the partial derivatives in $(1,2)$

$$f(x,y) \approx 3 + \frac{1}{3}(x-1) + 2(y-2) + \frac{4}{27}(x-1)^2 - \frac{2}{9}(x-1)(y-2) + \frac{1}{3}(y-2)^2.$$

This is the required second-degree Taylor polynomial for f near $(1,2)$. Therefore,

$$\begin{aligned}
\sqrt{1.02^2+1.97^3} &= f(1+0.02, 2-0.03) \\
&\approx 3 + \frac{1}{3}(0.02) + 2(-0.03) + \frac{4}{27}(0.02)^2 - \frac{2}{9}(0.02)(-0.03) + \frac{1}{3}(-0.03)^2 \\
&\approx 2.9471593.
\end{aligned}$$

It can be verified that the true value is 2.9471636, so our approximation is accurate to six significant figures.

Clearly, in line with what we devised for functions of one variable, the Taylor series expansion of a function of two variables is given by

$$f(x,y) = \sum_{i=0}^{+\infty} \frac{1}{i!}\left(\frac{\partial}{\partial x}(x-x_0) + \frac{\partial}{\partial y}(y-y_0)\right)^i f(x_0,y_0),$$

and likewise a Maclaurin series expansion can be formulated.

Example 16.35

Find a second-order Taylor series expansion of the function

$$f(x,y) = e^x \ln(1+y),$$

around the point $(0,0)$.

Solution

In order to compute a second-order Taylor series expansion we first compute the necessary partial derivatives and evaluate these derivatives at the origin:

Derivative function	Derivative at $(0,0)$
$f_x(x,y) = e^x \ln(1+y)$	$f_x(0,0) = 0$
$f_y(x,y) = \dfrac{e^x}{1+y}$	$f_y(0,0) = 1$
$f_{xx}(x,y) = e^x \ln(1+y)$	$f_{xx}(0,0) = 0$
$f_{xy}(x,y) = f_{yx} = \dfrac{e^x}{1+y}$	$f_{xy}(0,0) = 1$
$f_{yy}(x,y) = -\dfrac{e^x}{(1+y)^2}$	$f_{yy}(0,0) = -1$

Relying on Taylor's theorem, this leads to

$$\begin{aligned} f(x,y) &= 0+0(x-0)+1(y-0)+\frac{1}{2}\Big(0(x-0)^2+2(x-0)(y-0)+(-1)(y-0)^2\Big)+\cdots \\ &= y+xy-\frac{y^2}{2}+\cdots \end{aligned}.$$

Finally, we have

$$e^x \ln(1+y) = y+xy-\frac{y^2}{2}+\cdots$$

for $y > -1$.

Of course, Taylor series expansions may be extended to functions of n variables.

16.9 Extreme values

16.10 Extreme values

Given a function $z = f(x,y)$, we are often interested in points where z takes on the largest or smallest values. For instance, if z represents a cost function, we would likely want to know what (x,y) values minimize the cost. If z represents the ratio of a volume to surface area, we would likely want to know where z is greatest. This leads to the following definition.

Definitie 16.22 (Relative and absolute extreme)
Let $z = f(x,y)$ be defined on a set S containing the point $P = (x_0, y_0)$.

1. If $f(x_0,y_0) \geq f(x,y)$ for all (x,y) in S, then f has an **absolute maximum** (*globaal maximum*) at P.

   If $f(x_0,y_0) \leq f(x,y)$ for all (x,y) in S, then f has an **absolute minimum** (*globaal minimum*) at P.

2. If there is an open disk D containing P such that $f(x_0,y_0) \geq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative maximum** (*lokaal maximum*) at P.

   If there is an open disk D containing P such that $f(x_0,y_0) \leq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative minimum** (*lokaal minimum*) at P.

3. If f has an absolute maximum or minimum at P, then f has an **absolute extrema** at P.

If f has a relative maximum or minimum at P, then f has a **relative extrema** at P.

If f has a relative or absolute maximum at $P = (x_0, y_0)$, it means that every curve on the surface of f through P will also have a relative or absolute maximum at P. Recalling what we learned in Section 10.1, the slopes of the tangent lines to these curves at P must be 0 or undefined. Since directional derivatives are computed using f_x and f_y, we are led to the following definition and theorem.

Definitie 16.23 (Critical point)
Let $z = f(x,y)$ be continuous on a set S. A **critical point** (*kritisch punt*) $P = (x_0, y_0)$ of f is a point in S such that, at P,

$$f_x(x_0, y_0) = 0 \text{ and } f_y(x_0, y_0) = 0$$

Besides, when $f_x(x_0, y_0)$ and/or $f_y(x_0, y_0)$ is undefined, we call $P = (x_0, y_0)$ a **singular point**, just as we did within the framework of functions of one variable (Definition 10.4).

Theorem 16.15 (Critical and singular points and relative extrema)
Let $z = f(x,y)$ be defined on an open set S containing $P = (x_0, y_0)$. If f has a relative extrema at P, then P is a critical or singular point of f.

Therefore, to find relative extrema, we find the critical and singular points of f and determine which correspond to relative maxima, relative minima, or neither. The following examples demonstrates this.

Example 16.36

Let $f(x,y) = -\sqrt{x^2+y^2} + 2$. Find the relative extrema of f. The surface of f is graphed in Figure 16.21 along with the point $(0,0,2)$.

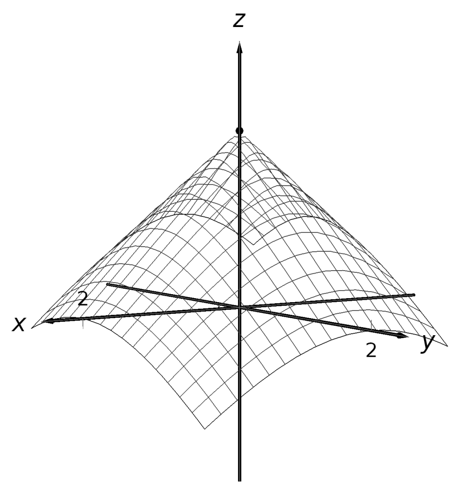


Figure 16.21: The surface in Example 16.36 with its absolute maximum indicated.

Solution

We start by computing the partial derivatives of f:

$$f_x(x,y) = \frac{-x}{\sqrt{x^2+y^2}} \quad \text{and} \quad f_y(x,y) = \frac{-y}{\sqrt{x^2+y^2}}.$$

It is clear that $f_x = 0$ when $x = 0$ and $y \neq 0$, and that $f_y = 0$ when $y = 0$ and $x \neq 0$. At $(0,0)$, both

f_x and f_y are not 0, but rather undefined. The point $(0,0)$ is hence a singular point, though. The graph in Figure 16.21 shows that this point is the absolute maximum of f.

Example 16.37

Let $f(x,y) = x^3 - 3x - y^2 + 4y$. Find the relative extrema of f.

Solution

Once again we start by finding the partial derivatives of f:

$$f_x(x,y) = 3x^2 - 3 \qquad \text{and} \qquad f_y(x,y) = -2y + 4.$$

Each is always defined. Setting each equal to 0 and solving for x and y, we find

$$\begin{aligned} f_x(x,y) = 0 &\quad \Rightarrow \quad x = \pm 1 \\ f_y(x,y) = 0 &\quad \Rightarrow \quad y = 2. \end{aligned}$$

We have two critical points: $(-1,2)$ and $(1,2)$, while there are no singular points. To determine if they correspond to a relative maximum or minimum, we consider the graph of f in Figure 16.22.

The critical point $(-1,2)$ clearly corresponds to a relative maximum. However, the critical point at $(1,2)$ is neither a maximum nor a minimum, displaying a different, interesting characteristic.

If one walks parallel to the y-axis towards this critical point, then this point becomes a relative maximum along this path. But if one walks towards this point parallel to the x-axis, this point becomes a relative minimum along this path. A point that seems to act as both a maximum and a minimum is a saddle point. A formal definition follows.

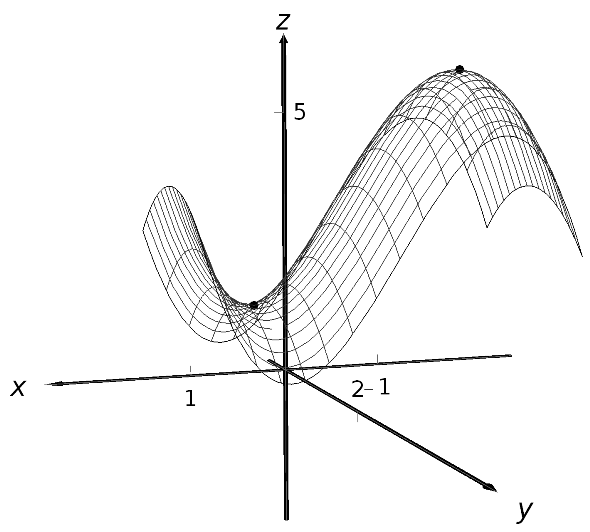


Figure 16.22: The surface in Example 16.37 with both critical points marked.

Definitie 16.24 (Saddle point)
Let $P = (x_0, y_0)$ be in the domain of f where $f_x = 0$ and $f_y = 0$ at P. We say P is a **saddle point** (*zadelpunt*) of f if, for every open disk D containing P, there are points (x_1, y_1) and (x_2, y_2) in D such that $f(x_0,y_0) > f(x_1,y_1)$ and $f(x_0,y_0) < f(x_2,y_2)$.

At a saddle point, the instantaneous rate of change in all directions is 0 and there are points nearby with z-values both less than and greater than the z-value of the saddle point.

Before Example 16.37 we mentioned the need for a test to differentiate between relative maxima and minima. We now recognize that our test also needs to account for saddle points. To do so, we consider the second partial derivatives of f. Recall that with single variable functions, such as $y = f(x)$, if $f''(c) > 0$, then if f is concave up at c, and if $f'(c) = 0$, then f has a relative minimum at $x = c$. Note that at a saddle point, it seems the graph is both concave up and concave down, depending on which direction you are considering.

It would be nice if the following were true:

$$\begin{aligned} f_{xx} \text{ and } f_{yy} > 0 \quad &\Rightarrow \quad \text{relative minimum,} \\ f_{xx} \text{ and } f_{yy} < 0 \quad &\Rightarrow \quad \text{relative maximum,} \\ f_{xx} \text{ and } f_{yy} \text{ have opposite signs} \quad &\Rightarrow \quad \text{saddle point.} \end{aligned}$$

However, this is not the case. Functions f exist where f_{xx} and f_{yy} are both positive but a saddle point still exists. In such a case, while the concavity in the x-direction is up (i.e., $f_{xx} > 0$) and the concavity in the y-direction is also up (i.e., $f_{yy} > 0$), the concavity switches somewhere in between the x- and y-directions.

To account for this, consider

$$D = f_{xx}f_{yy} - f_{xy}f_{yx}.$$

Since f_{xy} and f_{yx} are equal when continuous (refer back to Theorem 16.2), we can rewrite this as $D = f_{xx}f_{yy} - f_{xy}^2$. D can be used to test whether the concavity at a point changes depending on direction. If $D > 0$, the concavity does not switch (i.e., at that point, the graph is concave up or down in all directions). If $D < 0$, the concavity does switch. If $D = 0$, our test fails to determine whether concavity switches or not. We state the use of D in the following theorem.

Theorem 16.16 (Second derivative test)
Let R be an open set on which a function $z = f(x, y)$ and all its first and second partial derivatives are defined, let $P = (x_0, y_0)$ be a critical point of f in R, and let

$$D(x_0, y_0) = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0).$$

1. *If $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then f has a relative minimum at P.*
2. *If $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then f has a relative maximum at P.*
3. *If $D < 0$, then f has a saddle point at P.*
4. *If $D = 0$, the test is inconclusive.*

We practice this test with the function in the previous example, where we visually determined we had a relative maximum and a saddle point.

Example 16.38

Let $f(x, y) = x^3 - 3x - y^2 + 4y$ as in Example 16.37. Determine whether the function has a relative minimum, maximum, or saddle point at each critical point.

Solution

We determined previously that the critical points of f are $(-1, 2)$ and $(1, 2)$. To use the second derivative test, we must find the second partial derivatives of f:

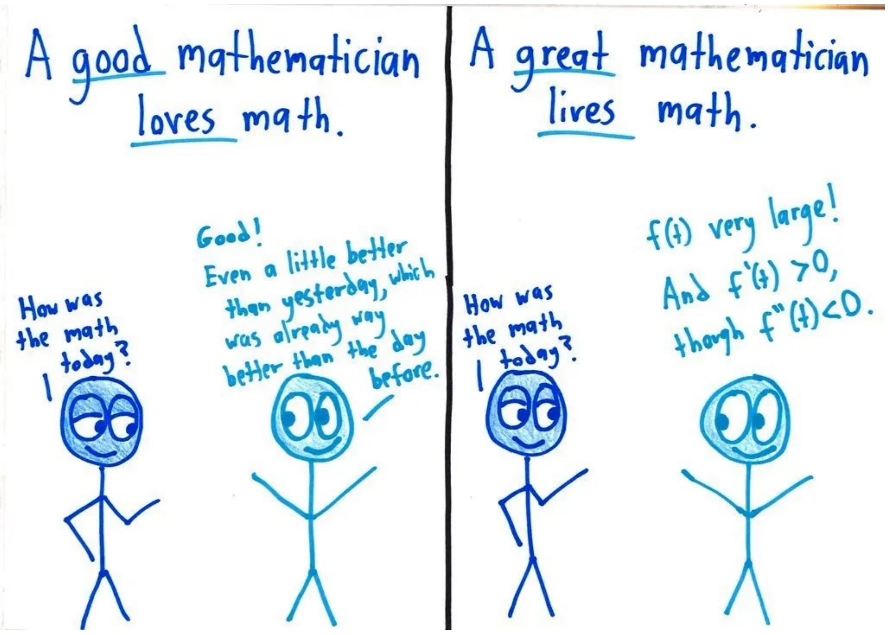
$$f_{xx} = 6x; \qquad f_{yy} = -2; \qquad f_{xy} = 0.$$

Thus $D(x,y) = -12x$.

At $(-1,2)$: $D(-1,2) = 12 > 0$, and $f_{xx}(-1,2) = -6$. By the second derivative test, f has a relative maximum at $(-1,2)$.

At $(1,2)$: $D(1,2) = -12 < 0$. The second derivative test states that f has a saddle point at $(1,2)$.

The second derivative test confirmed what we determined visually.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

16.11 Exercices

Introduction to multivariable functions

Assignment 16.1 — Find the domain and image of the functions below.

(a) $f(x,y) = \dfrac{\ln(x)}{\sin(y)}$

(b) $f(x,y) = \dfrac{2+\arcsin(y)}{\ln(2x)}$

(c) $f(x,y) = \sqrt{1-x^2-y^2}$

(d) $f(x,y) = \sin(xy)$

(e) $f(x,y) = |x|-|y|$

(f) $f(x,y) = \dfrac{1}{x-y}$

(g) $f(x,y) = \ln^{-1}(x^2+y^2-3)$

(h) $f(x,y) = \pi - \arcsin(x^2+2y^2)$

Assignment 16.2 — Sketch some level curves for the functions below

(a) $f(x,y) = \dfrac{x^2}{y}$

(b) $f(x,y) = \dfrac{y}{x^2+y^2}$

(c) $f(x,y) = \sqrt{\dfrac{1}{y}-x^2}$

Assignment 16.3 — In each case, describe the graph of the function $f(x,y)$ whose level curves $f(x,y)=C$ are represented in Figure 16.23 (a) and (b).

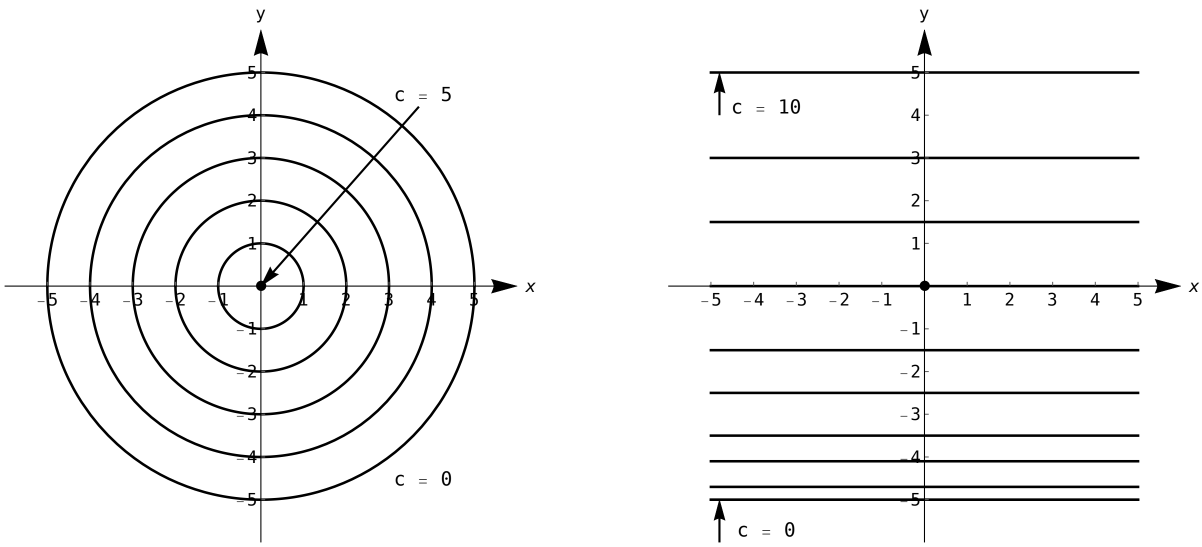


Figure 16.23: Level curves $f(x,y)=C$ of the function $f(x,y)$ from Exercise 3.

Partial derivatives

Assignment 16.4 — Find $f_x(0,0)$ and $f_y(0,0)$ for the functions below, if they exist, by using definition: 16.8.

$$f(x,y) = \begin{cases} (x^3+y)\sin\left(\dfrac{1}{x^2+y^2}\right), & \text{if } (x,y) \neq (0,0), \\ 0, & \text{if } (x,y) = (0,0). \end{cases}$$

Find $f_x(x,y)$. Is it continuous in $(0,0)$?

Assignment 16.5 — Find the partial derivatives of the first and second order of the functions below, and at the specified point.

(a) $f(x,y,z) = \arctan(x+y+z)$ $(1,0,0)$

(b) $f(x,y,z) = x^2+3y^2+6z^2-2xy+6xz+7yz+4x-3y+7$ $(0,0,0)$

(c) $f(x,y,z) = \sqrt{xy+z^2}$ $(1,1,1)$

(d) $f(x,y,z) = xe^{xy+z}$ $(1,0,0)$

(e) $f(x,y,z) = e^{x+y^2+z^3}$ $(0,0,0)$

(f) $f(x,y,z) = x\sin(y)+y\ln(z)$ $(1,\pi,1)$

(g) $f(x,y,z) = (xy)^z + z^{xy}$ $(1,1,1)$

Assignment 16.6 — Let $z = \ln\left(\sqrt{x^2+y^2}\right)$. Then, prove

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = 1.$$

Assignment 16.7 — Let

$$z = e^{\frac{x}{y}}\sin\left(\frac{x}{y}\right) + e^{\frac{y}{x}}\cos\left(\frac{y}{x}\right).$$

Then, prove

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = 0.$$

Assignment 16.8 — Prove that $z = f(x)g(y)$ satisfies

$$z\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial z}{\partial x}\frac{\partial z}{\partial y}.$$

Assignment 16.9 — Let $z = A\sin(a\lambda y + \varphi)\sin(\lambda x)$. Then, prove that

$$\frac{\partial^2 z}{\partial y^2} = a^2 \frac{\partial^2 z}{\partial x^2}.$$

Assignment 16.10 — Prove that the function $u(x,y,t) = t^{-1}e^{-(x^2+y^2)/(4t)}$ is a solution of the two-dimensional heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Assignment 16.11 — Let$u = x^2y + y^2z + z^2x$. Then, prove

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} = (x+y+z)^2.$$

Assignment 16.12 — Consider the Van der Waals equation

$$f(p,V,n,T) = \left(p + \frac{n^2 a}{V^2}\right)(V - nb) - nRT = 0,$$

with a, b and R constant. Find

(a) $\left(\frac{\partial V}{\partial T}\right)_{p,n}$

(b) $\left(\frac{\partial V}{\partial p}\right)_{T,n}$

(c) $\left(\frac{\partial p}{\partial T}\right)_{V,n}$

(d) $\left(\frac{\partial p}{\partial V}\right)_{T,n}$

The notation $\left(\frac{\partial f}{\partial x}\right)_{y,z}$ indicates that y and z are considered constants when differentiating with respect to x.

Assignment 16.13 — A function $f(x,y)$ is called **homogeneous in degree** n if $f(tx,ty) = t^n f(x,y)$, with $t > 0$.

(a) Show that the following functions are homogeneous in degree n.

a) $f(x,y) = x^2y - 2y^3$, with $n = 3$

b) $f(x,y) = \sqrt{x^2+y^2}$, with $n = 1$

c) $f(x,y) = \dfrac{5}{(x^2+2y^2)^2}$, with $n = -4$

(b) Show that for homogeneous functions of degree n it holds that

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial x} = nf(x,y).$$

Check this for the functions from (a).

Total differential and differentiability

Assignment 16.14 — Find the partial derivatives of the first and second order of the functions below, and at the specified point. Also, find the corresponding total differentials of first and second order.

(a) $f(x,y) = x^2 + 2xy + y^2 - 2x + 3y - 7$ $\quad (1,2)$

(b) $f(x,y) = x^2y^5 + xy^2 + x^3y$ $\quad (3,1)$

(c) $f(x,y) = \dfrac{xy}{x^2+y^2}$ $\quad (1,1)$

(d) $f(x,y) = x^y$ $\quad (1,1)$

(e) $f(x,y) = \ln(2x - 3y)$ $\quad (2,1)$

(f) $f(x,y) = \dfrac{e^y}{x}$ $\quad (1,1)$

(g) $f(x,y) = e^{\frac{y}{x}}$ $\quad (1,1)$

(h) $f(x,y) = \cos(3x + 2y)$ $\quad (0,\pi)$

(i) $f(x,y) = \arctan(x + y)$ $\quad (1,0)$

(j) $f(x,y) = x\sinh(y) + y\cosh(x)$ $\quad (0,0)$

(k) $f(x,y) = (2x + y)^{x+3y}$ $\quad (0,1)$

Assignment 16.15 — The sides of a rectangular box are measured to the nearest 1% of their length. What is the estimated maximum error rate of

(a) the volume of the box,

(b) the area of the sides of the box,

(c) the length of the diagonal of the box?

The multivariable chain rule and implicit function theorem

Assignment 16.16 — Let $z = xy$ with $x = \dfrac{1}{t}$ and $y = t^2$. Find $\frac{dz}{dt}$.

Assignment 16.17 — Find u_t if $u = \sqrt{x^2 + y^2}$ with $x = e^{st}$ and $y = 1 + s^2\cos(t)$.

Assignment 16.18 — Let $z = x^2 + 2xy + 4y^2$ met $y = e^{ax}$. Find $\frac{dz}{dx}$.

Assignment 16.19 — Let $z = f(u,v)$ with $u = u(x,y)$ and $v = v(x,y)$. Find z_x and z_y.

(a) $z = \ln\left(u^2 + v^2\right)$ with $\begin{cases} u = x + 2y + 1 \\ v = 3x - y - 1 \end{cases}$

(b) $z = \cosh\left(u^2 - v^2\right)$ with $\begin{cases} u = 2x - 3y \\ v = 3x - 4y \end{cases}$

(c) $z = \dfrac{v}{u}$ with $\begin{cases} u = \sin\left(x^2 - y^2\right) \\ v = e^{xy} \end{cases}$

(d) $z = ve^{uv}$ with $\begin{cases} u = x^2 y \\ v = xy^2 \end{cases}$

(e) $z = u^v$ with $\begin{cases} u = x^2 + y^2 \\ v = xy \end{cases}$

Assignment 16.20 — Let $z = f(x, y)$ with $x = 2s + 3t$ and $y = 3s - 2t$. Find

$$\frac{\partial^2 z}{\partial s^2}, \quad \frac{\partial^2 z}{\partial s \partial t} \quad \text{and} \quad \frac{\partial^2 z}{\partial t^2}.$$

Assignment 16.21 — Let $x = e^s \cos(t)$, $y = e^s \sin(t)$ en $z = u(x, y) = v(s, t)$. Then prove that

$$\frac{\partial^2 z}{\partial s^2} + \frac{\partial^2 z}{\partial t^2} = (x^2 + y^2)\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right).$$

Assignment 16.22 — Let $u(x, y) = r^2 \ln(r)$, with $r^2 = x^2 + y^2$. Verify that u is a biharmonic function by proving that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = 0.$$

Assignment 16.23 — Consider the function $w = f(x, y, z)$ with $x = g(s)$, $y = h(s, t)$ en $z = k(t)$. Find an expression for $\frac{\partial w}{\partial t}$.

Assignment 16.24 — Consider the function $z = g(x, y)$ with $y = f(x)$ and $x = h(u, v)$. Find an expression for $\frac{\partial z}{\partial u}$.

Assignment 16.25 — Consider the function $w = f(x, y)$ with $x = g(r, s)$, $y = h(r, t)$, $r = k(s, t)$ and $s = m(t)$. Find an expression for $\frac{dw}{dt}$.

Assignment 16.26 — In each case, find the derivative for the given equation. Which condition(s) for the variables will ensure the existence of the indicated function?

(a) $xy^3 + x^4 y = 2$ defines x as a function of y. Determine $\frac{dx}{dy}$.

(b) $z^2 + xy^3 = \dfrac{xz}{y}$ defines z as a function of x and y. Determine $\frac{\partial z}{\partial y}$.

(c) $e^{yz} - x^2 z \ln(y) = \pi$ defines y as a function of x and z. Determine $\frac{\partial y}{\partial z}$.

Assignment 16.27 — The implicit equation $x^2 + y^2 = r^2$ defines two explicit function: $y = f_1(x)$ and $y = f_2(x)$.

(a) Prove that the equation of the tangent to the graph at $P = (x_1, y_1)$ is $xx_1 + yy_1 = r^2$.

(b) Find the slope of that tangent and show that the tangent is perpendicular to the line segment $[OP]$.

Assignment 16.28 — Does the equation $(x^2 + y^2 + 2z^2)^{1/2} = \cos(z)$ have a unique solution y as a function of x and z close to $(0, 1, 0)$? Does there exist a unique solution for z as a function of x and y?

Assignment 16.29 — Assume that $F(x, y) = 0$ is of class C^1 and that $F(0, 0) = 0$. Which conditions for F ensure that $F(F(x, y), y) = 0$ can be solved to y as a C^1 function of x near $(0, 0)$?

Directional derivatives

Assignment 16.30 — Determine the degree of change for the function $f(x, y) = x^2 y$ at $(-1, -1)$ in the direction of $\vec{\boldsymbol{v}} = (1, 2)$.

Assignment 16.31 — Let $f(x, y) = x + y^2 - 3xy + 5y - 1$. In which direction at$(1, 1)$ does the function value change the most?

Assignment 16.32 — Determine the directional derivative $D_{\hat{\boldsymbol{u}}} f(P)$ if $f(x, y, z) = xy^2 z^3$, $P = (1, 1, 1)$ and $\hat{\boldsymbol{u}}$ are perpendicular to the plane $x^4 + 2y^4 + 2z^4 = 5$ in P.

Assignment 16.33 — We consider

$$G(x, y, z) = z \ln\left(\frac{xz + 1}{y + 2}\right)$$

and $\vec{\boldsymbol{a}} = (1, 2, 1)$ en $\vec{\boldsymbol{b}} = (5, 4, -3)$. Find the directional derivative $D_{\hat{\boldsymbol{u}}} G(\vec{\boldsymbol{a}})$ where $\hat{\boldsymbol{u}}$ has the same direction as the vector pointing from $\vec{\boldsymbol{a}}$ to $\vec{\boldsymbol{b}}$.

Assignment 16.34 — A racehorse lives in a valley whose shape is described by $f(x, y) = 3x^2 + y^2$. The horse is on a racetrack that is the collection of points for which $x^2 + y^2 = 1$. The racehorse's wants to escape from the valley and take the steepest possible path uphill from the racecourse. At what point should the horse escape and in which direction should it move uphill from that point?

Tangent lines, normal lines, and tangent planes

Assignment 16.35 — Find the equations of the tangent plane and the normal to the graph of the given function at the given point.

(a) $f(x, y) = x^2 - 2xy + y^2 - x + 2y$ in $(1, -1)$

(b) $2x^2 + 4yz - 5z^2 = -10$ in $(3, -1, 2)$

(c) $f(x, y) = \dfrac{2xy}{x^2 + y^2}$ in $(0, 2)$

(d) $f(x,y) = 2\sin(x)\cos(y)$ in $\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$

Assignment 16.36 — Determine the distance from the point $(1,1,0)$ to the circular paraboloid $z = x^2 + y^2$.

Assignment 16.37 — Find an equation of the tangent plane to the level surface of the function $f(x,y,z) = \cos(x+2y+3z)$ at $(\pi/2, \pi, \pi)$.

Taylor series expansions

Assignment 16.38 — For the functions below, determine a Taylor series of second order at the specified point.

(a) $f(x,y) = xe^{xy+y}$ $(1,1)$

(b) $f(x,y) = x\ln(y)$ $(0,1)$

(c) $f(x,y) = xy + \ln(xy)$ $(1,1)$

(d) $f(x,y) = x\sin(y)$ $(1,0)$

(e) $f(x,y) = xy\cos(x+y)$ $\left(0, \frac{\pi}{2}\right)$

(f) $f(x,y) = \arctan\left(\frac{x}{y}\right)$ $(0,1)$

(g) $f(x,y) = \sin(xe^y)$ $(0,0)$

(h) $f(x,y) = \frac{\sin(x)}{y}$ $\left(\frac{\pi}{2}, 1\right)$

Assignment 16.39 — Give an approximation for

(a) $\arctan\left(\frac{1.02}{0.95}\right)$,

(b) $\sqrt{3.99 \times 4.02}$.

Extreme values

Assignment 16.40 — Determine and classify the critical points of the given functions.

(a) $f(x,y) = x^2+y^2-3xy$

(b) $f(x,y) = xy$

(c) $f(x,y) = y\sqrt{x}-xy+y^2$

(d) $f(x,y) = (2x^2-y)(2-y)$

(e) $f(x,y) = x^4+y^4-2x^2+4xy-2y^2$

(f) $f(x,y) = x^3-3xy^2$

(g) $f(x,y) = (x-y)^4+(y-1)^4$

(h) $f(x,y) = x+y\sin(x)$

(i) $f(x,y) = (x+y)e^{-xy}$

(j) $f(x,y) = x^2+y^2$

(k) $f(x,y) = \cos(x+y)$

(l) $f(x,y) = x^2+2y^2-4x+4y$

(m) $f(x,y) = xy-x+y$

(n) $f(x,y) = x^3+y^3-3xy$

(o) $f(x,y) = x^4+y^4-4xy$

(p) $f(x,y) = \dfrac{x}{y}+\dfrac{8}{x}-y$

(q) $f(x,y) = x\sin(y)$

(r) $f(x,y) = \cos(x)+\cos(y)$

(s) $f(x,y) = x^2ye^{-(x^2+y^2)}$

(t) $f(x,y) = \dfrac{xy}{2+x^4+y^4}$

(u) $f(x,y) = xe^{-x^3+y^3}$

(v) $f(x,y) = \dfrac{x^2}{x^2+y^2}$

(w) $f(x,y) = \dfrac{xy}{x^2+y^2}$

(x) $f(x,y,z) = xy+x^2z-x^2-y-z^2$

Assignment 16.41 — Determine the maximum and minimum of the function $f(x,y,z) = x+y^2z$ under the conditions $y^2+z^2=2$ and $z=x$.

Assignment 16.42 — The strength of a wooden beam with rectangular cross-section and given length is directly proportional to the product of its width x and the square of its height y (take as a proportionality factor 1). Find the dimensions (x and y) of the strongest beam that can be sawn from a tree trunk, of circular cross-section and diameter $\sqrt{3}$.

Assignment 16.43 — Find the dimensions of an open (no top surface) beam-shaped box that has a volume of 32 dm^3, but a minimum sheath area.

Assignment 16.44 — Find the positive numbers a, b and c for which the sum is 30 and ab^2c^3 is maximal.

Assignment 16.45 — The material to make the bottom of a beam-shaped box is twice as expensive per unit area as the material used for the lid and side walls. Find the dimensions of a box of volume 12 m^2 for which the material cost is minimal.

Assignment 16.46 — A metal surface whose shape can be described by $4x^2+y^2+4z^2=16$ is heated. After one hour the temperature at (x,y,z) equals:

$$T(x,y,z) = 8x^2+4yz-16z+600.$$

Find the warmest point on the surface if we restrict ourselves to $x>0$.

Assignment 16.47 — An open beam-shaped cardboard box is reinforced at the ribs of the bottom and the sides with tape. One has 96 cm of tape available. What are the dimensions of the box with the largest possible volume?

Extreme values

Assignment 16.48 — Find the extreme value(s) of the functions below constrained by the specified region.

(a) $f(x,y) = x - x^2 + y^2$ constrained by the rectangle $0 \le x \le 2, 0 \le y \le 1$

(b) $f(x,y) = xy - 2x$ constrained by the rectangle $-1 \le x \le 1, 0 \le y \le 1$

(c) $f(x,y) = xy - x^3y^2$ constrained by the square $0 \le x \le 1, 0 \le y \le 1$

(d) $f(x,y) = xy(1-x-y)$ constrained by the triangle with vertices $(0,0)$, $(0,1)$ and $(1,0)$

(e) $f(x,y) = \sin x \cos y$ constrained by the triangle with vertices $(0,0)$, $(0,2\pi)$ and $(2\pi,0)$

(f) $f(x,y) = x^2ye^{-(x+y)}$ constrained by the triangle with vertices $(0,0)$, $(0,4)$ and $(4,0)$

Assignment 16.49 — Find the maximum value of the function f constrained by the specified area.

(a) $f(x,y) = 2x + 7y$ constrained by the region where $x + 2y \le 6$, $2x + y \le 6$, $x \ge 0$ and $y \ge 0$

(b) $f(x,y) = 2x + 3y$ constrained by the region where $x + 2y \le 12$, $4x + y \le 12$, $y \le 5$, $x \ge 0$ and $y \ge 0$

Assignment 16.50 — Find the minimal value of $f(x,y,z) = 2x + 3y + 4z$ constrained by the region where $x + y \ge 2$, $y + z \ge 2$, $x + z = 2$, $x \ge 0$, $y \ge 0$ and $z \ge 0$.

Assignment 16.51 — A textile manufacturer produces two types of fabric made from wool, cotton and polyester. The luxury fabric consists of 20 percent wool, 50 percent cotton, and 30 percent polyester and is sold at $3 per kilogram. The standard fabric consists of 10 percent wool, 40 percent cotton and 50 percent polyester and is sold at $2 per kilogram. There is 2000 kg of wool in stock, 6000 kg of cotton and also 6000 kg of polyester. How many kilograms of fabric should the manufacturer produce to maximize its profit?

Assignment 16.52 — A circuit board manufacturer has a stock of 200 resistors, 120 transistors, and 150 capacitors. The manufacturer is asked to produce two types of circuits A and B. Type A requires 20 resistors, 10 transistors, and 10 capacitors, while type B requires 10 resistors, 20 transistors, and 30 capacitors. The profit on the former is 5 €and on the latter 12 €. How many circuits of each type should be produced to have maximum profit? What is that profit then?

Assignment 16.53 — A radio station conducted a survey on the ratings for 3 types of radio programs: pop music, oldies and information programs. Broadcasting pop music costs €6 per hour and is rated 10, whereas broadcasting oldies costs €3 per hour and is rated 15. Information programs cost €2 per hour and are rated 5. The budget per day is at most €60, for which the station should broadcast exactly 20 hours per day with at least 4 hours of pop music. Find the ratio of different programs that maximizes the rating.

Assignment 16.54 — A plot of 10-hectare is divided into zones in which 6 detached houses per hectare, 8 semi-detached buildings per hectare or 12 apartments per hectare can be built. The entire plot should be built on. The plot's owner can make a profit of €40,000 per detached house, €20,000 per semi-detached house and €16,000 per apartment. In total, he would like to have at least as many apartments as houses (detached and half open together). How many buildings of each type should be built to maximize profit?

Assignment 16.55 — Karl Lagerfeld has 230 m of fabric that he plans to use to its full potential. He wants to use it to make a maximum of 20 suits, 30 jackets and 40 pants. For a suit he needs 6 m of fabric, for a coat 3 m and for pants 2 m. For each suit, coat and trouser he has a respective profit of €20,€14 and €12. How many pieces of each product must Karl make to maximize his profit?

Assignment 16.56 — A businesswoman wants to buy coffee, tea and sugar together for no more than €340. Coffee costs €3 per kg, tea €3.5 per kg and sugar €4 per kg. The businesswoman can buy no more than 75 kg of coffee, 75 kg of tea and 90 kg of sugar per day. She wants to buy exactly 90 kg all together per day, which is the maximum carrying capacity of her van. Suppose the businesswoman makes a profit of €1 per kg of coffee, €2 per kg of tea and €1.5 per kg of sugar. For which combination of coffee, tea and sugar does she makes most profit?

Assignment 16.57 — A tennis ball manufacturer produces three types of tennis balls: Silver, Yellow and Gold. To practise at Roland Garros, the organization wants to order exactly 1800 tennis balls, of which at least 200 must be Silver balls, 300 Yellow balls, and at most 1200 Gold balls. The number of Silver balls may at most be double of the number of Yellow balls. The manufacturer sells the Silver balls at €2 each, the Yellow balls cost €2.20 each and the Gold balls cost only €1 each. Based on these conditions and prices, the Roland Garros organization would like to place an order at the lowest possible cost. How many balls of each type should the organization order? What will the organization spend in total?

Revieuw exercises

Assignment 16.58 — Let $f(x,y) = \ln(x^2+y^2)$ and $P = (1,-2)$. Find

(a) the gradient of this function at P,

(b) an equation of the tangent plane to the graph of f at P,

(c) an equation of the tangent at P to the level curve of f through P.

Assignment 16.59 — The temperature $T(x,y)$ across the xy-plane is given by $T(x,y) = x^2 - 2y^2$.

(a) Draw a contour plot for T.

(b) In which direction should an ant move in $(2,-1)$ to cool down as quickly as possible?

(c) Along which curve through $(2,-1)$ should the ant move to experience maximum cooling?

God does not care about our mathematical difficulties. He integrates empirically.

— Albert Einstein —

17

Multiple integrals

The previous chapter introduced multivariable functions and we applied concepts of differential calculus to these functions. We learned how we can view a function of two variables as a surface in space, and learned how partial derivatives convey information about how the surface is changing in any direction.

In this chapter we apply techniques of integral calculus to multivariable functions. In Chapter 12 we learned how the definite integral of a single variable function gave us area under the curve. In this chapter we will see that integration applied to a multivariable function gives us volume under a surface. And just as we learned applications of integration beyond finding areas, we will find applications of integration in this chapter beyond finding volume.

17.1 Iterated integrals and area

17.1.1 Iterated integrals

In Chapter 16 we found that it was useful to differentiate functions of several variables with respect to one variable, while treating all the other variables as constants or coefficients. We can integrate functions of several variables in a similar way. For instance, if we are told that $f_x(x,y) = 2xy$, we can treat y as staying constant and integrate to obtain $f(x,y)$:

$$\begin{aligned} f(x,y) &= \int f_x(x,y)\ dx \\ &= \int 2xy\ dx \\ &= x^2y + C. \end{aligned}$$

Make a careful note about the constant of integration, C. This "constant" is something with a derivative of 0 with respect to x, so it could be any expression that contains only constants and functions of y. For instance, if $f(x,y) = x^2y + \sin(y) + y^3 + 17$, then $f_x(x,y) = 2xy$. To signify that C is actually a function of y, we write:

$$f(x,y) = \int f_x(x,y)\ dx = x^2y + C(y).$$

Using this process we can even evaluate definite integrals. For instance, to evaluate the integral

$$\int_1^{2y} 2xy\ dx.$$

We find the indefinite integral as before, then apply the fundamental theorem of calculus to evaluate the definite integral:

$$\begin{aligned}\int_1^{2y} 2xy\ dx &= x^2y\Big|_1^{2y}\\ &= (2y)^2y - (1)^2y\\ &= 4y^3 - y.\end{aligned}$$

We can also integrate with respect to y. In general,

$$\int_{h_1(y)}^{h_2(y)} f_x(x,y)\ dx = f(x,y)\Big|_{h_1(y)}^{h_2(y)} = f\big(h_2(y),y\big) - f\big(h_1(y),y\big),$$

and

$$\int_{g_1(x)}^{g_2(x)} f_y(x,y)\ dy = f(x,y)\Big|_{g_1(x)}^{g_2(x)} = f\big(x,g_2(x)\big) - f\big(x,g_1(x)\big).$$

Note that when integrating with respect to x, the bounds are functions of y (of the form $x = h_1(y)$ and $x = h_2(y)$) and the final result is also a function of y. When integrating with respect to y, the bounds are functions of x (of the form $y = g_1(x)$ and $y = g_2(x)$) and the final result is a function of x.

When evaluating $\int_1^{2y} 2xy\ dx$, we integrated a function with respect to x and ended up with a function of y. We can integrate this as well. This process is known as iterated integration, or **multiple integration** (*meervoudige integratie*). Of course, when considering

$$\int_{x_1}^{x_2}\int_{g_1(x)}^{g_2(x)} f(x,y)dy\ dx$$

x should have constant bounds, whereas y may have variable ones, and vice versa when considering

$$\int_{y_1}^{y_2}\int_{h_1(y)}^{h_2(y)} f(x,y)dx\ dy$$

Example 17.1

Evaluate

$$\int_1^2 \left(\int_1^x \left(5x^3y^{-3} + 6y^2\right) dy \right) dx.$$

Solution

We follow a standard order of operations and perform the operations inside parentheses first.

$$\begin{aligned}
\int_1^2 \left(\int_1^x \left(5x^3y^{-3} + 6y^2\right) dy \right) dx &= \int_1^2 \left(\frac{5x^3y^{-2}}{-2} + \frac{6y^3}{3} \right) \Bigg|_1^x dx \\
&= \int_1^2 \left(\frac{5x^3}{-2x^2} + \frac{6x^3}{3} - \frac{5x^3}{-2} - \frac{6}{3} \right) dx \\
&= \int_1^2 \left(-\frac{5x}{2} + 2x^3 + \frac{5x^3}{2} - 2 \right) dx \\
&= \int_1^2 \left(\frac{9x^3}{2} - \frac{5x}{2} - 2 \right) dx \\
&= \left(\frac{9}{8}x^4 - \frac{5}{4}x^2 - 2x \right) \Bigg|_1^2 \\
&= \frac{89}{8}
\end{aligned}$$

Note how the bounds of x were $x = 1$ to $x = 2$ and the final result was a number.

The previous example showed how we could perform something called an iterated integral; we do not yet know why we would be interested in doing so nor what the result, such as the number $89/8$, means. We will now investigate that.

17.1.2 Area of a plane region

Consider the plane region R bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, shown in Figure 17.1(a). We learned in Section 13.1 that the area of R is given by

$$\int_a^b \left(g_2(x) - g_1(x)\right) dx.$$

We can view the expression $\left(g_2(x) - g_1(x)\right)$ as

$$\left(g_2(x) - g_1(x)\right) = \int_{g_1(x)}^{g_2(x)} 1\ dy = \int_{g_1(x)}^{g_2(x)} dy,$$

meaning we can express the area of R as an iterated integral:

$$\text{area} = \int_a^b \big(g_2(x) - g_1(x)\big)\,dx = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} dy\right) dx = \int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx.$$

In short: a certain iterated integral can be viewed as giving the area of a plane region.

A region R could also be defined by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, as shown in Figure 17.1(b). Using a process similar to that above, we have

$$\int_c^d \int_{h_1(y)}^{h_2(y)} dx\,dy.$$

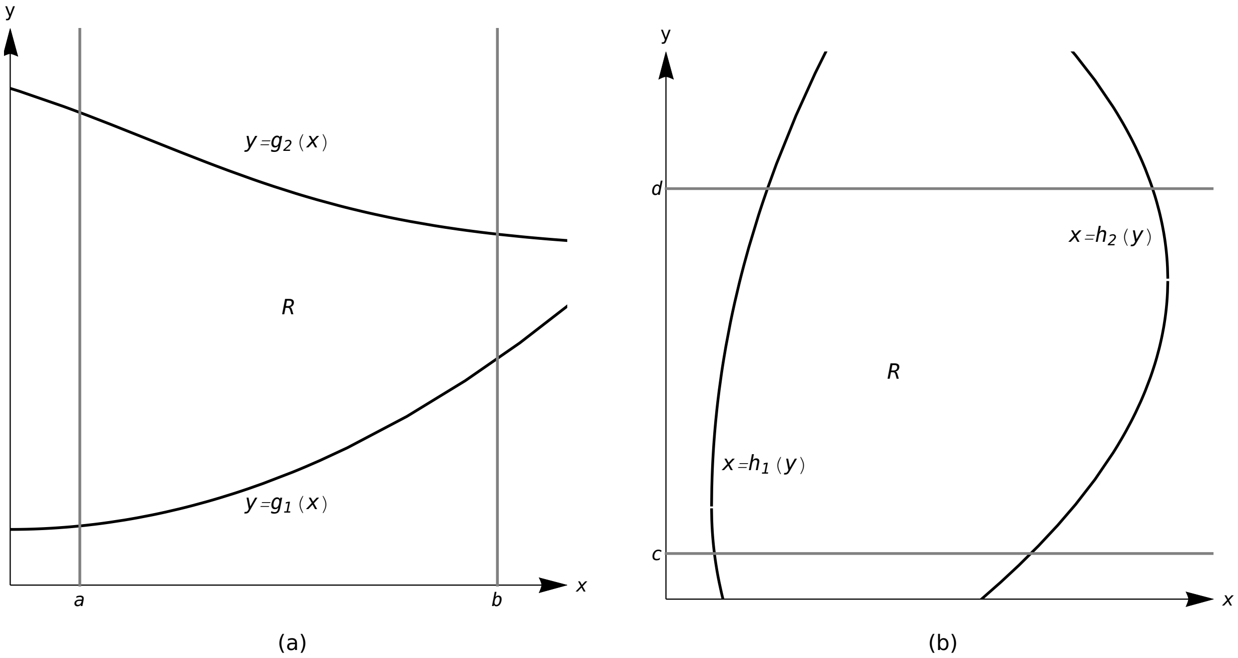


Figure 17.1: Calculating the area of a plane region R with iterated integrals.

We state this formally in a theorem.

Theorem 17.1 (Area of a plane region)

1. *Let R be a plane region bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous functions on $[a,b]$. The **area** A **of** R is*

$$A = \int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx.$$

2. *Let R be a plane region bounded by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are*

continuous functions on $[c,d]$. *The* ***area*** A ***of*** R *is*

$$A = \int_c^d \int_{h_1(y)}^{h_2(y)} dx\, dy.$$

The following examples should help us understand this theorem.

Example 17.2

Find the area A of the triangle with vertices at $(1,1)$, $(3,1)$ and $(5,5)$, as shown in Figure 17.2(a).

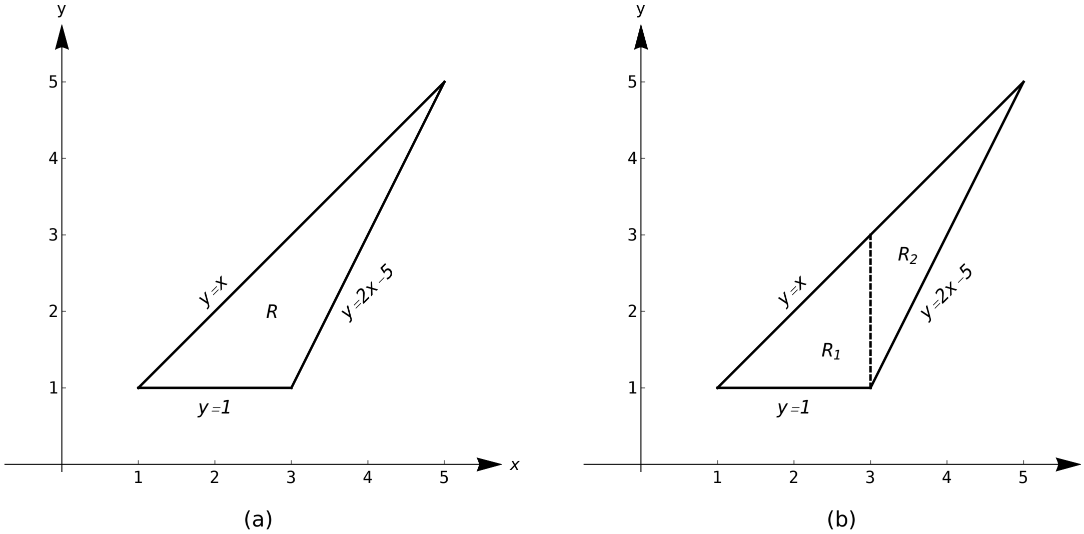


Figure 17.2: Calculating the area of a triangle with iterated integrals in Example 17.2 by using constant bounds on y (a) and x (b)

Solution

The triangle is bounded by the lines as shown in Figure 17.2(a). Choosing to integrate with respect to x first gives that x is bounded by $x = y$ to $x = \frac{y+5}{2}$, while y is bounded by $y = 1$ to $y = 5$, i.e. the bounds with respect to y are fixed. Recall that since x-values increase from left to right, the leftmost curve, $x = y$, is the lower bound and the rightmost curve, $x = (y+5)/2$, is the upper bound. The area is

$$\begin{aligned} A &= \int_1^5 \int_y^{\frac{y+5}{2}} dx\, dy \\ &= \int_1^5 (x)\Big|_y^{\frac{y+5}{2}} dy \\ &= \int_1^5 \left(-\frac{1}{2}y + \frac{5}{2}\right) dy \\ &= \left(-\frac{1}{4}y^2 + \frac{5}{2}y\right)\Bigg|_1^5 \\ &= 4. \end{aligned}$$

We can also find the area by integrating with respect to y first, which implies that the bounds on y are variable and dependent on x, whereas the one on x are fixed. In this situation, though, there is not one function that defines the lower bound on the interval $[1,5]$. Essentially, we have two functions that act as the lower bound for the region R, $y=1$ and $y=2x-5$. In this way, the region R is split into two smaller regions, namely R_1 and R_2 (Figure 17.2(b)). This requires us to use two iterated integrals. Note how the x-bounds are different for each integral:

$$\begin{aligned} A &= \int_1^3 \int_1^x 1 \, dy \, dx + \int_3^5 \int_{2x-5}^x 1 \, dy \, dx \\ &= \int_1^3 y \Big|_1^x \, dx + \int_3^5 y \Big|_{2x-5}^x \, dx \\ &= \int_1^3 (x-1) \, dx + \int_3^5 (-x+5) \, dx \\ &= 2+2 \\ &= 4. \end{aligned}$$

As expected, we get the same answer both ways.

Example 17.3

Find the area of the region enclosed by $y=2x$ and $y=x^2$, as shown in Figure 17.3.

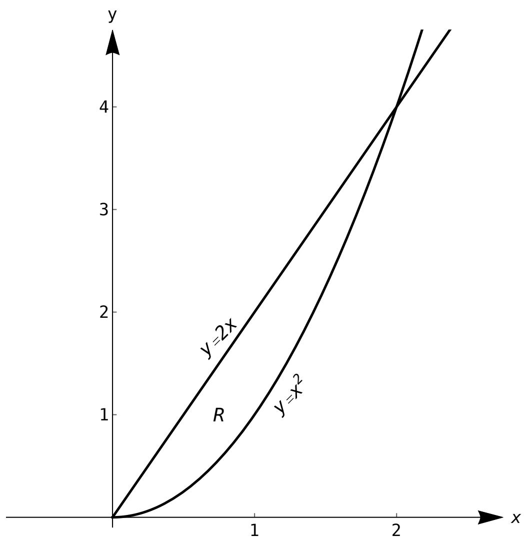


Figure 17.3: Calculating the area of a plane region with iterated integrals in Example 17.3.

Solution

Once again we will find the area of the region using both orders of integration. We can approach this problem in two ways, either by choosing fixed bounds for x and variable ones - depending on x - for y or by choosing fixed bounds for y and variable ones for x, which then depend on y. So, in the former case, we first integrate with respect to y and then with respect to x:

$$\int_0^2 \int_{x^2}^{2x} 1\, dy\, dx = \int_0^2 (2x - x^2)\, dx = \left(x^2 - \frac{1}{3}x^3\right)\Big|_0^2 = \frac{4}{3}.$$

In the latter case, we, however, do exactly the opposite:

$$\int_0^4 \int_{y/2}^{\sqrt{y}} 1\, dx\, dy = \int_0^4 \left(\sqrt{y} - \frac{y}{2}\right) dy = \left(\frac{2}{3}y^{3/2} - \frac{1}{4}y^2\right)\Big|_0^4 = \frac{4}{3}.$$

In each of the previous examples, we have been given a region R and found the bounds needed to find the area of R using both orders of integration. We integrated using both orders of integration to demonstrate their equality.

We now approach the skill of describing a region using both orders of integration from a different perspective. Instead of starting with a region and creating iterated integrals, we will start with an iterated integral and rewrite it in the other integration order. To do so, we will need to understand the region over which we are integrating.

The simplest of all cases is when both integrals are bound by constants. The region described by these bounds is a rectangle, and so:

$$\int_a^b \int_c^d 1\, dy\, dx = \int_c^d \int_a^b 1\, dx\, dy.$$

When the inner integral's bounds are not constants, it is generally very useful to sketch the bounds to determine what the region we are integrating over looks like. From the sketch we can then rewrite the integral with the other order of integration.

Example 17.4

Change the order of integration of

$$\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\, dx\, dy.$$

Solution

We sketch the region described by the bounds to help us change the integration order. x is bounded below and above (i.e., to the left and right) by $x = y^2/4$ and $x = (y+4)/2$ respectively, and y is bounded between 0 and 4. Graphing the previous curves, we find the region R to be that shown in Figure 17.4(a).

To change the order of integration, we need to give x fixed bounds. The figure makes it clear that there are two lower bounds for y: $y = 0$ on $0 \le x \le 2$, and $y = 2x - 4$ on $2 \le x \le 4$, thereby splitting R in two smaller regions R_1 and R_2 (Figure 17.4(b)). Thus we need two double integrals. The upper bound for each is $y = 2\sqrt{x}$. Thus we have

$$\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\, dx\, dy = \int_0^2 \int_0^{2\sqrt{x}} 1\, dy\, dx + \int_2^4 \int_{2x-4}^{2\sqrt{x}} 1\, dy\, dx.$$

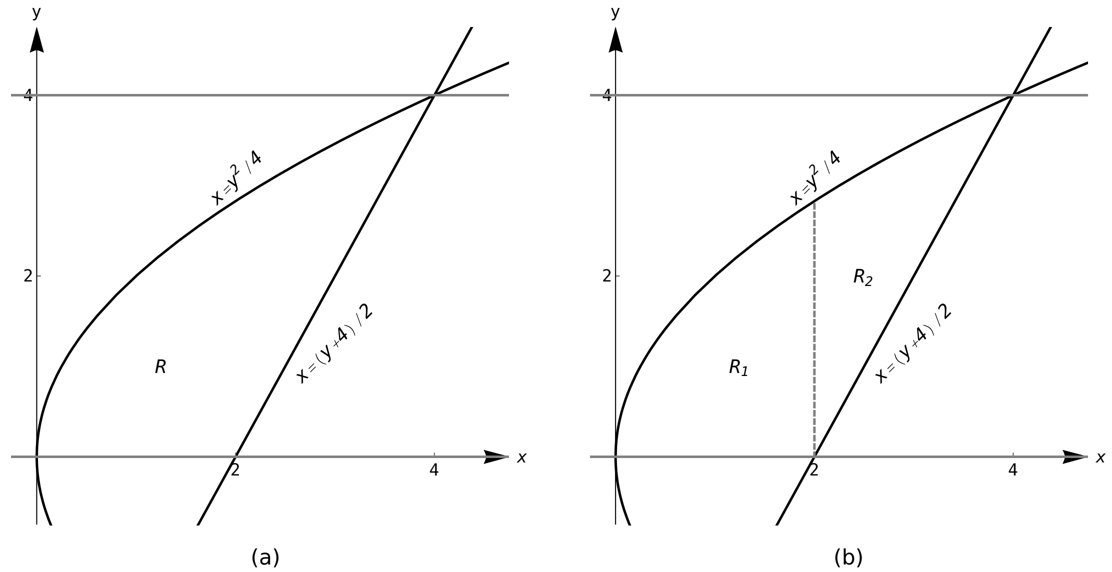


Figure 17.4: Drawing the region determined by the bounds of integration in Example 17.4.

This section has introduced a new concept, the iterated integral. We developed one application for iterated integration: area between curves. However, this is not new, for we already know how to find areas bounded by curves. In the next section we apply iterated integration to solve problems we currently do not know how to handle.

17.2 Double integration and volume

17.2.1 Definition

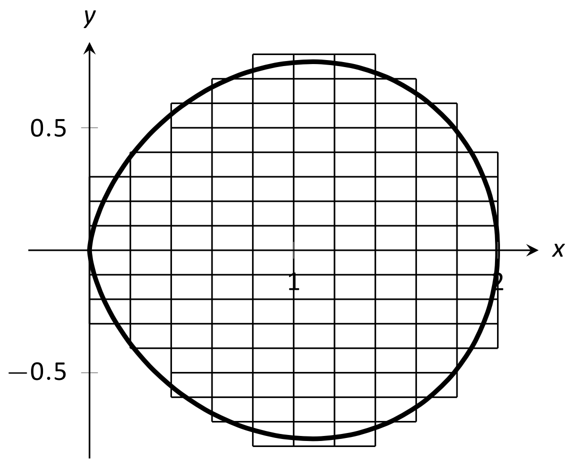
The definite integral of f over $[a,b]$, $\int_a^b f(x)\ dx$, was introduced as the signed area under the curve. We approximated the value of this area by first subdividing $[a,b]$ into n subintervals, where the i^{th} subinterval has length Δx_i, and letting c_i be any value in the i^{th} subinterval. We formed rectangles that approximated part of the region under the curve with width Δx_i, height $f(c_i)$, and hence with area $f(c_i)\Delta x_i$. Summing all the rectangle's areas gave an approximation of the definite integral, and Theorem 12.4 stated that

$$\int_a^b f(x)\ dx = \lim_{\mathcal{L}\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i,$$

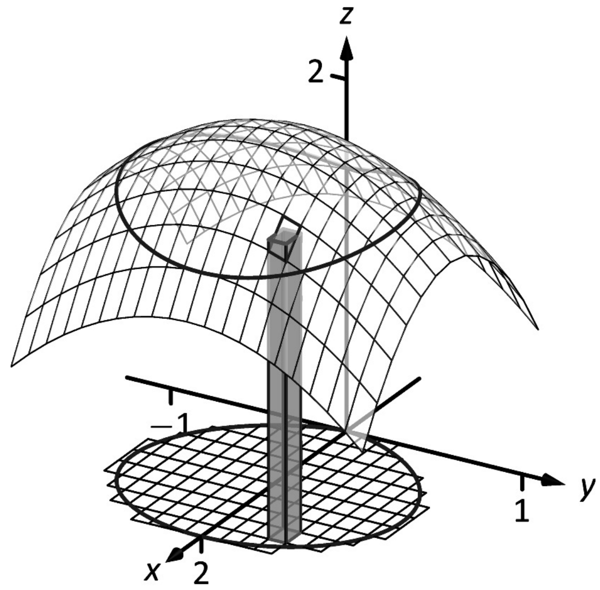
connecting the area under the curve with sums of the areas of rectangles. Recall that $\mathcal{L}$ is the size of the partition, which is the length of the largest subinterval of the partition, i.e. $\mathcal{L} = \max_i (\Delta x_i)$.

We use a similar approach in this section to find volume under a surface. Let R be a closed, bounded region in the xy-plane and let $z = f(x,y)$ be a continuous function defined on R. We wish to find the signed volume under the surface of f over R. We use the term signed volume to denote that space above the xy-plane, under f, will have a positive volume; space above f and under the xy-plane will have a negative volume, similar to the notion of signed area used before.

We start by partitioning R into n rectangular subregions as shown in Figure 17.5(a). For simplicity's sake, we let all widths be Δx and all heights be Δy. Note that the sum of the areas of the rectangles is not equal to the area of R, but rather is a close approximation. Arbitrarily number the rectangles 1 through n, and pick a point (x_i, y_i) in the i^{th} subregion.



(a)



(b)

Figure 17.5: Developing a method for finding signed volume under a surface.

The volume of the rectangular solid whose base is the i^{th} subregion and whose height is $f(x_i, y_i)$ is $V_i = f(x_i, y_i)\Delta x\Delta y$. Such a solid is shown in Figure 17.5(b). Note how this rectangular solid only approximates the true volume under the surface; part of the solid is above the surface and part is below.

For each subregion R_i used to approximate R, create the rectangular solid with base area $\Delta x\Delta y$ and height $f(x_i, y_i)$. The sum of all rectangular solids is

$$\sum_{i=1}^{n} f(x_i, y_i)\Delta x\Delta y.$$

This approximates the signed volume under f over R. As we have done before, to get a better approximation we can use more rectangles to approximate the region R.

In general, each rectangle could have a different width Δx_j and height Δy_k, giving the i^{th} rectangle an area $\Delta A_i = \Delta x_j\Delta y_k$ and the i^{th} rectangular solid a volume of $f(x_i, y_i)\Delta A_i$. Let now $\mathcal{A}$ denote the length of the longest diagonal of all rectangles in the subdivision of R; $\mathcal{A} \to 0$ means each rectangle's width and height are both approaching 0. If f is a continuous function, as $\mathcal{A}$ shrinks (and hence $n \to +\infty$) the summation $\sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$ approximates the signed volume better and better.

When adding up the volumes of rectangular solids over a partition of a region R, as done in Figure 17.5(b), one could first add up the volumes across each row (one type of sum), then add these totals together (another sum), as in

$$\sum_{j=1}^{n}\sum_{i=1}^{m} f(x_i, y_j)\Delta x_i\Delta y_j.$$

One can rewrite this as

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} f(x_i, y_j)\Delta x_i\right)\Delta y_j.$$

The summation inside the parenthesis indicates the sum of (heights × widths), which gives an area; multiplying these areas by the thickness Δy_j gives a volume. The illustration in Figure 17.5(b) relates to this understanding.

This all leads us to a definition.

Definitie 17.1 (Double integral and signed volume)
Let $z = f(x,y)$ be a continuous function defined over a closed, bounded region R in the xy-plane. The **signed volume** V **under** f **over** R is denoted by the **double integral** (*dubbelintegraal*)

$$V = \iint_R f(x,y)\;dA.$$

Alternate notations for the double integral are

$$\iint_R f(x,y)\;dA = \iint_R f(x,y)\;dx\;dy = \iint_R f(x,y)\;dy\;dx.$$

We can find the signed volume by considering increasingly smaller, so more, rectangles; that is by considering the limit

$$\lim_{\mathcal{A}\to 0}\sum_{i=1}^{n} f(x_i,y_i)\Delta A_i.$$

In this limiting situation, it holds that

$$V = \iint_R f(x,y)\;dA = \lim_{\mathcal{A}\to 0}\sum_{i=1}^{n} f(x_i,y_i)\Delta A_i. \tag{17.1}$$

Note that this equation does not specify the partition of the region R, so any partitioning where the diagonal of each rectangle shrinks to 0 results in the same answer. This does not offer a very satisfying way of computing volume, though. Our experience has shown that evaluating the limits of sums can be tedious. We seek a more direct method.

Recall Theorem 13.1. This stated that if $A(x)$ gives the cross-sectional area of a solid at x, then $\int_a^b A(x)\;dx$ gave the volume of that solid over $[a,b]$. Consider Figure 17.6, where a surface $z = f(x,y)$ is drawn over a region R. Fixing a particular x-value, we can consider the area under f over R where x has that fixed value. That area can be found with a definite integral, namely

$$A(x) = \int_{g_1(x)}^{g_2(x)} f(x,y)\;dy.$$

Remember that though the integrand contains x, we are viewing x as fixed. Also note that the bounds of integration are functions of x: the bounds depend on the value of x. As $A(x)$ is a cross-sectional area function, we can find the signed volume V under f by integrating it:

$$V = \int_a^b A(x)\;dx = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x,y)\;dy\right) dx = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\;dy\;dx.$$

This gives a concrete method for finding signed volume under a surface. We could do a similar procedure where we started with y fixed, resulting in an iterated integral with the order of integration $dx\;dy$. The following theorem states that both methods give the same result, which is the value of the double integral. It is such an important theorem it has a name associated with it.

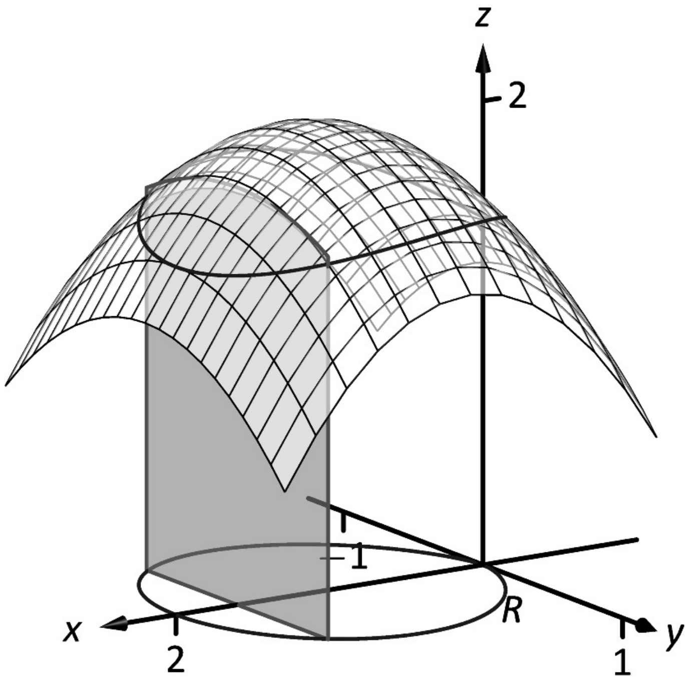


Figure 17.6: Finding volume under a surface by sweeping out a cross–sectional area.

Theorem 17.2 (Fubini's theorem)
Let R be a closed, bounded region in the xy-plane and let $z = f(x,y)$ be a continuous function on R.

1. *If R is bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous functions on $[a,b]$, then*
$$\iint_R f(x,y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\, dy\, dx.$$

2. *If R is bounded by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are continuous functions on $[c,d]$, then*
$$\iint_R f(x,y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\, dx\, dy.$$

Proof We state this theorem without proof because it is rather technical, but it is important to realize that it is valid only if
$$\iint_R \left|f(x,y)\right|\, dA < +\infty\,,$$
which is fulfilled if R is a closed, bounded region on which $z = f(x,y)$ is a continuous function. □

Example 17.5

Let $f(x,y) = xy + e^y$. Find the signed volume under f on the region R, which is the rectangle with corners $(3,1)$ and $(4,2)$ pictured in Figure 17.7, using both orders of integration.

Solution

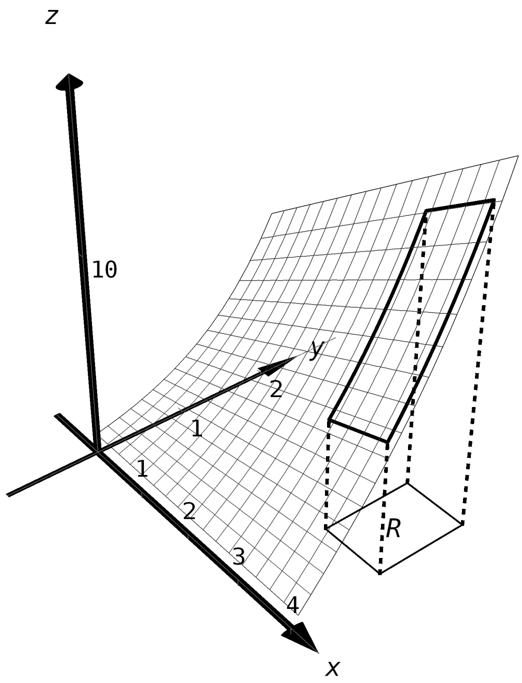


Figure 17.7: Finding the signed volume under a surface in Example 17.5.

We wish to evaluate $\iint_R \left(xy + e^y\right)\, dA$. As R is a rectangle, the bounds are easily described as $3 \le x \le 4$ and $1 \le y \le 2$.

Using the order $dy\ dx$:

$$\begin{aligned}
\iint_R \left(xy + e^y\right)\, dA &= \int_3^4 \int_1^2 \left(xy + e^y\right)\, dy\ dx \\
&= \int_3^4 \left(\frac{1}{2}xy^2 + e^y\right)\Bigg|_1^2\ dx \\
&= \int_3^4 \left(\frac{3}{2}x + e^2 - e\right)\, dx \\
&= \left(\frac{3}{4}x^2 + \left(e^2 - e\right)x\right)\Bigg|_3^4 \\
&= \frac{21}{4} + e^2 - e \approx 9.92.
\end{aligned}$$

Now we check the validity of Fubini's theorem by using the order $dx\ dy$:

$$\iint_R \left(xy + e^y\right)\, dA = \int_1^2 \int_3^4 \left(xy + e^y\right)\, dx\ dy$$

$$= \int_1^2 \left(\frac{1}{2}x^2y + xe^y\right)\Bigg|_3^4 \, dy$$

$$= \int_1^2 \left(\frac{7}{2}y + e^y\right) dy$$

$$= \left(\frac{7}{4}y^2 + e^y\right)\Bigg|_1^2$$

$$= \frac{21}{4} + e^2 - e \approx 9.92.$$

Both orders of integration return the same result, as expected.

Example 17.6

Evaluate

$$\iint_R \left(3xy - x^2 - y^2 + 6\right) dA,$$

where R is the triangle bounded by $x = 0$, $y = 0$ and $x/2 + y = 1$, as shown in Figure 17.8.

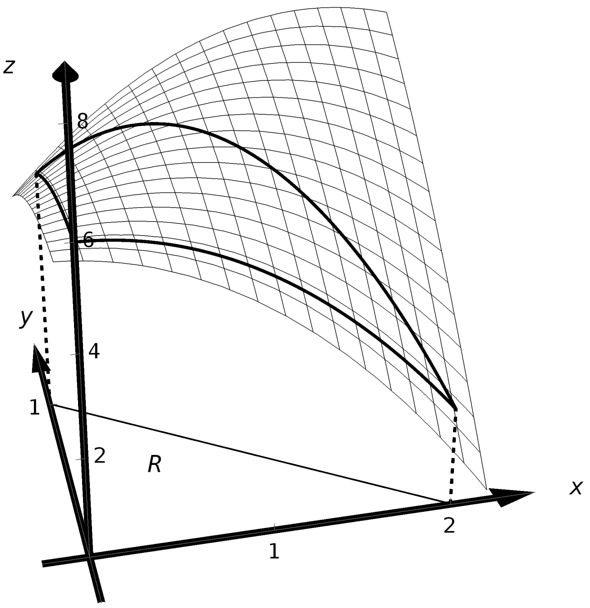


Figure 17.8: Finding the signed volume under a surface in Example 17.6.

Solution

While it is not specified which order we are to use, we will evaluate the double integral using both orders to help drive home the point that it does not matter which order we use.

Using the order $dy\,dx$: The bounds on y go from curve to curve, i.e., $0 \leq y \leq 1 - x/2$, and the bounds on x go from point to point, i.e., $0 \leq x \leq 2$.

$$\iint_R (3xy - x^2 - y^2 + 6)\,dA = \int_0^2 \int_0^{-\frac{x}{2}+1} (3xy - x^2 - y^2 + 6)\,dy\,dx$$
$$= \int_0^2 \left(\frac{3}{2}xy^2 - x^2y - \frac{1}{3}y^3 + 6y\right)\Bigg|_0^{-\frac{x}{2}+1} dx$$
$$= \int_0^2 \left(\frac{11}{12}x^3 - \frac{11}{4}x^2 - x + \frac{17}{3}\right) dx$$
$$= \left(\frac{11}{48}x^4 - \frac{11}{12}x^3 - \frac{1}{2}x^2 + \frac{17}{3}x\right)\Bigg|_0^2$$
$$= \frac{17}{3} \approx 5.6.$$

Now lets consider the order $dx\,dy$. Here x goes from curve to curve, $0 \leq x \leq 2 - 2y$, and y goes from point to point, $0 \leq y \leq 1$:

$$\iint_R (3xy - x^2 - y^2 + 6)\,dA = \int_0^1 \int_0^{2-2y} (3xy - x^2 - y^2 + 6)\,dx\,dy$$
$$= \int_0^1 \left(\frac{3}{2}x^2y - \frac{1}{3}x^3 - xy^2 + 6x\right)\Bigg|_0^{2-2y} dy$$
$$= \int_0^1 \left(\frac{32}{3}y^3 - 22y^2 + 2y + \frac{28}{3}\right) dy$$
$$= \left(\frac{8}{3}y^4 - \frac{22}{3}y^3 + y^2 + \frac{28}{3}y\right)\Bigg|_0^1$$
$$= \frac{17}{3} \approx 5.6.$$

We obtained the same result using both orders of integration.

17.2.2 Properties

Note how in these two examples that the bounds of integration depend only on R; the bounds of integration have nothing to do with $f(x,y)$. Moreover, let f and g be continuous functions over a closed, bounded plane region R, and let c be a constant, then we have the following properties, in line with the ones of single integrals.

- Constant multiple rule:
$$\iint_R cf(x,y)\,dA = c\iint_R f(x,y)\,dA.$$
- Sum/Difference rule:
$$\iint_R \big(f(x,y) \pm g(x,y)\big)\,dA = \iint_R f(x,y)\,dA \pm \iint_R g(x,y)\,dA$$

- If $f(x,y) \geq 0$ on R, then
$$\iint_R f(x,y)\ dA \geq 0.$$
- If $f(x,y) \geq g(x,y)$ on R, then
$$\iint_R f(x,y)\ dA \geq \iint_R g(x,y)\ dA.$$
- Let R be the union of two nonoverlapping regions, $R = R_1 \cup R_2$. Then
$$\iint_R f(x,y)\ dA = \iint_{R_1} f(x,y)\ dA + \iint_{R_2} f(x,y)\ dA.$$

Actually, since this property is intuitively clear, we relied already in Examples 17.2 and 17.4. Of course, this property generalizes to n nonoverlapping regions.

Example 17.7

Let $f(x,y) = \sin(x)\cos(y)$ and R be the triangle with vertices $(-1,0)$, $(1,0)$ and $(0,1)$ (see Figure 17.9). Evaluate the double integral $\iint_R f(x,y)\ dA$.

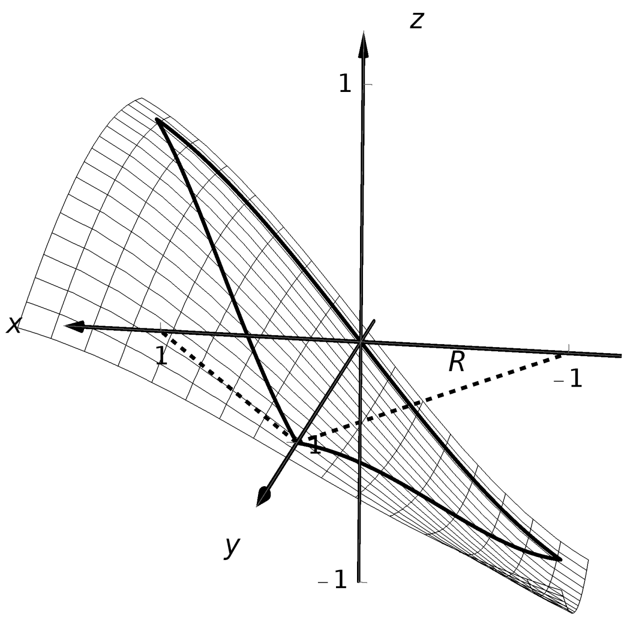


Figure 17.9: Finding the signed volume under a surface in Example 17.7.

Solution

If we attempt to integrate using an iterated integral with the order $dy\ dx$, note how there are two upper bounds on R meaning we will need to use two iterated integrals. We would need to split the triangle into two regions along the y-axis.

Instead, let us use the order $dx\ dy$. The curves bounding x are $y-1 \leq x \leq 1-y$; the bounds on y are $0 \leq y \leq 1$. This gives us:

$$\iint_R f(x,y)\ dA = \int_0^1 \int_{y-1}^{1-y} \sin(x)\cos(y)\ dx\ dy$$

$$= \int_0^1 \Big(-\cos(x)\cos(y) \Big)\Big|_{y-1}^{1-y} \, dy$$

$$= \int_0^1 \cos(y)\Big(-\cos(1-y) + \cos(y-1) \Big) \, dy.$$

Recall that the cosine function is an even function; that is, $\cos(x) = \cos(-x)$. Therefore, from the last integral above, we have $\cos(y-1) = \cos(1-y)$. Thus the integrand simplifies to 0, and we have

$$\iint_R f(x,y) \, dA = \int_0^1 0 \, dy = 0.$$

It turns out that over R, there is just as much volume above the xy-plane as below (look again at Figure 17.9), giving a final signed volume of 0.

In the previous section we practised changing the order of integration of a given iterated integral, where the region R was not explicitly given. Changing the bounds of an integral is more than just an test of understanding. Rather, there are cases where integrating in one order is really hard, if not impossible, whereas integrating with the other order is feasible.

Example 17.8

Rewrite the iterated integral

$$\int_0^3 \int_y^3 e^{-x^2} \, dx \, dy$$

with the order $dy\, dx$. Comment on the feasibility to evaluate each integral.

Solution

Once again we make a sketch of the region over which we are integrating to facilitate changing the order. The bounds on x are from $x = y$ to $x = 3$; the bounds on y are from $y = 0$ to $y = 3$. These curves are sketched in Figure 17.10(a), enclosing the region R.

To change the bounds, note that the curves bounding y are $y = 0$ up to $y = x$; the triangle is enclosed between $x = 0$ and $x = 3$. Thus the new bounds of integration are $0 \le y \le x$ and $0 \le x \le 3$, giving the iterated integral

$$\int_0^3 \int_0^x e^{-x^2} \, dy \, dx.$$

How easy is it to evaluate each iterated integral? Consider the order of integrating $dx\, dy$, as given in the original problem. The first indefinite integral we need to evaluate is $\int e^{-x^2} \, dx$; we have stated before that this integral cannot be evaluated in terms of elementary functions. We are stuck.

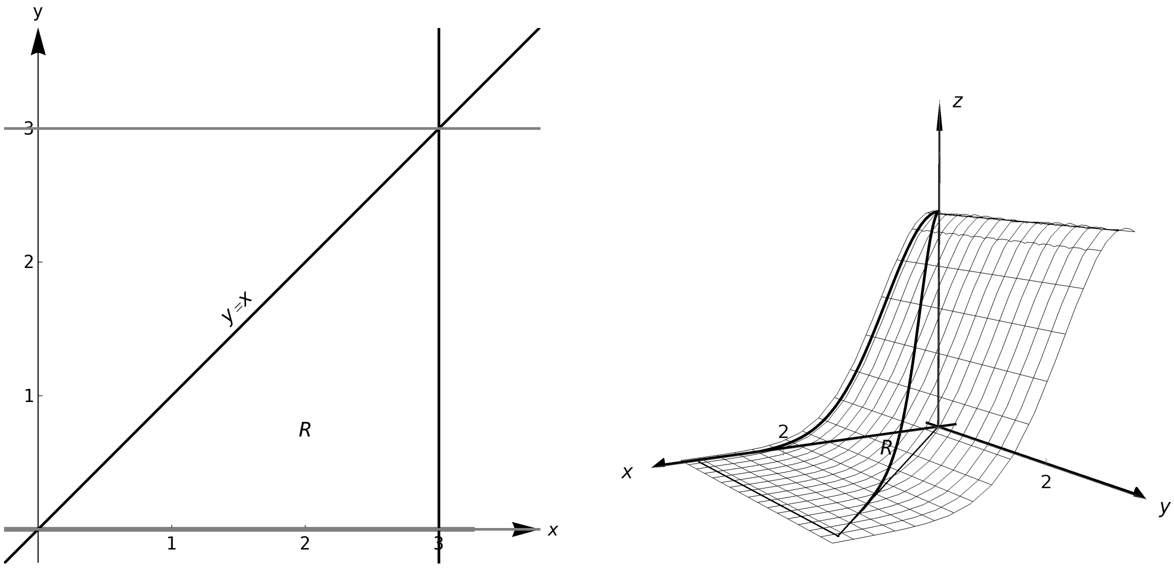
Changing the order of integration makes a big difference here. In the second iterated integral, we are faced with $\int e^{-x^2} \, dy$; integrating with respect to y gives us $ye^{-x^2} + C$, and the first definite integral evaluates to

$$\int_0^x e^{-x^2} \, dy = xe^{-x^2}.$$

Thus

$$\int_0^3 \int_0^x e^{-x^2}\, dy\, dx = \int_0^3 x e^{-x^2}\, dx.$$

This last integral is easy to evaluate with substitution, giving a final answer of $(1-e^{-9})/2 \approx 0.5$. Figure 17.10(b) shows the surface over R. In short, evaluating one iterated integral is impossible; the other iterated integral is relatively simple.



(a) (b)

Figure 17.10: Determining the region R determined by the bounds of integration (a) and the surface f over its region R in Example 17.8.

Using double integrals we can also determine the average value of $z = f(x,y)$ over a region R. This is nothing but the volume under f over R divided by the area of R; that is

$$\text{average value of } f \text{ on } R = \frac{\iint_R f(x,y)\, dA}{\iint_R dA}. \tag{17.2}$$

Example 17.9

Find the average value of $f(x,y) = 4-y$ over the region R, which is bounded by the parabolas $y^2 = 4x$ and $x^2 = 4y$ (see Figure 17.11).

Solution

Graphing each curve can help us find their points of intersection. Solving analytically, the second equation tells us that $y = x^2/4$. Substituting this value in for y in the first equation gives us $x^4/16 = 4x$. Solving for x:

$$\begin{aligned} \frac{x^4}{16} &= 4x \\ \Leftrightarrow \quad x^4 - 64x &= 0 \end{aligned}$$

$$\Leftrightarrow \quad x(x^3-64)=0$$
$$\Leftrightarrow \quad x=0 \;\vee\; x=4.$$

Thus we have found analytically what was easy to approximate graphically: the regions intersect at $(0,0)$ and $(4,4)$, as shown in Figure 17.11.

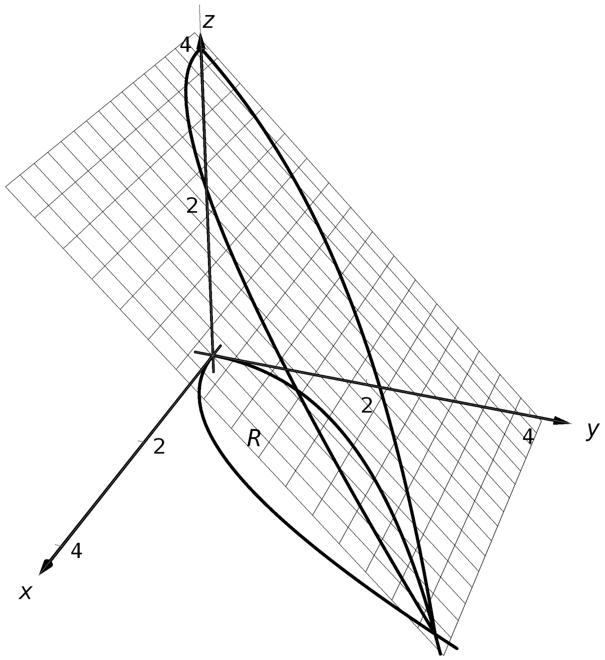


Figure 17.11: Finding the signed volume under a surface in Example 17.9.

We now choose an order of integration: $dy\,dx$ or $dx\,dy$? Either order works; since the integrand does not contain x, choosing $dx\,dy$ might be simpler – at least, the first integral is very simple.

Thus we have the following curve to curve, point to point bounds: $y^2/4 \le x \le 2\sqrt{y}$, and $0 \le y \le 4$.

$$\begin{aligned}
\iint_R (4-y)\,dA &= \int_0^4 \int_{y^2/4}^{2\sqrt{y}} (4-y)\,dx\,dy \\
&= \int_0^4 (4-y) \int_{y^2/4}^{2\sqrt{y}} dx\,dy \\
&= \int_0^4 \big(x(4-y)\big)\Big|_{y^2/4}^{2\sqrt{y}} dy \\
&= \int_0^4 \left[\left(2\sqrt{y}-\frac{y^2}{4}\right)(4-y)\right] dy = \int_0^4 \left(\frac{y^3}{4}-y^2-2y^{3/2}+8y^{1/2}\right) dy \\
&= \left(\frac{y^4}{16}-\frac{y^3}{3}-\frac{4y^{5/2}}{5}+\frac{16y^{3/2}}{3}\right)\Bigg|_0^4
\end{aligned}$$

$$= \frac{176}{15} \approx 11.73.$$

We now find the area of R by computing $\iint_R dA$:

$$\iint_R dA = \int_0^4 \int_{y^2/4}^{2\sqrt{y}} dx\, dy = \frac{16}{3}.$$

Dividing the volume under the surface by the area gives the average value:

$$\text{average value of } f \text{ on } R = \frac{176/15}{16/3} = \frac{11}{5} = 2.2.$$

While the surface covers z-values from $z = 0$ to $z = 4$, the average z-value on R is 2.2.

Our new understanding of double integrals allows us to revisit what we did in the previous section. Given a region R in the plane, we computed $\iint_R 1\, dA$; again. Our understanding at the time was that we were finding the area of R. However, we can now view the function $z = 1$ as a surface, a flat surface with constant z-value of 1. The double integral $\iint_R 1\, dA$ finds the volume, under $z = 1$, over R. We were actually computing the volume of a solid, though we interpreted the number as an area.

17.2.3 Differentiation under the integral sign

Sometimes we will run into the derivative of an integral of a function of two variables, where, however, one of the variables is considered as a parameter. For instance, we may encounter something like

$$\int_{a(x)}^{b(x)} f(x,t)\, dt\,,$$

where $-\infty < a(x), b(x) < +\infty$. Even though the way how to proceed with this is very intuitive, the formal proof is rather involved, especially for the general case with variable limits. We first state the result as a theorem.

Theorem 17.3 (Leibniz integral rule)
Let $f(x,t)$ be a function such that both $f(x,t)$ and its partial derivative $f_x(x,t)$ are continuous in some region of the xt-plane, including $a(x) \le t \le b(x), x_0 \le x \le x_1$. Also suppose that the functions $a(x)$ and $b(x)$ are both continuous and both have continuous derivatives for $x_0 \le x \le x_1$. Then, for $x_0 \le x \le x_1$, we have that

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)\, dt\right) = f\big(x, b(x)\big)\frac{d}{dx}b(x) - f\big(x, a(x)\big)\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)\, dt\,.$$

Proof The general form of this theorem can be derived as a consequence of the basic form of leibniz's integral rule with constant limits of integration, the multivariable chain rule, and the fundamental theorem of calculus. So, let us first prove the simpler case with constant limits of integration, i.e.

$$\frac{d}{dx}\left(\int_a^b f(x,t)\, dt\right) = \int_a^b \frac{\partial}{\partial x} f(x,t)\, dt.$$

Essentially, in an attempt to get a derivative in the right-hand side of the above equation, as is the case in its left-hand side, we may rewrite it according to the fundamental theorem of calculus (Section 12.3) as

$$\frac{d}{dx}\left(\int_A^x\left(\int_a^b \frac{\partial}{\partial s} f(s,t)\,dt\right)ds\right),$$

because we have function of one variable s within the inner parentheses. Subsequently, we rely on Fubini's theorem to interchange the order of integration

$$\begin{aligned}
\frac{d}{dx}\left(\int_A^x\left(\int_a^b \frac{\partial}{\partial s} f(s,t)\,dt\right)ds\right) &= \frac{d}{dx}\left(\int_a^b\left(\int_A^x \frac{\partial}{\partial s} f(s,t)\ ds\right)dt\right),\\
&= \frac{d}{dx}\int_a^b \big(f(x,t)-f(A,t)\big)\ dt\,,\\
&= \frac{d}{dx}\int_a^b f(x,t)\ dt - \frac{d}{dx}\int_a^b f(A,t)\ dt\,,\\
&= \frac{d}{dx}\int_a^b f(x,t)\ dt\,,
\end{aligned}$$

To prove the general case with variable limits, suppose now f is defined in a rectangle in the xt-plane, for $x \in [x_0, x_1]$ and $t \in [t_0, t_1]$. Also, assume f and the partial derivative f_x are both continuous functions on this rectangle. Suppose a, b are differentiable real valued functions defined on $[x_0, x_1]$ with values in $[t_0, t_1]$ (i.e for every $x \in [x_0, x_1], a(x), b(x) \in [t_0, t_1]$). Now, set

$$F(x,y) = \int_{t_0}^{y} f(x,t)\,dt$$

for $x \in [x_0, x_1]$ and $y \in [t_0, t_1]$ and

$$G(x) = \int_{a(x)}^{b(x)} f(x,t)\,dt$$

for $x \in [x_0, x_1]$. Then, by properties of definite integrals, we can write

$$\begin{aligned}
G(x) &= \int_{t_0}^{b(x)} f(x,t)\,dt - \int_{t_0}^{a(x)} f(x,t)\,dt\\
&= F(x,b(x)) - F(x,a(x)).
\end{aligned}$$

Since the partial derivatives of F are given by

$$\frac{\partial F}{\partial x}(x,y) = \int_{t_0}^{y} \frac{\partial f}{\partial x}(x,t)dt$$

and

$$\frac{\partial F}{\partial y}(x,y) = f(x,y)$$

and recalling f_x is continuous, its integral is also a continuous function. Moreover, since f is also continuous, these two results show that both partial derivatives of F are continuous. Since continuity of partial derivatives implies differentiability of the function, F is differentiable. So, since the functions F, a, b are all differentiable, by the multivariable chain rule, it follows that G is differentiable, and its derivative is given by the formula:

$$G'(x) = \left(\frac{\partial F}{\partial x}\big(x, b(x)\big) + \frac{\partial F}{\partial y}\big(x, b(x)\big)\, b'(x)\right) - \left(\frac{\partial F}{\partial x}\big(x, a(x)\big) + \frac{\partial F}{\partial y}\big(x, a(x)\big)\, a'(x)\right).$$

Now, note that for every $x \in [x_0, x_1]$, and for every $y \in [t_0, t_1]$, we have that

$$\frac{\partial F}{\partial x}(x,y) = \int_{t_0}^{y} \frac{\partial f}{\partial x}(x,t)dt\,;$$

because when taking the partial derivative with respect to x of F, we are keeping y fixed in the expression

$$\int_{t_0}^{y} f(x,t)\,dt\,.$$

Consequently, the basic form of Leibniz's integral rule with constant limits of integration applies. Next, by the first fundamental theorem of calculus, we have that

$$\frac{\partial F}{\partial y}(x,y) = f(x,y)\,;$$

because when taking the partial derivative with respect to y of F, the first variable x is fixed, so the fundamental theorem can indeed be applied. Substituting these results into the equation for $G'(x)$ above gives:

$$\begin{aligned} G'(x) &= \left(\int_{t_0}^{b(x)} \frac{\partial f}{\partial x}(x,t)dt + f\big(x, b(x)\big)\, b'(x)\right) - \left(\int_{t_0}^{a(x)} \frac{\partial f}{\partial x}(x,t)dt + f\big(x, a(x)\big)\, a'(x)\right) \\ &= f\big(x, b(x)\big)\, b'(x) - f\big(x, a(x)\big)\, a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f}{\partial x}(x,t)dt, \end{aligned}$$

as desired. □

Example 17.10

Using Leibniz integral rule, we easily find that, for instance,

$$\begin{aligned}\frac{d}{dx}\left(\int_{\sin(x)}^{\cos(x)} \cosh(t^2)\,dt\right) &= \cosh\left(\cos^2(x)\right)\frac{d}{dx}(\cos(x)) - \cosh\left(\sin^2(x)\right)\frac{d}{dx}(\sin(x)) + \int_{\sin(x)}^{\cos(x)} \frac{\partial}{\partial x}\left(\cosh(t^2)\right)dt \\ &= \cosh\left(\cos^2(x)\right)(-\sin(x)) - \cosh\left(\sin^2(x)\right)(\cos(x)) + 0 \\ &= -\cosh\left(\cos^2(x)\right)\sin(x) - \cosh\left(\sin^2(x)\right)\cos(x).\end{aligned}$$

Still, the principle of differentiating under the integral sign is typically used to evaluate a definite integral in cases where one cannot easily find an antiderivative.

Example 17.11

Compute the definite integral

$$\int_0^1 \frac{t^3-1}{\ln(t)}\,dt\,.$$

Solution

This integral appears resistant to standard integration techniques such as integration by parts, substitution, and so on. Hence, we would like to use differentiation under the integral sign to compute it. But how can we choose a function to differentiate under the integral sign? The appearance of $\ln(t)$ in the denominator of the integrand is of course quite unwelcome, and we would like to get rid of it. Luckily, we know

$$\frac{d(t^x)}{dx} = t^x \ln(t)$$

so differentiating the numerator with respect to the exponent seems to be what we would like to do. Accordingly, we define a function

$$g(x) = \int_0^1 \frac{t^x-1}{\ln(t)}\,dt\,.$$

In this notation, the integral we wish to evaluate is $g(3)$. Observe that the given integral has been recast as member of a family of definite integrals indexed by the variable x.

By Leibniz integral rule, we compute

$$\frac{dg(x)}{dx} = \int_0^1 \frac{\partial}{\partial x}\left(\frac{t^x-1}{\ln(t)}\right)dt = \int_0^1 \frac{t^x\ln(t)}{\ln(t)}\,dt = \left.\frac{t^{x+1}}{x+1}\right|_0^1 = \frac{1}{x+1}\,.$$

It follows that $g(x) = \ln|x+1| + C$ for some constant C. To determine this constant C, note that $g(0) = 0$, so $g(0) = 0 = \ln|1| + C$, from which we conclude that $C = 0$. Hence, $g(x) = \ln|x+1|$ for all x such that the integral exists. In particular, $g(3) = \ln(4)$.

17.2.4 Double integration with polar coordinates

Some regions R are easy to describe using rectangular coordinates – that is, with equations of the form $y = f(x)$, $x = a$, etc. However, some regions are easier to handle if we represent their boundaries with polar equations of the form $r = f(\theta)$, $\theta = \alpha$, etc.

The basic form of the double integral is $\iint_R f(x,y)\ dA$. We interpret this integral as follows: over the region R, sum up lots of products of heights (given by $f(x_i, y_i)$) and areas (given by ΔA_i). That is, dA represents a little bit of area. In rectangular coordinates, we can describe a small rectangle as having area $dx\ dy$ or $dy\ dx$ – the area of a rectangle is simply (length×width) – a small change in x times a small change in y. Thus we replace dA in the double integral with $dx\ dy$ or $dy\ dx$.

Now consider representing a region R with polar coordinates. Consider Figure 17.12(a). Let R be the region in the first quadrant bounded by the curve. We can approximate this region using the natural shape of polar coordinates: portions of sectors of circles. In the figure, one such region is shaded, shown again in Figure 17.12(b).

As the area of a sector of a circle with radius r, subtended by an angle θ, is $A = \frac{1}{2}r^2\theta$, we can find the area of the shaded region. The whole sector has area $\frac{1}{2}r_2^2\Delta\theta$, whereas the smaller, unshaded sector has area $\frac{1}{2}r_1^2\Delta\theta$. The area of the shaded region is the difference of these areas:

$$\Delta A_i = \frac{1}{2}r_2^2\Delta\theta - \frac{1}{2}r_1^2\Delta\theta = \frac{1}{2}\big(r_2^2 - r_1^2\big)\Delta\theta = \frac{r_2 + r_1}{2}\big(r_2 - r_1\big)\Delta\theta.$$

Note that $(r_2 + r_1)/2$ is just the average of the two radii.

To approximate the region R, we use many such subregions; doing so shrinks the difference $r_2 - r_1$ between radii to 0 and shrinks the change in angle $\Delta\theta$ also to 0. We represent these infinitesimal changes in radius and angle as dr and $d\theta$, respectively. Finally, as dr is small, $r_2 \approx r_1$, and so $(r_2 + r_1)/2 \approx r_1$. Thus, when dr and $d\theta$ are small,

$$\Delta A_i \approx r_i\ dr\ d\theta.$$

Taking a limit, where the number of subregions goes to infinity and both $r_2 - r_1$ and $\Delta\theta$ go to 0, we get

$$dA = r\ dr\ d\theta.$$

So to evaluate $\iint_R f(x,y)\ dA$, replace dA with $r\ dr\ d\theta$. Convert the function $z = f(x,y)$ to a function with polar coordinates with the substitutions $x = r\cos(\theta)$, $y = r\sin(\theta)$. Finally, find bounds $g_1(\theta) \le r \le g_2(\theta)$ and $\alpha \le \theta \le \beta$ that describe R. Consequently, if $z = f(x,y)$ is a continuous function defined over a closed, bounded region R in the xy-plane, where R is bounded by the polar equations $\alpha \le \theta \le \beta$ and $g_1(\theta) \le r \le g_2(\theta)$, then

$$\iint_R f(x,y)\ dA = \int_\alpha^\beta \int_{g_1(\theta)}^{g_2(\theta)} f\big(r\cos(\theta), r\sin(\theta)\big)\ r\ dr\ d\theta. \tag{17.3}$$

Examples will help us understand this.

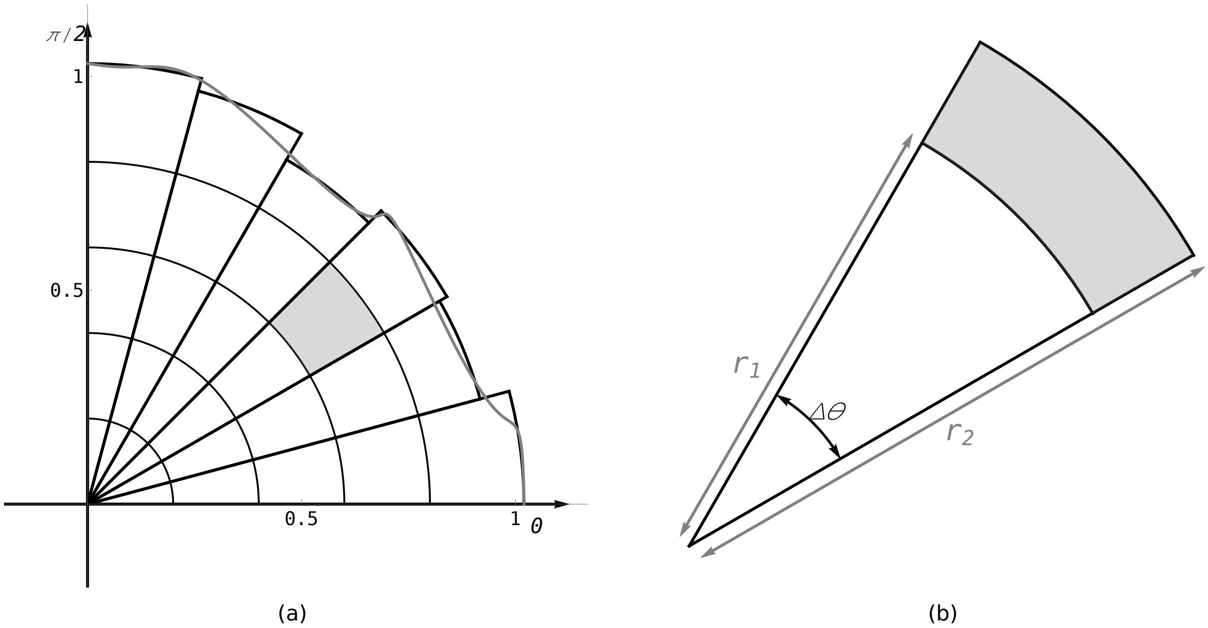


Figure 17.12: Approximating a region R with portions of sectors of circles.

Example 17.12

Find the signed volume under the plane $z = 4 - x - 2y$ over the disk bounded by the circle with equation $x^2 + y^2 = 1$.

Solution

The bounds of the integral are determined solely by the region R over which we are integrating. In this case, it is a disk with boundary $x^2 + y^2 = 1$. We need to find polar bounds for this region. Bounds for this disk are $0 \le r \le 1$ and $0 \le \theta \le 2\pi$.

We replace $f(x, y)$ with $f\big(r\cos(\theta), r\sin(\theta)\big)$. That means we make the following substitutions:

$$4 - x - 2y \quad \Rightarrow \quad 4 - r\cos(\theta) - 2r\sin(\theta).$$

Finally, we replace dA in the double integral with $r\ dr\ d\theta$. This gives the final iterated integral, which we evaluate:

$$\begin{aligned}
\iint_R f(x, y)\ dA &= \int_0^{2\pi}\int_0^1 \big(4 - r\cos(\theta) - 2r\sin(\theta)\big)r\ dr\ d\theta \\
&= \int_0^{2\pi}\int_0^1 \Big(4r - r^2\big(\cos(\theta) - 2\sin(\theta)\big)\Big)\ dr\ d\theta \\
&= \int_0^{2\pi} \left(2r^2 - \frac{1}{3}r^3\big(\cos(\theta) - 2\sin(\theta)\big)\right)\Bigg|_0^1\ d\theta \\
&= \int_0^{2\pi} \left(2 - \frac{1}{3}\big(\cos(\theta) - 2\sin(\theta)\big)\right)\ d\theta \\
&= \left(2\theta - \frac{1}{3}\big(\sin(\theta) + 2\cos(\theta)\big)\right)\Bigg|_0^{2\pi}
\end{aligned}$$

$$= 4\pi \approx 12.566.$$

The surface and region R are shown in Figure 17.13.

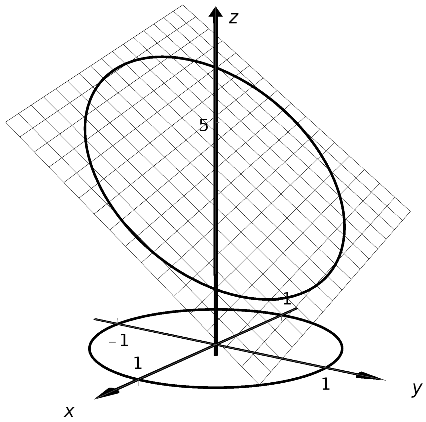


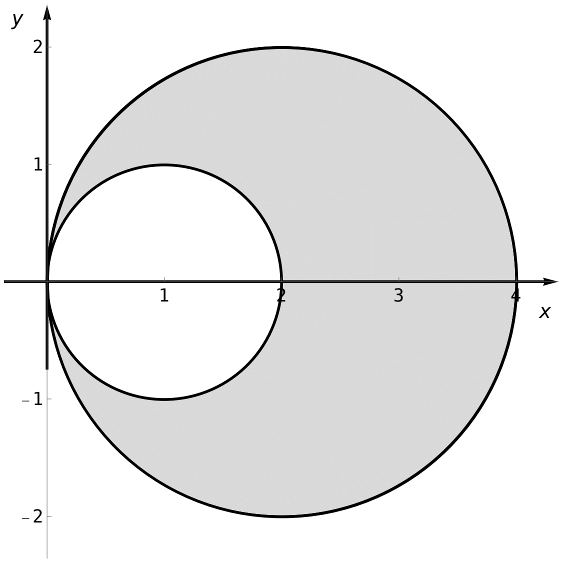
Figure 17.13: Evaluating a double integral with polar coordinates in Example 17.12.

Example 17.13

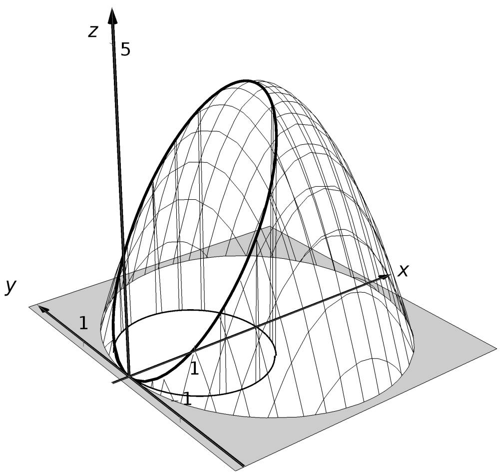
Find the volume under the paraboloid $z = 4-(x-2)^2-y^2$ over the region bounded by the circles $(x-1)^2+y^2=1$ and $(x-2)^2+y^2=4$.

Solution

At first glance, this seems like a very hard volume to compute as the region R (shown in Figure 17.14(a)) has a hole in it, cutting out a strange portion of the surface, as shown in Figure 17.14(b). However, by describing R in terms of polar equations, the volume is not very difficult to compute.



(a)



(b)

Figure 17.14: Showing the region R (a) and surface (b) used in Example 17.13

The circle $(x-1)^2+y^2=1$ has polar equation $r = 2\cos(\theta)$, while the circle $(x-2)^2+y^2=4$ has polar equation $r=4\cos(\theta)$. We may trace out semicircles on the interval $0 \le \theta \le \pi/2$. The bounds on r are $2\cos(\theta) \le r \le 4\cos(\theta)$. Replacing x with $r\cos(\theta)$ in the integrand, along with replacing y with $r\sin(\theta)$ and noting that we should add a factor 2 to account for the entire volume, prepares us to evaluate the double integral $\iint_R f(x,y)\,dA$:

$$\begin{aligned}
\iint_R f(x,y)\,dA &= 2\int_0^{\pi/2}\int_{2\cos(\theta)}^{4\cos(\theta)} \Big(4-\big(r\cos(\theta)-2\big)^2-\big(r\sin(\theta)\big)^2\Big) r\,dr\,d\theta \\
&= 2\int_0^{\pi/2}\int_{2\cos(\theta)}^{4\cos(\theta)} \big(-r^3+4r^2\cos(\theta)\big)\,dr\,d\theta \\
&= 2\int_0^{\pi/2} \left(-\frac{1}{4}r^4+\frac{4}{3}r^3\cos(\theta)\right)\Bigg|_{2\cos(\theta)}^{4\cos(\theta)} d\theta \\
&= 2\int_0^{\pi/2} \Bigg(\bigg[-\frac{1}{4}\Big(256\cos^4(\theta)\Big)+\frac{4}{3}\Big(64\cos^4(\theta)\Big)\bigg]- \\
&\qquad \bigg[-\frac{1}{4}\Big(16\cos^4(\theta)\Big)+\frac{4}{3}\Big(8\cos^4(\theta)\Big)\bigg]\Bigg)\,d\theta \\
&= 2\int_0^{\pi/2} \frac{44}{3}\cos^4(\theta)\,d\theta.
\end{aligned}$$

To integrate $\cos^4(\theta)$, rewrite it as $\cos^2(\theta)\cos^2(\theta)$ and employ the power-reducing formula twice:

$$\begin{aligned}
\cos^4(\theta) &= \cos^2(\theta)\cos^2(\theta) \\
&= \frac{1}{2}\big(1+\cos(2\theta)\big)\frac{1}{2}\big(1+\cos(2\theta)\big) \\
&= \frac{1}{4}\big(1+2\cos(2\theta)+\cos^2(2\theta)\big) \\
&= \frac{1}{4}\Big(1+2\cos(2\theta)+\frac{1}{2}\big(1+\cos(4\theta)\big)\Big) \\
&= \frac{3}{8}+\frac{1}{2}\cos(2\theta)+\frac{1}{8}\cos(4\theta).
\end{aligned}$$

Picking up from where we left off above, we have

$$\begin{aligned}
\iint_R f(x,y)\,dA &= 2\int_0^{\pi/2} \frac{44}{3}\cos^4(\theta)\,d\theta \\
&= 2\int_0^{\pi/2} \frac{44}{3}\left(\frac{3}{8}+\frac{1}{2}\cos(2\theta)+\frac{1}{8}\cos(4\theta)\right)d\theta \\
&= \frac{88}{3}\left(\frac{3}{8}\theta+\frac{1}{4}\sin(2\theta)+\frac{1}{32}\sin(4\theta)\right)\Bigg|_0^{\pi/2} \\
&= \frac{11}{2}\pi \approx 17.279.
\end{aligned}$$

Note that the double integral would have been much harder to evaluate had we used rectangular coordinates.

Example 17.14

Find the volume under the surface $f(x,y) = \left(x^2+y^2+1\right)^{-1}$ over the sector of the circle with radius a centered at the origin in the first quadrant, as shown in Figure 17.15.

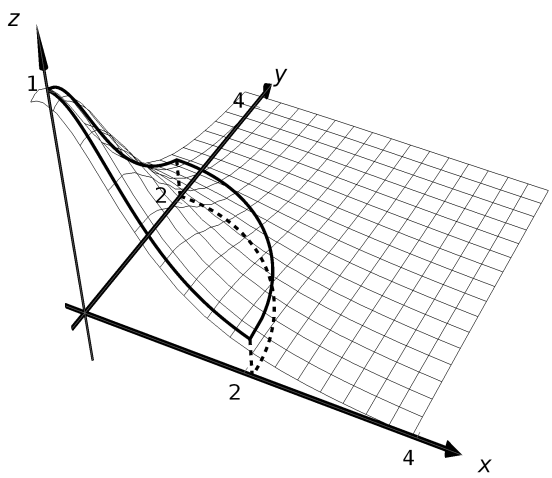


Figure 17.15: The surface and region R used in Example 17.14.

Solution

The region R we are integrating over is a circle with radius a, restricted to the first quadrant. Thus, in polar, the bounds on R are $0 \le r \le a$, $0 \le \theta \le \pi/2$. The integrand is rewritten in polar coordinates as

$$\frac{1}{x^2+y^2+1} \quad \Rightarrow \quad \frac{1}{r^2\cos^2(\theta)+r^2\sin^2(\theta)+1} = \frac{1}{r^2+1}.$$

We find the volume as follows:

$$\begin{aligned}
\iint_R f(x,y)\, dA &= \int_0^{\pi/2}\int_0^{a} \frac{r}{r^2+1}\, dr\, d\theta \\
&= \int_0^{\pi/2} \frac{1}{2}\big(\ln|r^2+1|\big)\Big|_0^{a}\, d\theta \\
&= \int_0^{\pi/2} \frac{1}{2}\ln(a^2+1)\, d\theta \\
&= \left(\frac{1}{2}\ln(a^2+1)\theta\right)\Bigg|_0^{\pi/2}
\end{aligned}$$

$$= \frac{\pi}{4} \ln(a^2 + 1).$$

Figure 17.15 shows that f shrinks to near 0 very quickly. Regardless, as a grows, so does the volume, without bound.

Example 17.15

Find the volume of a sphere with radius a.

Solution

The sphere of radius a, centred at the origin, has equation $x^2 + y^2 + z^2 = a^2$; solving for z, we have

$$z = \pm\sqrt{a^2 - x^2 - y^2},$$

where the half solution $z > 0$ gives the upper half of a sphere. We wish to find the volume under this top half, then double it to find the total volume.

The region we need to integrate over is the disk of radius a, centred at the origin. Polar bounds for this equation are $0 \le r \le a$, $0 \le \theta \le 2\pi$.

All together, the volume of a sphere with radius a is:

$$\begin{aligned} 2\iint_R \sqrt{a^2 - x^2 - y^2}\, dA &= 2\int_0^{2\pi}\int_0^{a} \sqrt{a^2 - \Big(r\cos(\theta)\Big)^2 - \Big(r\sin(\theta)\Big)^2}\, r\, dr\, d\theta \\ &= 2\int_0^{2\pi}\int_0^{a} r\sqrt{a^2 - r^2}\, dr\, d\theta. \end{aligned}$$

We can evaluate this inner integral with substitution. With $u = a^2 - r^2$, $du = -2r\, dr$. The new bounds of integration are $u(0) = a^2$ to $u(a) = 0$. Thus we have:

$$\begin{aligned} &= \int_0^{2\pi}\int_{a^2}^{0} \left(-u^{1/2}\right) du\, d\theta \\ &= \int_0^{2\pi} \left(-\frac{2}{3}u^{3/2}\right)\Bigg|_{a^2}^{0} d\theta \\ &= \int_0^{2\pi} \left(\frac{2}{3}a^3\right) d\theta \\ &= \left(\frac{2}{3}a^3\theta\right)\Bigg|_0^{2\pi} \\ &= \frac{4}{3}\pi a^3. \end{aligned}$$

Generally, the formula for the volume of a sphere with radius r is given as $4\pi r^3/3$; we have justified this formula with our calculation.

We have used iterated integrals to find areas of plane regions and volumes under surfaces. Just as a single integral can be used to compute much more than area under the curve, iterated integrals can be used to compute much more than we have thus far seen. The next two sections show two, among many, applications of iterated integrals.

17.3 Centre of mass

We have used iterated integrals to find areas of plane regions and signed volumes under surfaces. Here, we will apply iterated integrals to compute the **mass** and **centre of mass** (*massamiddelpunt*) of planar regions.

17.3.1 Mass and weight

Consider a thin sheet of material with constant thickness and finite area. Mathematicians (and physicists and engineers) call such a sheet a lamina. So consider a lamina, as shown in Figure 17.16(a), with the shape of some planar region R, as shown in Figure 17.16(b).

We can write a simple double integral that represents the mass of the lamina: $\iint_R \, dm$, where dm means a little mass. That is, the double integral states the total mass of the lamina can be found by summing up lots of little masses over R. To evaluate this double integral, partition R into n subregions as we have done in the past. The i^{th} subregion has area ΔA_i. A fundamental property of mass is that "mass=density×area". If the lamina has a constant density δ, then the mass of this i^{th} subregion is $\Delta m_i = \delta \Delta A_i$. That is, we can compute a small amount of mass by multiplying a small amount of area by the density.

If density is variable, with density function $\delta = \delta(x, y)$, then we can approximate the mass of the i^{th} subregion of R by multiplying ΔA_i by $\delta(x_i, y_i)$, where (x_i, y_i) is a point in that subregion. That is, for a small enough subregion of R, the density across that region is almost constant.

Note that mass and weight are different measures. Since they are scalar multiples of each other, it is often easy to treat them as the same measure. Here, we effectively treat them as the same, as our technique for finding mass is the same as for finding weight. The density functions used will simply have different units.

The total mass M of the lamina is approximately the sum of approximate masses of subregions:

$$M \approx \sum_{i=1}^{n} \Delta m_i = \sum_{i=1}^{n} \delta(x_i, y_i)\Delta A_i.$$

Taking the limit as the size of the subregions shrinks to 0 gives us the actual mass; that is, integrating $\delta(x, y)$ over R gives the mass of the lamina:

$$M = \iint_R \, dm = \iint_R \delta(x, y) \, dA. \tag{17.4}$$

Example 17.16

Find the mass of a square lamina, represented by the unit square with lower lefthand corner at the origin, with variable density $\delta(x, y) = (x + y + 2)\text{g/cm}^2$.

Solution

The variable density δ, in this example, is very uniform, giving a density of 3 in the centre of the square and changing linearly. A graph of $\delta(x, y)$ can be seen in Figure 17.17; notice how same amount of density is above $z = 3$ as below. We'll comment on the significance of this momentarily.

The mass M is found by integrating $\delta(x, y)$ over R. The order of integration is not important; we

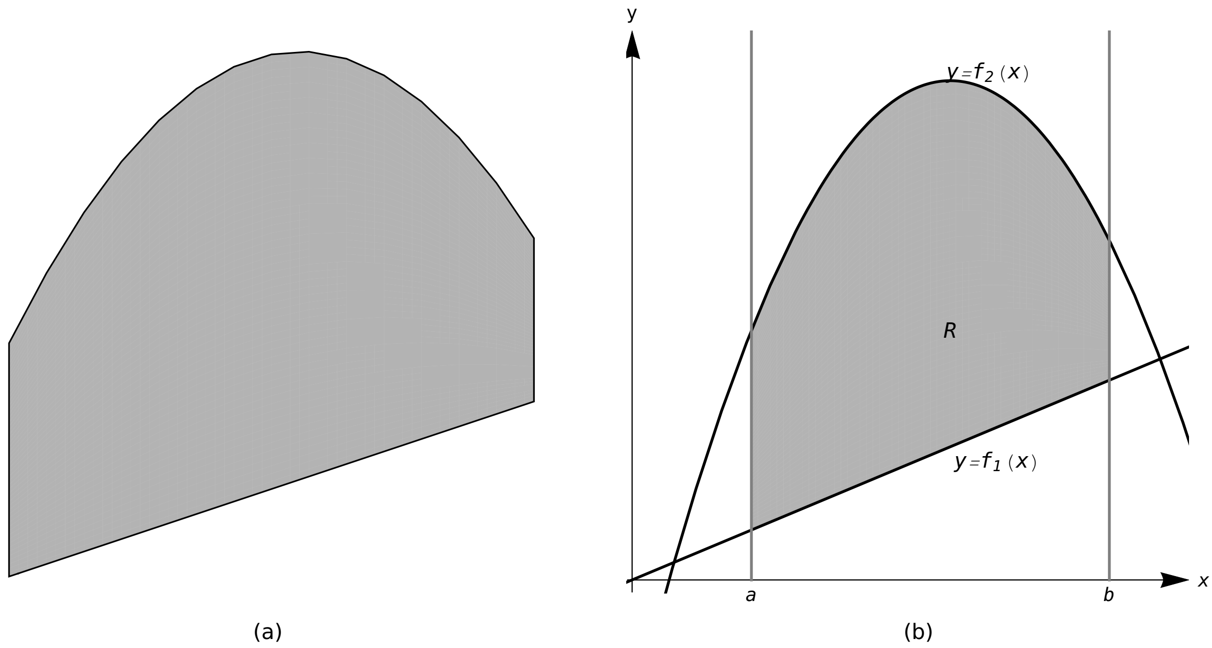


Figure 17.16: Illustrating the concept of a lamina.

choose $dx\ dy$ arbitrarily. Thus:

$$\begin{aligned} M = \iint_R (x+y+2)\ dA &= \int_0^1 \int_0^1 (x+y+2)\ dx\ dy \\ &= \int_0^1 \left(\frac{1}{2}x^2 + x(y+2) \right)\Bigg|_0^1 dy \\ &= \int_0^1 \left(\frac{5}{2} + y \right)\ dy \\ &= \left(\frac{5}{2}y + \frac{1}{2}y^2 \right)\Bigg|_0^1 = 3\text{g}. \end{aligned}$$

It turns out that since the density of the lamina is so uniformly distributed above and below $z = 3$ that the mass of the lamina is the same as if it had a constant density of 3. The density function $\delta = 3\text{g/cm}^2$ and the one from this example are graphed in Figure 17.17, which illustrates this concept.

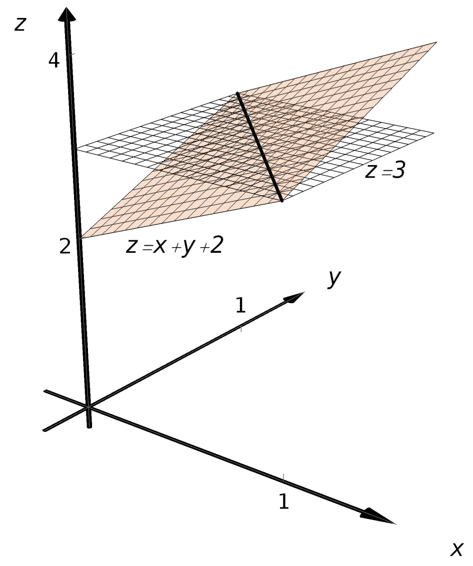


Figure 17.17: Graphing the density functions $\delta = 3\text{g/cm}^2$ and $\delta(x,y) = (x+y+2)\text{g/cm}^2$.

Example 17.17

Find the weight of the lamina represented by the disk with radius 2cm, centred at the origin, with density function $\delta(x,y) = (x^2+y^2+1)\text{g/cm}^2$. Compare this to the weight of the lamina with the same shape and density $\delta(x,y) = (2\sqrt{x^2+y^2}+1)\text{g/cm}^2$.

Solution

A direct application of Equation (17.4) states that the weight of the lamina is $\iint_R \delta(x,y)\,dA$. Since our lamina is in the shape of a circle, it makes sense to approach the double integral using polar coordinates.

The density function $\delta(x,y) = x^2+y^2+1$ becomes

$$\delta(r,\theta) = (r\cos(\theta))^2 + (r\sin(\theta))^2 + 1 = r^2+1.$$

The circle is bounded by $0 \le r \le 2$ and $0 \le \theta \le 2\pi$. Thus the weight W is:

$$\begin{aligned} W &= \int_0^{2\pi}\int_0^2 (r^2+1)r\,dr\,d\theta \\ &= \int_0^{2\pi} \left(\frac{1}{4}r^4 + \frac{1}{2}r^2\right)\Big|_0^2\,d\theta \\ &= \int_0^{2\pi} 6\,d\theta \\ &= 12\pi \approx 37.70\text{g}. \end{aligned}$$

Now compare this with the density function $\delta(x,y) = 2\sqrt{x^2+y^2}+1$. Converting this to polar coordinates gives

$$\delta(r,\theta) = 2\sqrt{\Big(r\cos(\theta)\Big)^2 + \Big(r\sin(\theta)\Big)^2} + 1 = 2r + 1.$$

Thus the weight W is:

$$\begin{aligned} W &= \int_0^{2\pi}\int_0^{2} (2r+1)r\ dr\ d\theta \\ &= \int_0^{2\pi} \left(\frac{2}{3}r^3 + \frac{1}{2}r^2\right)\Big|_0^2 d\theta \\ &= \int_0^{2\pi} \left(\frac{22}{3}\right)\ d\theta \\ &= \frac{44}{3}\pi \approx 46.08\text{g}. \end{aligned}$$

One would expect different density functions to return different weights, as we have here. The density functions were chosen, though, to be similar: each gives a density of 1 at the origin and a density of 5 at the outside edge of the circle, as seen in Figure 17.18.

Notice how $x^2 + y^2 + 1 \le 2\sqrt{x^2+y^2} + 1$ over the circle; this results in less weight.

Plotting the density functions can be useful as our understanding of mass can be related to our understanding of volume under a surface. We interpreted $\iint_R f(x,y)\ dA$ as giving the volume under f over R; we can understand $\iint_R \delta(x,y)\ dA$ in the same way. The volume under δ over R is actually mass; by compressing the volume under δ onto the xy-plane, we get more mass in some areas than others – i.e., areas of greater density.

Knowing the mass of a lamina is one of several important measures. Another is the centre of mass, which we discuss next.

17.3.2 Centre of mass

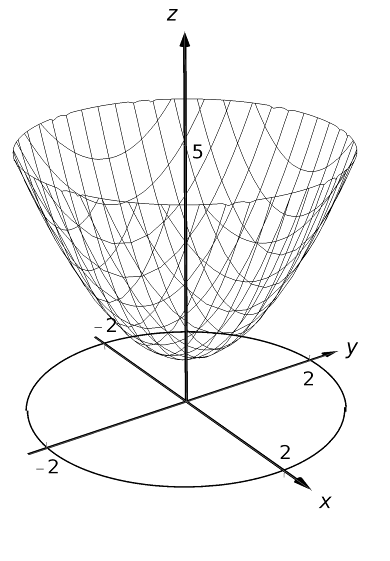
Consider a disk of radius 1 with uniform density. It is common knowledge that the disk will balance on a point if the point is placed at the centre of the disk. What if the disk does not have a uniform density? Through trial-and-error, we should still be able to find a spot on the disk at which the disk will balance on a point. This balance point is referred to as the **centre of mass** (*massamiddelpunt*), or **centre of gravity** (*zwaartepunt*). It is though all the mass is centred there. In fact, if the disk has a mass of 3kg, the disk will behave physically as though it were a point mass of 3kg located at its centre of mass. For instance, the disk will naturally spin with an axis through its centre of mass.

We find the centre of mass based on the principle of a weighted average. Consider a college class in which your homework average is 90%, your test average is 73%, and your final exam grade is an 85%. Experience tells us that our final grade is not the *average* of these three grades: that is, it is not:

$$\frac{0.9 + 0.73 + 0.85}{3} \approx 0.837 = 83.7\%.$$

That is, you are probably not pulling a B in the course. Rather, your grades are weighted. Let us say the homework is worth 10% of the grade, tests are 60% and the exam is 30%. Then your final grade is:

$$(0.1)(0.9) + (0.6)(0.73) + (0.3)(0.85) = 0.783 = 78.3\%.$$



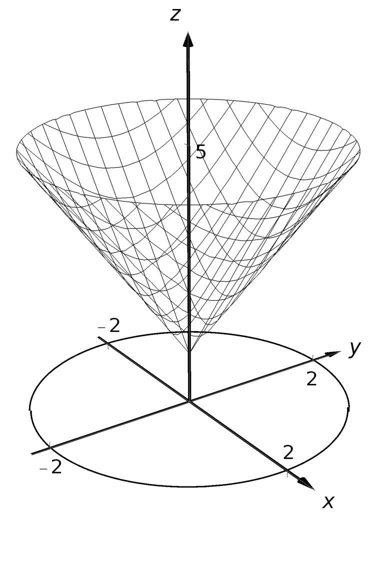


Figure 17.18: Graphing the density functions $\delta(x,y)=x^2+y^2+1$ (a) and $\delta(x,y)=2\sqrt{x^2+y^2}+1$ (b).

Each grade is multiplied by a weight.

In general, given values $x_1, x_2, \ldots, x_n$ and weights $w_1, w_2, \ldots, w_n$, the weighted average of the n-values is

$$\frac{\displaystyle\sum_{i=1}^{n} w_i x_i}{\displaystyle\sum_{i=1}^{n} w_i}.$$

How this relates to centre of mass is given in the following definition.

Definitie 17.2 (Centre of mass of a discrete linear system)
Let point masses $m_1, m_2, \ldots, m_n$ be distributed along the x-axis at locations $x_1, x_2, \ldots, x_n$, respectively. The **centre of mass** $\overline{x}$ of the system is located at

$$\overline{x} = \frac{\displaystyle\sum_{i=1}^{n} m_i x_i}{\displaystyle\sum_{i=1}^{n} m_i}.$$

In a discrete system (i.e., mass is located at individual points, not along a continuum) we find the centre of mass by dividing the mass into a **moment** (*moment*) of the system. In general, a moment is a weighted measure of distance from a particular point or line. In the case described by Definition 17.2, we are finding a weighted measure of distances from the y-axis, so we refer to this as **the moment about the y-axis** (*moment om de y-as*), represented by M_y. Letting M be the total mass of the system, we have $\overline{x} = M_y/M$.

We can extend the concept of the centre of mass of discrete points along a line to the centre of mass of discrete points in the plane rather easily. To do so, we define some terms then give a theorem.

Definitie 17.3 (Moments about the x- and y- axes)
Let point masses m_1, $m_2, \ldots, m_n$ be located at points (x_1, y_1), $(x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane.

1. The **moment about the x-axis**, M_x, is

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

2. The **moment about the y-axis**, M_y, is

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

We now define the centre of mass of discrete points in the plane.

Definitie 17.4 (Centre of mass of a discrete planar system)
Let point masses m_1, $m_2, \ldots, m_n$ be located at points (x_1, y_1), $(x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane, and let $M = \sum_{i=1}^{n} m_i$.

The **centre of mass** of the system is at $(\overline{x}, \overline{y})$, where

$$\overline{x} = \frac{M_y}{M} \quad \text{and} \quad \overline{y} = \frac{M_x}{M}.$$

Example 17.18

Let point masses of 1kg, 2kg and 5kg be located at points $(2,0)$, $(1,1)$ and $(3,1)$, respectively, and are connected by thin rods of negligible weight. Find the centre of mass of the system.

Solution

We follow Definitions 17.4 and 17.3 to find M, M_x and M_y:

$M = 1 + 2 + 5 = 8$kg.

$$\begin{aligned} M_x &= \sum_{i=1}^{n} m_i y_i \\ &= 1(0) + 2(1) + 5(1) \\ &= 7. \end{aligned} \qquad \begin{aligned} M_y &= \sum_{i=1}^{n} m_i x_i \\ &= 1(2) + 2(1) + 5(3) \\ &= 19. \end{aligned}$$

Thus the centre of mass is

$$(\overline{x}, \overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M} \right) = \left(\frac{19}{8}, \frac{7}{8} \right) = (2.375, 0.875),$$

illustrated in Figure 17.19.

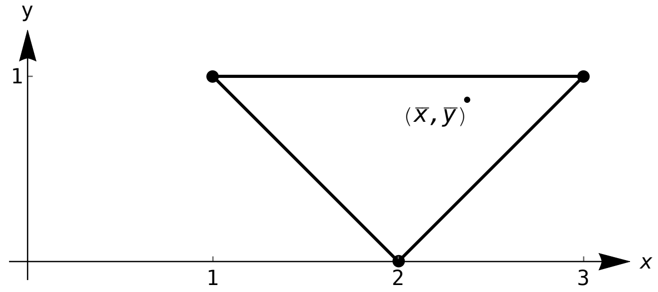


Figure 17.19: Illustrating the centre of mass of a discrete planar system in Example 17.18.

We finally arrive at our true goal of this section: finding the centre of mass of a lamina with variable density. While the above measurement of centre of mass is interesting, it does not directly answer more realistic situations where we need to find the centre of mass of a contiguous region. However, understanding the discrete case allows us to approximate the centre of mass of a planar lamina; using calculus, we can refine the approximation to an exact value.

We begin by representing a planar lamina with a region R in the xy-plane with density function $\delta(x, y)$. Partition R into n subdivisions, each with area ΔA_i. As done before, we can approximate the mass of the i^{th} subregion with $\delta(x_i, y_i)\Delta A_i$, where (x_i, y_i) is a point inside the i^{th} subregion. We can approximate the moment of this subregion about the y-axis with $x_i\delta(x_i, y_i)\Delta A_i$ – that is, by multiplying the approximate mass of the region by its approximate distance from the y-axis. Similarly, we can approximate the moment about the x-axis with $y_i\delta(x_i, y_i)\Delta A_i$. By summing over all subregions, we have:

$$\begin{aligned}
\text{mass: } M &\approx \sum_{i=1}^{n} \delta(x_i, y_i)\Delta A_i\,, \qquad \text{(as seen before)}\\
\text{moment about the } x\text{-axis: } M_x &\approx \sum_{i=1}^{n} y_i\delta(x_i, y_i)\Delta A_i\,,\\
\text{moment about the } y\text{-axis: } M_y &\approx \sum_{i=1}^{n} x_i\delta(x_i, y_i)\Delta A_i\,.
\end{aligned}$$

By taking limits, where size of each subregion shrinks to 0 in both the x- and y- directions, we arrive at the double integrals given in the following definition.

Definitie 17.5 (Centre of mass of a planar lamina)
Let a planar lamina be represented by a closed, bounded region R in the xy-plane with density function $\delta(x, y)$. Then we can infer the following information about the lamina:

1. The **mass** (*massa*) of a planar lamina is $M = \displaystyle\iint_R \delta(x, y)\, dA$.
2. The **moment about the x-axis** is $M_x = \displaystyle\iint_R y\delta(x, y)\, dA$.
3. The **moment about the y-axis** is $M_y = \displaystyle\iint_R x\delta(x, y)\, dA$.

4. The **centre of mass** (*massamiddelpunt*) of the object is
$$(\overline{x},\overline{y})=\left(\frac{M_y}{M},\frac{M_x}{M}\right).$$

We practice finding centres of mass by revisiting some of the lamina used previously in this section when finding mass. We will just set up the integrals needed to compute M, M_x and M_y and leave the details of the integration to the reader.

Example 17.19

Find the centre of mass of a square lamina, represented by the unit square with lower lefthand corner at the origin (see Figure 17.17), with variable density $\delta(x,y)=(x+y+2)$g/cm^2. This is the lamina from Example 17.16.

Solution

We follow Theorem 17.5, to find M, M_x and M_y:

$$M=\iint_R (x+y+2)\ dA=\int_0^1\int_0^1 (x+y+2)\ dx\ dy=3\text{g}.$$

$$M_x=\iint_R y(x+y+2)\ dA=\int_0^1\int_0^1 y(x+y+2)\ dx\ dy=\frac{19}{12}.$$

$$M_y=\iint_R x(x+y+2)\ dA=\int_0^1\int_0^1 x(x+y+2)\ dx\ dy=\frac{19}{12}.$$

Thus the centre of mass is

$$(\overline{x},\overline{y})=\left(\frac{M_y}{M},\frac{M_x}{M}\right)=\left(\frac{19}{36},\frac{19}{36}\right)\approx(0.528,0.528).$$

While the mass of this lamina is the same as the lamina in the previous example, the greater density found with greater x- and y-values pulls the centre of mass from the centre slightly towards the upper righthand corner.

Example 17.20

Find the centre of mass of the lamina represented by the circle with radius 2cm, centred at the origin, with density function $\delta(x,y)=(x^2+y^2+1)$g/cm^2. This is one of the lamina used in Example 17.17.

Solution

As done in Example 17.17, it is best to describe R using polar coordinates. Thus when we compute M_y, we will integrate not $x\delta(x,y)=x(x^2+y^2+1)$, but rather $\big(r\cos(\theta)\big)\delta(r\cos(\theta),r\sin(\theta))$ $=\big(r\cos(\theta)\big)\big(r^2+1\big)$. We compute M, M_x and M_y:

$$M=\int_0^{2\pi}\int_0^2 (r^2+1)r\ dr\ d\theta=12\pi\approx 37.7\text{g},$$

$$M_x = \int_0^{2\pi}\int_0^{2} (r\sin(\theta))(r^2+1)r\,dr\,d\theta = 0\,,$$

$$M_y = \int_0^{2\pi}\int_0^{2} (r\cos(\theta))(r^2+1)r\,dr\,d\theta = 0\,.$$

Since R and the density of R are both symmetric about the x- and y-axes, it should come as no big surprise that the moments about each axis is 0. Thus the centre of mass is $(\overline{x},\overline{y}) = (0,0)$.

17.4 Surface area

In Section 13.4 we used definite integrals to compute the arc length of plane curves of the form $y = f(x)$. We later extended these ideas to compute the arc length of plane curves defined by parametric or polar equations.

The natural extension of the concept of arc length over an interval to surfaces is surface area over a region. For that purpose, consider the surface $z = f(x,y)$ over a region R in the xy-plane, shown in Figure 17.20(a). Because of the domed shape of the surface, the surface area will be greater than that of the area of the region R. We can find this area using the same basic technique we have used over and over: we'll make an approximation, then using limits, we'll refine the approximation to the exact value.

As done to find the volume under a surface or the mass of a lamina, we subdivide R into n subregions. Here we subdivide R into rectangles, as shown in the figure. One such subregion is outlined in the figure, where the rectangle has dimensions Δx_i and Δy_i, along with its corresponding region on the surface.

In Figure 17.20(b), we zoom in on this portion of the surface. When Δx_i and Δy_i are small, the function is approximated well by the tangent plane at any point (x_i, y_i) in this subregion, which is graphed in part (b). In fact, the tangent plane approximates the function so well that in this figure, it is virtually indistinguishable from the surface itself! Therefore we can approximate the surface area S_i of this region of the surface with the area T_i of the corresponding portion of the tangent plane.

This portion of the tangent plane is a parallelogram, defined by sides $\vec{u}$ and $\vec{v}$, as shown. One of the applications of the cross product from Section 6.6 is that the area of this parallelogram is $\|\vec{u}\times\vec{v}\|$. So, once we can determine $\vec{u}$ and $\vec{v}$, we can determine the area.

$\vec{u}$ is tangent to the surface in the direction of x, therefore, from Section 16.7, $\vec{u}$ is parallel to $\big\langle 1, 0, f_x(x_i,y_i)\big\rangle$. The x-displacement of $\vec{u}$ is Δx_i, so we know that $\vec{u} = \Delta x_i\big\langle 1,0,f_x(x_i,y_i)\big\rangle$. Similar logic shows that $\vec{v} = \Delta y_i\big\langle 0,1,f_y(x_i,y_i)\big\rangle$.Thus:

$$\begin{aligned}\text{surface area } S_i &\approx \text{area of } T_i\\ &= \|\vec{u}\times\vec{v}\|\\ &= \Big\|\Delta x_i\big\langle 1,0,f_x(x_i,y_i)\big\rangle \times \Delta y_i\big\langle 0,1,f_y(x_i,y_i)\big\rangle\Big\|\\ &= \sqrt{1+f_x(x_i,y_i)^2+f_y(x_i,y_i)^2}\,\Delta x_i\Delta y_i.\end{aligned}$$

Note that $\Delta x_i\Delta y_i = \Delta A_i$, the area of the i^{th} subregion.

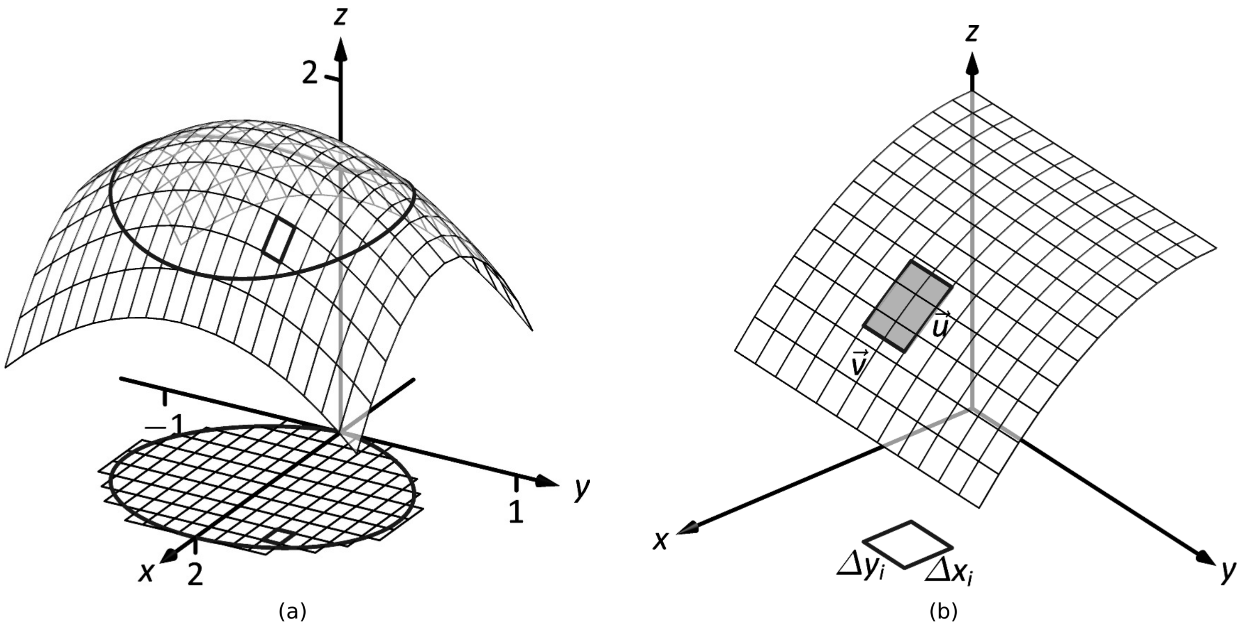


Figure 17.20: Developing a method of computing surface area.

Summing up all n of the approximations to the surface area gives

$$\text{surface area over } R \approx \sum_{i=1}^{n} \sqrt{1+f_x(x_i,y_i)^2+f_y(x_i,y_i)^2}\Delta A_i.$$

Once again take a limit as all of the Δx_i and Δy_i shrink to 0; this leads to a double integral:

$$SA = \iint_R \sqrt{1+f_x(x,y)^2+f_y(x,y)^2}\ dA. \qquad (17.5)$$

We use this definition to compute surface areas of known surfaces.

Example 17.21

Find the surface area of the sphere with radius a centred at the origin, whose top hemisphere has equation $f(x,y)=\sqrt{a^2-x^2-y^2}$.

Solution

We start by computing partial derivatives and find

$$f_x(x,y)=\frac{-x}{\sqrt{a^2-x^2-y^2}} \quad \text{and} \quad f_y(x,y)=\frac{-y}{\sqrt{a^2-x^2-y^2}}.$$

As our function f only defines the top upper hemisphere of the sphere, we double our surface area result to get the total area:

$$\begin{aligned} SA &= 2\iint_R \sqrt{1+f_x(x,y)^2+f_y(x,y)^2}\ dA \\ &= 2\iint_R \sqrt{1+\frac{x^2+y^2}{a^2-x^2-y^2}}\ dA. \end{aligned}$$

The region R that we are integrating over is bounded by the circle, centered at the origin, with radius a: $x^2+y^2=a^2$. Because of this region, we are likely to have greater success with our integration by converting to polar coordinates. Using the substitutions $x = r\cos(\theta)$, $y = r\sin(\theta)$, $dA = r\,dr\,d\theta$ and bounds $0 \le \theta \le 2\pi$ and $0 \le r \le a$, we have:

$$\begin{aligned} SA &= 2\int_0^{2\pi}\int_0^a \sqrt{1+\frac{r^2\cos^2(\theta)+r^2\sin^2(\theta)}{a^2-r^2\cos^2(\theta)-r^2\sin^2(\theta)}}\; r\,dr\,d\theta \\ &= 2\int_0^{2\pi}\int_0^a r\sqrt{1+\frac{r^2}{a^2-r^2}}\; dr\,d\theta \\ &= 2\int_0^{2\pi}\int_0^a r\sqrt{\frac{a^2}{a^2-r^2}}\; dr\,d\theta. \qquad (17.6) \end{aligned}$$

Apply substitution $u = a^2 - r^2$ and integrate the inner integral, giving

$$\begin{aligned} &= 2\int_0^{2\pi} a^2\; d\theta \\ &= 4\pi a^2. \end{aligned}$$

Our work confirms the known formula.

Note that the inner integral in Equation (17.6) is an improper integral, as its is not defined at $r = a$. To properly evaluate this integral, one must use the techniques of Section 12.5. Since the resulting improper integral does converge, the surface area is accurately computed.

Example 17.22

Find the area of the surface $f(x,y) = x^2 - 3y + 3$ over the region R bounded by $-x \le y \le x$, $0 \le x \le 4$, as pictured in Figure 17.21.

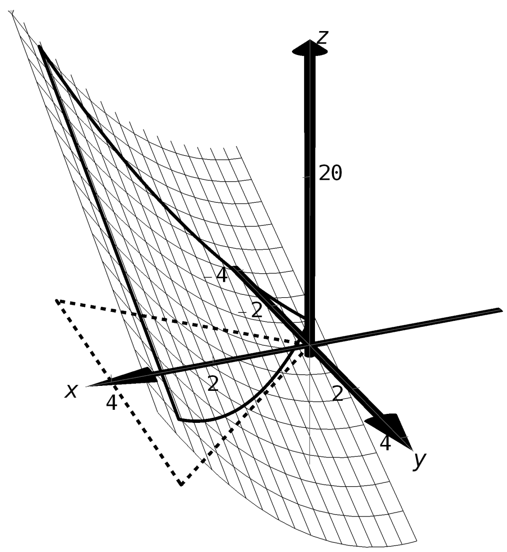


Figure 17.21: Graphing the surface in Example 17.22.

Solution

It is straightforward to compute $f_x(x,y) = 2x$ and $f_y(x,y) = -3$. Thus the surface area is described by the double integral

$$\iint_R \sqrt{1+(2x)^2+(-3)^2}\ dA = \iint_R \sqrt{10+4x^2}\ dA.$$

As with integrals describing arc length, double integrals describing surface area are in general hard to evaluate directly because of the square root. This particular integral can be easily evaluated, though, with judicious choice of our order of integration.

Integrating with order $dx\ dy$ requires us to evaluate $\int \sqrt{10+4x^2}\ dx$. This can be done, though it involves the goniometric substitution $2x = \sqrt{10}\tan(t)$. Integrating with order $dy\ dx$ has as its first integral $\int \sqrt{10+4x^2}\ dy$, which is easy to evaluate: it is simply $y\sqrt{10+4x^2}+C$. So we proceed with the order $dy\ dx$.

$$\begin{aligned} SA &= \iint_R \sqrt{10+4x^2}\ dA \\ &= \int_0^4 \int_{-x}^{x} \sqrt{10+4x^2}\ dy\ dx \\ &= \int_0^4 \left(y\sqrt{10+4x^2}\right)\Big|_{-x}^{x}\ dx \\ &= \int_0^4 2x\sqrt{10+4x^2}\ dx \end{aligned}$$

Apply substitution with $u = 10+4x^2$:

$$\begin{aligned} SA &= \left(\frac{1}{6}\left(10+4x^2\right)^{3/2}\right)\Big|_0^4 \\ &= \frac{1}{3}\left(37\sqrt{74}-5\sqrt{10}\right) \approx 100.825 \text{ units}^2. \end{aligned}$$

So while the region R over which we integrate has an area of 16 units^2, the surface has a much greater area as its z-values change dramatically over R.

In practice, technology helps greatly in the evaluation of such integrals. High powered computer algebra systems can compute integrals that are difficult, or at least time consuming, by hand, and can at least produce very accurate approximations with numerical methods. In general, just knowing how to set up the proper integrals brings one very close to being able to compute the needed value. Most of the work is actually done in just describing the region R in terms of polar or rectangular coordinates. Once this is done, technology can usually provide a good answer.

17.5 Triple integration

17.5.1 Volume between surfaces

We learned in Section 17.2 how to compute the signed volume V under a surface $z = f(x,y)$ over a region R. It follows that if $f(x,y) \geq g(x,y)$ on R, then the volume between $f(x,y)$ and $g(x,y)$ on R is

$$V = \iint_R f(x,y)\, dA - \iint_R g(x,y)\, dA = \iint_R \big(f(x,y) - g(x,y)\big)\, dA.$$

Consider, for instance, the volume of the space region bounded by the planes $z = 3x + y - 4$, $z = 8 - 3x - 2y$, $x = 0$ and $y = 0$. In Figure 17.22(a) the planes are drawn; in Figure 17.22(b), only the defined region is given. The region R over which we will integrate is bounded by $x = 0$, $y = 0$ and the line of intersection of the planes $z = 3x + y - 4$ and $z = 8 - 3x - 2y$, which is $y = 4 - 2x$. So, we find the volume by evaluating the integral

$$\int_0^2 \int_0^{4-2x} \big(8 - 3x - 2y - (3x + y - 4)\big)\, dy\, dx = 16.$$

Note how we can rewrite the integrand as an integral, much as we did in Section 17.1:

$$8 - 3x - 2y - (3x + y - 4) = \int_{3x+y-4}^{8-3x-2y} dz.$$

Thus we can rewrite the double integral that finds volume as

$$\int_0^2 \int_0^{4-2x} \big(8 - 3x - 2y - (3x + y - 4)\big)\, dy\, dx = \int_0^2 \int_0^{4-2x} \left(\int_{3x+y-4}^{8-3x-2y} dz \right) dy\, dx.$$

This no longer looks like a double integral, but more like a triple integral.

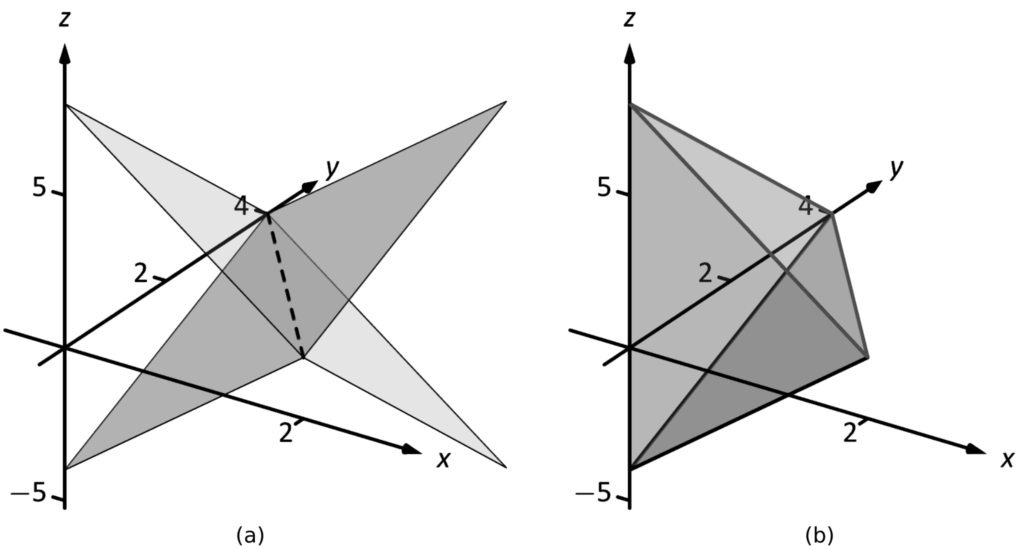


Figure 17.22: Finding the volume between $z = 3x + y - 4$, $z = 8 - 3x - 2y$, $x = 0$ and $y = 0$.

To formally find the volume of a closed, bounded region D in space, such as the one shown in Figure 17.23(a), we start with an approximation. Break D into n rectangular solids; the solids near the boundary of D may possibly not include portions of D and/or include extra space. In Figure 17.23(b), we zoom in on a portion of the boundary of D to show a rectangular solid that contains space not in D; as this is an approximation of the volume, this is acceptable and this error will be reduced as we shrink the size of our solids.

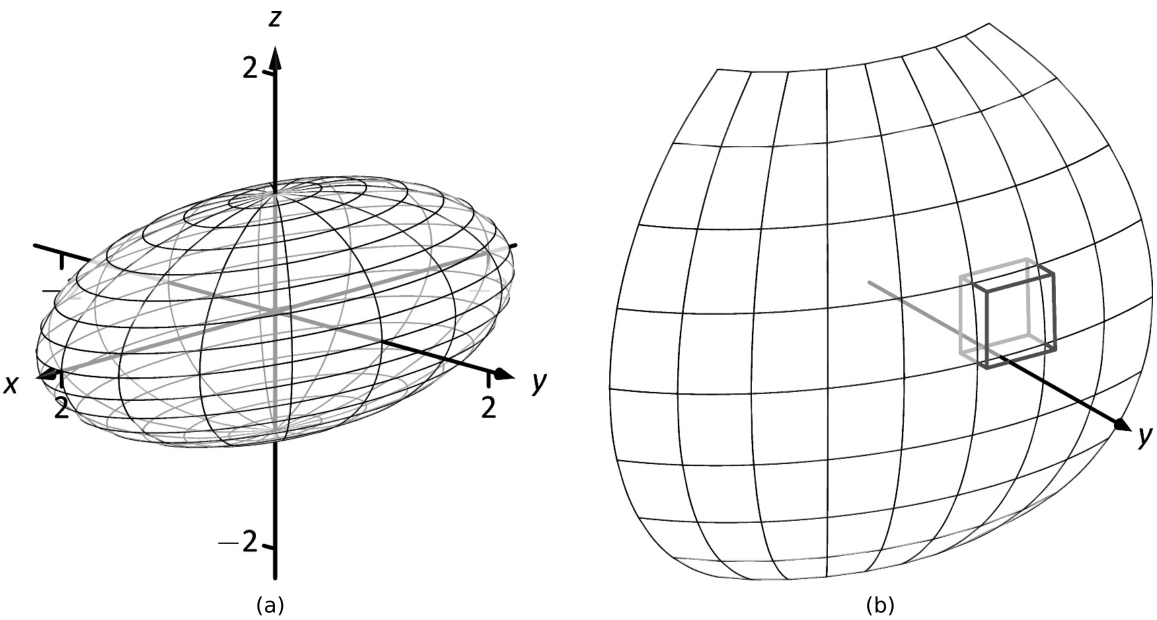


Figure 17.23: Approximating the volume of a region D in space.

The volume ΔV_i of the i^{th} solid D_i is $\Delta V_i = \Delta x_i \Delta y_i \Delta z_i$, where Δx_i, Δy_i and Δz_i give the dimensions of the rectangular solid in the x-, y- and z- directions, respectively. By summing up the volumes of all n solids, we get an approximation of the volume V of D:

$$V \approx \sum_{i=1}^{n} \Delta V_i = \sum_{i=1}^{n} \Delta x_i \Delta y_i \Delta z_i.$$

Let $\mathcal{D}$ represent the length of the longest diagonal of rectangular solids in the subdivision of D. As $\mathcal{D} \to 0$, the volume of each solid goes to 0, as do each of Δx_i, Δy_i and Δz_i, for all i. Our calculus experience tells us that taking a limit as $\mathcal{D} \to 0$ turns our approximation of V into an exact calculation of V. Before we state this result in a theorem, we use a definition to define some terms.

Definitie 17.6 (Triple integrals, iterated integration (Part I))
Let D be a closed, bounded region in space. Let a and b be real numbers, let $g_1(x)$ and $g_2(x)$ be continuous functions of x, and let $f_1(x, y)$ and $f_2(x, y)$ be continuous functions of x and y.

1. The volume V of D is denoted by a **triple integral** (*drievoudige integraal*)

$$V = \iiint_D dV.$$

2. The iterated integral

$$\int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\, dy\, dx$$

is evaluated as

$$\int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\ dy\ dx = \int_a^b \int_{g_1(x)}^{g_2(x)} \left(\int_{f_1(x,y)}^{f_2(x,y)} dz \right) dy\ dx.$$

Our informal understanding of the notation $\iiint_D \ dV$ is sum up lots of little volumes over D, analogous to our understanding of $\iint_R \ dA$ and $\iint_R \ dm$.

We now state the major theorem of this section.

Theorem 17.4 (Triple integration (Part I))

Let D be a closed, bounded region in space and let ΔD be any subdivision of D into n rectangular solids, where the i^{th} subregion D_i has dimensions $\Delta x_i \times \Delta y_i \times \Delta z_i$ and volume ΔV_i.

1. *The volume V of D is*

$$V = \iiint_D \ dV = \lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} \Delta V_i = \lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} \Delta x_i \Delta y_i \Delta z_i.$$

2. *If D is defined as the region bounded by the planes $x = a$ and $x = b$, the cylinders $y = g_1(x)$ and $y = g_2(x)$, and the surfaces $z = f_1(x, y)$ and $z = f_2(x, y)$, where $a < b$, $g_1(x) \le g_2(x)$ and $f_1(x, y) \le f_2(x, y)$ on D, then*

$$\iiint_D \ dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\ dy\ dx.$$

3. *V can be determined using iterated integration with other orders of integration (there are 6 total), as long as D is defined by the region enclosed by a pair of planes, a pair of cylinders, and a pair of surfaces.*

We evaluated the area of a plane region R by iterated integration, where the bounds were from curve to curve, then from point to point. Theorem 17.4 allows us to find the volume of a space region with an iterated integral with bounds from surface to surface, then from curve to curve, then from point to point. In the iterated integral

$$\int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} dz\ dy\ dx,$$

the bounds $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$ define a region R in the xy-plane over which the region D exists in space. However, these bounds are also defining surfaces in space; $x = a$ is a plane and $y = g_1(x)$ is a cylinder. The combination of these six surfaces enclose, and define, D.

Examples will help us understand triple integration, including integrating with various orders of integration.

Example 17.23

Find the volume of the space region in the first octant bounded by the plane $z = 2 - y/3 - 2x/3$, shown in Figure 17.24(a), using the order of integration $dz\ dy\ dx$. Set up the triple integrals that give the volume in the other five orders of integration.

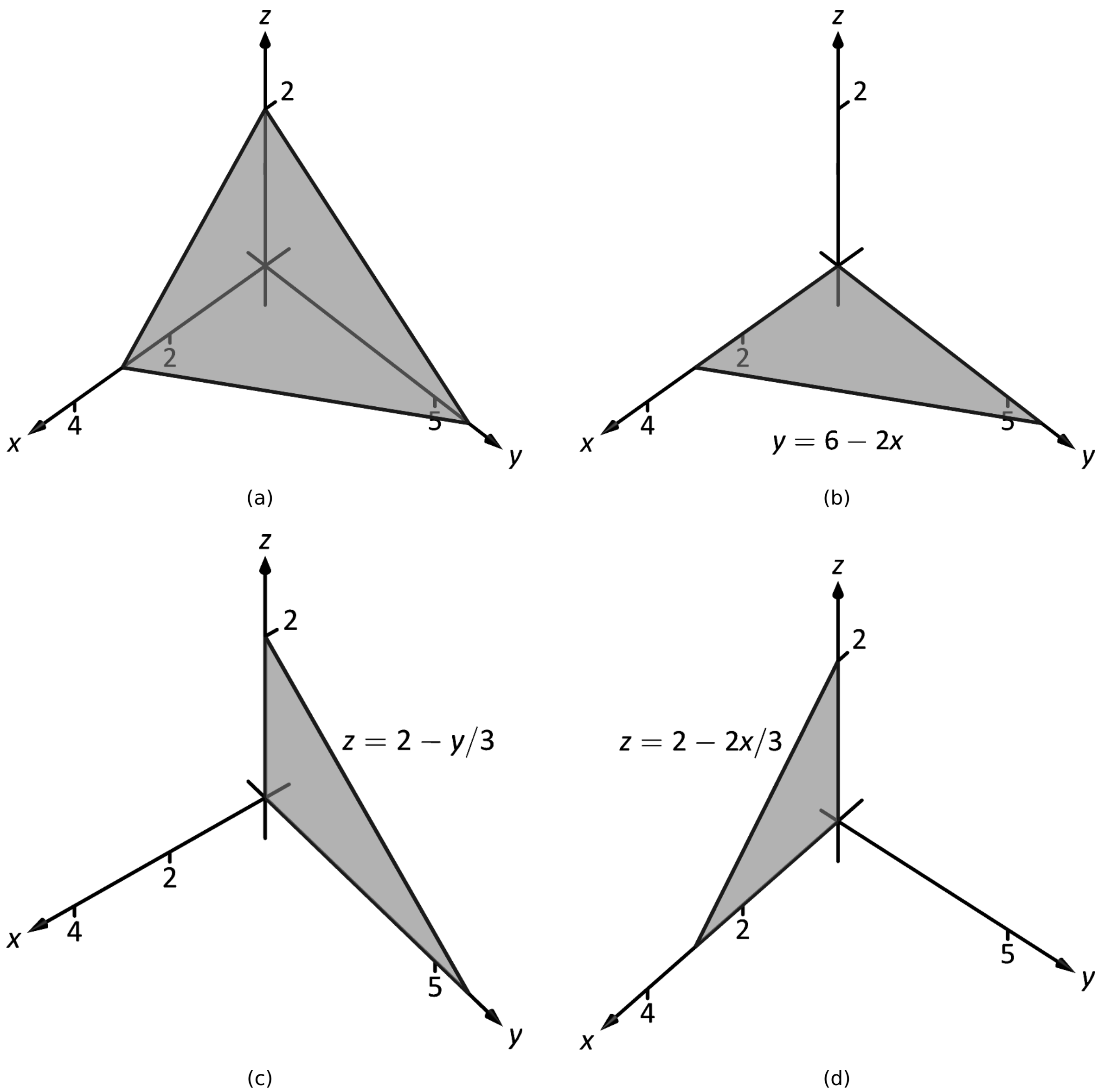


Figure 17.24: The region D used in Example 17.23 in (a); the region found by projecting D onto the xy- (b), yz- (c) and xz- (d) plane.

Solution

Starting with the order of integration $dz\ dy\ dx$, we need to first find bounds on z. The region D is bounded below by the plane $z=0$ because we are restricted to the first octant and above by $z=2-y/3-2x/3$; $0 \leq z \leq 2-y/3-2x/3$.

To find the bounds on y and x, we project the region onto the xy-plane, giving the triangle shown in Figure 17.24(b). We define that region R, in the integration order of $dy\ dx$, with bounds $0 \leq y \leq 6-2x$ and $0 \leq x \leq 3$. Thus the volume V of the region D is:

$$\begin{aligned} V &= \iiint_D dV \\ &= \int_0^3 \int_0^{6-2x} \int_0^{2-y/3-2x/3} dz\ dy\ dx \end{aligned}$$

$$= \int_0^3 \int_0^{6-2x} \left(\int_0^{2-y/3-2x/3} dz \right) dy\, dx$$

$$= \int_0^3 \int_0^{6-2x} (z)\Big|_0^{2-y/3-2x/3} dy\, dx$$

$$= \int_0^3 \int_0^{6-2x} \left(2 - \frac{1}{3}y - \frac{2}{3}x\right) dy\, dx.$$

From this step on, we are evaluating a double integral as done many times before. We skip these steps and give the final volume

$$V = 6 \text{ units}^3.$$

The order $dz\, dx\, dy$:

Now consider the volume using the order of integration $dz\, dx\, dy$. The bounds on z are the same as before, $0 \le z \le 2 - y/3 - 2x/3$. Projecting the space region on the xy-plane as shown in Figure 17.24(b), we now describe this triangle with the order of integration $dx\, dy$. This gives bounds $0 \le x \le 3 - y/2$ and $0 \le y \le 6$. Thus the volume is given by the triple integral

$$V = \int_0^6 \int_0^{3-y/2} \int_0^{2-y/3-2x/3} dz\, dx\, dy.$$

The order $dx\, dy\, dz$:

Following our surface to surface strategy, we need to determine the x-surfaces that bound our space region. To do so, approach the region from behind, in the direction of increasing x. The first surface we hit as we enter the region is the yz-plane, defined by $x = 0$. We come out of the region at the plane $z = 2 - y/3 - 2x/3$; solving for x, we have $x = 3 - y/2 - 3z/2$. Thus the bounds on x are: $0 \le x \le 3 - y/2 - 3z/2$.

Now project the space region onto the yz-plane, as shown in Figure 17.24(c). We need to find bounds on this region with the order $dy\, dz$. The curves that bound y are $y = 0$ and $y = 6 - 3z$; the *points* that bound z are 0 and 2. Thus the triple integral giving volume is:

$$V = \int_0^2 \int_0^{6-3z} \int_0^{3-y/2-3z/2} dx\, dy\, dz.$$

The order $dx\, dz\, dy$:

The x-bounds are the same as the order above. We now consider the triangle in Figure 17.24(c) and describe it with the order $dz\, dy$: $0 \le z \le 2 - y/3$ and $0 \le y \le 6$. Thus the volume is given by:

$$V = \int_0^6 \int_0^{2-y/3} \int_0^{3-y/2-3z/2} dx\, dz\, dy.$$

The order $dy\, dz\, dx$:

We now need to determine the y-surfaces that determine our region. Approaching the space region from behind and moving in the direction of increasing y, we first enter the region at $y=0$, and exit along the plane $z=2-y/3-2x/3$. Solving for y, this plane has equation $y=6-2x-3z$. Thus y has bounds $0\le y\le 6-2x-3z$.

Now project the region onto the xz-plane, as shown in Figure 17.24(d). The curves bounding this triangle are $z=0$ and $z=2-2x/3$; x is bounded by the points $x=0$ to $x=3$. Thus the triple integral giving volume is:

$$V=\int_0^3\int_0^{2-2x/3}\int_0^{6-2x-3z} dy\,dz\,dx.$$

The order $dy\,dx\,dz$:

The y-bounds are the same as in the order above. We now determine the bounds of the triangle in Figure 17.24(d) using the order $dy\,dx\,dz$. x is bounded by $x=0$ and $x=3-3z/2$; z is bounded between $z=0$ and $z=2$. This leads to the triple integral:

$$V=\int_0^2\int_0^{3-3z/2}\int_0^{6-2x-3z} dy\,dx\,dz.$$

This problem was long, but hopefully useful, demonstrating how to determine bounds with every order of integration to describe the region D. In practice, we only need one, but being able to do them all gives us flexibility to choose the order that suits us best.

In the previous example, we collapsed the surface into the xy-, xz-, and yz- planes as we determined the curve to curve, point to point bounds of integration. Since the surface was a triangular portion of a plane, this projecting was simple, but this of course is not always the case.

Example 17.24

Set up the triple integrals that find the volume of the space region D bounded by the surfaces $x^2+y^2=1$, $z=0$ and $z=-y$, as shown in Figure 17.25(a), with the order of integration $dz\,dy\,dx$.

Solution

The region D is bounded below by the plane $z=0$ and above by the plane $z=-y$. The cylinder $x^2+y^2=1$ does not offer any bounds in the z-direction, as that surface is parallel to the z-axis. Thus $0\le z\le -y$.

Projecting the region into the xy-plane, we get part of the disk bounded by the circle with equation $x^2+y^2=1$ as shown in Figure 17.25(b). As a function of x, this half circle has equation $y=-\sqrt{1-x^2}$. Thus y is bounded below by $-\sqrt{1-x^2}$ and above by $y=0$: $-\sqrt{1-x^2}\le y\le 0$. The x-bounds of the half circle are $-1\le x\le 1$. All together, the bounds of integration and triple integral are as follows:

$$V=\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\int_0^{-y} dz\,dy\,dx.$$

We evaluate this triple integral:

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}\int_0^{-y} dz\,dy\,dx=\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0}(z)\Big|_0^{-y}\,dy\,dx$$

$$= \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{0} (-y)\, dy\, dx$$

$$= \int_{-1}^{1} \left(-\frac{1}{2} y^2 \right) \Bigg|_{-\sqrt{1-x^2}}^{0} dx$$

$$= \int_{-1}^{1} \frac{1}{2}(1-x^2)\, dx$$

$$= \frac{1}{2} \left(x - \frac{1}{3} x^3 \right) \Bigg|_{-1}^{1}$$

$$= \frac{2}{3} \text{ units}^3.$$

The careful reader might have noticed that the region D is symmetric with respect to the yz-planes so that also its projection onto the xy-plane is symmetric with respect to the y-axis. Exploiting this symmetry allows us to compute the volume alternatively as

$$V = 2 \int_{0}^{1} \int_{-\sqrt{1-x^2}}^{0} \int_{0}^{-y} dz\, dy\, dx.$$

We leave it up to the reader to verify that sticking to the orders of integration $dy\ dx\ dz$ and $dx\ dz\ dy$ requires setting up more complicated integrals.

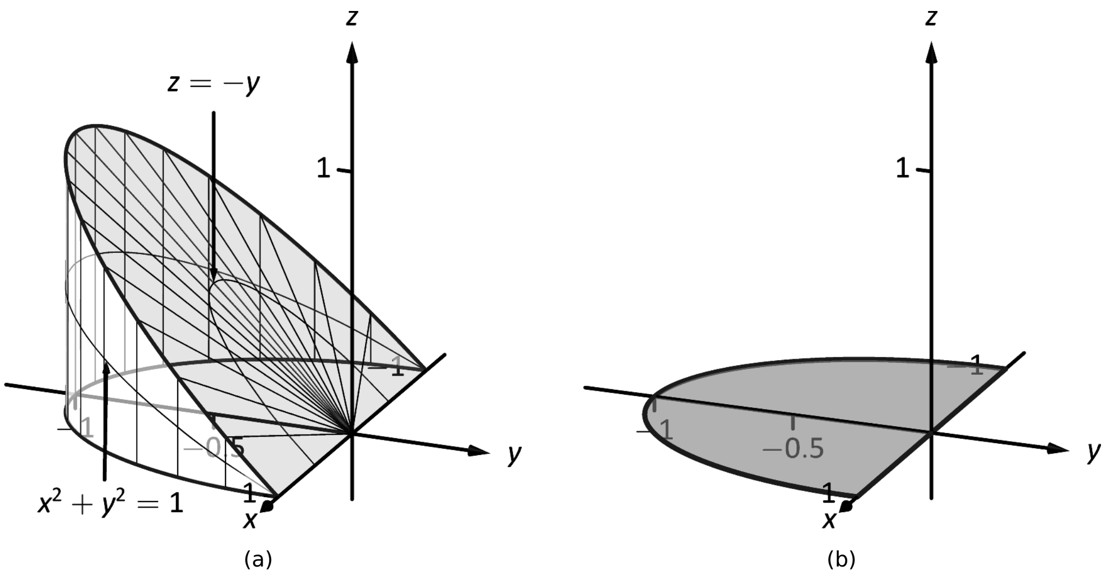


Figure 17.25: The region D used in Example 17.24 in (a); the region found by projecting D onto the xy-plane (b).

The following theorem states two things that should make common sense to us. First, using the triple integral to find volume of a region D should always return a positive number; we are computing volume here, not signed volume. Secondly, to compute the volume of a complicated region, we could break it up into subregions and compute the volumes of each subregion separately, summing them later to find the total volume.

Theorem 17.5 (Properties of triple integrals)
Let D be a closed, bounded region in space, and let D_1 and D_2 be non-overlapping regions such that $D = D_1 \cup D_2$.

1. $\displaystyle\iiint_D dV \geq 0$
2. $\displaystyle\iiint_D dV = \iiint_{D_1} dV + \iiint_{D_2} dV.$

Example 17.25

Set up a triple integral that gives the volume of the space region D bounded by $z = 6 - 2x^2 - y^2$ and $z = 2x^2 + 2$. These surfaces are plotted in Figure 17.26(a) and (b), respectively; the region D is shown in part (c) of the figure.

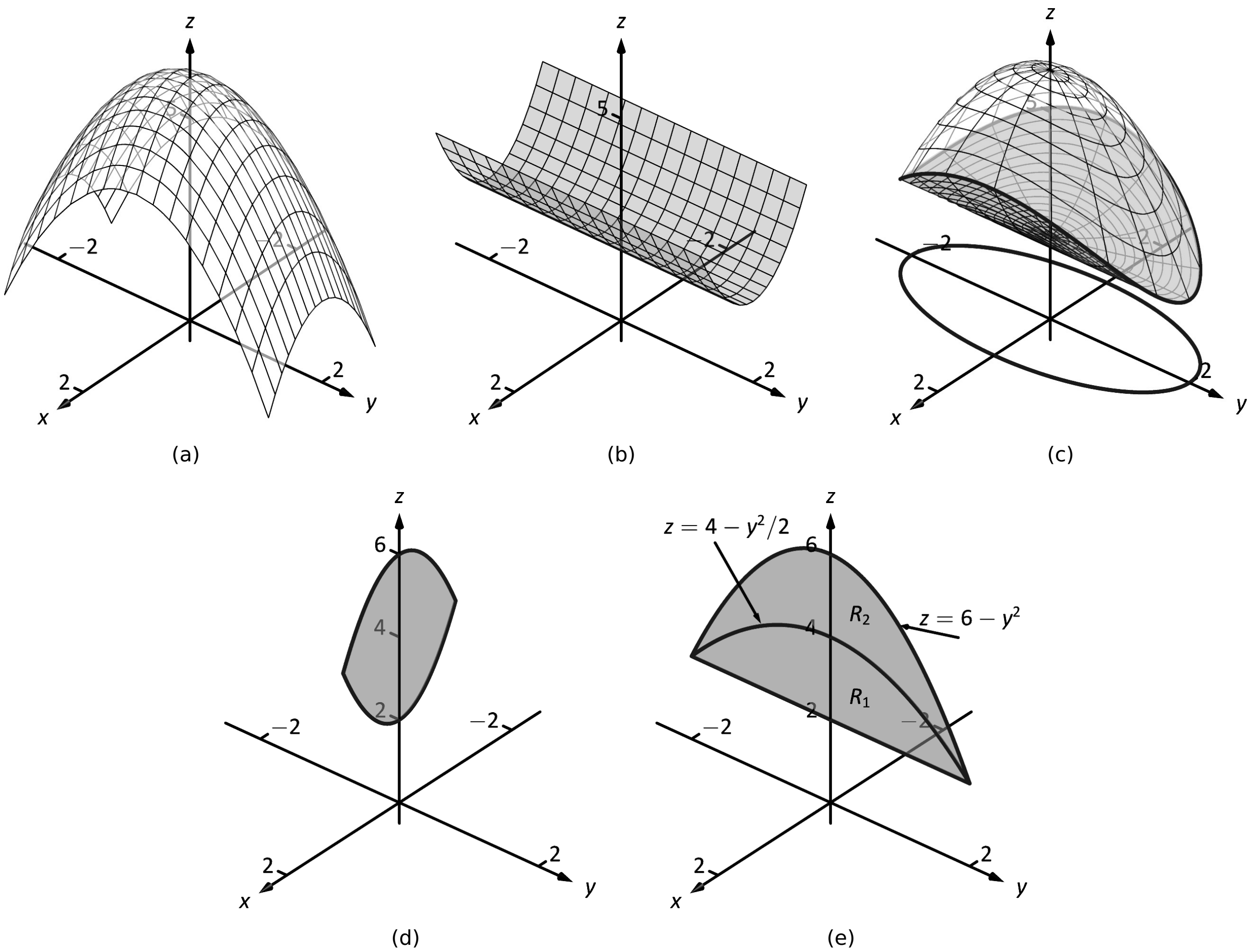


Figure 17.26: The region D in Example 17.25 is bounded by the surfaces shown in (a) and (b); D is shown in (c).

Solution

The main point of this example is this: integrating with respect to z first is rather straightforward; integrating with respect to x first is not.

The bounds on z are clearly $2x^2 + 2 \leq z \leq 6 - 2x^2 - y^2$. Projecting D onto the xy-plane gives the ellipse shown in Figure 17.26(c). The equation of this ellipse is found by setting the two surfaces equal to each other:

$$2x^2+2=6-2x^2-y^2 \quad \Leftrightarrow \quad 4x^2+y^2=4 \quad \Leftrightarrow \quad x^2+\frac{y^2}{4}=1.$$

We can describe this ellipse with the bounds

$$-\sqrt{4-4x^2} \leq y \leq \sqrt{4-4x^2} \quad \text{and} \quad -1 \leq x \leq 1.$$

Thus the triple integral giving volume is:

$$V=\int_{-1}^{1}\int_{-\sqrt{4-4x^2}}^{\sqrt{4-4x^2}}\int_{2x^2+2}^{6-2x^2-y^2} dz\ dy\ dx.$$

We leave it to the reader to confirm that sticking to the orders of integration *dy dz dx* and *dx dz dy* requires setting up more complicated integrals.

If all one wanted to do in Example 17.25 was find the volume of the region D, one would have likely stopped at the first integration setup (with order *dz dy dx*) and computed the volume from there. However, we included the other two methods 1) to show that it could be done, messy or not, and 2) because sometimes we have to use a less desirable order of integration in order to actually integrate.

17.5.2 Functions of three variables

There are uses for triple integration beyond merely finding volume, just as there are uses for integration beyond area under the curve. These uses start with understanding how to integrate functions of three variables, which is effectively no different than integrating functions of two variables. This leads us to a definition.

Definitie 17.7 (Iterated integration, (Part II))
Let D be a closed, bounded region in space, over which $g_1(x)$, $g_2(x)$, $f_1(x,y)$, $f_2(x,y)$ and $h(x,y,z)$ are all continuous, and let a and b be real numbers. Then, we may write

$$\int_a^b\int_{g_1(x)}^{g_2(x)}\int_{f_1(x,y)}^{f_2(x,y)} h(x,y,z)\ dz\ dy\ dx = \int_a^b\int_{g_1(x)}^{g_2(x)}\left(\int_{f_1(x,y)}^{f_2(x,y)} h(x,y,z)\ dz\right) dy\ dx.$$

But what does a triple integral of a function of three variables mean? We build up this understanding in a way very similar to how we have understood integration and double integration.

Let $h(x,y,z)$ be a continuous function of three variables, defined over some space region D. We can partition D into n rectangular–solid subregions, each with dimensions $\Delta x_i \times \Delta y_i \times \Delta z_i$. Let (x_i,y_i,z_i) be some point in the i^{th} subregion, and consider the product $h(x_i,y_i,z_i)\Delta x_i\Delta y_i\Delta z_i$. It is the product of a function value (that is the $h(x_i,y_i,z_i)$ part) and a small volume ΔV_i (that's the $\Delta x_i\Delta y_i\Delta z_i$ part). One of the simplest understanding of this type of product is when h describes the density of an object, for then $h \times \text{volume} = \text{mass}$.

We can sum up all n products over D. Again letting $\mathcal{D}$ represent the length of the longest diagonal of the n rectangular solids in the partition, we can take the limit of the sums of products as $\mathcal{D} \to 0$. That is, we can find

$$S = \lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta V_i = \lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta x_i \Delta y_i \Delta z_i.$$

While this limit has lots of interpretations depending on the function h, in the case where h describes density, S is the total mass of the object described by the region D.

We now use the above limit to define the triple integral, give a theorem that relates triple integrals to iterated iteration, followed by the application of triple integrals to find the centres of mass of solid objects.

Definitie 17.8 (Triple integral)
Let $w = h(x, y, z)$ be a continuous function over a closed, bounded region D in space, and consider any partition of D into n rectangular solids with volume ΔV_i. The triple integral of h over D is

$$\iiint_D h(x, y, z)\ dV = \lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta V_i.$$

The following theorem assures us that the above limit exists for continuous functions h and gives us a method of evaluating the limit.

Theorem 17.6 (Triple integration (Part II))
Let $w = h(x, y, z)$ be a continuous function over a closed, bounded region D in space, and let ΔD be any partition of D into n rectangular solids with volume V_i.

1. *The limit $\lim_{\mathcal{D}\to 0} \sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta V_i$ exists.*
2. *If D is defined as the region bounded by the planes $x = a$ and $x = b$, the cylinders $y = g_1(x)$ and $y = g_2(x)$, and the surfaces $z = f_1(x, y)$ and $z = f_2(x, y)$, where $a < b$, $g_1(x) \leq g_2(x)$ and $f_1(x, y) \leq f_2(x, y)$ on D, then*

$$\iiint_D h(x, y, z)\ dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{f_1(x,y)}^{f_2(x,y)} h(x, y, z)\ dz\ dy\ dx.$$

Actually, we can view the sum

$$\sum_{i=1}^{n} h(x_i, y_i, z_i)\Delta x_i \Delta y_i \Delta z_i$$

as a triple sum,

$$\sum_{k=1}^{p} \sum_{j=1}^{n} \sum_{i=1}^{m} h(x_i, y_j, z_k)\Delta x_i \Delta y_j \Delta z_k,$$

which we evaluate as

$$\sum_{k=1}^{p} \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} h(x_i, y_j, z_k)\Delta x_i \right) \Delta y_j \right) \Delta z_k.$$

Here we fix a k-value, which establishes the z-height of the rectangular solids on one level of all the rectangular solids in the space region D. The inner double summation adds up all the volumes of the rectangular solids on this level, while the outer summation adds up the volumes of each level.

This triple summation understanding leads to the $\iiint_D$ notation of the triple integral, as well as the method of evaluation shown in Theorem 17.6.

We now apply triple integration to find the centres of mass of solid objects.

17.5.3 Mass and centre of mass

One may wish to review Section 17.3 for a reminder of the relevant terms and concepts.

Definitie 17.9 (Mass, centre of mass of solids)
Let a solid be represented by a closed, bounded region D in space with variable density function $\delta(x,y,z)$.

1. The **mass** (*massa*) of the object is $M = \iiint_D \, dm = \iiint_D \delta(x,y,z)\, dV$.
2. The **moment about the** yz**-plane** is $M_{yz} = \iiint_D x\delta(x,y,z)\, dV$.
3. The **moment about the** xz**-plane** is $M_{xz} = \iiint_D y\delta(x,y,z)\, dV$.
4. The **moment about the** xy**-plane** is $M_{xy} = \iiint_D z\delta(x,y,z)\, dV$.
5. The **centre of mass** (*massamiddelpunt*) of the object is

$$(\overline{x},\overline{y},\overline{z}) = \left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right).$$

Example 17.26

Find the centre of mass of the solid represented by the region bounded by the planes $z=0$ and $z=-y$ and the cylinder $x^2+y^2=1$, shown in Figure 17.25(a), with density function $\delta(x,y,z)=10+x^2+5y-5z$. This space region was used in Example 17.24.

Solution

As we start, consider the density function. It is symmetric about the yz- plane, and the farther one moves from this plane, the denser the object is. The symmetry indicates that $\overline{x}$ should be 0.

As one moves away from the origin in the y- or z-directions, the object becomes less dense, though there is more volume in these regions.

Though none of the integrals needed to compute the centre of mass are particularly hard, they do require a number of steps. We emphasize here the importance of knowing how to set up the proper integrals; in complex situations we can appeal to technology for a good approximation, if not the exact answer. We use the order of integration $dz\ dy\ dx$, using the bounds found in Example 17.24. As these are the same for all four triple integrals, we explicitly show the bounds only for M:

$$\begin{aligned} M &= \iiint_D \left(10+x^2+5y-5z\right) dV \\ &= \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{0} \int_{0}^{-y} \left(10+x^2+5y-5z\right) dz\, dy\, dx \end{aligned}$$

$$= \frac{64}{5} - \frac{15\pi}{16} \approx 3.855,$$

$$M_{yz} = \iiint_D x\left(10 + x^2 + 5y - 5z\right)\, dV = 0,$$

$$M_{xz} = \iiint_D y\left(10 + x^2 + 5y - 5z\right)\, dV$$

$$= 2 - \frac{61\pi}{48} \approx -1.99,$$

$$M_{xy} = \iiint_D z\left(10 + x^2 + 5y - 5z\right)\, dV$$

$$= \frac{61\pi}{96} - \frac{10}{9} \approx 0.885.$$

Note how $M_{yz} = 0$, as expected. The centre of mass is

$$\left(\overline{x}, \overline{y}, \overline{z}\right) = \left(0, \frac{-1.99}{3.855}, \frac{0.885}{3.855}\right) \approx \left(0, -0.516, 0.230\right).$$

As stated before, there are many uses for triple integration beyond finding volume. When $h(x, y, z)$ describes a rate of change function over some space region D, then

$$\iiint_D h(x, y, z)\ dV$$

gives the total change over D. Our one specific example of this was computing mass; a density function is simply a rate of mass change per volume function. Integrating density gives total mass.

While knowing how to integrate is important, it is arguably much more important to know how to set up integrals. It takes skill to create a formula that describes a desired quantity; modern technology is very useful in evaluating these formulas quickly and accurately.

In the next section, we learn about two new coordinate systems that allow us to integrate over closed regions in space more easily than when using rectangular coordinates.

17.6 Triple integration with cylindrical and spherical coordinates

Just as polar coordinates gave us a new way of describing curves in the plane, in this section we will see how cylindrical and spherical coordinates give us new ways of describing surfaces and regions in space.

17.6.1 Cylindrical coordinates

In short, cylindrical coordinates can be thought of as a combination of the polar and rectangular coordinate systems. One can identify a point (x_0, y_0, z_0), given in rectangular coordinates, with the point (r_0, θ_0, z_0), given in cylindrical coordinates, where the z-value in both systems is the same, and the point (x_0, y_0) in the xy-plane is identified with the polar point $P = (r_0, \theta_0)$; see Figure 17.27. So that each point in space that does not lie on the z-axis is defined uniquely, we will restrict $r \geq 0$ and $0 \leq \theta \leq 2\pi$.

We use the identity $z = z$ along with the identities in Theorem 11.1 to convert between the rectangular coordinate (x, y, z) and the cylindrical coordinate (r, θ, z). More specifically, we have from rectangular to cylindrical:

$$x = r\cos(\theta) \quad y = r\sin(\theta) \quad \text{and} \quad z = z\,. \tag{17.7}$$

and from cylindrical to rectangular:

$$r = \sqrt{x^2 + y^2}, \quad \tan(\theta) = \frac{y}{x} \quad \text{and} \quad z = z\,, \tag{17.8}$$

Note that our rectangular to polar conversion formulas (Eq. (17.8)) used $r^2 = x^2 + y^2$, allowing for negative r-values. Since we now restrict $r \geq 0$, we can use $r = \sqrt{x^2 + y^2}$.

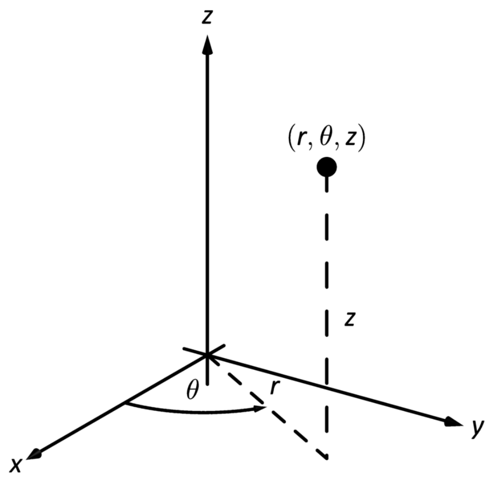


Figure 17.27: Illustrating the principles behind cylindrical coordinates.

Setting each of r, θ and z equal to a constant defines a surface in space, as illustrated in the following example.

Example 17.27

Describe the surfaces $r = 1$, $\theta = \pi/3$ and $z = 2$, given in cylindrical coordinates.

Solution

The equation $r = 1$ describes all points in space that are 1 unit away from the z-axis. This surface is a cylinder of radius 1, centred on the z-axis.

The equation $\theta = \pi/3$ describes the plane formed by extending the line $\theta = \pi/3$, as given by polar coordinates in the xy-plane, parallel to the z-axis.

The equation $z = 2$ describes the plane of all points in space that are 2 units above the xy-plane. This plane is described by $z = 2$ in rectangular coordinates. All three surfaces are graphed in Figure 17.28. Note how their intersection uniquely defines the point $P = (1, \pi/3, 2)$.

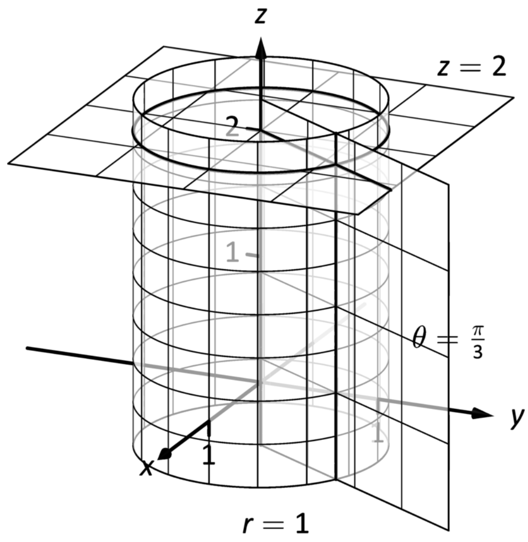


Figure 17.28: Graphing the surfaces in cylindrical coordinates from Example 17.27.

Cylindrical coordinates are useful when describing certain domains in space, allowing us to evaluate triple integrals over these domains more easily than if we used rectangular coordinates.

Theorem 17.6 shows how to evaluate $\iiint_D h(x,y,z)\ dV$ using rectangular coordinates. In that evaluation, we use $dV = dz\,dy\,dx$ (or one of the other five orders of integration). Recall how, in this order of integration, the bounds on y are curve to curve and the bounds on x are point to point: these bounds describe a region R in the xy-plane. We could describe R using polar coordinates as done in Section 17.2.4. In that section, we saw how we used $dA = r\,dr\,d\theta$ instead of $dA = dy\,dx$.

Considering the above thoughts, we have $dV = dz\big(r\,dr\,d\theta\big) = r\,dz\,dr\,d\theta$. We set bounds on z as surface to surface as done in the previous section, and then use curve to curve and point to point bounds on r and θ, respectively. Finally, using the identities given above, we change the integrand $h(x,y,z)$ to $h(r,\theta,z)$.

This process should sound plausible; the following theorem states it is truly a way of evaluating a triple integral.

Theorem 17.7 (Triple integration in cylindrical coordinates)
Let $w = h(r,\theta,z)$ be a continuous function on a closed, bounded region D in space, bounded in cylindrical coordinates by $\alpha \le \theta \le \beta$, $g_1(\theta) \le r \le g_2(\theta)$ and $f_1(r,\theta) \le z \le f_2(r,\theta)$. Then

$$\iiint_D h(r,\theta,z)\ dV = \int_\alpha^\beta \int_{g_1(\theta)}^{g_2(\theta)} \int_{f_1(r,\theta)}^{f_2(r,\theta)} h(r,\theta,z) r\,dz\,dr\,d\theta.$$

Example 17.28

Find the mass of the solid represented by the region in space bounded by $z = \sqrt{4 - x^2 - y^2} + 3$, $z = 0$ and the cylinder $x^2 + y^2 = 4$ (Figure 17.29), with density function $\delta(x,y,z) = x^2 + y^2 + z + 1$, using a triple integral in cylindrical coordinates. Distances are measured in centimetres and density is measured in g/cm^3.

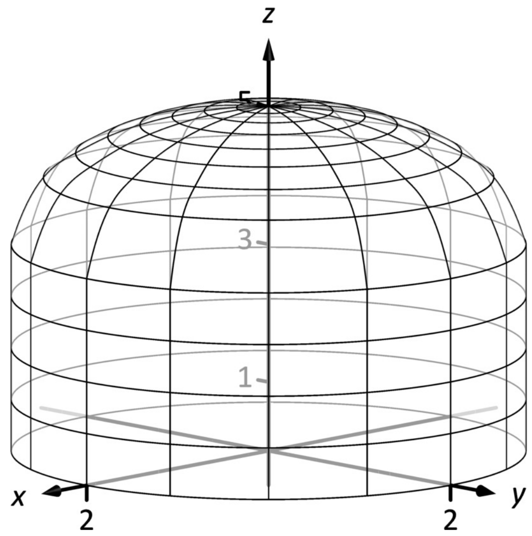


Figure 17.29: Visualizing the solid used in Example 17.28.

Solution

We begin by describing this region of space with cylindrical coordinates. The plane $z = 0$ is left unchanged; with the identity $r = \sqrt{x^2+y^2}$, we convert the hemisphere of radius 2 to the equation $z = \sqrt{4-r^2}+3$; the cylinder $x^2+y^2 = 4$ is converted to $r^2 = 4$, or, more simply, $r = 2$. We also convert the density function: $\delta(r,\theta,z) = r^2+z+1$. To describe this solid with the bounds of a triple integral, we bound z with $0 \le z \le \sqrt{4-r^2}+3$; we bound r with $0 \le r \le 2$; we bound θ with $0 \le \theta \le 2\pi$.

Using Definition 17.9 and Theorem 17.7, we have the mass of the solid is

$$\begin{aligned} M = \iiint_D \delta(x,y,z)\, dV &= \int_0^{2\pi}\int_0^2\int_0^{\sqrt{4-r^2}+3} \big(r^2+z+1\big) r\, dz\, dr\, d\theta \\ &= \int_0^{2\pi}\int_0^2 \left((r^3+4r)\sqrt{4-r^2}+\frac{5}{2}r^3+\frac{19}{2}r\right) dr\, d\theta \\ &= \frac{1318\pi}{15} \approx 276.04 \text{ g}, \end{aligned}$$

where we leave the details of the remaining double integral to the reader.

Example 17.29

Find the centre of mass of the solid with constant density whose base can be described by the polar curve $r = \cos(3\theta)$ and whose top is defined by the plane $z = 1-x+0.1y$, where distances are measured in metres, as seen in Figure 17.30.

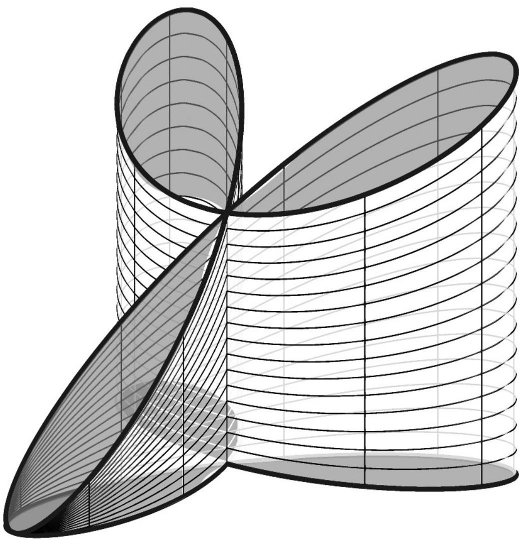


Figure 17.30: Visualizing the solid used in Example 17.29.

Solution

We convert the equation of the plane to use cylindrical coordinates: $z = 1 - r\cos(\theta) + 0.1r\sin(\theta)$. Thus the region in space is bounded by $0 \le z \le 1 - r\cos(\theta) + 0.1r\sin(\theta)$, $0 \le r \le \cos(3\theta)$, $0 \le \theta \le \pi$. Note that the rose curve $r = \cos(3\theta)$ is traced out once on $[0, \pi]$.

Since density is constant, we set $\delta = 1$ and finding the mass is equivalent to finding the volume of the solid. We set up the triple integral to compute this but do not evaluate it; we leave it to the reader to confirm it evaluates to:

$$M = \iiint_D \delta\, dV = \int_0^{\pi} \int_0^{\cos(3\theta)} \int_0^{1-r\cos(\theta)+0.1r\sin(\theta)} r\, dz\, dr\, d\theta \approx 0.785.$$

From Definition 17.9 we set up the triple integrals to compute the moments about the three coordinate planes. The computation of each is left to the reader (using Mathematica is recommended):

$$M_{yz} = \iiint_D x\,\delta\, dV = \int_0^{\pi} \int_0^{\cos(3\theta)} \int_0^{1-r\cos(\theta)+0.1r\sin(\theta)} (r\cos(\theta))r\, dz\, dr\, d\theta = -0.147,$$

$$M_{xz} = \iiint_D y\,\delta\, dV = \int_0^{\pi} \int_0^{\cos(3\theta)} \int_0^{1-r\cos(\theta)+0.1r\sin(\theta)} (r\sin(\theta))r\, dz\, dr\, d\theta = 0.015,$$

$$M_{xy} = \iiint_D z\,\delta\, dV = \int_0^{\pi} \int_0^{\cos(3\theta)} \int_0^{1-r\cos(\theta)+0.1r\sin(\theta)} (z)r\, dz\, dr\, d\theta = 0.467.$$

The centre of mass, in rectangular coordinates, is located at $(-0.147/0.785, 0.015/0.785, 0.467/0.785)$.

17.6.2 Spherical coordinates

In short, spherical coordinates can be thought of as a double application of the polar coordinate system. In spherical coordinates, a point P is identified with (ρ, θ, φ), where ρ is the distance from the origin to

P, θ is the same angle as would be used to describe P in the cylindrical coordinate system, and φ is the angle between the positive z-axis and the ray from the origin to P; see Figure 17.31. So that each point in space that does not lie on the z-axis is defined uniquely, we will restrict $\rho \geq 0$, $0 \leq \theta \leq 2\pi$ and $0 \leq \varphi \leq \pi$. The symbol ρ is the Greek letter rho. Traditionally it is used in the spherical coordinate system, while r is used in the polar and cylindrical coordinate systems.

For comprehensiveness, we list the conversions to/from our three spatial coordinate systems.

Rectangular and Cylindrical

$$\begin{gathered} r^2 = x^2 + y^2, \quad \tan(\theta) = \frac{y}{x}, \quad z = z \\ x = r\cos(\theta), \quad y = r\sin(\theta), \quad z = z \end{gathered} \tag{17.9}$$

Rectangular and Spherical

$$\begin{gathered} \rho = \sqrt{x^2 + y^2 + z^2}, \quad \tan(\theta) = \frac{y}{x}, \quad \cos(\varphi) = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \\ x = \rho\sin(\varphi)\cos(\theta), \quad y = \rho\sin(\varphi)\sin(\theta), \quad z = \rho\cos(\varphi) \end{gathered} \tag{17.10}$$

Cylindrical and Spherical

$$\begin{gathered} \rho = \sqrt{r^2 + z^2}, \quad \theta = \theta, \quad \tan(\varphi) = \frac{r}{z} \\ r = \rho\sin(\varphi), \quad \theta = \theta, \quad z = \rho\cos(\varphi) \end{gathered} \tag{17.11}$$

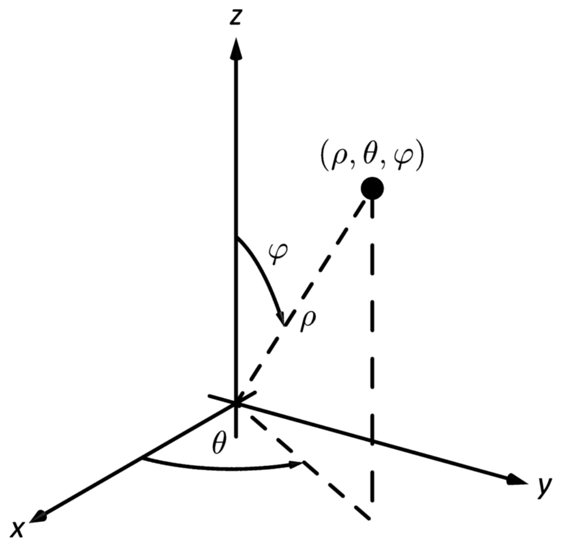


Figure 17.31: Illustrating the principles behind spherical coordinates.

Example 17.30

Describe the surfaces $\rho = 1$, $\theta = \pi/3$ and $\varphi = \pi/6$, given in spherical coordinates.

Solution

The equation $\rho = 1$ describes all points in space that are 1 unit away from the origin: this is the sphere of radius 1, centred at the origin.

The equation $\theta = \pi/3$ describes the same surface in spherical coordinates as it does in cylindrical coordinates: beginning with the line $\theta = \pi/3$ in the xy-plane as given by polar coordinates, extend the line parallel to the z-axis, forming a plane.

The equation $\varphi = \pi/6$ describes all points P in space where the ray from the origin to P makes an angle of $\pi/6$ with the positive z-axis. This describes a cone, with the positive z-axis its axis of symmetry, with point at the origin.

All three surfaces are graphed in Figure 17.32. Note how their intersection uniquely defines the point $P = (1, \pi/3, \pi/6)$.

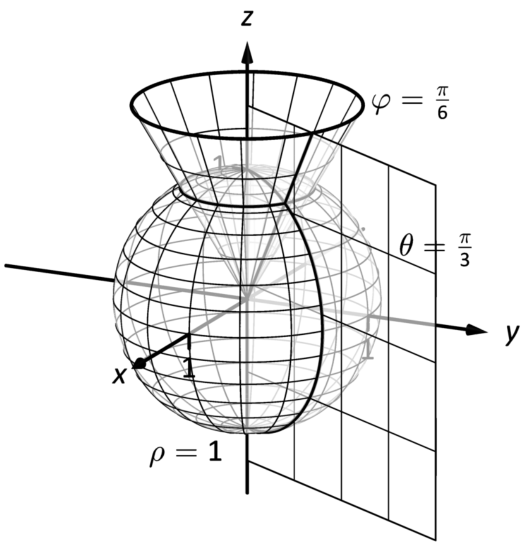


Figure 17.32: Graphing the surfaces in spherical coordinates from Example 17.30.

Spherical coordinates are useful when describing certain domains in space, allowing us to evaluate triple integrals over these domains more easily than if we used rectangular coordinates or cylindrical coordinates. The crux of setting up a triple integral in spherical coordinates is appropriately describing the small amount of volume, dV, used in the integral.

Considering Figure 17.33, we can make a small spherical wedge by varying ρ, θ and φ each a small amount, $\Delta\rho$, $\Delta\theta$ and $\Delta\varphi$, respectively. This wedge is approximately a rectangular solid when the change in each coordinate is small, giving a volume of about

$$\Delta V \approx \Delta\rho \;\times\; \rho\Delta\varphi \;\times\; \rho\sin(\varphi)\Delta\theta.$$

Given a region D in space, we can approximate the volume of D with many such wedges. As the size of each of $\Delta\rho$, $\Delta\theta$ and $\Delta\varphi$ goes to zero, the number of wedges increases to infinity and the volume of D is more accurately approximated, giving

$$dV = d\rho \;\times\; \rho\, d\varphi \;\times\; \rho\sin(\varphi)d\theta = \rho^2\sin(\varphi)\; d\rho\; d\theta\; d\varphi.$$

Again, this development of dV should sound reasonable, and the following theorem states it is the appropriate manner by which triple integrals are to be evaluated in spherical coordinates. Note that it is generally most intuitive to evaluate the triple integral in Theorem 17.8 by integrating with respect to ρ first; it often does not matter whether we next integrate with respect to θ or φ. As the bounds for these variables are usually constants in practice, it generally is a matter of preference.

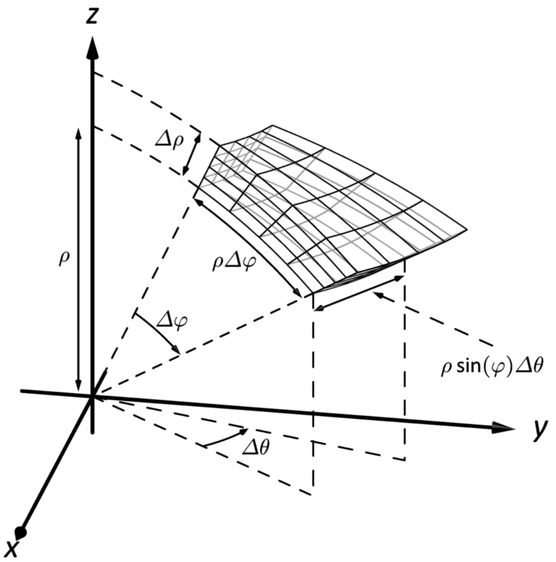


Figure 17.33: Approximating the volume of a standard region in space using spherical coordinates.

Theorem 17.8 (Triple integration in spherical coordinates)
Let $w = h(\rho, \theta, \varphi)$ be a continuous function on a closed, bounded region D in space, bounded in spherical coordinates by $\alpha_1 \leq \varphi \leq \alpha_2$, $\beta_1 \leq \theta \leq \beta_2$ and $f_1(\theta, \varphi) \leq \rho \leq f_2(\theta, \varphi)$. Then

$$\iiint_D h(\rho, \theta, \varphi)\, dV = \int_{\alpha_1}^{\alpha_2} \int_{\beta_1}^{\beta_2} \int_{f_1(\theta,\varphi)}^{f_2(\theta,\varphi)} h(\rho, \theta, \varphi) \rho^2 \sin(\varphi)\, d\rho\, d\theta\, d\varphi.$$

Example 17.31

Let D be the region in space bounded by the sphere, centred at the origin, of radius r. Use a triple integral in spherical coordinates to find the volume V of D.

Solution

The sphere of radius r, centred at the origin, has equation $\rho = r$. To obtain the full sphere, the bounds on θ and φ are $0 \leq \theta \leq 2\pi$ and $0 \leq \varphi \leq \pi$. This leads us to:

$$\begin{aligned} V &= \iiint_D dV \\ &= \int_0^{\pi} \int_0^{2\pi} \int_0^{r} \left(\rho^2 \sin(\varphi)\right) d\rho\, d\theta\, d\varphi \\ &= \int_0^{\pi} \int_0^{2\pi} \left(\frac{1}{3}\rho^3 \sin(\varphi)\right)\Bigg|_0^r d\theta\, d\varphi \\ &= \int_0^{\pi} \int_0^{2\pi} \left(\frac{1}{3} r^3 \sin(\varphi)\right) d\theta\, d\varphi \end{aligned}$$

$$= \int_0^{\pi} \left(\frac{2\pi}{3} r^3 \sin(\varphi) \right) \, d\varphi$$

$$= \left(-\frac{2\pi}{3} r^3 \cos(\varphi) \right) \Bigg|_0^{\pi}$$

$$= \frac{4\pi}{3} r^3,$$

the familiar formula for the volume of a sphere. Note how the integration steps were easy, not using square roots nor integration steps such as substitution.

Example 17.32

Find the centre of mass of the solid with constant density enclosed above by $\rho = 4$ and below by $\varphi = \pi/6$, as illustrated in Figure 17.34.

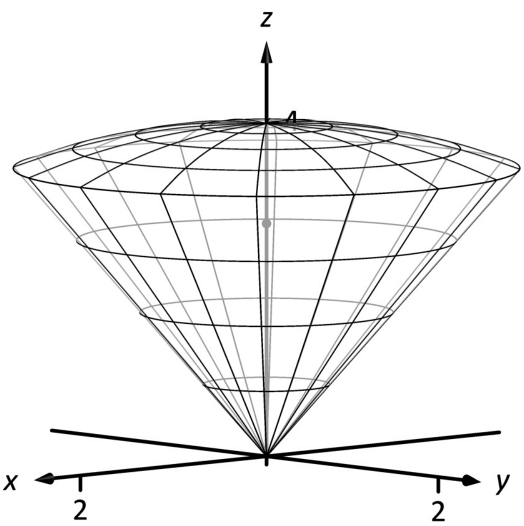


Figure 17.34: Graphing the solid, and its centre of mass, from Example 17.32.

Solution

We will set up the four triple integrals needed to find the centre of mass (i.e., to compute M, M_{yz}, M_{xz} and M_{xy}) and leave it to the reader to evaluate each integral. Because of symmetry, we expect the x- and y- coordinates of the centre of mass to be 0.

While the surfaces describing the solid are given in the statement of the problem, to describe the full solid D, we use the following bounds: $0 \le \rho \le 4$, $0 \le \theta \le 2\pi$ and $0 \le \varphi \le \pi/6$. Since density δ is constant, we assume $\delta = 1$.

The mass of the solid is:

$$M = \iiint_D dm = \iiint_D dV$$

$$= \int_0^{\pi/6} \int_0^{2\pi} \int_0^{4} \left(\rho^2 \sin(\varphi) \right) d\rho \, d\theta \, d\varphi$$

$$= \frac{64}{3}(2-\sqrt{3})\pi \approx 17.958.$$

To compute M_{yz}, the integrand is x; we use $x = \rho \sin(\varphi)\cos(\theta)$. This gives:

$$\begin{aligned} M_{yz} &= \iiint_D x\,dm \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big((\rho\sin(\varphi)\cos(\theta))\rho^2\sin(\varphi)\big)\,d\rho\,d\theta\,d\varphi \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big(\rho^3\sin^2(\varphi)\cos(\theta)\big)\,d\rho\,d\theta\,d\varphi = 0, \end{aligned}$$

which we expected as we expect $\overline{x} = 0$. Actually, by rewriting the above integral as

$$\begin{aligned} M_{yz} &= \int_0^{\pi/6} \sin^2(\varphi) \int_0^{2\pi} \cos(\theta) \int_0^4 \rho^3\,d\rho\,d\theta\,d\varphi \\ &= \int_0^{\pi/6} \sin^2(\varphi)d\varphi \int_0^{2\pi} \cos(\theta)d\theta \int_0^4 \rho^3\,d\rho \end{aligned}$$

it is immediately clear that it yields 0 due to the integral of the cosine function over its fundamental period that pops up.

To compute M_{xz}, the integrand is y. So, we use $y = \rho\sin(\varphi)\sin(\theta)$. This gives:

$$\begin{aligned} M_{xz} &= \iiint_D y\,dm \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big((\rho\sin(\varphi)\sin(\theta))\rho^2\sin(\varphi)\big)\,d\rho\,d\theta\,d\varphi \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big(\rho^3\sin^2(\varphi)\sin(\theta)\big)\,d\rho\,d\theta\,d\varphi = 0, \end{aligned}$$

which we also expected as we expect $\overline{y} = 0$.

To compute M_{xy}, the integrand is z. So, we use $z = \rho\cos(\varphi)$. This gives:

$$\begin{aligned} M_{xy} &= \iiint_D z\,dm \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big((\rho\cos(\varphi))\rho^2\sin(\varphi)\big)\,d\rho\,d\theta\,d\varphi \\ &= \int_0^{\pi/6}\int_0^{2\pi}\int_0^4 \big(\rho^3\cos(\varphi)\sin(\varphi)\big)\,d\rho\,d\theta\,d\varphi \end{aligned}$$

$$= 16\pi \approx 50.266.$$

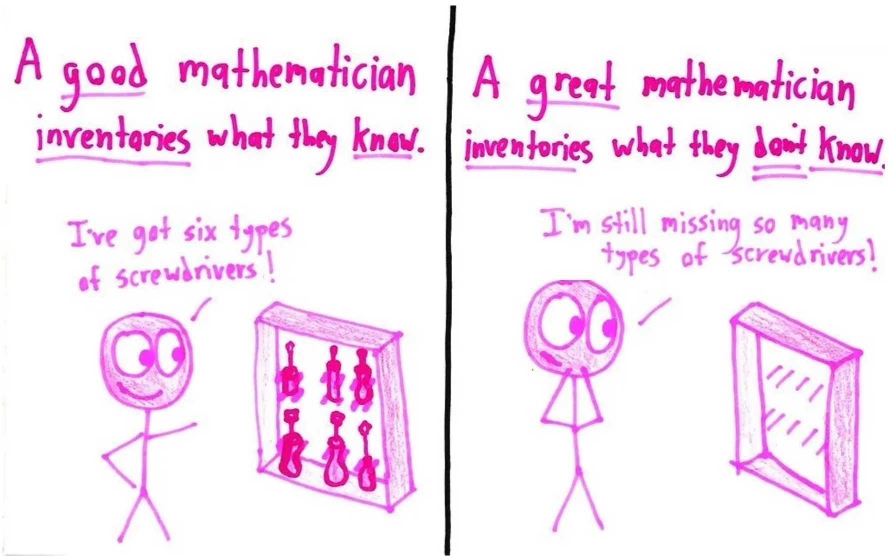
Thus the centre of mass is $(0, 0, M_{xy}/M) \approx (0, 0, 2.799)$, as indicated in Figure 17.34.

This section has provided a brief introduction into two new coordinate systems useful for identifying points in space. Each can be used to define a variety of surfaces in space beyond the canonical surfaces graphed as each system was introduced.

However, the usefulness of these coordinate systems does not lie in the variety of surfaces that they can describe nor the regions in space these surfaces may enclose. Rather, cylindrical coordinates are mostly used to describe cylinders and spherical coordinates are mostly used to describe spheres. These shapes are of special interest in the sciences, especially in physics, and computations on/inside these shapes is difficult using rectangular coordinates. For instance, in the study of electricity and magnetism, one often studies the effects of an electrical current passing through a wire; that wire is essentially a cylinder, described well by cylindrical coordinates.

This chapter investigated the natural follow–on to partial derivatives: iterated integration. We learned how to use the bounds of a double integral to describe a region in the plane using both rectangular and polar coordinates, then later expanded to use the bounds of a triple integral to describe a region in space. We used double integrals to find volumes under surfaces, surface area, and the centre of mass of lamina; we used triple integrals as an alternate method of finding volumes of space regions and also to find the centre of mass of a region in space.

Integration does not stop here. We could continue to iterate our integrals, next investigating quadruple integrals whose bounds describe a region in 4–dimensional space. We can also look back to regular integration where we found the area under a curve in the plane. A natural analogue to this is finding the area under a curve, where the curve is in space, not in a plane. These are just two of many avenues to explore under the heading of integration.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

17.7 Exercises

Iterated integrals and area

Assignment 17.1 — Evaluate the double integrals below.

(a) $$\int_1^2 \int_y^{3y} (x+y)\, dx\, dy$$

(b) $$\int_{-1}^2 \int_{2x^2-2}^{x^2+x} x\, dy\, dx$$

(c) $$\int_0^{\pi} \int_0^{\cos(\theta)} r\sin(\theta)\, dr\, d\theta$$

(d) $$\int_0^{\frac{\pi}{2}} \int_2^{4\cos(\theta)} r^3\, dr\, d\theta$$

(e) $$\int_0^1 \int_0^1 \frac{x^2}{1+y^2}\, dy\, dx$$

(f) $$\int_{-1}^2 \int_0^1 (xy^2+x^2y)\, dx\, dy$$

(g) $$\int_0^1 \int_{x^2}^{2-x^2} \sqrt{x}y\, dy\, dx$$

(h) $$\int_0^{\pi/4} \int_{\sin(x)}^{\cos(x)} xy\, dy\, dx$$

Assignment 17.2 — Identify the boundaries for the double integrals below and evaluate.

(a) $\iint_R dA$ with R the region in the first quadrant between $y^2 = x^3$ and $y = x$

(b) $\iint_R x^2\, dA$ with R the region in the first quadrant between $xy = 16$, $y = x$, $y = 0$ and $x = 8$

(c) $\iint_R y\, dA$ with R the region enclosed by $y = x^2$ and $y = x^3$

(d) $\iint_R \frac{1}{\sqrt{2y-y^2}}\, dA$ with R the region enclosed by $x^2 = 4-2y$ in the first quadrant

(e) $\iint_R e^{\frac{x}{y}}\, dA$ with R the region bounded by $y^2 = x$, $x = 0$ and $y = 1$

(f) $\iint_R e^{-(x+y)}\, dA$ with $R = \left\{(x,y)\,\middle|\, \frac{1}{2} \le x \le 1, y \ge 0, x \ge y\right\}$

(g) $\iint_R dA$ with $R = \left\{(x,y)\,\middle|\, a-y \le x \le \sqrt{a^2-y^2}, 0 \le y \le a\right\}$

(h) $\iint_R \frac{dA}{x+y}$ with R the triangle with vertices $(0,0)$, $(1,0)$ and $(1,1)$

Assignment 17.3 — Reverse the order of integration in the integrals below.

(a) $\displaystyle\int_0^3 \int_1^{\sqrt{4-y}} f(x,y)\,dx\,dy$

(b) $\displaystyle\int_0^1 \int_{\arccos(y)}^{\frac{\pi}{2}} f(x,y)\,dx\,dy$

(c) $\displaystyle\int_{-6}^{2} \int_{\frac{x^2}{4}}^{3-x} f(x,y)\,dy\,dx$

(d) $\displaystyle\int_{\frac{a}{2}}^{a} \int_0^{\sqrt{2ax-x^2}} f(x,y)\,dy\,dx$ with $a > 0$

Assignment 17.4 — Using a double integral, find the area of the regions below.

(a) the region bounded by $y^2 = 10x + 25$ and $y^2 = -6x + 9$

(b) the region bounded by $x^2 + y^2 = 2x$, $x^2 + y^2 = 4x$, $y = x$ and $y = 0$

(c) the region bounded by $r\cos(\theta) = 1$ and $r = 2$ (the region that does not contain $r = 0$)

(d) the region bounded by $r = 1 + \cos(\theta)$ and $r = \cos(\theta)$

Double integration and volume

Assignment 17.5 — Find the volume of the solids below using Cartesian coordinates.

(a) the solid enclosed by the three coordinate planes and the surfaces $x^2 + 4y^2 = 4$ and $y^2 + 2z = 4$

(b) the solid enclosed by the three coordinate planes and the surfaces $z = 1 - x^2 - y^2$ and $x + y \leq 1$

(c) the solid enclosed by the coordinate planes and the surfaces $y = x^2$, $x = y^2$, $z = 0$ en $z = 12 + y - x^2$

(d) the frustum with side faces $x = 0$, $x = 2$, $y = 0$, $y = 1$, base $z = 0$ and top face $-x + y + 2z = 4$

(e) the solid enclosed by $x^2 + y^2 = 1$ and $x^2 + z^2 = 1$

Assignment 17.6 — Find the volume of the solids below using polar coordinates

(a) the solid within $x^2 + y^2 = 1$, cut off by $x^2 + y^2 + z^2 = 4$

(b) the solid inside $x^2 + y^2 = 2az\,(a > 0)$, cut off by $x^2 + y^2 + z^2 = 3a^2$

(c) the solid enclosed by $x^2 + y^2 + z^2 = a^2$ and $x^2 + y^2 = ax$

(d) the solid enclosed by $x^2 + y^2 = 2y$ and $z^2 = y$

Differentiation under the integral sign

Assignment 17.7 — Find the derivatives below.

(a) $\dfrac{d}{dt}\left(\displaystyle\int_0^1 \sin(x-t)\,dx\right)$

(b) $\dfrac{d}{dt}\left(\displaystyle\int_1^{t^2} \frac{e^{tx}}{x}\,dx\right)$

(c) $\dfrac{d}{dx}\left(\displaystyle\int_{\frac{1}{x}}^{\frac{2}{x}} \frac{\sin(tx)}{t}\,dt\right)$

Assignment 17.8 — Evaluate the following integrals.

(a) $F(\alpha) = \displaystyle\int_0^{+\infty} e^{-x}\frac{\sin(\alpha x)}{x}\,dx$

(b) $F(a) = \displaystyle\int_0^{+\infty} \frac{e^{-ax}\sin(x)}{x}\,dx$

(c) $F(a) = \displaystyle\int_0^{1} \frac{x^a-1}{\ln(x)}\,dx$

Centre of mass

Assignment 17.9 — Find the center of mass of the considered region with the given mass density.

(a) the flat region enclosed by$y = \sin(x)$ en $y = 0$ $(0 \le x \le \pi)$ with mass density $\delta(x,y) = ky$.

(b) the flat region enclosed by $y^2 = 4x+4$ and $y^2 = -2x+4$ with mass density $\delta = 1$

(c) a triangular plate defined by the region enclosed by the x- and y-axis and the line $2x+3y = 12$ with mass density $\delta = 1$

Surface area

Assignment 17.10 — Find the surface area of the solids below.

(a) the part of $x^2+y^2 = 3z^2$, that lies above the xy-plane and inside $x^2+y^2 = 4y$

(b) the part of $x^2+z^2 = 16$, that lies inside $x^2+y^2 = 16$

(c) the part of $z^2 = 4x$, that lies inside $y^2 = 4x$ and $x \le 1$

(d) the part of $z = 1-x^2-y^2$, that lies inside $x^2+y^2 = 1$

(e) the part of $x^2+y^2+z^2 = a^2$ cut off by $x^2+4y^2 = a^2$.

Triple integration

Assignment 17.11 — Evaluate the triple integrals below.

(a) $\displaystyle\iiint_D xyz\; dV$ with D the solid enclosed by the coordinate planes and the plane $x+y+z=1$

(b) $\displaystyle\iiint_D (x+z)\; dV$ with $D = \left\{(x,y,z)\,\middle|\, 0 \le x, 0 \le y, 0 \le z, x^2+y^2+z^2 \le 1\right\}$

(c) $\displaystyle\iiint_D xy\; dV$ with $D = \left\{(x,y,z)\,\middle|\, 0 \le x, 0 \le y, 0 \le z \le 1, x^2+y^2 \le 1\right\}$

(d) $\displaystyle\iiint_D \left(x^2+y^2+z^2\right)^{3/2}\; dV$ with $D = \left\{(x,y,z)\,\middle|\, x^2+y^2+z^2 \le 1\right\}$

Triple integration with cylindrical and spherical coordinates

Assignment 17.12 — Find the volume of the solids below using the most appropriate coordinate system.

(a) the solid enclosed by $z = 10 - x^2 - y^2$ and $z = 2(x^2+y^2-1)$

(b) the solid inside the ellipsoid $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$

(c) the solid above the xy-plane, under $x^2+y^2+2z=16$ and within $x^2+y^2=4$

(d) the solid inside $3(x^2+y^2)=z^2$, cut off by $x^2+y^2+z^2=a^2$

(e) the solid enclosed by $x^2+y^2+z^2=4$ and $x^2+y^2=3z$

Review exercises

Assignment 17.13 — Find the volume of a building that has its roof at height $z = \arctan\left(yx^{-1}\right)$, the xy-plane as a base and shape defined by $x^2+y^2=1$, with $x \ge 0$ and $y \ge 0$. What is the area of the roof?

Assignment 17.14 — At the campsite of a festival, there are many tents. These have a square base (in the xy-plane) and are enclosed by two parabolic cylinders $ax^2+z=a$ and $ay^2+z=a$. What is the volume of such a tent?

Assignment 17.15 — Find the center of mass of the solids below with the given mass density.

(a) the upper part of the cone $x^2+y^2=z^2$ above $z=h$ $(h>0)$ and with mass density $\delta(x,y,z) = 1 - \dfrac{z}{h}$

(b) the solid enclosed by $z = x^2 + y^2$ and $z = 4$ and with mass density $\delta(x, y, z) = k\sqrt{x^2 + y^2}$.

(c) the surface $x^2 + y^2 - z^2 = 1$ between $z = 0$ and $z = 2$ and with mass density $\delta = 1$

Calculus required continuity, and continuity was supposed to require the infinitely little; but nobody could discover what the infinitely little might be.

— Bertrand Russell —

18

Vector calculus

18.1 Line integrals over a scalar field

This section explores completely different relationships between vectors and integration. These relationships will enable us to compute the work done by a magnetic field in moving an object along a path and find how much air moves through an oddly–shaped screen in space, among other things.

18.1.1 Definition

Consider the surface and curve shown in Figure 18.1(a). The surface is given by

$$f(x,y) = 1 - \cos(x)\sin(y).$$

The dashed curve lies in the xy-plane and is the familiar $y = x^2$ parabola from $-1 \leq x \leq 1$; we will call this curve C. The curve drawn with a solid line in the graph is the curve in space that lies on our surface with x- and y- values that lie on C.

The question we want to answer is this: what is the area that lies below the curve drawn with the solid line? In other words, what is the area of the region above C and under the surface f? This region is shown in Figure 18.1(b). We suspect the answer can be found using an integral, but before trying to figure out what that integral is, let us first try to approximate its value.

In Figure 18.1(c), four rectangles have been drawn over the curve C. The bottom corners of each rectangle lie on C, and each rectangle has a height given by the function $f(x,y)$ for some (x,y) pair along C between the rectangle's bottom corners. As we know how to find the area of each rectangle, we are able to approximate the area above C and under f. Clearly, our approximation will be an approximation. The heights of the rectangles do not match exactly with the surface f, nor does the base of each rectangle follow perfectly the path of C.

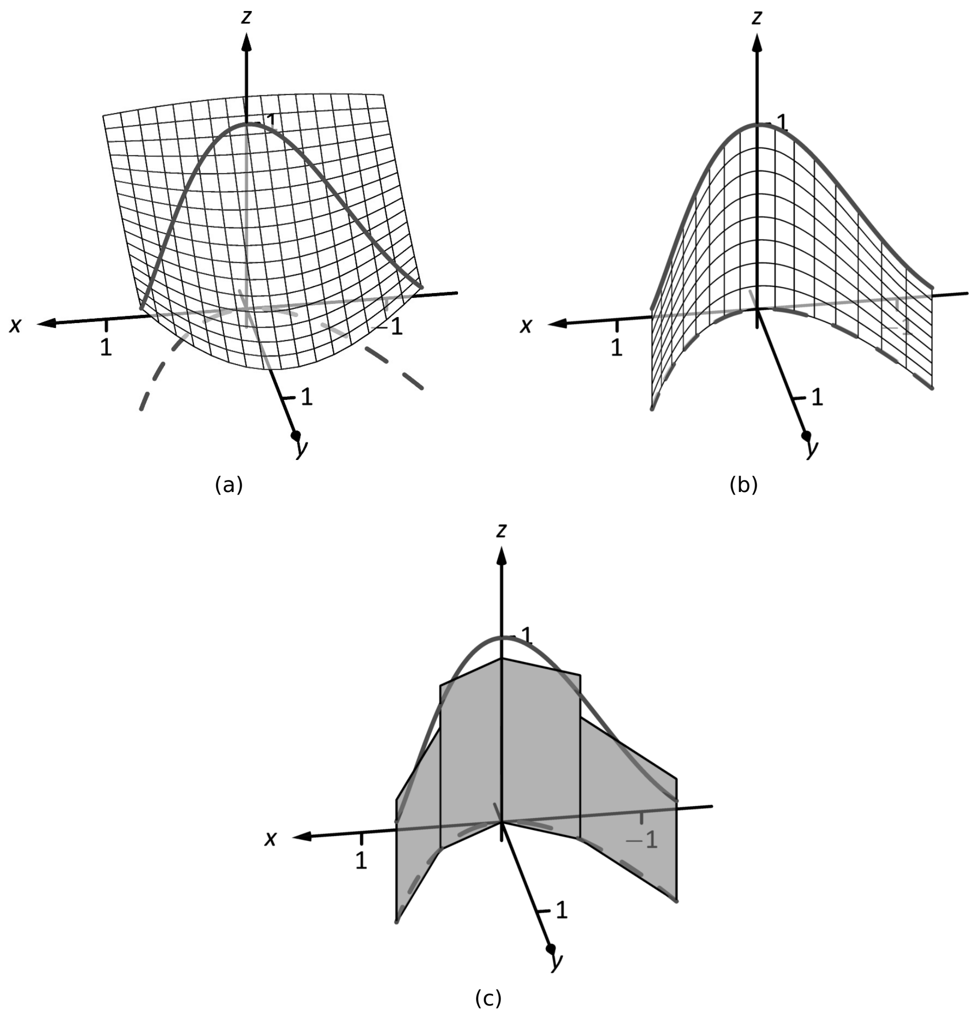


Figure 18.1: Finding area under a curve in space.

In typical calculus fashion, our approximation can be improved by using more rectangles. The sum of the areas of these rectangles gives an approximate value of the true area above C and under f. As the area of each rectangle is height × width, we assert that the

$$\text{area above } C \approx \sum(\text{heights} \times \text{widths}).$$

When first learning of the integral, and approximating areas with (heights × widths), the width was a small change in x: dx. That will not suffice in this context. Rather, each width of a rectangle is actually approximating the arc length of a small portion of C. In Section 15.4, we used s to represent the arc length parameter of a curve. Hence, a small amount of arc length will thus be represented by ds.

The height of each rectangle will be determined in some way by the surface f. If we parametrize C by s, an s-value corresponds to an (x, y) pair that lies on the parabola C. Since f is a function of x and y, and x and y are functions of s, we can say that f is a function of s. Given a value s, we can compute $f(s)$ and find a height. Thus

$$\begin{aligned}
\text{area under } f \text{ and above } C &\approx \sum(\text{heights} \times \text{widths});\\
\text{area under } f \text{ and above } C &= \lim_{\mathcal{L}\to 0}\sum f(c_i)\Delta s_i\\
&= \int_C f(s)\, ds. \qquad (18.1)
\end{aligned}$$

Here we have introduce a new notation, the integral symbol with a subscript of C. It is reminiscent of our usage of $\iint_R$. Using the train of thought found in the Integration Review preceding this section, we interpret $\int_C f(s)\ ds$ as meaning sum up, along a curve C, function values $f(s) \times$ small arc lengths. It is understood here that s represents the arc length parameter.

All this leads us to a definition. The integral found in Equation 18.1 is called a **line integral**. We formally define it below.

Definitie 18.1 (Line integral over a scalar field)
Let C be a smooth curve parametrized by s, the arc length parameter, and let f be a continuous function of s. A **line integral** (*lijnintegraal*) is an integral of the form

$$\int_C f(s)\ ds = \lim_{\mathcal{L}\to 0} \sum_{i=1}^{n} f(c_i)\Delta s_i,$$

where $s_1 < s_2 < \cdots < s_n$ is any partition of the s-interval over which C is defined, c_i is any value in the i^{th} subinterval, Δs_i is the width of the i^{th} subinterval, and $\mathcal{L}$ is the length of the longest subinterval in the partition.

Note that Definition 18.1 uses the term scalar field which has not yet been defined. Its meaning is discussed in when it is compared to a vector field. Besides, when C is a closed curve, i.e., a curve that ends at the same point at which it starts, we use

$$\oint_C f(s)\ ds$$

instead of

$$\int_C f(s)\ ds.$$

The definition of the line integral does not specify whether C is a curve in the plane or space (or hyperspace), as the definition holds regardless. For now, however, we will assume C lies in the xy-plane.

Actually, this definition of the line integral does not really say anything new. If C is a curve and s is the arc length parameter of C on $a \le s \le b$, then

$$\int_C f(s)\ ds = \int_a^b f(s)\ ds.$$

The real difference with this integral from the standard $\int_a^b f(x)\ dx$ we used in the past is that of context. Our previous integrals naturally summed up values over an interval on the x-axis, whereas now we are summing up values over a curve. If we can parametrize the curve with the arc length parameter, we can evaluate the line integral just as before. Unfortunately, parametrizing a curve in terms of the arc length parameter is usually very difficult, so we must develop a method of evaluating line integrals using a different parametrization.

Given a curve C, find any parametrization of C: $x = g(t)$ and $y = h(t)$, for continuous functions g and h, where $a \le t \le b$. We can represent this parametrization with a vector–valued function, $\vec{r}(t) = \big(g(t), h(t)\big)$.

In Section 15.4, we defined the arc length parameter as

$$s(t) = \int_0^t \left\|\vec{r}\,'(u)\right\|\,du.$$

By the fundamental theorem of calculus, $ds = \left\|\vec{r}\,'(t)\right\|\,dt$. We can substitute the right hand side of this equation for ds in the line integral definition. Moreover, we can view f as being a function of x and y since it is a function of s. Thus $f(s) = f(x,y) = f\big(g(t),h(t)\big)$. This gives us a concrete way to evaluate a line integral:

$$\int_C f(s)\,ds = \int_a^b f\big(g(t),h(t)\big)\,\left\|\vec{r}\,'(t)\right\|\,dt.$$

We restate this as a theorem for its n–dimensional analogue.

Theorem 18.1 (Evaluating a line integral over a scalar field)
Let C be a curve parametrized by $\vec{r}(t) = \big(g_1(t), g_2(t), \ldots, g_n(t)\big)$, $a \le t \le b$, where g_i is continuously differentiable, and let $z = f(\mathbf{x})$, where f is continuous over C. Then

$$\int_C f(s)\,ds = \int_a^b f\big(g_1(t), g_2(t), \ldots, g_n(t)\big)\,\left\|\vec{r}\,'(t)\right\|\,dt.$$

To be clear, the first point of Theorem 18.1 can be used to find the area under a surface $z = f(x,y)$ and above a curve C. We will later give an understanding of the line integral when C is a curve in space.

Let us do an example where we actually compute an area.

Example 18.1

Find the area under the surface $f(x,y) = \cos(x) + \sin(y) + 2$ over the curve C, which is the segment of the line $y = 2x + 1$ on $-1 \le x \le 1$, as shown in Figure 18.2.

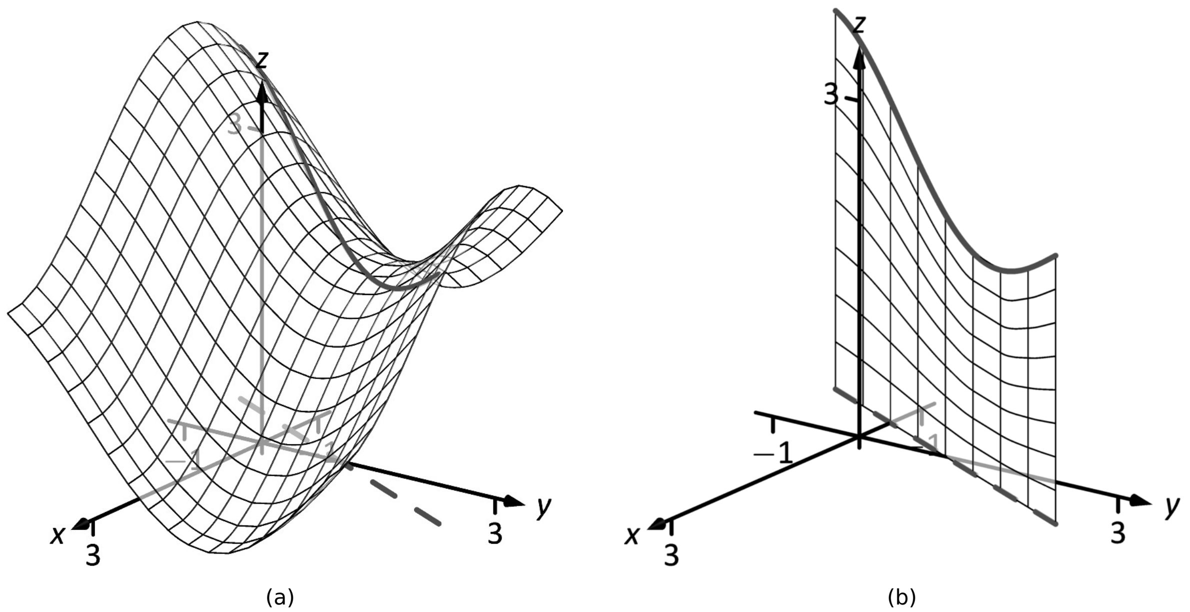


Figure 18.2: Finding area under a curve in Example 18.1.

Solution

Our first step is to represent C with a vector–valued function. Since C is a simple line, and we have a explicit relationship between y and x (namely, that $y = 2x+1$), we can let $x = t$, $y = 2t+1$, and write $\vec{r}(t) = (t, 2t+1)$ for $-1 \le t \le 1$.

We find the values of f over C as

$$f(x,y) = f(t, 2t+1) = \cos(t) + \sin(2t+1) + 2.$$

We also need $\|\vec{r}\,'(t)\|$; with $\vec{r}\,'(t) = (1,2)$, we have $\|\vec{r}\,'(t)\| = \sqrt{5}$. Thus $ds = \sqrt{5}\ dt$.

The area we seek is

$$\begin{aligned}\int_C f(s)\ ds &= \int_{-1}^{1} \big(\cos(t) + \sin(2t+1) + 2\big)\sqrt{5}\ dt \\ &= \sqrt{5}\left(\sin(t) - \frac{1}{2}\cos(2t+1) + 2t\right)\Bigg|_{-1}^{1} \\ &\approx 14.418 \text{ units}^2.\end{aligned}$$

We now consider the example that introduced this section.

Example 18.2

Find the area under $f(x,y) = 1 - \cos(x)\sin(y)$ and over the parabola $y = x^2$, from $-1 \le x \le 1$.

Solution

We parametrize our curve C as $\vec{r}(t) = \left(t, t^2\right)$ for $-1 \le t \le 1$; we find $\|\vec{r}\,'(t)\| = \sqrt{1+4t^2}$, so $ds = \sqrt{1+4t^2}\ dt$.

Replacing x and y with their respective functions of t, we have

$$f(x,y) = f(t, t^2) = 1 - \cos(t)\sin\left(t^2\right).$$

Thus the area under f and over C is found to be

$$\int_C f(s)\ ds = \int_{-1}^{1} \left(1 - \cos(t)\sin\left(t^2\right)\right)\sqrt{1+4t^2}\ dt.$$

This integral is impossible to evaluate using the techniques developed in this text. We resort to a numerical approximation; accurate to two places after the decimal, we find the area is 2.17.

Note how in each of the previous examples we are effectively finding area under a curve, just as we did when first learning of integration. We have used the phrase area over a curve C and under a surface, but that is because of the important role C plays in the integral. The figures show how the curve C defines another curve on the surface $z = f(x,y)$, and we are finding the area under that curve.

18.1.2 Properties

Many properties of line integrals can be inferred from general integration properties. For instance, if k is a scalar, then
$$\int_C kf(s)ds = k\int_C f(s)ds,$$
and similarly
$$\int_C \big(f(s)+g(s)\big)\ ds = \int_C f(s)\ ds + \int_C g(s)\ ds,$$
where f and g are continuous functions of s.

One property in particular of line integrals is worth noting. If C is a curve composed of subcurves C_1 and C_2, where they share only one point in common (see Figure 18.3(a)), then the line integral over C is the sum of the line integrals over C_1 and C_2:

$$\int_C f(s)\ ds = \int_{C_1} f(s)\ ds + \int_{C_2} f(s)\ ds.$$

This property allows us to evaluate line integrals over some curves C that are not smooth. Note how in Figure 18.3(b) the curve is not smooth at D, so by our definition of the line integral we cannot evaluate $\int_C f(s)ds$. However, one can evaluate line integrals over C_1 and C_2 and their sum will be the desired quantity. A curve C that is composed of two or more smooth curves is said to be piecewise smooth. In this section, any statement that is made about smooth curves also holds for piecewise smooth curves.

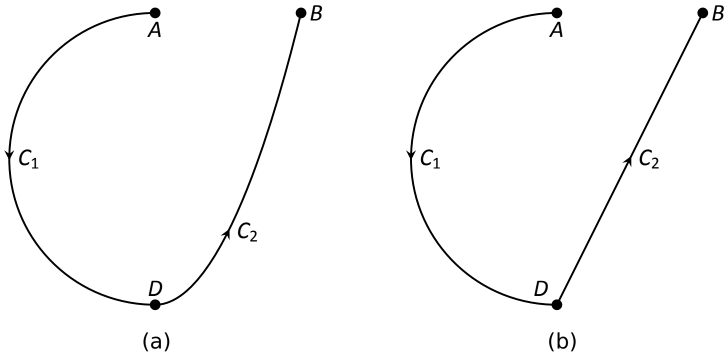


Figure 18.3: Illustrating properties of line integrals.

18.1.3 Centre of mass

Let a curve C (either in the plane or in space) represent a thin wire with variable density $\delta(s)$. We can approximate the mass of the wire by dividing the wire (i.e., the curve) into small segments of length Δs_i and assume the density is constant across these small segments. The mass of each segment is density of the segment × its length; by summing up the approximate mass of each segment we can approximate the total mass:
$$\text{total mass of wire } \approx \sum_i \delta(s_i)\Delta s_i.$$

By taking the limit as the length of the segments approaches 0, we have the definition of the line integral as seen in Definition 18.1. When learning of the line integral, we let $f(s)$ represent a height; now we let $f(s) = \delta(s)$ represent a density. We can extend this understanding of computing mass to also compute the centre of mass of a thin wire. We give the relevant formulas in the next definition, followed by an example.

Definitie 18.2 (Mass and centre of mass of a thin wire)
Let a thin wire lie along a smooth curve C with continuous density function $\delta(s)$, where s is the arc length parameter.

1. The **mass** (*massa*) of the thin wire is $M = \int_C \delta(s)\ ds$.
2. The **moment about the** yz**-plane** is $M_{yz} = \int_C x\delta(s)\ ds$.
3. The **moment about the** xz**-plane** is $M_{xz} = \int_C y\delta(s)\ ds$.
4. The **moment about the** xy**-plane** is $M_{xy} = \int_C z\delta(s)\ ds$.
5. The **centre of mass** (*massamiddelpunt*) of the wire is

$$(\overline{x}, \overline{y}, \overline{z}) = \left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right).$$

Example 18.3

A thin wire follows the path $\vec{r}(t) = \big(1+\cos(t), 1+\sin(t), 1+\sin(2t)\big)$, $0 \leq t \leq 2\pi$. The density of the wire is determined by its position in space: $\delta(x,y,z) = y+z$ g/cm. The wire is shown in Figure 18.4, where a light colour indicates low density and a dark colour represents high density. Find the mass and centre of mass of the wire.

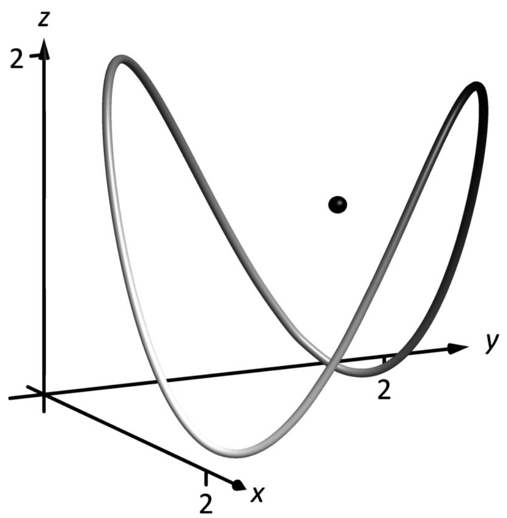


Figure 18.4: Finding the mass of a thin wire in Example 18.3.

Solution

We compute the density of the wire as

$$\delta(x,y,z) = \delta\big(1+\cos(t), 1+\sin(t), 1+\sin(2t)\big) = 2+\sin(t)+\sin(2t).$$

We compute ds as

$$ds = \left\|\vec{r}\,'(t)\right\|\ dt = \sqrt{\sin^2(t)+\cos^2(t)+4\cos^2(2t)}\ dt = \sqrt{1+4\cos^2(2t)}\ dt.$$

Thus the mass is

$$M = \oint_C \delta(s)\ ds = \int_0^{2\pi} \left(2+\sin(t)+\sin(2t)\right)\sqrt{1+4\cos^2(2t)}\ dt \approx 21.08 \text{ g}\,.$$

We compute the moments about the coordinate planes:

$$M_{yz} = \oint_C x\delta(s)\ ds = \int_0^{2\pi} \left(1+\cos(t)\right)\left(2+\sin(t)+\sin(2t)\right)\sqrt{1+4\cos^2(2t)}\ dt \approx 21.08\,,$$

$$M_{xz} = \oint_C y\delta(s)\ ds = \int_0^{2\pi} \left(1+\sin(t)\right)\left(2+\sin(t)+\sin(2t)\right)\sqrt{1+4\cos^2(2t)}\ dt \approx 26.35\,,$$

$$M_{xy} = \oint_C z\delta(s)\ ds = \int_0^{2\pi} \left(1+\sin(2t)\right)\left(2+\sin(t)+\sin(2t)\right)\sqrt{1+4\cos^2(2t)}\ dt \approx 25.40.$$

Thus the center of mass of the wire is located at

$$(\overline{x},\overline{y},\overline{z}) = \left(\frac{M_{yz}}{M},\frac{M_{xz}}{M},\frac{M_{xy}}{M}\right) \approx (1, 1.25, 1.20),$$

as indicated by the dot in Figure 18.4. Note how in this example, the curve C is "centered" about the point $(1,1,1)$, though the variable density of the wire pulls the center of mass out along the y- and z-axes.

In the following, we investigate a new mathematical object, the vector field, after which we increase our understanding of integration in the context of vector fields.

18.2 Vector fields

18.2.1 Definition

We have studied functions, where the input of such functions is a point and the output is a number. We could also create functions where the input is a point, but the output is a vector. For instance, we could create the following function: $\vec{\boldsymbol{F}}(x,y) = (x+y, x-y)$, where $\vec{\boldsymbol{F}}(2,3) = (5,-1)$. We are to think of $\vec{\boldsymbol{F}}$ assigning the vector $(5,-1)$ to the point $(2,3)$; in some sense, the vector $(5,-1)$ lies at the point $(2,3)$.

Such functions are extremely useful in any context where magnitude and direction are important. For instance, we could create a function $\vec{\boldsymbol{F}}$ that represents the electromagnetic force exerted at a point by a electromagnetic field, or the velocity of air as it moves across an airfoil.

Because these functions are so important, we need to formally define them.

Definitie 18.3 (Vector field)

1. A **vector field in the plane** (*vectorveld in het vlak*) is a function $\vec{\boldsymbol{F}}(x,y)$ whose domain is a subset of $\mathbb{R}^2$ and whose output is a two–dimensional vector:

$$\vec{\boldsymbol{F}}(x,y) = \big(M(x,y), N(x,y)\big)\,.$$

2. A **vector field in n-dimensional space** (*vectorveld in de n-dimensionale ruimte*) is a function $\vec{\boldsymbol{F}}(\mathbf{x})$ whose domain is a subset of $\mathbb{R}^n$ and whose output is a n–dimensional vector:

$$\vec{\boldsymbol{F}}(\mathbf{x}) = \big(M_1(\mathbf{x}), M_2(\mathbf{x}), \ldots, M_n(\mathbf{x})\big).$$

This definition may seem odd at first, as a special type of function is called a field. However, as the function determines a field of vectors, we can say the field is defined by the function, and thus the field is a function.

When graphing a vector field in the plane, the general idea is to draw the vector $\vec{\boldsymbol{F}}(x,y)$ at the point (x,y). For instance, using $\vec{\boldsymbol{F}}(x,y) = (x+y, x-y)$ as before, at $(1,1)$ we would draw $(2,0)$.

In Figure 18.5(a), one can see that the vector $(2,0)$ is drawn starting from the point $(1,1)$. A total of 8 vectors are drawn, with the x- and y-values of $-1,0,1$. In many ways, the resulting graph is a mess. In Figure 18.5(b), the same field is redrawn with each vector $\vec{\boldsymbol{F}}(x,y)$ drawn centred on the point (x,y). This makes for a better looking image, though when one vector intersects another, the image looks cluttered.

A common way to address this problem is limit the length of each arrow, and represent long vectors with thick arrows, as done in Figure 18.5(c). Usually we do not use a graph of a vector field to determine exactly the magnitude of a particular vector. Rather, we are more concerned with the relative magnitudes of vectors: which are bigger than others? Thus limiting the length of the vectors is not problematic. Mathematica obviously allows us to plot many vectors in a vector field nicely; in Figure 18.5(d), we see the same vector field drawn with using the Mathematica command `VectorPlot`, and finally get a clear picture of how this vector field behaves.If this vector field represented the velocity of air moving across a flat surface, we could see that the air tends to move either to the upper–right or lower–left, and moves very slowly near the origin. We can similarly plot vector fields in space, though the plots get very busy very quickly.

18.2.2 The del operator

Often, we will drop the of x, y and z portions of the notation in Definition 18.3 and refer to vector fields in the plane and in space as

$$\vec{\boldsymbol{F}} = (M,N) \quad \text{and} \quad \vec{\boldsymbol{F}} = (M,N,P),$$

respectively, as this shorthand is quite convenient.

Another item of notation will become useful: the del operator. Recall in Section 16.6 how we used the symbol $\vec{\nabla}$ to represent the gradient of a function of two variables. We now define $\vec{\nabla}$ to be the del operator. It is a vector whose components are partial derivative operations.

In the plane,

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right);$$

in space,

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right).$$

With this definition of $\vec{\nabla}$, we can better understand the gradient $\vec{\nabla} f$. As f returns a scalar, the properties of scalar and vector multiplication gives

$$\vec{\nabla} f = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) f = \left(\frac{\partial}{\partial x} f, \frac{\partial}{\partial y} f\right) = \left(f_x, f_y\right).$$

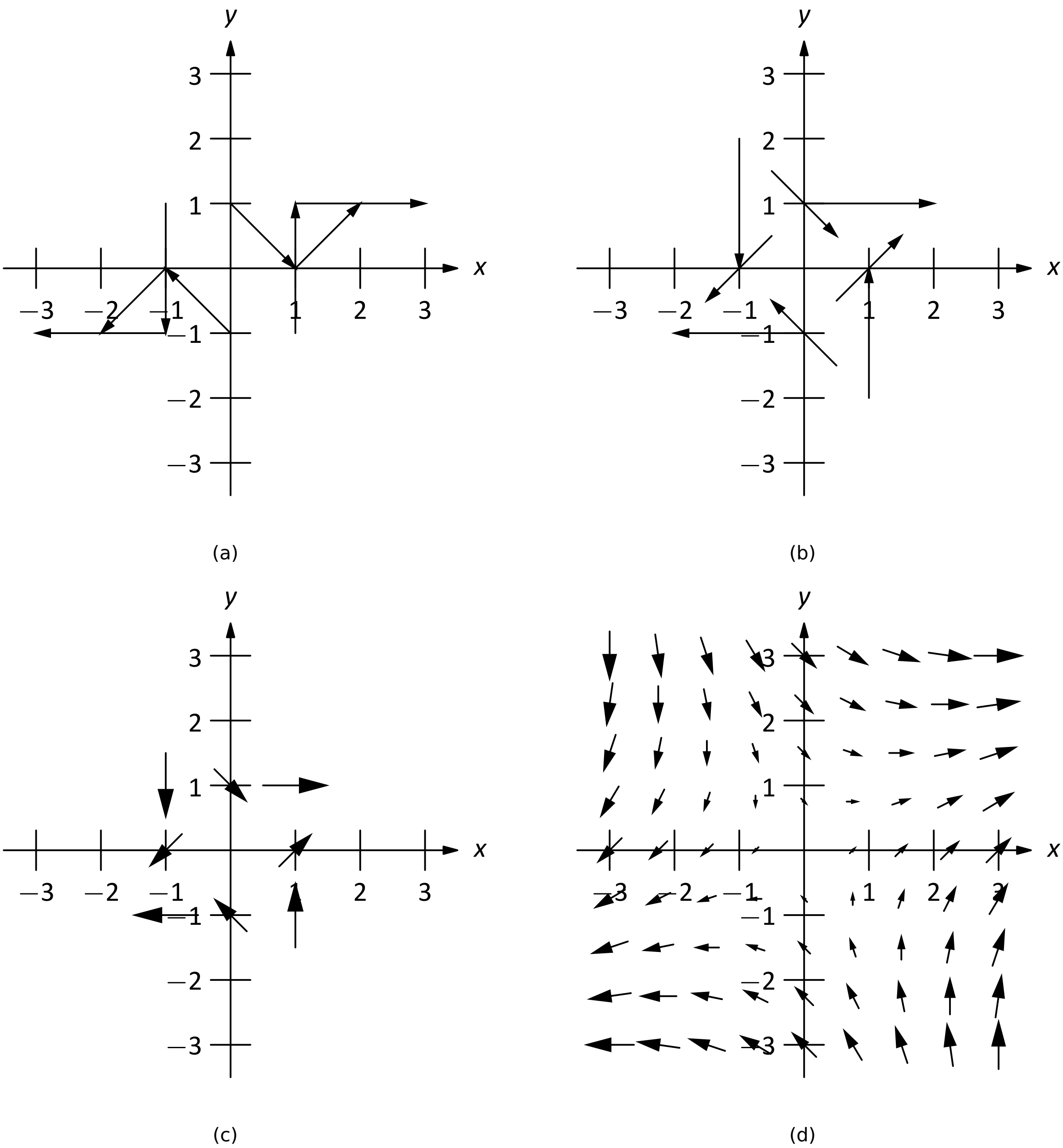


Figure 18.5: Demonstrating methods of graphing vector fields.

Now apply the del operator $\vec{\nabla}$ to vector fields. Let $\vec{F} = \left(x+\sin(y), y^2+z, x^2\right)$. We can use vector operations and find the dot product of $\vec{\nabla}$ and $\vec{F}$:

$$\begin{aligned}\vec{\nabla}\cdot\vec{F} &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\cdot\left(x+\sin(y), y^2+z, x^2\right)\\ &= \frac{\partial}{\partial x}\big(x+\sin(y)\big)+\frac{\partial}{\partial y}\left(y^2+z\right)+\frac{\partial}{\partial z}\left(x^2\right)\\ &= 1+2y.\end{aligned}$$

We can also compute their cross products:

$$\vec{\nabla}\times\vec{F} = \left(\frac{\partial}{\partial y}\left(x^2\right)-\frac{\partial}{\partial z}\left(y^2+z\right), \frac{\partial}{\partial z}\big(x+\sin(y)\big)-\frac{\partial}{\partial x}\left(x^2\right), \frac{\partial}{\partial x}\left(y^2+z\right)-\frac{\partial}{\partial y}\big(x+\sin(y)\big)\right)$$

$$= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x+\sin(y) & y^2+z & x^2 \end{vmatrix}$$

$$= \big(-1, -2x, -\cos(y)\big).$$

As we next learn about properties of vector fields, we will see how these dot and cross products with the del operator are quite useful.

18.2.3 Divergence and curl

Two properties of vector fields will prove themselves to be very important: divergence and curl. Each is a special "derivative" of a vector field; that is, each measures an instantaneous rate of change of a vector field.

If the vector field represents the velocity of a fluid or gas, then the divergence of the field is a measure of the compressibility of the fluid. If the divergence is negative at a point, it means that the fluid is compressing: more fluid is going into the point than is going out. If the divergence is positive, it means the fluid is expanding: more fluid is going out at that point than going in. A divergence of zero means the same amount of fluid is going in as is going out. If the divergence is zero at all points, we say the field is incompressible.

It turns out that the proper measure of divergence is simply $\vec{\nabla}\cdot\vec{F}$, as stated in the following definition.

Definitie 18.4 (Divergence)
The **divergence of a vector field $\vec{F}$** (*divergentie van een vectorveld*) is

$$\operatorname{div}\vec{F} = \vec{\nabla}\cdot\vec{F}.$$

- In the plane, with $\vec{F} = (M, N)$: $\operatorname{div}\vec{F} = M_x + N_y$.
- In space, with $\vec{F} = (M, N, P)$: $\operatorname{div}\vec{F} = M_x + N_y + P_z$.

Curl is a measure of the spinning action of the field. Let $\vec{F}$ represent the flow of water over a flat surface. If a small round cork were held in place at a point in the water, would the water cause the cork to spin? No spin corresponds to zero curl; counterclockwise spin corresponds to positive curl and clockwise spin corresponds to negative curl.

In space, things are a bit more complicated. Again let $\vec{F}$ represent the flow of water, and imagine suspending a tennis ball in one location in this flow. The water may cause the ball to spin along an axis. If so, the curl of the vector field is a vector (not a scalar, as before), parallel to the axis of rotation, following a right hand rule: when the thumb of one's right hand points in the direction of the curl, the ball will spin in the direction of the curling fingers of the hand.

In space, it turns out the proper measure of curl is $\vec{\nabla}\times\vec{F}$, as stated in the following definition. To find the curl of a planar vector field $\vec{F} = (M, N)$, embed it into space as $\vec{F} = (M, N, 0)$ and apply the cross product definition. Since M and N are functions of just x and y (and not z), all partial derivatives with respect to z become 0 and the result is simply $\big(0, 0, N_x - M_y\big)$. The third component is the measure of curl of a planar vector field.

Definitie 18.5 (Curl)

- Let $\vec{F} = (M, N)$ be a vector field in the plane. The **curl of $\vec{F}$** (*rotatie of rotor van* $\vec{F}$) is

$$\operatorname{curl} \vec{F} = N_x - M_y.$$

- Let $\vec{F} = (M, N, P)$ be a vector field in space. The **curl of $\vec{F}$** (*rotatie of rotor van* $\vec{F}$) is

$$\operatorname{curl} \vec{F} = \vec{\nabla} \times \vec{F} = \big(P_y - N_z,\, M_z - P_x,\, N_x - M_y\big).$$

We adopt the convention of referring to curl as $\vec{\nabla} \times \vec{F}$, regardless of whether $\vec{F}$ is a vector field in two or three dimensions.

We now practice computing these quantities.

Example 18.4

For each of the planar vector fields given below, view its graph and try to visually determine if its divergence and curl are 0. Then compute the divergence and curl.

1. $\vec{F} = (y, 0)$ (see Figure 18.6(a))
2. $\vec{F} = (-y, x)$ (see Figure 18.6(b))
3. $\vec{F} = (x, y)$ (see Figure 18.6(c))
4. $\vec{F} = \big(\cos(y), \sin(x)\big)$ (see Figure 18.6(d))

Solution

1. The arrow sizes are constant along any horizontal line, so if one were to draw a small box anywhere on the graph, it would seem that the same amount of fluid would enter the box as exit. Therefore it seems the divergence is zero; it is, as

$$\operatorname{div} \vec{F} = \vec{\nabla} \cdot \vec{F} = M_x + N_y = \frac{\partial}{\partial x}(y) + \frac{\partial}{\partial y}(0) = 0.$$

   At any point on the x-axis, arrows above it move to the right and arrows below it move to the left, indicating that a cork placed on the axis would spin clockwise. A cork placed anywhere above the x-axis would have water above it moving to the right faster than the water below it, also creating a clockwise spin. A clockwise spin also appears to be created at points below the x-axis. Thus it seems the curl should be negative (and not zero). Indeed, it is:

$$\operatorname{curl} \vec{F} = \vec{\nabla} \times \vec{F} = N_x - M_y = \frac{\partial}{\partial x}(0) - \frac{\partial}{\partial y}(y) = -1.$$

2. It appears that all vectors that lie on a circle of radius r, centered at the origin, have the same length (and indeed this is true). That implies that the divergence should be zero: draw any box on the graph, and any fluid coming in will lie along a circle that takes the same amount of fluid out. Indeed, the divergence is zero, as

$$\operatorname{div} \vec{F} = \vec{\nabla} \cdot \vec{F} = M_x + N_y = \frac{\partial}{\partial x}(-y) + \frac{\partial}{\partial y}(x) = 0.$$

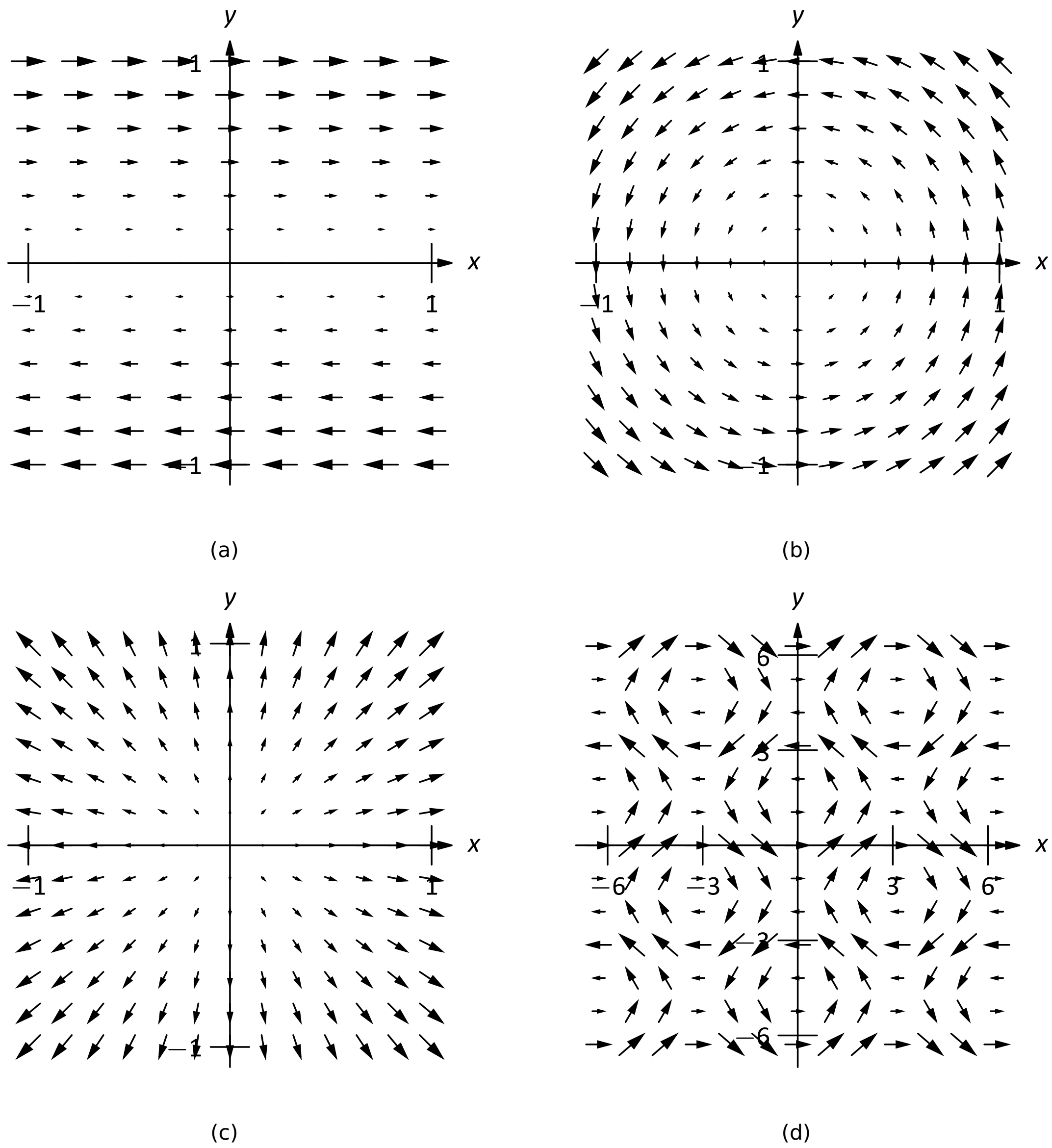


Figure 18.6: The vector fields in parts 1 (a), 2 (b), 3 (c) and 4 (d) in Example 18.4.

Clearly this field moves objects in a circle, but would it induce a cork to spin? It appears that yes, it would: place a cork anywhere in the flow, and the point of the cork closest to the origin would feel less flow than the point on the cork farthest from the origin, which would induce a counterclockwise flow. Indeed, the curl is positive:

$$\operatorname{curl}\vec{F} = \vec{\nabla}\times\vec{F} = N_x - M_y = \frac{\partial}{\partial x}(x) - \frac{\partial}{\partial y}(-y) = 1-(-1) = 2.$$

Since the curl is constant, we conclude the induced spin is the same no matter where one is in this field.

3. At the origin, there are many arrows pointing out but no arrows pointing in. We conclude that at the origin, the divergence must be positive (and not zero). If one were to draw a box anywhere in the field, the edges farther from the origin would have larger arrows passing through them than the edges close to the origin, indicating that more is going from a point

than going in. This indicates a positive (and not zero) divergence. This is correct:

$$\operatorname{div}\vec{F} = \vec{\nabla}\cdot\vec{F} = M_x + N_y = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) = 1 + 1 = 2.$$

One may find this curl to be harder to determine visually than previous examples. One might note that any arrow that induces a clockwise spin on a cork will have an equally sized arrow inducing a counterclockwise spin on the other side, indicating no spin and no curl. This is correct, as

$$\operatorname{curl}\vec{F} = \vec{\nabla}\times\vec{F} = N_x - M_y = \frac{\partial}{\partial x}(y) - \frac{\partial}{\partial y}(x) = 0.$$

4. One might find this divergence hard to determine visually as large arrows appear in close proximity to small arrows, each pointing in different directions. Instead of trying to rationalize a guess, we compute the divergence:

$$\operatorname{div}\vec{F} = \vec{\nabla}\cdot\vec{F} = M_x + N_y = \frac{\partial}{\partial x}\big(\cos(y)\big) + \frac{\partial}{\partial y}\big(\sin(x)\big) = 0.$$

Perhaps surprisingly, the divergence is 0.

With all the loops of different directions in the field, one is apt to reason the curl is variable. Indeed, it is:

$$\operatorname{curl}\vec{F} = \vec{\nabla}\times\vec{F} = N_x - M_y = \frac{\partial}{\partial x}\big(\sin(x)\big) - \frac{\partial}{\partial y}\big(\cos(y)\big) = \cos(x) + \sin(y).$$

Depending on the values of x and y, the curl may be positive, negative, or zero.

Example 18.5

The force of gravity between two objects is inversely proportional to the square of the distance between the objects. Locate a point mass at the origin. Create a vector field $\vec{F}$ that represents the gravitational pull of the point mass at any point (x, y, z). Find the divergence and curl of this field.

Solution

The point mass pulls toward the origin, so at (x, y, z), the force will pull in the direction of $(-x, -y, -z)$. To get the proper magnitude, it will be useful to find the unit vector in this direction. Dividing by its magnitude, we have

$$\vec{u} = \left(\frac{-x}{\sqrt{x^2+y^2+z^2}}, \frac{-y}{\sqrt{x^2+y^2+z^2}}, \frac{-z}{\sqrt{x^2+y^2+z^2}}\right).$$

The magnitude of the force is inversely proportional to the square of the distance between the two points. Letting k be the constant of proportionality, we have the magnitude as

$$\frac{k}{x^2+y^2+z^2}.$$

Multiplying this magnitude by the unit vector above, we have the desired vector field:

$$\vec{F} = \left(\frac{-kx}{(x^2+y^2+z^2)^{3/2}}, \frac{-ky}{(x^2+y^2+z^2)^{3/2}}, \frac{-kz}{(x^2+y^2+z^2)^{3/2}}\right).$$

We leave it to the reader to confirm that div $\vec{F} = 0$ and curl $\vec{F} = \vec{0}$.

The analogous planar vector field is given in Figure 18.7. Note how all arrows point to the origin, and the magnitude gets very small when far from the origin.

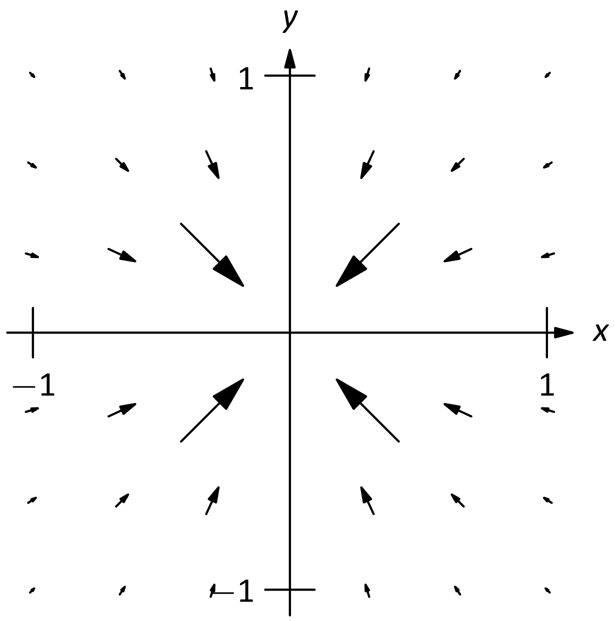


Figure 18.7: A vector field representing a planar gravitational force.

A function $z = f(x,y)$ naturally induces a vector field, $\vec{F} = \vec{\nabla} f = \left(f_x, f_y\right)$. Given what we learned of the gradient in Section 16.6, we know that the vectors of $\vec{F}$ point in the direction of greatest increase of f. Because of this, f is said to be the **potential function of $\vec{F}$**. Vector fields that are the gradient of potential functions will play an important role in the remainder of this section.

The last part of this section applies calculus to vector fields. A common application is this: let $\vec{F}$ be a vector field representing a force (hence it is called a force field) and let a particle move along a curve C under the influence of this force. What work is performed by the field on this particle? The solution lies in correctly applying the concepts of line integrals in the context of vector fields.

18.3 Line Integrals over vector fields

18.3.1 Definition

Suppose a particle moves along a curve C under the influence of an electromagnetic force described by a vector field $\vec{F}$. Since a force is inducing motion, work is performed. How can we calculate how much work is performed?

Recall that when moving in a straight line, if $\vec{F}$ represents a constant force and $\vec{d}$ represents the direction and length of travel, then work is simply $W = \vec{F} \cdot \vec{d}$. However, we generally want to be able to calculate work even if $\vec{F}$ is not constant and C is not a straight line.

As we have practised many times before, we can calculate work by first approximating, then refining our approximation through a limit that leads to integration.

Assume as we did at the beginning of this section, C can be parametrized by the arc length parameter s. Over a short piece of the curve with length ds, the curve is approximately straight and our force is approximately constant. The straight–line direction of this short length of curve is given by $\widehat{\boldsymbol{T}}$, the unit tangent vector; let $\vec{\boldsymbol{d}} = \widehat{\boldsymbol{T}}\,ds$, which gives the direction and magnitude of a small section of C. Thus work over this small section of C is $\vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{d}} = \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\,ds$.

Summing up all the work over these small segments gives an approximation of the work performed. By taking the limit as ds goes to zero, and hence the number of segments approaches infinity, we can obtain the exact amount of work. Hence, we see that

$$W = \int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\,ds,$$

is a line integral.

This line integral is beautiful in its simplicity, yet is not so useful in making actual computations (largely because the arc length parameter is so difficult to work with). To compute actual work, we need to parametrize C with another parameter t via a vector–valued function $\vec{\boldsymbol{r}}(t)$. Since $ds = \left\|\vec{\boldsymbol{r}}\,'(t)\right\|\,dt$ and $\widehat{\boldsymbol{T}} = \vec{\boldsymbol{r}}\,'(t)/\left\|\vec{\boldsymbol{r}}\,'(t)\right\|$, we get

$$W = \int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\;ds = \int_C \vec{\boldsymbol{F}} \cdot \frac{\vec{\boldsymbol{r}}\,'(t)}{\left\|\vec{\boldsymbol{r}}\,'(t)\right\|}\;\left\|\vec{\boldsymbol{r}}\,'(t)\right\|\;dt = \int_C \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{r}}\,'(t)\;dt = \int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \tag{18.2}$$

where the final integral uses the differential $d\vec{\boldsymbol{r}}$ for $\vec{\boldsymbol{r}}\,'(t)\,dt$.

These integrals are known as line integrals over vector fields. By contrast, the line integrals we dealt with earlier are sometimes referred to as line integrals over scalar fields (Definition 18.1). Just as a vector field is defined by a function that returns a vector, a scalar field is a function that returns a scalar, such as $z = f(x, y)$.

We formally define this line integral, then give examples and applications.

Definitie 18.6 (Line integral over a vector field)
Let $\vec{\boldsymbol{F}}$ be a vector field with continuous components defined on a smooth curve C, parametrized by $\vec{\boldsymbol{r}}(t)$, and let $\widehat{\boldsymbol{T}}$ be the unit tangent vector of $\vec{\boldsymbol{r}}(t)$. The **line integral over $\vec{\boldsymbol{F}}$** along C is

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\;ds.$$

In Definition 18.6, note how the dot product $\vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}$ is just a scalar. Therefore, this new line integral is really just a special kind of line integral. Indeed, letting $f(s) = \vec{\boldsymbol{F}}(s) \cdot \widehat{\boldsymbol{T}}(s)$, the right–hand side simply becomes $\int_C f(s)\;ds$, and we can use the corresponding techniques to evaluate the integral. This is summarized in the following theorem.

Theorem 18.2 (Evaluating a line integral over a vector field)
Let $\vec{\boldsymbol{F}}$ be a vector field with continuous components defined on a smooth curve C, parametrized by $\vec{\boldsymbol{r}}(t)$, $a \le t \le b$, where $\vec{\boldsymbol{r}}$ is continuously differentiable. Then

$$\int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\;ds = \int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_a^b \vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}(t)\big) \cdot \vec{\boldsymbol{r}}\,'(t)\;dt.$$

This theorem indicates that we can use any continuously differentiable parametrization $\vec{\boldsymbol{r}}(t)$ of C that preserves the orientation of C: there isn't a right one. In practice, choose one that seems easy to work with.

Note that the above definition and theorem implicitly evaluate $\vec{\boldsymbol{F}}$ along the curve C, which is parametrized by $\vec{\boldsymbol{r}}(t)$. For instance, if $\vec{\boldsymbol{F}} = (x+y, x-y)$ and $\vec{\boldsymbol{r}}(t) = \left(t^2, \cos(t)\right)$, then evaluating $\vec{\boldsymbol{F}}$ along C means substituting the x- and y-components of $\vec{r}(t)$ in for x and y, respectively, in $\vec{F}$. Therefore, along C, $\vec{F}=(x+y,x-y)=\left(t^2+\cos(t),t^2-\cos(t)\right)$. Since we are substituting the output of $\vec{r}(t)$ for the input of $\vec{F}$, we write this as $\vec{F}\left(\vec{r}(t)\right)$. This is a slight abuse of notation as technically the input of $\vec{F}$ is to be a point, not a vector, but this shorthand is useful.

Example 18.6

Two particles move from $(0,0)$ to $(1,1)$ under the influence of the force field $\vec{F}=(x,x+y)$. One particle follows C_1, the line $y=x$; the other follows C_2, the curve $y=x^4$, as shown in Figure 18.8. Force is measured in Newtons and distance is measured in meters. Find the work performed by each particle.

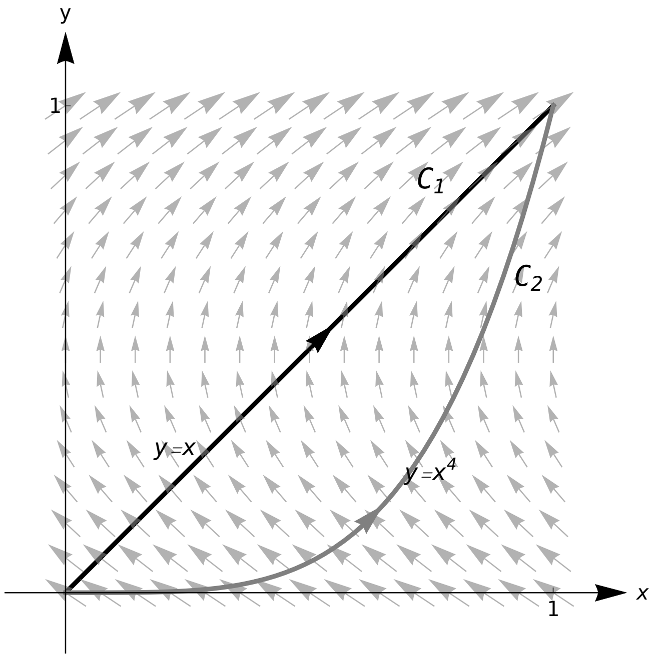


Figure 18.8: Paths through a vector field in Example 18.6.

Solution

To compute work, we need to parametrize each path. We use $\vec{r}_1(t)=(t,t)$ to parametrize $y=x$, and let $\vec{r}_2(t)=\left(t,t^4\right)$ parametrize $y=x^4$; for each, $0\le t\le 1$.

Along the straight–line path, $\vec{F}\left(\vec{r}_1(t)\right)=(x,x+y)=(t,t+t)=(t,2t)$. We find $\vec{r}\,'_1(t)=(1,1)$. The integral that computes work is:

$$\begin{aligned}\int_{C_1}\vec{F}\cdot d\vec{r}&=\int_0^1 (t,2t)\cdot(1,1)\ dt\\&=\int_0^1 3t\ dt\\&=\left(\frac{3}{2}t^2\right)\Big|_0^1=1.5\text{ joules.}\end{aligned}$$

Along the curve $y = x^4$, we have that

$$\vec{F}\big(\vec{r}_2(t)\big) = (x, x+y) = \left(t, t+t^4\right).$$

We find $\vec{r}\,'_2(t) = \left(1, 4t^3\right)$. The work performed along this path is

$$\begin{aligned}\int_{C_2} \vec{F}\cdot d\vec{r} &= \int_0^1 \left(t, t+t^4\right)\cdot\left(1, 4t^3\right)\, dt\\ &= \int_0^1 \left(t+4t^4+4t^7\right) dt\\ &= \left(\frac{1}{2}t^2+\frac{4}{5}t^5+\frac{1}{2}t^8\right)\Bigg|_0^1 = \frac{9}{5} \text{ joules.}\end{aligned}$$

Note how differing amounts of work are performed along the different paths. This should not be too surprising: the force is variable, one path is longer than the other, etc.

Example 18.7

Two particles move from $(-1, 1)$ to $(1, 1)$ under the influence of a force field $\vec{F} = (y, x)$. One moves along the curve C_1, the parabola defined by $y = 2x^2 - 1$. The other particle moves along the curve C_2, the bottom half of the circle defined by $x^2 + (y-1)^2 = 1$, as shown in Figure 18.9. Force is measured in Newton and distances are measured in meters. Find the work performed by moving each particle along its path.

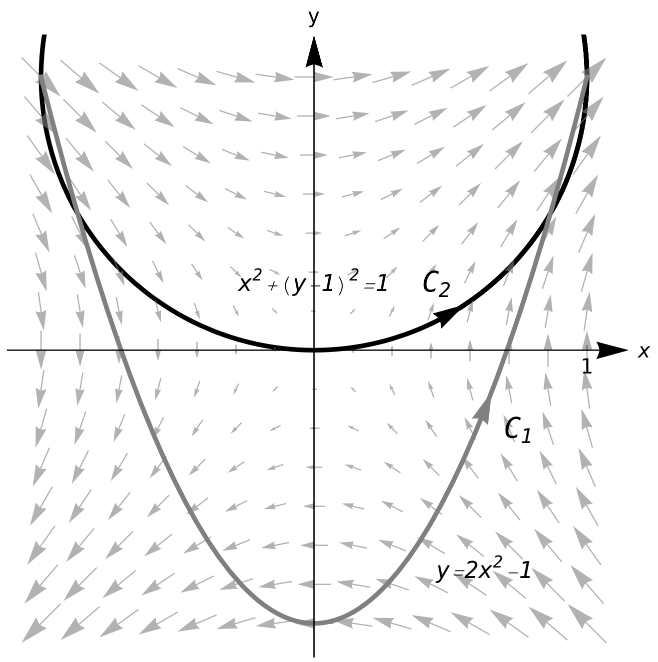


Figure 18.9: Paths through a vector field in Example 18.7.

Solution

We start by parametrizing C_1: the parametrization $\vec{r}_1(t) = \left(t, 2t^2-1\right)$ is straightforward, giving

$\vec{r}\,'_1 = (1, 4t)$. On C_1, $\vec{F}\big(\vec{r}_1(t)\big) = (y, x) = \big(2t^2 - 1, t\big)$.

Computing the work along C_1, we have:

$$\begin{aligned}\int_{C_1} \vec{F} \cdot d\vec{r}_1 &= \int_{-1}^{1} \big(2t^2 - 1, t\big) \cdot (1, 4t)\ dt \\ &= \int_{-1}^{1} \big(2t^2 - 1 + 4t^2\big)\, dt = 2 \text{ joules.}\end{aligned}$$

For C_2, it is probably simplest to parametrize the half circle using sine and cosine. Recall that $\vec{r}(t) = \big(\cos(t), \sin(t)\big)$ is a parametrization of the unit circle on $0 \le t \le 2\pi$; we add 1 to the second component to shift the circle up one unit, then restrict the domain to $\pi \le t \le 2\pi$ to obtain only the lower half, giving $\vec{r}_2(t) = \big(\cos(t), \sin(t) + 1\big)$, $\pi \le t \le 2\pi$, and hence $\vec{r}\,'_2(t) = \big(-\sin(t), \cos(t)\big)$ and $\vec{F}\big(\vec{r}_2(t)\big) = (y, x) = \big(\sin(t) + 1, \cos(t)\big)$.

Computing the work along C_2, we have:

$$\begin{aligned}\int_{C_2} \vec{F} \cdot d\vec{r}_2 &= \int_{\pi}^{2\pi} \big(\sin(t) + 1, \cos(t)\big) \cdot \big(-\sin(t), \cos(t)\big)\ dt \\ &= \int_{\pi}^{2\pi} \big(-\sin^2(t) - \sin(t) + \cos^2(t)\big)\, dt = 2 \text{ joules.}\end{aligned}$$

Note how the work along C_1 and C_2 in this example is the same.

18.3.2 Properties

Line integrals over vector fields share the same properties as line integrals over scalar fields, with one important distinction. The orientation of the curve C matters with line integrals over vector fields, whereas it did not matter with line integrals over scalar fields.

It is relatively easy to see why. Let C be the unit circle. The area under a surface over C is the same whether we traverse the circle in a clockwise or counterclockwise fashion, hence the line integral over a scalar field on C is the same irrespective of orientation. On the other hand, if we are computing work done by a force field, direction of travel definitely matters. Opposite directions create opposite signs when computing dot products, so traversing the circle in opposite directions will create line integrals that differ by a factor of -1.

In summary, we have the following properties of line integrals over vector fields.

1. Let $\vec{F}$ and $\vec{G}$ be vector fields with continuous components defined on a smooth curve C, parametrized by $\vec{r}(t)$, and let k_1 and k_2 be scalars. Then

$$\int_C \big(k_1\vec{F} + k_2\vec{G}\big) \cdot d\vec{r} = k_1 \int_C \vec{F} \cdot d\vec{r} + k_2 \int_C \vec{G} \cdot d\vec{r}.$$

2. Let C be piecewise smooth, composed of smooth components C_1 and C_2. Then

$$\int_C \vec{F} \cdot d\vec{r} = \int_{C_1} \vec{F} \cdot d\vec{r} + \int_{C_2} \vec{F} \cdot d\vec{r}.$$

3. Let C^* be the curve C with opposite orientation, parametrized by $\vec{\boldsymbol{r}}^*$. Then

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = -\int_{C^*} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}^*.$$

We demonstrate using these properties in the following example.

Example 18.8

Let $\vec{\boldsymbol{F}} = \big(3(y-1/2), 1\big)$ and let C be the path that starts at $(0,0)$, goes to $(1,1)$ along the curve $y = x^3$, then returns to $(0,0)$ along the line $y = x$, as shown in Figure 18.10. Evaluate $\oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$.

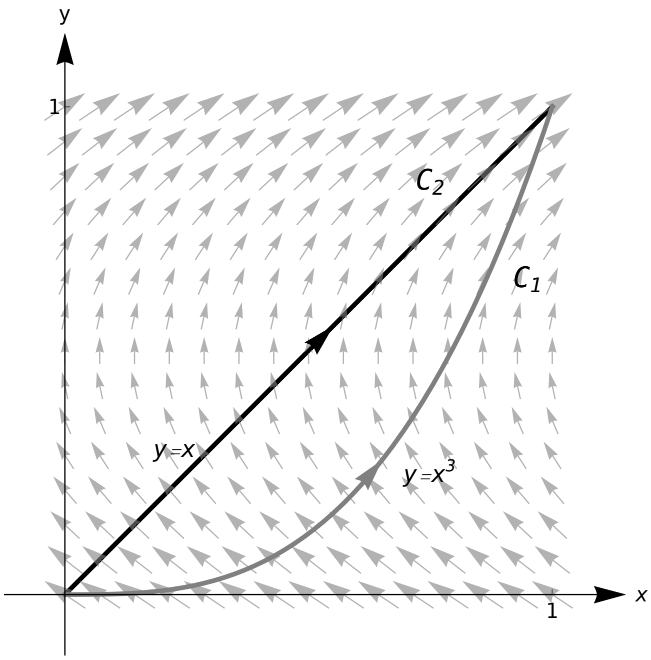


Figure 18.10: The vector field and curve in Example 18.8.

Solution

As C is piecewise smooth, we break it into two components C_1 and C_2, where C_1 follows the curve $y = x^3$ and C_2 follows the curve $y = x$.

We parametrize C_1 with $\vec{\boldsymbol{r}}_1(t) = \big(t, t^3\big)$ on $0 \le t \le 1$, with $\vec{\boldsymbol{r}}\,'_1(t) = \big(1, 3t^2\big)$. We will use $\vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}_1(t)\big) = \big(3(t^3 - 1/2), 1\big)$.

While we always have unlimited ways in which to parametrize a curve, there are two direct methods to choose from when parametrizing C_2. The parametrization $\vec{\boldsymbol{r}}_2(t) = (t, t)$, $0 \le t \le 1$ traces the correct line segment but with the wrong orientation. Relying on property (3) of line integrals over vector fields, we can use this parametrization and negate the result.

Another choice is to use the techniques of Section 7.1 to create the line with the orientation we desire. We wish to start at $(1,1)$ and travel in the $\vec{\boldsymbol{d}} = (-1,-1)$ direction for one length of $\vec{\boldsymbol{d}}$, giving equation $\vec{\boldsymbol{\ell}}(t) = (1,1) + t(-1,-1) = (1-t, 1-t)$ on $0 \le t \le 1$.

Either choice is fine; we choose $\vec{\boldsymbol{r}}_2(t)$ to practice using line integral properties. We find $\vec{\boldsymbol{r}}\,'_2(t) = (1,1)$ and $\vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}_2(t)\big) = \big(3(t-1/2), 1\big)$.

Evaluating the line integral:

$$\begin{aligned}\oint_C \vec{F}\cdot d\vec{r} &= \int_{C_1} \vec{F}\cdot d\vec{r}_1 - \int_{C_2} \vec{F}\cdot d\vec{r}_2\\ &= \int_0^1 \left(3\left(t^3-\frac{1}{2}\right),1\right)\cdot\left(1,3t^2\right)dt - \int_0^1 \left(3\left(t-\frac{1}{2}\right),1\right)\cdot(1,1)\,dt\\ &= \int_0^1 \left(3t^3+3t^2-\frac{3}{2}\right)\,dt - \int_0^1 \left(3t-\frac{1}{2}\right)\,dt\\ &= \frac{1}{4}-1=-\frac{3}{4}.\end{aligned}$$

If we interpret this integral as computing work, the negative work implies that the motion is mostly against the direction of the force, which seems plausible when we look at Figure 18.10.

18.3.3 The fundamental theorem of line integrals

We are preparing to make important statements about the value of certain line integrals over special vector fields. Before we can do that, we need to define some terms that describe the domains over which a vector field is defined.

A region in the plane is **connected** (*samenhangend*) if any two points in the region can be joined by a piecewise smooth curve that lies entirely in the region. In Figure 18.11(a), sets R_1 and R_2 are connected; set R_3 is not connected, though it is composed of two connected subregions.

A region is **simply connected** (*enkelvoudig samenhangend*) if every simple closed curve that lies entirely in the region can be continuously deformed (shrunk) to a single point without leaving the region. (A curve is **simple** if it does not cross itself.) In Figure 18.11(a), only set R_1 is simply connected. Region R_2 is not simply connected as any closed curve that goes around the "hole" in R_2 cannot be continously shrunk to a single point. As R_3 is not even connected, it cannot be simply connected, though again it consists of two simply connected subregions.

We have applied these terms to regions of the plane, but they can be extended intuitively to domains in space (and hyperspace). In Figure 18.11(b), the domain bounded by the sphere (at left) and the domain with a subsphere removed (at right) are both simply connected. Any simple closed path that lies entirely within these domains can be continuously deformed into a single point. In Figure 18.11(c), neither domain is simply connected. At left, the ball has a hole that extends its length and the pictured closed path cannot be deformed to a point. At right, two paths are illustrated on the torus that cannot be shrunk to a point.

We will use the terms connected and simply connected in subsequent definitions and theorems.

Recall how in Example 18.7 particles moved from $A=(-1,1)$ to $B=(1,1)$ along two different paths, wherein the same amount of work was performed along each path. It turns out that regardless of the choice of path from A to B, the amount of work performed under the field $\vec{F}=(y,x)$ is the same. Since our expectation is that differing amounts of work are performed along different paths, we give such special fields a name.

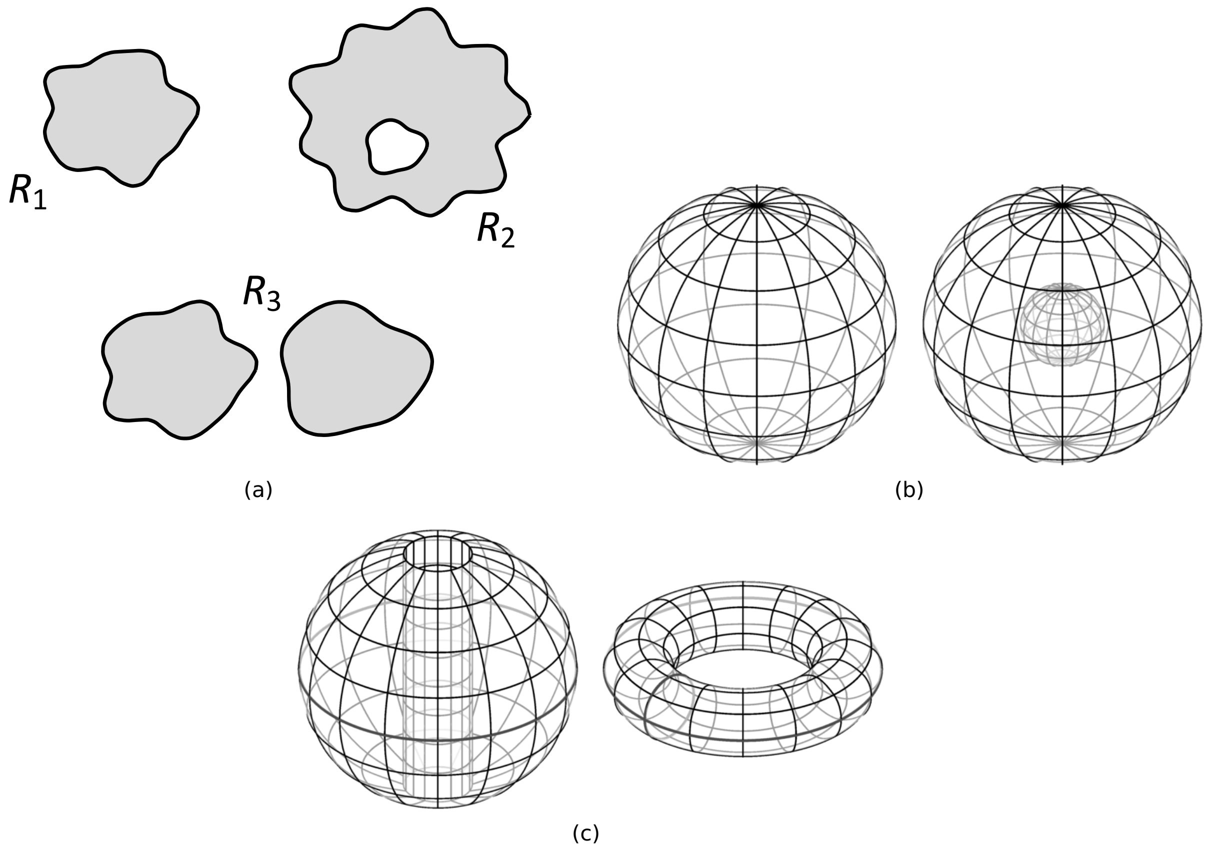


Figure 18.11: Different types of regions (a): R_1 is simply connected; R_2 is connected, but not simply connected; R_3 is not connected; simply connected domains (b) and not simply connected domains in (c).

Definitie 18.7 (Conservative field)
Let $\vec{F}$ be a vector field defined on an open, connected domain D containing points A and B. If the line integral $\int_C \vec{F} \cdot d\vec{r}$ has the same value for all choices of paths C starting at A and ending at B, and parametrized by $\vec{r}(t)$, then

- $\vec{F}$ is a **conservative field** (*conservatief vectorveld*) and
- The line integral $\int_C \vec{F} \cdot d\vec{r}$ is path independent and can be written as

$$\int_C \vec{F} \cdot d\vec{r} = \int_A^B \vec{F} \cdot\, d\vec{r}.$$

How can we tell if a field is conservative? To show a field $\vec{F}$ is conservative using the definition, we need to show that all line integrals from points A to B have the same value. It is equivalent to show that all line integrals over closed paths C are 0. Each of these tasks are generally nontrivial.

There is, however, a simpler method. Consider the surface defined by $z = f(x,y) = xy$. We can compute the gradient of this function: $\vec{\nabla} f = \left(f_x, f_y\right) = (y,x)$. Note that this is the field from Example 18.7, which we have claimed is conservative. We will soon give a theorem that states that a field $\vec{F}$ is conservative if, and only if, it is the gradient of some scalar function f. To show $\vec{F}$ is conservative, we need to determine whether or not $\vec{F} = \vec{\nabla} f$ for some function f. To recognize the special relationship between $\vec{F}$ and f in this situation, f is given a name.

Definitie 18.8 (Potential function)
Let f be a differentiable function defined on a domain D (e.g., $z = f(x, y)$ or $w = f(x, y, z)$) and let $\vec{\boldsymbol{F}} = \vec{\nabla} f$, the gradient of f. Then f is a **potential function** (*potentiaalfunctie*) of $\vec{\boldsymbol{F}}$.

We now state the fundamental theorem of line integrals, also known as the gradient theorem, which connects conservative fields and path independence to fields with potential functions.

Theorem 18.3 (Fundamental theorem of line integrals)
Let $\vec{\boldsymbol{F}}$ be a vector field whose components are continuous on a connected domain D, let A and B be any points in D, and let C be any path in D starting at A at $t = a$, ending at B at $t = b$ and parametrized by $\vec{\boldsymbol{r}}(t)$ such that $\vec{\boldsymbol{r}}(a) = A$ and $\vec{\boldsymbol{r}}(b) = B$.

1. *$\vec{\boldsymbol{F}}$ is conservative if and only if there exists a differentiable function f such that $\vec{\boldsymbol{F}} = \vec{\nabla} f$.*
2. *If $\vec{\boldsymbol{F}}$ is conservative, then*

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_A^B \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = f\left(\vec{\boldsymbol{r}}(b)\right) - f\left(\vec{\boldsymbol{r}}(a)\right) = f(B) - f(A).$$

Note that we did not specify the number of variables for the function since it is really immaterial to the theorem. The theorem will hold regardless of the number of variables in the function.

Proof For the purpose of the proof we will assume that we are working in three dimensions, but it can be done in any dimension. Let us then start by just computing the line integral.

$$\begin{aligned} \int_C \vec{\nabla} f \cdot d\vec{\boldsymbol{r}} &= \int_a^b \vec{\nabla} f\left(\vec{\boldsymbol{r}}(t)\right) \cdot \vec{\boldsymbol{r}}'(t)\, dt \\ &= \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt \end{aligned}$$

Now, at this point we can use the chain rule to simplify the integrand as follows,

$$\begin{aligned} \int_C \vec{\nabla} f \cdot d\vec{\boldsymbol{r}} &= \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt \\ &= \int_a^b \frac{d}{dt}\left[f\left(\vec{\boldsymbol{r}}(t)\right)\right] dt \end{aligned}$$

□

To finish this off we just need to use the fundamental theorem of calculus for single integrals.

$$\int_C \vec{\nabla} f \cdot d\vec{\boldsymbol{r}} = f\left(\vec{\boldsymbol{r}}(b)\right) - f\left(\vec{\boldsymbol{r}}(a)\right) = f(B) - f(A)$$

Once again considering Example 18.7, we have $A = (-1, 1)$, $B = (1, 1)$ and $\vec{\boldsymbol{F}} = (y, x)$. In that example, we evaluated two line integrals from A to B and found the value of each was 2. Note that $f(x, y) = xy$ is a potential function for $\vec{\boldsymbol{F}}$. Following the fundamental theorem of line integrals, consider $f(B) - f(A)$:

$$f(B) - f(A) = f(1, 1) - f(-1, 1) = 1 - (-1) = 2,$$

the same value given by the line integrals.

We practice using this theorem again in the next example.

Example 18.9

Let $\vec{F} = \big(3x^2y+2x, x^3+1\big)$, $A = (0,1)$ and $B = (1,4)$. Use the first part of the fundamental theorem of line integrals to show that $\vec{F}$ is conservative, then choose any path from A to B and confirm the second part of the theorem.

Solution

To show $\vec{F}$ is conservative, we need to find $z = f(x,y)$ such that $\vec{F} = \vec{\nabla}f = \big(f_x, f_y\big)$. That is, we need to find f such that $f_x = 3x^2y+2x$ and $f_y = x^3+1$. As all we know about f are its partial derivatives, we recover f by integration:

$$\int \frac{\partial f}{\partial x}\, dx = f(x,y)+K_1(y).$$

Note how the constant of integration $K_1(y)$ is more than just a constant: it is anything that acts as a constant when taking a derivative with respect to x. Any function that is a function of y (containing no x's) acts as a constant when deriving with respect to x.

Integrating f_x in this example gives:

$$\int \frac{\partial f}{\partial x}\, dx = \int (3x^2y+2x)\, dx = x^3y+x^2+K_1(y).$$

Likewise, integrating f_y with respect to y gives:

$$\int \frac{\partial f}{\partial y}\, dy = \int (x^3+1)\, dy = x^3y+y+K_2(x).$$

These two results should be equal with appropriate choices of $K_2(x)$ and $K_1(y)$:

$$x^3y+x^2+K_1(y) = x^3y+y+K_2(x) \quad \Rightarrow \quad K_2(x) = x^2 \quad \text{and} \quad K_1(y) = y.$$

We find $f(x,y) = x^3y+x^2+y$, a potential function of $\vec{F}$.

By the fundamental theorem of line integrals, regardless of the path from A to B,

$$\begin{aligned}\int_A^B \vec{F}\cdot d\vec{r} &= f(B)-f(A)\\ &= f(1,4)-f(0,1)\\ &= 9-1 = 8.\end{aligned}$$

To illustrate the validity of the Fundamental Theorem, we pick a path from A to B. The line between these two points would be simple to construct; we choose a slightly more complicated path by choosing the parabola $y = x^2+2x+1$. This leads to the parametrization $\vec{r}(t) = \big(t, t^2+2t+1\big)$, $0 \le t \le 1$, with $\vec{r}\,'(t) = (1, 2t+2)$. Thus

$$\begin{aligned}\int_C \vec{F}\cdot d\vec{r} &= \int_C \vec{F}\big(\vec{r}(t)\big)\cdot\vec{r}\,'(t)\, dt\\ &= \int_0^1 \Big(3t^2(t^2+2t+1)+2t, t^3+1\Big)\cdot(1, 2t+2)\, dt\end{aligned}$$

$$= \int_0^1 \left(5t^4 + 8t^3 + 3t^2 + 4t + 2\right)\, dt$$
$$= \left(t^5 + 2t^4 + t^3 + 2t^2 + 2t\right)\Big|_0^1$$
$$= 8,$$

which matches our previous result.

The fundamental theorem of line integrals states that we can determine whether or not $\vec{\boldsymbol{F}}$ is conservative by determining whether or not $\vec{\boldsymbol{F}}$ has a potential function. This can be difficult. A simpler method exists if the domain of $\vec{\boldsymbol{F}}$ is simply connected, which is a reasonable requirement. We state this simpler method as a theorem.

Theorem 18.4 (Curl of conservative fields)
Let $\vec{\boldsymbol{F}}$ be a vector field whose components are continuous on a simply connected domain D in the plane or in space. Then $\vec{\boldsymbol{F}}$ is conservative if and only if $\operatorname{curl}\vec{\boldsymbol{F}} = 0$ *or* $\vec{\boldsymbol{0}}$*, respectively.*

In Example 18.9, we showed that $\vec{\boldsymbol{F}} = \left(3x^2y + 2x, x^3 + 1\right)$ is conservative by finding a potential function for $\vec{\boldsymbol{F}}$. Using the above theorem, we can show that $\vec{\boldsymbol{F}}(M,N)$ is conservative much more easily by computing its curl:

$$\operatorname{curl}\vec{\boldsymbol{F}} = N_x - M_y = 3x^2 - 3x^2 = 0.$$

18.4 Green's theorem and the divergence theorem

18.4.1 Flow and flux

Line integrals over vector fields have the natural interpretation of computing work when $\vec{\boldsymbol{F}}$ represents a force field. It is also common to use vector fields to represent velocities. In these cases, the line integral $\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ is said to represent **flow** (*stroming*).

Let the vector field $\vec{\boldsymbol{F}} = (1, 0)$ represent the velocity of water as it moves across a smooth surface, depicted in Figure 18.12. A line integral over C will compute how much water is moving *along* the path C.

In the figure, all of the water above C_1 is moving along that curve, whereas none of the water above C_2 is moving along that curve (the curve and the flow of water are at right angles to each other). Because C_3 has nonzero horizontal and vertical components, some of the water above that curve is moving along the curve.

When C is a closed curve, we call flow **circulation** (*circulatie*), represented by $\oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$.

The opposite of flow is **flux**, a measure of how much water is moving *across* the path C. If a curve represents a filter in flowing water, flux measures how much water will pass through the filter. Considering again Figure 18.12, we see that a screen along C_1 will not filter any water as no water passes across that curve. Because of the nature of this field, C_2 and C_3 each filter the same amount of water per second.

The terms flow and flux are used apart from velocity fields, too. Flow is measured by $\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$, which is the same as $\int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\, ds$ by Definition 18.6. That is, flow is a summation of the amount of $\vec{\boldsymbol{F}}$ that is tangent to the curve C.

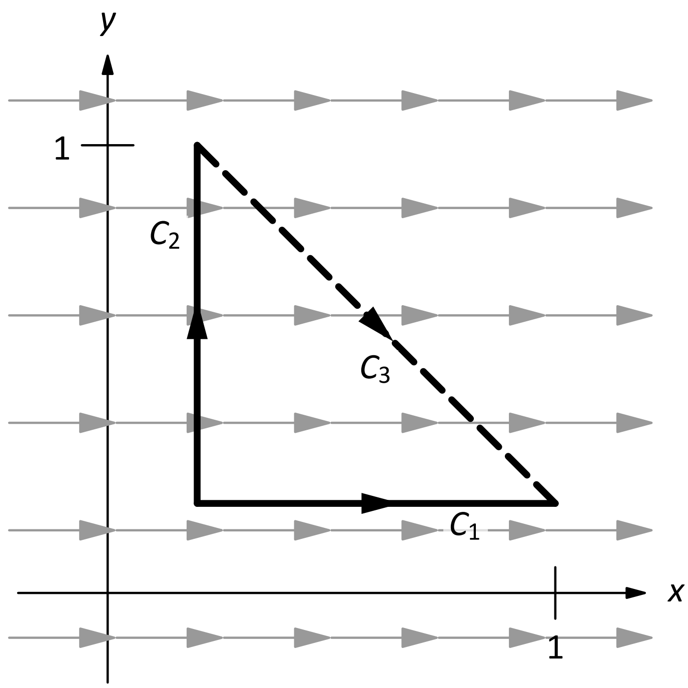


Figure 18.12: Illustrating the principles of flow and flux.

By contrast, flux is a summation of the amount of $\vec{F}$ that is *orthogonal* to the direction of travel. To capture this orthogonal amount of $\vec{F}$, we use $\int_C \vec{F}\cdot\hat{n}\ ds$ to measure flux, where $\hat{n}$ is a unit vector orthogonal to the curve C.

How is $\hat{n}$ determined? We'll later see that if C is a closed curve, we'll want $\hat{n}$ to point to the outside of the curve (measuring how much is going out). We'll also adopt the convention that closed curves should be traversed counterclockwise.

If C is a complicated closed curve, it can be difficult to determine what counterclockwise means. So we offer this definition: a closed curve is being traversed counterclockwise if the outside is to the right of the path and the inside is to the left.

When a curve C is traversed counterclockwise by $\vec{r}(t) = \big(f(t), g(t)\big)$, we rotate $\widehat{T}$ clockwise 90° to obtain $\hat{n}$:

$$\widehat{T} = \frac{\vec{r}\,'(t)}{\left\|\vec{r}\,'(t)\right\|} = \frac{\big(f'(t), g'(t)\big)}{\left\|\vec{r}\,'(t)\right\|} \quad \Rightarrow \quad \hat{n} = \frac{\big(g'(t), -f'(t)\big)}{\left\|\vec{r}\,'(t)\right\|}.$$

Letting $\vec{F} = (M, N)$, we calculate flux as:

$$\begin{aligned}\int_C \vec{F}\cdot\hat{n}\ ds &= \int_C \vec{F}\cdot\frac{\big(g'(t), -f'(t)\big)}{\left\|\vec{r}\,'(t)\right\|}\left\|\vec{r}\,'(t)\right\|\ dt\\ &= \int_C (M, N)\cdot\big(g'(t), -f'(t)\big)\ dt\\ &= \int_C \big(Mg'(t) - Nf'(t)\big)\ dt\\ &= \int_C Mg'(t)\ dt - \int_C Nf'(t)\ dt.\end{aligned}$$

As the x- and y-components of $\vec{r}(t)$ are $f(t)$ and $g(t)$ respectively, the differentials of x and y are $dx = f'(t)\ dt$ and $dy = g'(t)\ dt$. We can then write the above integrals as:

$$\int_C \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}\, ds = \int_C M\, dy - \int_C N\, dx.$$

The right-hand side is often written as one integral (not incorrectly, though somewhat confusingly, as this one integral has two "d's"):

$$\int_C \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}\, ds = \int_C \big(M\, dy - N\, dx\big).$$

We summarize the above in the following definition.

Definitie 18.9 (Flow, flux)
Let $\vec{\boldsymbol{F}} = \big(M(x,y), N(x,y)\big)$ be a vector field with continuous components defined on a smooth curve C, parametrized by $\vec{\boldsymbol{r}}(t) = \big(f(t), g(t)\big)$, let $\widehat{\boldsymbol{T}}$ be the unit tangent vector of $\vec{\boldsymbol{r}}(t)$, and let $\hat{\boldsymbol{n}}$ be the clockwise 90° rotation of $\widehat{\boldsymbol{T}}$.

- The **flow** (*stroming*) of $\vec{\boldsymbol{F}}$ along C is

$$\int_C \vec{\boldsymbol{F}} \cdot \widehat{\boldsymbol{T}}\, ds = \int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}.$$

- The **flux** (*flux*) of $\vec{\boldsymbol{F}}$ across C is

$$\int_C \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}\, ds = \int_C \big(M\, dy - N\, dx\big) = \int_C \big(M g'(t) - N f'(t)\big)\, dt.$$

This definition of flow also holds for curves in space, though it does not make sense to measure flux across a curve in space. Measuring flow is essentially the same as finding work performed by a force as done in the previous examples. Therefore we practice finding only flux in the following example.

Example 18.10

Curves C_1 and C_2 each start at $(1,0)$ and end at $(0,1)$, where C_1 follows the line $y = 1 - x$ and C_2 follows the unit circle, as shown in Figure 18.13(a). Find the flux across both curves for the vector fields $\vec{\boldsymbol{F}}_1 = (y, -x+1)$ and $\vec{\boldsymbol{F}}_2 = (-x, 2y - x)$.

Solution

We begin by finding parametrizations of C_1 and C_2. As done in Example 18.8, parametrize C_1 by creating the line that starts at $(1,0)$ and moves in the $(-1,1)$ direction:
$\vec{\boldsymbol{r}}_1(t) = (1,0) + t(-1,1) = (1-t, t)$, for $0 \le t \le 1$. We parametrize C_2 with the familiar $\vec{\boldsymbol{r}}_2(t) = \big(\cos(t), \sin(t)\big)$ on $0 \le t \le \pi/2$. For reference later, we give each function and its derivative below:

$$\vec{\boldsymbol{r}}_1(t) = (1-t, t), \quad \vec{\boldsymbol{r}}\,'_1(t) = (-1, 1).$$

$$\vec{\boldsymbol{r}}_2(t) = \big(\cos(t), \sin(t)\big), \quad \vec{\boldsymbol{r}}\,'_2(t) = \big(-\sin(t), \cos(t)\big).$$

When $\vec{\boldsymbol{F}}_1 = (y, -x+1)$ (as shown in Figure 18.13(a)), over C_1 we have $M = y = t$ and $N = -x + 1 = -(1-t) + 1 = t$. Using Definition 18.9, we compute the flux:

$$\int_{C_1} \vec{\boldsymbol{F}}_1 \cdot \hat{\boldsymbol{n}}\, ds = \int_{C_1} \big(M g'(t) - N f'(t)\big)\, dt$$

$$= \int_0^1 \big(t(1) - t(-1)\big)\ dt$$
$$= \int_0^1 2t\ dt$$
$$= 1.$$

Over C_2, we have $M = y = \sin(t)$ and $N = -x+1 = 1-\cos(t)$. Thus the flux across C_2 is:

$$\begin{aligned}\int_{C_2} \vec{\boldsymbol{F}}_1 \cdot \hat{\boldsymbol{n}}\ ds &= \int_{C_2} \big(Mg'(t) - Nf'(t)\big)\ dt \\ &= \int_0^{\pi/2} \big(\sin(t)\cos(t) - (1-\cos(t))(-\sin(t))\big)\ dt \\ &= \int_0^{\pi/2} \sin(t)\ dt \\ &= 1.\end{aligned}$$

Notice how the flux was the same across both curves. This won't hold true when we change the vector field.

When $\vec{\boldsymbol{F}}_2 = (-x, 2y-x)$ (as shown in Figure 18.13(b)), over C_1 we have $M = -x = t-1$ and $N = 2y-x = 2t-(1-t) = 3t-1$. Computing the flux across C_1:

$$\begin{aligned}\int_{C_1} \vec{\boldsymbol{F}}_2 \cdot \hat{\boldsymbol{n}}\ ds &= \int_{C_1} \big(Mg'(t) - Nf'(t)\big)\ dt \\ &= \int_0^1 \big((t-1)(1) - (3t-1)(-1)\big)\ dt \\ &= \int_0^1 (4t-2)\ dt \\ &= 0.\end{aligned}$$

Over C_2, we have $M = -x = -\cos(t)$ and $N = 2y-x = 2\sin(t) - \cos(t)$. Thus the flux across C_2 is:

$$\begin{aligned}\int_{C_2} \vec{\boldsymbol{F}}_2 \cdot \hat{\boldsymbol{n}}\ ds &= \int_{C_2} \big(Mg'(t) - Nf'(t)\big)\ dt \\ &= \int_0^{\pi/2} \big(-\cos(t)\cos(t) - (2\sin t - \cos(t))(-\sin(t))\big)\ dt \\ &= \int_0^{\pi/2} \big(2\sin^2(t) - \sin(t)\cos t - \cos^2(t)\big)\ dt \\ &= \frac{\pi}{4} - \frac{1}{2} \approx 0.285.\end{aligned}$$

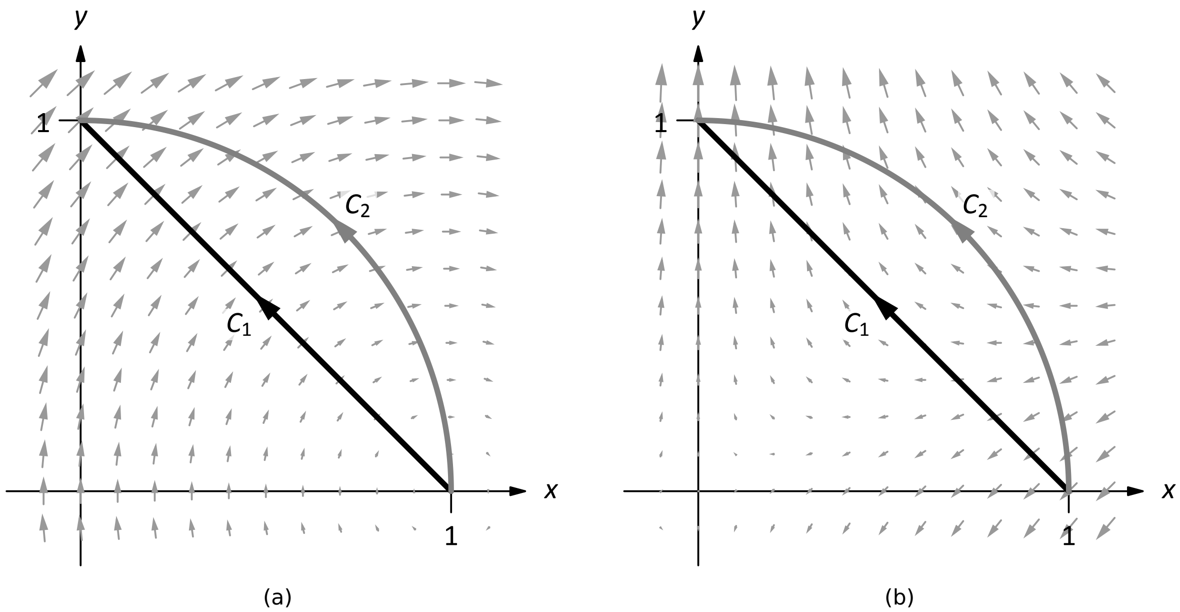


Figure 18.13: Illustrating the curves and vector fields in Example 18.10. In (a) the vector field is $\vec{F}_1$, and in (b) the vector field is $\vec{F}_2$.

We analyse the results of this example below.

In Example 18.10, we saw that the flux across the two curves was the same when the vector field was $\vec{F}_1 = (y, -x+1)$. This is not a coincidence. We show why they are equal in Example 18.14. In short, the reason is this: the divergence of $\vec{F}_1$ is 0, and when $\operatorname{div}\vec{F} = 0$, the flux across any two paths with common beginning and ending points will be the same.

We also saw in the example that the flux across C_1 was 0 when the field was $\vec{F}_2 = (-x, 2y-x)$. Flux measures "how much" of the field crosses the path from left to right (following the conventions established before). Positive flux means most of the field is crossing from left to right; negative flux means most of the field is crossing from right to left; zero flux means the same amount crosses from each side. When we consider Figure 18.13(b), it seems plausible that the same amount of $\vec{F}_2$ was crossing C_1 from left to right as from right to left.

18.4.2 Green's theorem

There is an important connection between the circulation around a closed region R and the curl of the vector field inside of R, as well as a connection between the flux across the boundary of R and the divergence of the field inside R. These connections are described by Green's theorem and the divergence theorem, respectively. We will explore each in turn.

Green's theorem states "the counterclockwise circulation around a closed region R is equal to the sum of the curls over R."

Theorem 18.5 (Green's theorem)
Let R be a closed, bounded region of the plane whose boundary C is composed of finitely many smooth curves, let $\vec{r}(t)$ be a counterclockwise parametrization of C, and let $\vec{F} = (M, N)$ where N_x and M_y are continuous over R. Then

$$\oint_C \vec{F} \cdot d\vec{r} = \iint_R \operatorname{curl} \vec{F} \, dA = \iint_R \left(N_x - M_y\right) \, dA.$$

Proof It suffices to demonstrate the theorem for rectangular regions in the xy-plane. The Riemann sum nature of the double integral will then guarantee the proof of the theorem for arbitrary regions, because a Riemann sum is technically a summation of the areas of arbitrarily small rectangles. As the proof is for a rectangle, the proof will work for arbitrary regions, which can be approximated by collections of ever smaller rectangles.

First of all, let us recall that $\vec{F} = (M, N)$ and that $d\vec{r} = (dx, dy)$, such that we can rewrite the contour integral as

$$\oint_C \left(M \, dx + N \, dy\right).$$

Let $R = \left\{(x, y) \,\middle|\, a \le x \le b, c \le y \le d\right\}$ be a rectangular region, and let the boundary C of R be oriented counterclockwise. We break this boundary into four pieces (Figure 18.14):

1. C_1, which runs from (a, c) to (b, c),
2. C_2, which runs from (b, c) to (b, d),
3. C_3, which runs from (b, d) to (a, d),
4. C_4, which runs from (a, d) to (a, c). □

Then,

$$\iint_R \frac{\partial N}{\partial x} \, dx \, dy \quad = \quad \int_c^d \int_a^b \frac{\partial N}{\partial x} \, dx \, dy$$

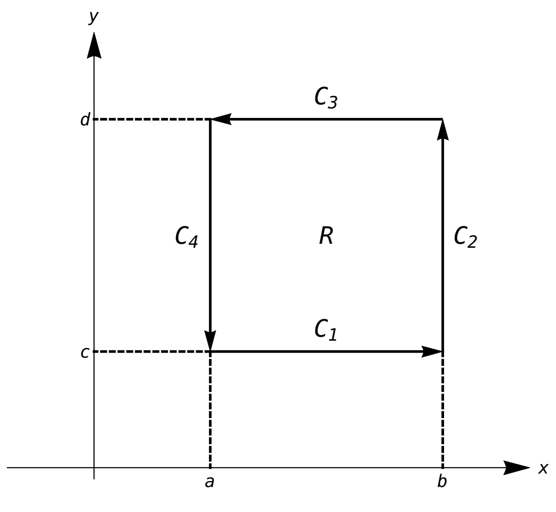


Figure 18.14: The rectangular region used in the proof of Green's theorem.

$$
\begin{aligned}
&= \int_c^d \big(N(b,y) - N(a,y)\big)\ dy \\
&= \int_c^d N(b,y)\ dy + \int_d^c N(a,y)\ dy \\
&= \int_{C_2} N\ dy + \int_{C_4} N\ dy\,.
\end{aligned}
$$

We note that y is constant along C_1 and C_3, so

$$\int_{C_1} N\ dy = \int_{C_3} N\ dy = 0\,.$$

Hence

$$
\begin{aligned}
\iint_R \frac{\partial N}{\partial x}\ dx\ dy = \int_{C_2} N\ dy + \int_{C_4} N\ dy &= \int_{C_1} N\ dy + \int_{C_2} N\ dy + \int_{C_3} N\ dy + \int_{C_4} N\ dy \\
&= \oint_C N\ dy\,.
\end{aligned}
$$

A similar argument demonstrates that:

$$\iint_R \frac{\partial M}{\partial y}\ dy\ dx = -\oint_C M\ dx\,.$$

Consequently, we have that:

$$\oint_C (Mdx + Ndy) = \iint_R \big(N_x - M_y\big)\ dA.$$

We will explore Green's theorem through an example.

Example 18.11

Let $\vec{\pmb{F}} = \big(\sin(x), \cos(y)\big)$ and let R be the region enclosed by the curve C parametrized by $\vec{\pmb{r}}(t) = \big(2\cos(t) + \cos(10t)/10, 2\sin(t) + \sin(10t)/10\big)$ on $0 \le t \le 2\pi$, as shown in Figure 18.15. Find the circulation around C.

Solution

Computing the circulation directly using the line integral looks difficult, as the integrand will include terms like $\sin\big(2\cos(t) + \cos(10t)/10\big)$.

Green's theorem states that $\oint_C \vec{\pmb{F}} \cdot d\vec{\pmb{r}} = \iint_R \operatorname{curl} \vec{\pmb{F}}\ dA$; since $\operatorname{curl} \vec{\pmb{F}} = 0$ in this example, the double integral is simply 0 and hence the circulation is 0.

Since $\operatorname{curl} \vec{\pmb{F}} = 0$, we can conclude that the circulation is 0 in two ways. One method is to employ Green's theorem as done above. The second way is to recognize that $\vec{\pmb{F}}$ is a conservative field, hence there is a function $z = f(x,y)$ wherein $\vec{\pmb{F}} = \vec{\nabla} f$. Let A be any point on the curve C; since C is closed, we can say that C begins and ends at A. By the fundamental theorem of line integrals, we have $\oint_C \vec{\pmb{F}} \cdot d\vec{\pmb{r}} = f(A) - f(A) = 0$.

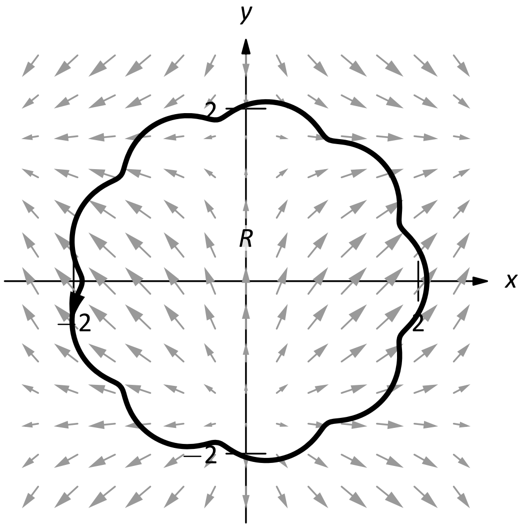


Figure 18.15: The vector field and planar region used in Example 18.11.

One can use Green's theorem to find the area of an enclosed region by integrating along its boundary. Let C be a closed curve, enclosing the region R, parametrized by $\vec{r}(t) = \big(f(t), g(t)\big)$. We know the area of R is computed by the double integral $\iint_R \, dA$, where the integrand is 1. By creating a field $\vec{F}$ where $\text{curl}\,\vec{F} = 1$, we can employ Green's theorem to compute the area of R as $\oint_C \vec{F} \cdot d\vec{r}$.

One is free to choose any field $\vec{F}$ to use as long as $\text{curl}\,\vec{F} = 1$. Common choices are $\vec{F} = \langle 0, x\rangle$, $\vec{F} = \langle -y, 0\rangle$ and $\vec{F} = \langle -y/2, x/2\rangle$. We demonstrate this below.

Example 18.12

Let C be the closed curve parametrized by $\vec{r}(t) = \left(t - t^3, t^2\right)$ on $-1 \le t \le 1$, enclosing the region R, as shown in Figure 18.16. Find the area of R.

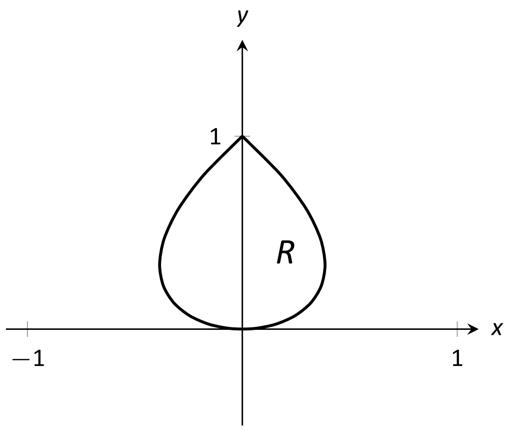


Figure 18.16: The region R, whose area is found in Example 18.12

Solution

We can choose any field $\vec{F}$, as long as $\text{curl}\,\vec{F} = 1$. We choose $\vec{F} = \langle -y, 0\rangle$. We also confirm (left to the reader) that $\vec{r}(t)$ traverses the region R in a counterclockwise fashion. Thus the area of R is:

$$A = \iint_R \, dA$$

$$= \oint_C \vec{F}\cdot d\vec{r}$$
$$= \int_{-1}^{1} \left(-t^2, 0\right)\cdot\left(1-3t^2, 2t\right)\, dt$$
$$= \int_{-1}^{1} \left(-t^2+3t^4\right)\, dt$$
$$= \frac{8}{15}.$$

18.4.3 The divergence theorem

Green's theorem makes a connection between the circulation around a closed region R and the sum of the curls over R. The divergence theorem (also known as Gauss's theorem) makes a somewhat opposite connection: the total flux across the boundary of R is equal to the sum of the divergences over R.

Theorem 18.6 (The divergence theorem (in the plane))
Let R be a closed, bounded region of the plane whose boundary C with unit normal vector $\hat{\boldsymbol{n}}$ is composed of finitely many smooth curves, let $\vec{r}(t)$ be a counterclockwise parametrization of C and let $\vec{F} = (M, N)$ where M_x and N_y are continuous over R. Then

$$\oint_C \vec{F}\cdot\hat{\boldsymbol{n}}\, ds = \iint_R \operatorname{div}\vec{F}\, dA.$$

A proof of this theorem will be given in Section 18.7.1, where we will deal with this theorem's counterpart in space.

Example 18.13

Let $\vec{F} = (x-y, x+y)$, let C be the circle of radius 2 centred at the origin and define R to be the interior of that circle, as shown in Figure 18.17. Verify the divergence theorem; that is, find the flux across C and show it is equal to the double integral of $\operatorname{div}\vec{F}$ over R.

Solution

We parametrize the circle in the usual way, with $\vec{r}(t) = \left(2\cos(t), 2\sin(t)\right)$, $0 \le t \le 2\pi$. The flux across C is

$$\oint_C \vec{F}\cdot\hat{\boldsymbol{n}}\, ds = \oint_C \left(M g'(t) - N f'(t)\right) dt$$
$$= \int_0^{2\pi} \Big(\left(2\cos(t)-2\sin(t)\right)\left(2\cos(t)\right) - \left(2\cos(t)+2\sin(t)\right)\left(-2\sin(t)\right)\Big)\, dt$$
$$= \int_0^{2\pi} 4\, dt = 8\pi.$$

We compute the divergence of $\vec{F}$ as $\operatorname{div}\vec{F} = M_x + N_y = 2$. Since the divergence is constant, we can compute the following double integral easily:

$$\iint_R \operatorname{div}\vec{F}\, dA = \iint_R 2\, dA = 2\iint_R dA = 2(\text{area of } R) = 8\pi,$$

which matches our previous result.

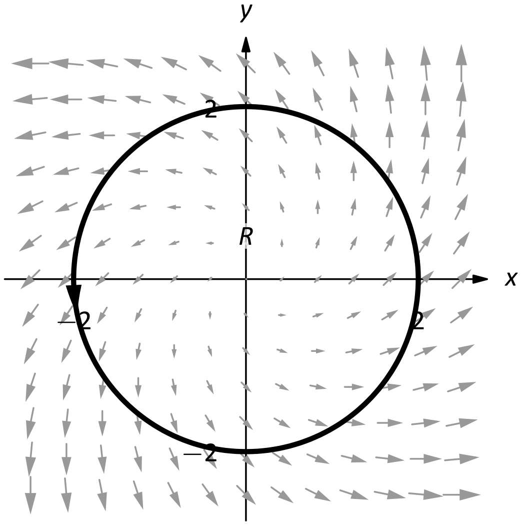


Figure 18.17: The region R used in Example 18.13.

Example 18.14

Let $\vec{F}$ be any field where $\operatorname{div}\vec{F} = 0$, and let C_1 and C_2 be any two nonintersecting paths, except that each begins at point A and ends at point B (see Figure 18.18). Show why the flux across C_1 and C_2 is the same.

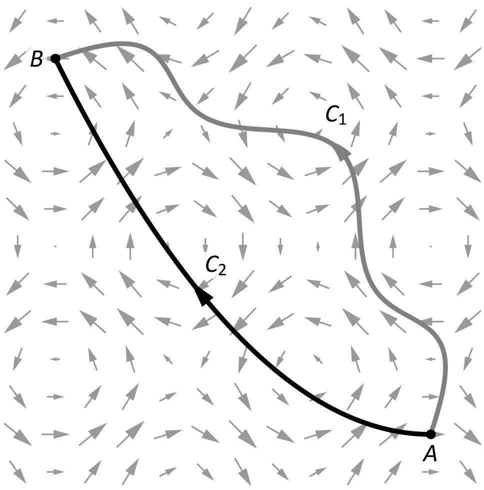


Figure 18.18: As used in Example 18.14, the vector field has a divergence of 0 and the two paths only intersect at their initial and terminal points.

Solution

By referencing Figure 18.18, we see we can make a closed path C that combines C_1 with C_2, where C_2 is traversed with its opposite orientation. We label the enclosed region R. Since $\operatorname{div}\vec{F} = 0$, the divergence Theorem states that

$$\oint_C \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds = \iint_R \operatorname{div}\vec{\boldsymbol{F}}\,dA = \iint_R 0\,dA = 0.$$

Recalling the properties of line integrals over vector fields, consider:

$$\begin{aligned} 0 &= \oint_C \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds \\ &= \int_{C_1} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds + \int_{C_2^*} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds \\ &= \int_{C_1} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds - \int_{C_2} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds. \end{aligned}$$

where C_2^* is the path C_2 traversed with opposite orientation. So:

$$\int_{C_1} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds = \int_{C_2} \left(\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\right)\,ds.$$

Thus the flux across each path is equal.

In this section, we have investigated flow and flux, quantities that measure interactions between a vector field and a planar curve. We can also measure flow along spatial curves, though as mentioned before, it does not make sense to measure flux across spatial curves.

It does, however, make sense to measure the amount of a vector field that passes across a surface in space – i.e, the flux across a surface. We will study this, though in the next section we first learn about a more powerful way to describe surfaces than using functions of the form $z = f(x,y)$.

18.5 Parameterized surfaces and surface area

18.5.1 Parameterized surfaces

Thus far we have focused mostly on 2-dimensional vector fields, measuring flow and flux along/across curves in the plane. Both Green's theorem and the divergence theorem make connections between planar regions and their boundaries. We now move our attention to 3-dimensional vector fields, considering both curves and surfaces in space.

We are accustomed to describing surfaces as functions of two variables, usually written as $z = f(x,y)$. For our coming needs, this method of describing surfaces will prove to be insufficient. Instead, we will parametrize our surfaces, describing them as the set of terminal points of some vector–valued function $\vec{\boldsymbol{r}}(u,v) = \big(f(u,v), g(u,v), h(u,v)\big)$. The bulk of this section is spent practicing the skill of describing a surface $\mathcal{S}$ using a vector–valued function. Once this skill is developed, we'll show how to find the surface area SA of a parametrically–defined surface $\mathcal{S}$, a skill needed in the remaining sections of this chapter.

Note that we distinguish a surface from its surface area by using a calligraphic S to denote a surface: $\mathcal{S}$. When writing this letter by hand, it may be useful to add serifs to the letter, such as: $\mathcal{S}$

Definitie 18.10 (Parametrized surface)
Let $\vec{\boldsymbol{r}}(u,v) = \big(f(u,v), g(u,v), h(u,v)\big)$ be a vector–valued function that is continuous and one to one on the interior of its domain R in the uv-plane. The set of all terminal points of $\vec{\boldsymbol{r}}$ (i.e., the

range of $\vec{r}$) is the **surface** (*oppervlak*) $\mathcal{S}$, and $\vec{r}$ along with its domain R form a **parametrization** (*parametrisering*) of $\mathcal{S}$.
This parametrization is **smooth** (*glad*) on R if $\vec{r}_u$ and $\vec{r}_v$ are continuous and $\vec{r}_u \times \vec{r}_v$ is never $\vec{0}$ on the interior of R.

Given a point (u_0, v_0) in the domain of a vector–valued function $\vec{r}$, the vectors $\vec{r}_u(u_0, v_0)$ and $\vec{r}_v(u_0, v_0)$ are tangent to the surface $\mathcal{S}$ at $\vec{r}(u_0, v_0)$ (a proof of this is developed later in this section). The definition of smoothness dictates that $\vec{r}_u \times \vec{r}_v \neq \vec{0}$; this ensures that neither $\vec{r}_u$ nor $\vec{r}_v$ are $\vec{0}$, nor are they ever parallel. Therefore smoothness guarantees that $\vec{r}_u$ and $\vec{r}_v$ determine a plane that is tangent to $\mathcal{S}$.

A surface $\mathcal{S}$ is said to be **orientable** (*oriënteerbaar*) if a field of normal vectors can be defined on $\mathcal{S}$ that vary continuously along $\mathcal{S}$. This definition may be hard to understand; it may help to know that orientable surfaces are often called two sided. A sphere is an orientable surface, and one can easily envision an inside and outside of the sphere. A paraboloid is orientable, where again one can generally envision inside and outside sides (or top and bottom sides) to this surface. Just about every surface that one can imagine is orientable, and we'll assume all surfaces we deal with in this text are orientable.

It is enlightening to examine a classic non-orientable surface: the Möbius band, shown in Figure 18.19. Vectors normal to the surface are given, starting at the point indicated in the figure. These normal vectors vary continuously as they move along the surface. Letting each vector indicate the top side of the band, we can easily see near any vector which side is the top. However, if as we progress along the band, we recognize that we are labelling both sides of the band as the top; in fact, there are not two sides to this band, but one. The Möbius band is a non-orientable surface.

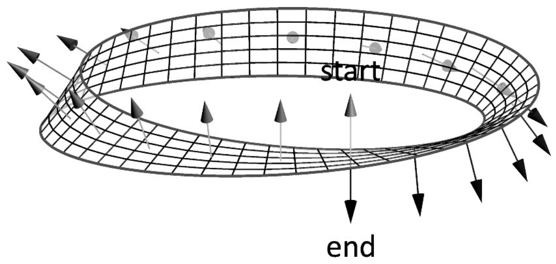


Figure 18.19: A Möbius band, a non-orientable surface.

We now practice parameterizing surfaces.

Example 18.15

Parametrize the surface $z = x^2 + 2y^2$ over the rectangular region R defined by $-3 \leq x \leq 3$, $-1 \leq y \leq 1$.

Solution

There is a straightforward way to parametrize a surface of the form $z = f(x, y)$ over a rectangular domain. We let $x = u$ and $y = v$, and let $\vec{r}(u, v) = \big(u, v, f(u, v)\big)$. In this instance, we have $\vec{r}(u, v) = \big(u, v, u^2 + 2v^2\big)$, for $-3 \leq u \leq 3$, $-1 \leq v \leq 1$. This surface is graphed in Figure 18.20.

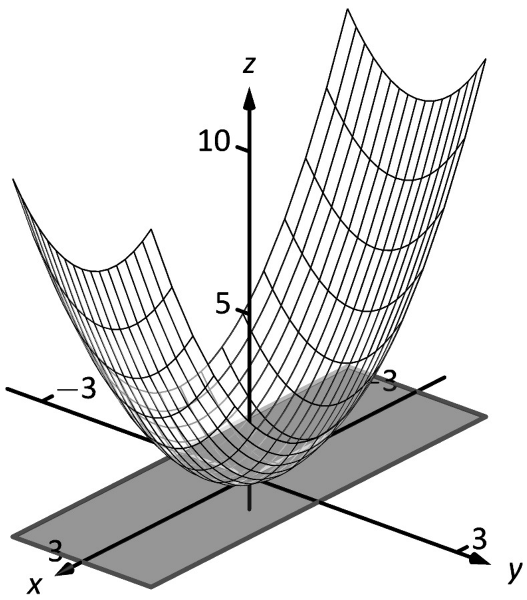


Figure 18.20: The surface parametrized in Example 18.15.

Example 18.16

Parametrize the surface $z = x^2 + 2y^2$ over the circular region R enclosed by the circle of radius 2 that is centred at the origin.

Solution

We can parametrize the circular boundary of R with the vector–valued function $\big(2\cos(u), 2\sin(u)\big)$, where $0 \le u \le 2\pi$. We can obtain the interior of R by scaling this function by a variable amount, i.e., by multiplying by v: $\big(2v\cos(u), 2v\sin(u)\big)$, where $0 \le v \le 1$.

It is important to understand the role of v in the above function. When $v = 1$, we get the boundary of R, a circle of radius 2. When $v = 0$, we simply get the point $(0, 0)$, the center of R (which can be thought of as a circle with radius of 0). When $v = 1/2$, we get the circle of radius 1 that is centred at the origin, which is the circle halfway between the boundary and the center. As v varies from 0 to 1, we create a series of concentric circles that fill out all of R.

Thus far, we have determined the x- and y-components of our parametrization of the surface: $x = 2v\cos(u)$ and $y = 2v\sin(u)$. We find the z-component simply by using $z = f(x, y) = x^2 + 2y^2$:

$$z = (2v\cos(u))^2 + 2(2v\sin(u))^2 = 4v^2\cos^2(u) + 8v^2\sin^2(u).$$

Thus $\vec{r}(u, v) = \Big(2v\cos(u), 2v\sin(u), 4v^2\cos^2(u) + 8v^2\sin^2(u)\Big)$, $0 \le u \le 2\pi$, $0 \le v \le 1$, which is graphed in Figure 18.21. The way that this graphic was generated highlights how the surface was parametrized. When viewing from above, one can see lines emanating from the origin; they represent different values of u as u sweeps from an angle of 0 up to 2π. One can also see concentric circles, each corresponding to a different value of v.

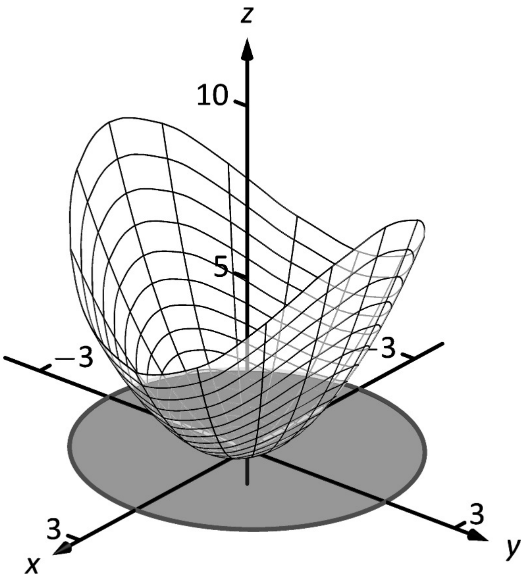


Figure 18.21: The surface parametrized in Example 18.16.

Examples 18.15 and 18.16 demonstrate an important principle when parameterizing surfaces given in the form $z = f(x,y)$ over a region R: if one can determine x and y in terms of u and v, then z follows directly as $z = f(x,y)$.

In the following example, we parametrize the same surface over triangular regions.

Example 18.17

Parametrize the surface $z = x^2 + 2y^2$ over the triangular region R enclosed by the lines $y = 3 - 2x/3$, $y = 1$ and $x = 0$ as shown in Figure 18.22(a).

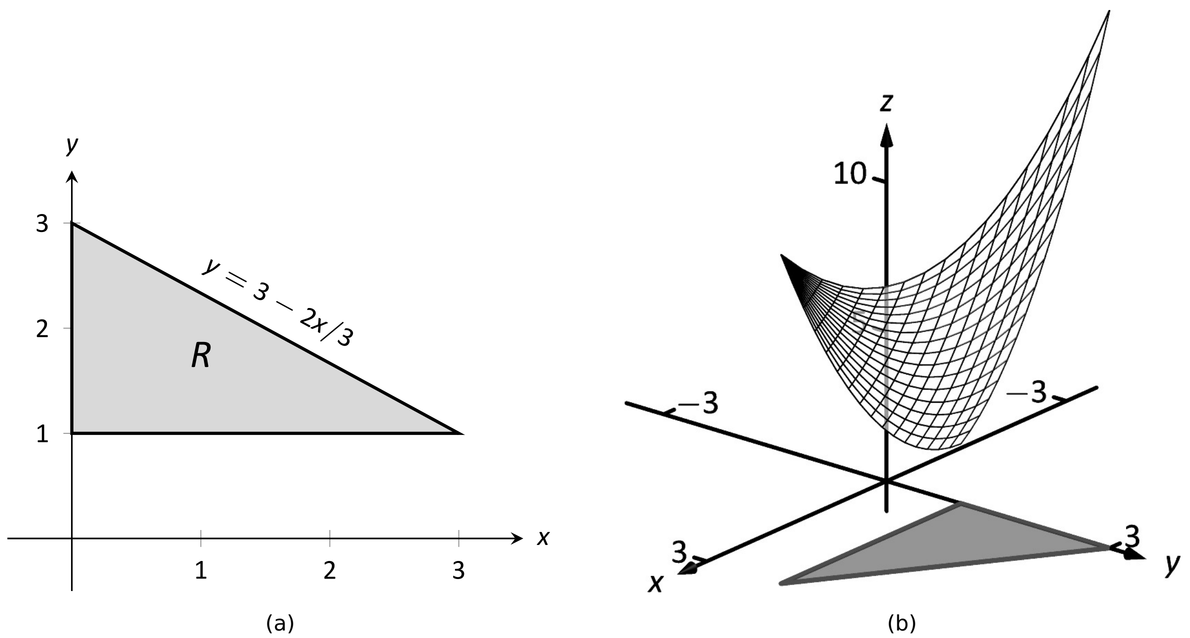


Figure 18.22: Part (a) shows a graph of the region R, and part (b) shows the surface over R, as defined in Example 18.17.

Solution

While the region R in this example is very similar to the region R in the previous example, and our method of parameterizing the surface is fundamentally the same, it will feel as though our answer is much different than before.

We begin with letting $x = u$, $0 \le u \le 3$. We may be tempted to let $y = v(3-2u/3)$, $0 \le v \le 1$, but this is incorrect. When $v = 1$, we obtain the upper line of the triangle as desired. However, when $v = 0$, the y-value is 0, which does not lie in the region R.

We will describe the general method of proceeding following this example. For now, consider $y = 1 + v(2-2u/3)$, $0 \le v \le 1$. Note that when $v = 1$, we have $y = 3-2u/3$, the upper line of the boundary of R. Also, when $v = 0$, we have $y = 1$, which is the lower boundary of R. With $z = x^2 + 2y^2$, we determine $\vec{r}(u,v) = \Big(u, 1+v(2-2u/3), u^2+2\big(1+v(2-2u/3)\big)^2\Big)$, $0 \le u \le 3$, $0 \le v \le 1$.

The surface is graphed in Figure 18.22(b).

Given a surface of the form $z = f(x,y)$, one can often determine a parametrization of the surface over a region R in a manner similar to determining bounds of integration over a region R. Using the techniques of Section 17.1, suppose a region R can be described by $a \le x \le b$, $g_1(x) \le y \le g_2(x)$, i.e., the area of R can be found using the iterated integral

$$\int_a^b \int_{g_1(x)}^{g_2(x)} dy\, dx.$$

When parameterizing the surface, we can let $x = u$, $a \le u \le b$, and we can let $y = g_1(u) + v\big(g_2(u) - g_1(u)\big)$, $0 \le v \le 1$. The parametrization of x is straightforward, but look closely at how y is determined. When $v = 0$, $y = g_1(u) = g_1(x)$. When $v = 1$, $y = g_2(u) = g_2(x)$.

As a specific example, consider the triangular region R from Example 18.17, shown in Figure 18.22(a). Using the techniques of Section 17.1, we can find the area of R as

$$\int_0^3 \int_1^{3-2x/3} dy\, dx.$$

Following the above discussion, we can set $x = u$, where $0 \le u \le 3$, and set $y = 1 + v\big(3-2u/3-1\big) = 1 + v(2-2u/3)$, $0 \le v \le 1$, as used in that example.

So, let a surface $\mathcal{S}$ be the graph of a function $z = f(x,y)$, where the domain of f is a closed, bounded region R in the xy-plane. Let R be bounded by $a \le x \le b$, $g_1(x) \le y \le g_2(x)$, i.e., the area of R can be found using the iterated integral $\int_a^b \int_{g_1(x)}^{g_2(x)} dy\, dx$, and let $h(u,v) = g_1(u) + v\big(g_2(u) - g_1(u)\big)$.

$\mathcal{S}$ can be parametrized as

$$\vec{r}(u,v) = \Big(u, h(u,v), f\big(u, h(u,v)\big)\Big), \quad a \le u \le b, \quad 0 \le v \le 1.$$

One can do a similar thing if R is bounded by $c \le y \le d$, $h_1(y) \le x \le h_2(y)$.

Example 18.18

Find a parametrization of the cylinder $x^2 + z^2/4 = 1$, where $-1 \le y \le 2$, as shown in Figure 18.23.

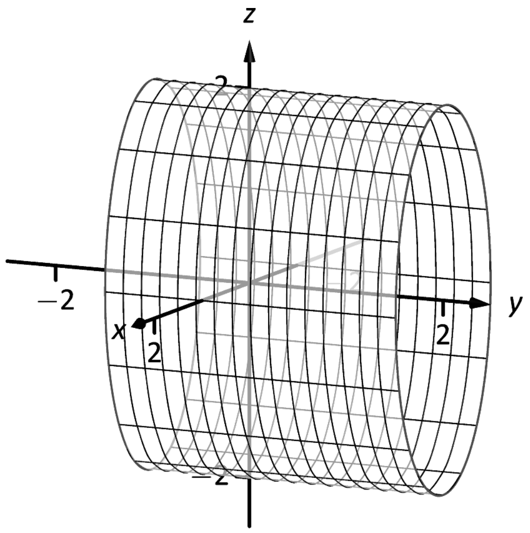


Figure 18.23: The cylinder parametrized in Example 18.18.

Solution

The equation $x^2 + z^2/4 = 1$ can be envisioned to describe an ellipse in the xz-plane; as the equation lacks a y-term, the equation describes a cylinder (recall Definition 7.4) that extends without bound parallel to the y-axis. This ellipse has a vertical major axis of length 4, a horizontal minor axis of length 2, and is centred at the origin. We can parametrize this ellipse using sines and cosines; our parametrization can begin with

$$\vec{r}(u,v) = \big(\cos(u), ???, 2\sin(u)\big), \quad 0 \le u \le 2\pi,$$

where we still need to determine the y-component.

While the cylinder $x^2 + z^2/4 = 1$ is satisfied by any y-value, the problem states that all y-values are to be between $y = -1$ and $y = 2$. Since the value of y does not depend at all on the values of x or z, we can use another variable, v, to describe y. Our final answer is

$$\vec{r}(u,v) = \big(\cos(u), v, 2\sin(u)\big), \quad 0 \le u \le 2\pi, \quad -1 \le v \le 2.$$

Example 18.19

Find a parametrization of the elliptic cone

$$z^2 = \frac{x^2}{4} + \frac{y^2}{9},$$

where $-2 \le z \le 3$, as shown in Figure 18.24(a).

Solution

One way to parametrize this cone is to recognize that given a z-value, the cross section of the cone at that z-value is an ellipse with equation

$$\frac{x^2}{(2z)^2} + \frac{y^2}{(3z)^2} = 1.$$

We can let $z = v$, for $-2 \le v \le 3$ and then parametrize the above ellipses using sines, cosines and v. We can parametrize the x-component of our surface with $x = 2z\cos(u)$ and the y-component with $y = 3z\sin(u)$, where $0 \le u \le 2\pi$. Putting all components together, we have

$$\vec{r}(u,v) = \big(2v\cos(u), 3v\sin(u), v\big), \quad 0 \le u \le 2\pi, \quad -2 \le v \le 3.$$

When v takes on negative values, the radii of the cross–sectional ellipses become negative, which can lead to some surprising results. Consider Figure 18.24(b), where the cone is graphed for $0 \le u \le \pi$. Because v is negative below the xy-plane, the radii of the cross–sectional ellipses are negative, and the opposite side of the cone is sketched below the xy-plane.

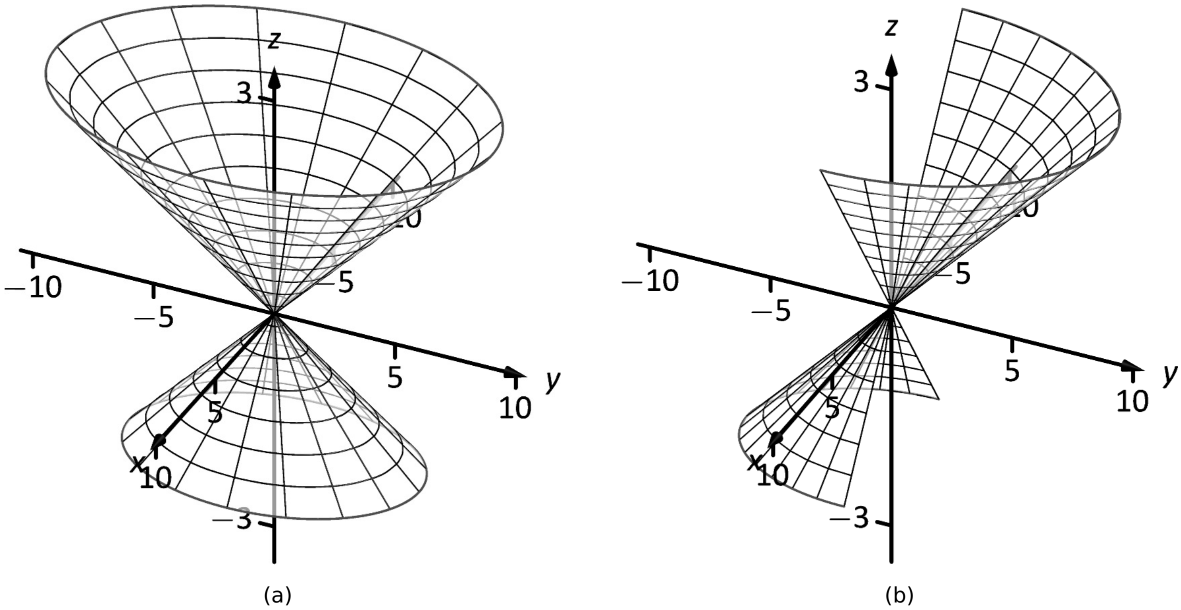


Figure 18.24: Part (a) shows the elliptic cone as described in Example 18.19 and part (b) shows the same elliptic cone with restricted domain.

Parametrization is a powerful way to represent surfaces. One of the advantages of the methods of parametrization described in this section is that the domain of $\vec{r}(u,v)$ is always a rectangle; that is, the bounds on u and v are constants. This will make some of our future computations easier to evaluate.

Just as we could parametrize curves in more than one way, there will always be multiple ways to parametrize a surface. Some ways will be more "natural" than others, but these other ways are not incorrect. Because technology is often readily available, it is often a good idea to check one's work by graphing a parametrization of a surface to check if it indeed represents what it was intended to.

18.5.2 Surface area

It will become important in the following sections to be able to compute the surface area of a surface $\mathcal{S}$ given a smooth parametrization $\vec{r}(u,v)$, $a \le u \le b$, $c \le v \le d$. Following the principles given in the integration review at the beginning of this chapter, we can say that the surface area of $\mathcal{S}$ is given by

$$SA = \iint_{\mathcal{S}} dS,$$

where dS represents a small amount of surface area. That is, to compute total surface area SA, add up lots of small amounts of surface area dS across the entire surface $\mathcal{S}$. The key to finding surface area is knowing how to compute dS. We begin by approximating.

In Section 17.4 we used the area of a plane to approximate the surface area of a small portion of a surface. We will do the same here.

Let R be the region of the uv-plane bounded by $a \leq u \leq b$, $c \leq v \leq d$ as shown in Figure 18.25(a). Partition R into rectangles of width $\Delta u = \frac{b-a}{n}$ and height $\Delta v = \frac{d-c}{n}$, for some n. Let $p = (u_0, v_0)$ be the lower left corner of some rectangle in the partition, and let m and q be neighbouring corners as shown.

The point with coordinates $\widetilde{P} = (u_0, v_0)$ maps to a point $P = \vec{\boldsymbol{r}}(u_0, v_0)$ on the surface $\mathcal{S}$, and the rectangle with corners $\widetilde{P}$, $\widetilde{M}$ and $\widetilde{Q}$ maps to some region (probably not rectangular) on the surface as shown in Figure 18.25(b), where $M = \vec{\boldsymbol{r}}(u_0 + \Delta u, v_0)$ and $Q = \vec{\boldsymbol{r}}(u_0, v_0 + \Delta v)$. We wish to approximate the surface area of this mapped region.

Let $\vec{\boldsymbol{u}} = M - P$ and $\vec{\boldsymbol{v}} = Q - P$. These two vectors form a parallelogram, illustrated in Figure 18.25(c), whose area approximates the surface area we seek. In this particular illustration, we can see that parallelogram does not particularly match well the region we wish to approximate, but that is acceptable; by increasing the number of partitions of R, Δu and Δv shrink and our approximations will become better.

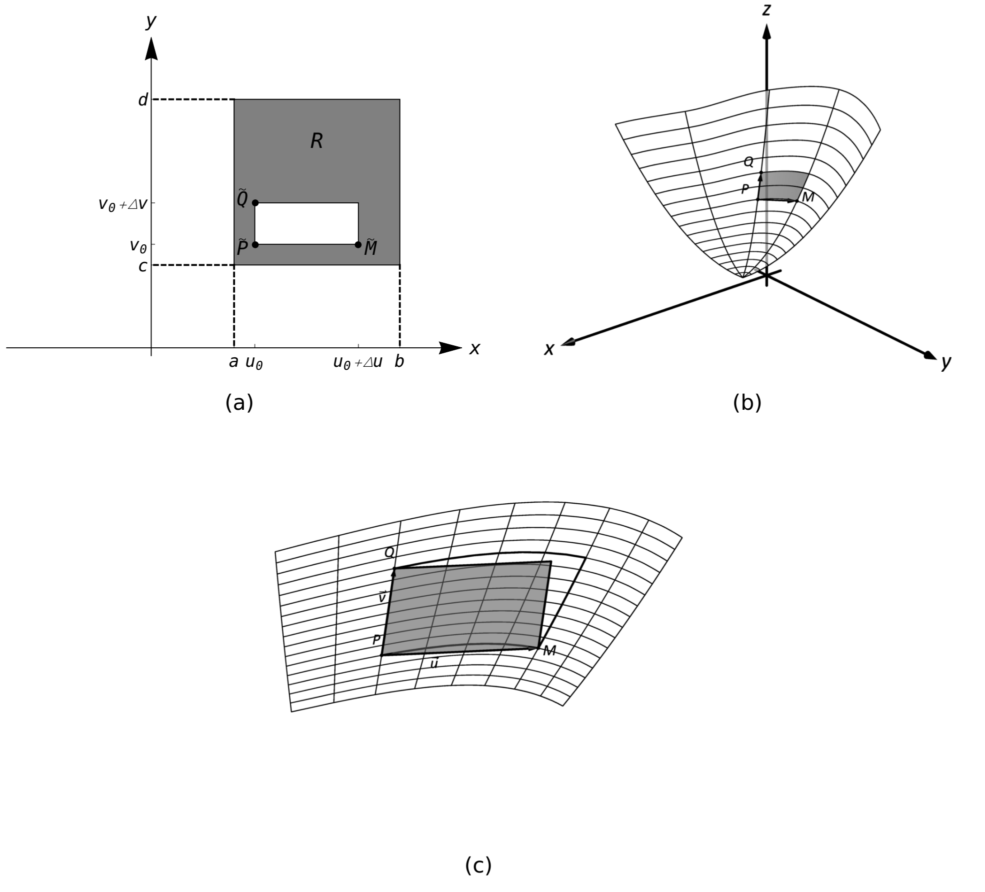


Figure 18.25: Illustrating the process of finding surface area by approximating with planes.

From Section 6.6 we know the area of this parallelogram is $\|\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}\|$. If we repeat this approximation process for each rectangle in the partition of R, we can sum the areas of all the parallelograms to get an approximation of the surface area SA:

$$SA \approx \sum_{j=1}^{n}\sum_{i=1}^{n} \left\|\vec{\boldsymbol{u}}_{i,j} \times \vec{\boldsymbol{v}}_{i,j}\right\|,$$

where $\vec{\boldsymbol{u}}_{i,j} = \vec{\boldsymbol{r}}(u_i+\Delta u, v_j) - \vec{\boldsymbol{r}}(u_i, v_j)$ and $\vec{\boldsymbol{v}}_{i,j} = \vec{\boldsymbol{r}}(u_i, v_j+\Delta v) - \vec{\boldsymbol{r}}(u_i, v_j)$.

From our previous calculus experience, we expect that taking a limit as $n \to +\infty$ will result in the exact surface area. However, the current form of the above double sum makes it difficult to realize what the result of that limit is. The following rewriting of the double summation will be helpful:

$$\begin{aligned}\sum_{j=1}^{n}\sum_{i=1}^{n} \left\|\vec{\boldsymbol{u}}_{i,j} \times \vec{\boldsymbol{v}}_{i,j}\right\| &= \sum_{j=1}^{n}\sum_{i=1}^{n} \left\|\big(\vec{\boldsymbol{r}}(u_i+\Delta u, v_j) - \vec{\boldsymbol{r}}(u_i, v_j)\big) \times \big(\vec{\boldsymbol{r}}(u_i, v_j+\Delta v) - \vec{\boldsymbol{r}}(u_i, v_j)\big)\right\| \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} \left\|\frac{\vec{\boldsymbol{r}}(u_i+\Delta u, v_j) - \vec{\boldsymbol{r}}(u_i, v_j)}{\Delta u} \times \frac{\vec{\boldsymbol{r}}(u_i, v_j+\Delta v) - \vec{\boldsymbol{r}}(u_i, v_j)}{\Delta v}\right\| \Delta u \Delta v.\end{aligned}$$

We now take the limit as $n \to +\infty$, forcing Δu and Δv to 0. As $\Delta u \to 0$,

$$\frac{\vec{\boldsymbol{r}}(u_i+\Delta u, v_j) - \vec{\boldsymbol{r}}(u_i, v_j)}{\Delta u} \quad \to \quad \vec{\boldsymbol{r}}_u(u_i, v_j),$$

and as $\Delta v \to 0$,

$$\frac{\vec{\boldsymbol{r}}(u_i, v_j+\Delta v) - \vec{\boldsymbol{r}}(u_i, v_j)}{\Delta v} \quad \to \quad \vec{\boldsymbol{r}}_v(u_i, v_j).$$

This limit process also demonstrates that $\vec{\boldsymbol{r}}_u(u,v)$ and $\vec{\boldsymbol{r}}_v(u,v)$ are tangent to the surface $\mathcal{S}$ at $\vec{\boldsymbol{r}}(u,v)$. We do not need this fact now, but it will be important in the next section.

Thus, in the limit, the double sum leads to a double integral:

$$\lim_{n\to+\infty} \sum_{j=1}^{n}\sum_{i=1}^{n} \left\|\vec{\boldsymbol{u}}_{i,j} \times \vec{\boldsymbol{v}}_{i,j}\right\| = \int_c^d \int_a^b \|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|\ du\ dv.$$

Theorem 18.7 (Surface area of parametrically defined surfaces)
Let $\vec{\boldsymbol{r}}(u,v)$ be a smooth parametrization of a surface $\mathcal{S}$ over a closed, bounded region R of the uv-plane.

- *The surface area differential dS is: $dS = \|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|\ dA$.*
- *The surface area SA of $\mathcal{S}$ is*

$$SA = \iint_{\mathcal{S}} dS = \iint_R \|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|\ dA.$$

Example 18.20

Using the parametrization found in Example 18.16, find the surface area of $z = x^2 + 2y^2$ over the circular disk of radius 2, centred at the origin.

Solution

In Example 18.16, we parametrized the surface as

$\vec{r}(u,v) = \Big(2v\cos(u), 2v\sin(u), 4v^2\cos^2(u) + 8v^2\sin^2(u)\Big)$, for $0 \le u \le 2\pi$, $0 \le v \le 1$. To find the surface area using Theorem 18.7, we need $\|\vec{r}_u \times \vec{r}_v\|$. We find:

$$\begin{aligned}
\vec{r}_u &= \Big(-2v\sin(u), 2v\cos(u), 8v^2\cos(u)\sin(u)\Big)\\
\vec{r}_v &= \Big(2\cos(u), 2\sin(u), 8v\cos^2(u) + 16v\sin^2(u)\Big)\\
&= \Big(2\cos(u), 2\sin(u), 16v - 8v\cos^2(u)\Big)\\
\Rightarrow\ \vec{r}_u \times \vec{r}_v &= \Big(16v^2\cos(u), 32v^2\sin(u), -4v\Big)\\
\Rightarrow\ \|\vec{r}_u \times \vec{r}_v\| &= \sqrt{256v^4\cos^2(u) + 1024v^4\sin^2(u) + 16v^2}\,.
\end{aligned}$$

Thus the surface area SA is

$$\begin{aligned}
\iint_{\mathcal{S}} dS &= \iint_R \|\vec{r}_u \times \vec{r}_v\|\ dA\\
&= \int_0^1\int_0^{2\pi} \sqrt{256v^4\cos^2(u) + 1024v^4\sin^2(u) + 16v^2}\ du\ dv \approx 53.59.
\end{aligned}$$

There is a lot of tedious work in the above calculations and the final integral is nontrivial. The use of a computer-algebra system is highly recommended.

In Section 18.1, we recalled the arc length differential $ds = \left\|\vec{r}\,'(t)\right\| dt$. In subsequent sections, we used that differential, but in most applications the $\left\|\vec{r}\,'(t)\right\|$ in the differential canceled out of the integrand (to our benefit, as integrating the square roots of functions is generally difficult). We will find a similar thing happens when we use the surface area differential dS in the following sections. That is, our main goal is not to be able to compute surface area; rather, surface area is a tool to obtain other quantities that are more important and useful. In our applications, we will use dS, but most of the time the $\|\vec{r}_u \times \vec{r}_v\|$ part will cancel out of the integrand, making the subsequent integration easier to compute.

18.6 Surface integrals

18.6.1 Definition

Consider a smooth surface $\mathcal{S}$ that represents a thin sheet of metal. How could we find the mass of this metallic object?

If the density of this object is constant, then we can find mass via mass$=$ density $\times$ surface area, and we could compute the surface area using the techniques of the previous section.

What if the density were not constant, but variable, described by a function $\delta(x,y,z)$? We can describe the mass using our general integration techniques as

$$\text{mass} = \iint_{\mathcal{S}} dm,$$

where dm represents a little bit of mass. That is, to find the total mass of the object, sum up lots of little masses over the surface.

How do we find the little bit of mass dm? On a small portion of the surface with surface area ΔS, the density is approximately constant, hence $dm \approx \delta(x,y,z)\Delta S$. As we use limits to shrink the size of ΔS to 0, we get $dm = \delta(x,y,z)dS$; that is, a little bit of mass is equal to a density times a small amount of surface area. Thus the total mass of the thin sheet is

$$\text{mass} = \iint_{\mathcal{S}} \delta(x,y,z)\; dS. \tag{18.3}$$

To evaluate the above integral, we would seek $\vec{r}(u,v)$, a smooth parametrization of $\mathcal{S}$ over a region R of the uv-plane. The density would become a function of u and v, and we would integrate $\iint_R \delta(u,v)\;\|\vec{r}_u \times \vec{r}_v\|\;dA$.

The integral in Equation (18.3) is a specific example of a more general construction defined below.

Definitie 18.11 (Surface integral)
Let $G(x,y,z)$ be a continuous function defined on a surface $\mathcal{S}$. The **surface integral** *(oppervlakte-integraal)* of G on $\mathcal{S}$ is

$$\iint_{\mathcal{S}} G(x,y,z)\; dS.$$

Surface integrals can be used to measure a variety of quantities beyond mass. If $G(x,y,z)$ measures the static charge density at a point, then the surface integral will compute the total static charge of the sheet. If G measures the amount of fluid passing through a screen (represented by $\mathcal{S}$) at a point, then the surface integral gives the total amount of fluid going through the screen.

Example 18.21

Find the mass of a thin sheet modeled by the plane $2x+y+z=3$ over the triangular region of the xy-plane bounded by the coordinate axes and the line $y=2-2x$, as shown in Figure 18.26, with density function $\delta(x,y,z)=x^2+5y+z$, where all distances are measured in cm and the density is given as g/cm^2.

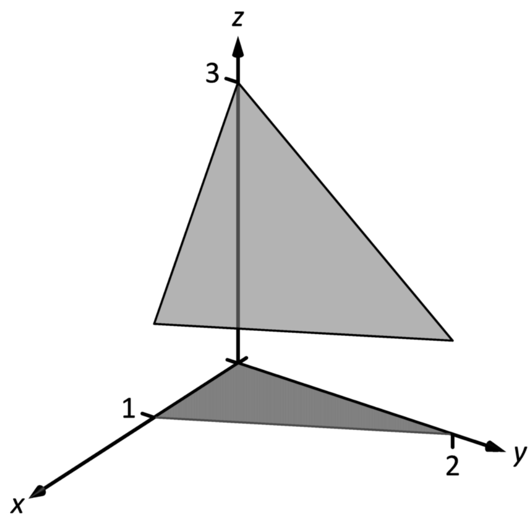


Figure 18.26: The surface whose mass is computed in Example 18.21.

Solution

We begin by parameterizing the planar surface $\mathcal{S}$. Using the techniques of the previous section, we can let $x=u$ and $y=v(2-2u)$, where $0\le u\le 1$ and $0\le v\le 1$. Solving for z in the equation of the plane, we have $z=3-2x-y$, hence $z=3-2u-v(2-2u)$, giving the parametrization $\vec{r}(u,v)=\big(u, v(2-2u), 3-2u-v(2-2u)\big)$.

We need $dS=\|\vec{r}_u\times\vec{r}_v\|\;dA$, so we need to compute $\vec{r}_u$, $\vec{r}_v$ and the norm of their cross product.

We leave it to the reader to confirm the following:

$$\vec{r}_u = (1, -2v, 2v-2), \quad \vec{r}_v = (0, 2-2u, 2u-2),$$

$$\vec{r}_u \times \vec{r}_v = (4-4u, 2-2u, 2-2u) \quad \text{and} \quad \|\vec{r}_u \times \vec{r}_v\| = 2\sqrt{6}\sqrt{(u-1)^2}.$$

We need to be careful to not simplify $\|\vec{r}_u \times \vec{r}_v\| = 2\sqrt{6}\sqrt{(u-1)^2}$ as $2\sqrt{6}(u-1)$; rather, it is $2\sqrt{6}|u-1|$. In this example, u is bounded by $0 \le u \le 1$, and on this interval $|u-1| = 1-u$. Thus $dS = 2\sqrt{6}(1-u)\ dA$.

The density is given as a function of x, y and z, for which we will substitute the corresponding components of $\vec{r}$:

$$\begin{aligned}
\delta(x,y,z) &= \delta\big(\vec{r}(u,v)\big) \\
&= u^2 + 5v(2-2u) + 3 - 2u - v(2-2u) \\
&= u^2 - 8uv - 2u + 8v + 3.
\end{aligned}$$

Thus the mass of the sheet is:

$$\begin{aligned}
M &= \iint_{\mathcal{S}} dm \\
&= \iint_R \delta\big(\vec{r}(u,v)\big)\ \|\vec{r}_u \times \vec{r}_v\|\ dA \\
&= \int_0^1 \int_0^1 \Big(u^2 - 8uv - 2u + 8v + 3\Big)\Big(2\sqrt{6}(1-u)\Big)\ du\ dv \\
&= \frac{31}{\sqrt{6}} \approx 12.66 \text{ g}.
\end{aligned}$$

18.6.2 Flux

Let a surface $\mathcal{S}$ lie within a vector field $\vec{F}$. One is often interested in measuring the flux of $\vec{F}$ across $\mathcal{S}$; that is, measuring how much of the vector field passes across $\mathcal{S}$. For instance, if $\vec{F}$ represents the velocity field of moving air and $\mathcal{S}$ represents the shape of an air filter, the flux will measure how much air is passing through the filter per unit time.

As flux measures the amount of $\vec{F}$ passing across $\mathcal{S}$, we need to find the amount of $\vec{F}$ orthogonal to $\mathcal{S}$. Similar to our measure of flux in the plane, this is equal to $\vec{F} \cdot \hat{n}$, where $\hat{n}$ is a unit vector normal to $\mathcal{S}$ at a point. We now consider how to find $\hat{n}$.

Given a smooth parametrization $\vec{r}(u,v)$ of $\mathcal{S}$, the work in the previous section showing the development of our method of computing surface area also shows that $\vec{r}_u(u,v)$ and $\vec{r}_v(u,v)$ are tangent to $\mathcal{S}$ at $\vec{r}(u,v)$. Thus $\vec{r}_u \times \vec{r}_v$ is orthogonal to $\mathcal{S}$, and we let

$$\hat{n} = \frac{\vec{r}_u \times \vec{r}_v}{\|\vec{r}_u \times \vec{r}_v\|},$$

which is a unit vector normal to $\mathcal{S}$ at $\vec{r}(u,v)$.

The measurement of flux across a surface is a surface integral; that is, to measure total flux we sum the product of $\vec{F} \cdot \hat{n}$ times a small amount of surface area: $\vec{F} \cdot \hat{n}\ dS$.

A nice thing happens with the actual computation of flux: the $\|\vec{r}_u \times \vec{r}_v\|$ terms go away. Consider:

$$\begin{aligned}\text{flux} &= \iint_{\mathcal{S}} \vec{F}\cdot\hat{n}\; dS\\ &= \iint_R \vec{F}\cdot\frac{\vec{r}_u\times\vec{r}_v}{\|\vec{r}_u\times\vec{r}_v\|}\;\|\vec{r}_u\times\vec{r}_v\|\; dA\\ &= \iint_R \vec{F}\cdot(\vec{r}_u\times\vec{r}_v)\; dA.\end{aligned}$$

Recall that R is the region in the uv-plane that we get by projecting $\mathcal{S}$ onto that plane. The above only makes sense if $\mathcal{S}$ is orientable; the normal vectors $\hat{n}$ must vary continuously across $\mathcal{S}$. We assume that $\hat{n}$ does vary continuously. If the parametrization $\vec{r}$ of $\mathcal{S}$ is smooth, then our above definition of $\hat{n}$ will vary continuously.

Definitie 18.12 (Flux across a surface)
Let $\vec{F}$ be a vector field with continuous components defined on an orientable surface $\mathcal{S}$ with normal vector $\hat{n}$. The **flux of $\vec{F}$ across** $\mathcal{S}$ is

$$\text{flux} = \iint_{\mathcal{S}} \vec{F}\cdot\hat{n}\; dS.$$

If $\mathcal{S}$ is parametrized by $\vec{r}(u,v)$, which is smooth on its domain R, then

$$\text{flux} = \iint_R \vec{F}\big(\vec{r}(u,v)\big)\cdot(\vec{r}_u\times\vec{r}_v)\; dA.$$

Since $\mathcal{S}$ is orientable, we adopt the convention of saying one passes from the back side of $\mathcal{S}$ to the front side when moving across the surface parallel to the direction of $\hat{n}$. Also, when $\mathcal{S}$ is closed, it is natural to speak of the regions of space inside and outside $\mathcal{S}$. We also adopt the convention that when $\mathcal{S}$ is a closed surface, $\hat{n}$ should point to the outside of $\mathcal{S}$. If $\hat{n} = \vec{r}_u \times \vec{r}_v$ points inside $\mathcal{S}$, use $\hat{n} = \vec{r}_v \times \vec{r}_u$ instead.

When the computation of flux is positive, it means that the field is moving from the back side of $\mathcal{S}$ to the front side; when flux is negative, it means the field is moving opposite the direction of $\hat{n}$, and is moving from the front of $\mathcal{S}$ to the back. When $\mathcal{S}$ is not closed, there is not a right and wrong direction in which $\hat{n}$ should point, but one should be mindful of its direction to make full sense of the flux computation.

We demonstrate the computation of flux, and its interpretation, in the following examples.

Example 18.22

Let $\mathcal{S}$ be the surface given in Example 18.21, where $\mathcal{S}$ is parametrized by $\vec{r}(u,v) = \big(u, v(2-2u), 3-2u-v(2-2u)\big)$ on $0 \le u \le 1$, $0 \le v \le 1$, and let $\vec{F} = (1, x, -y)$, as shown in Figure 18.27. Find the flux of $\vec{F}$ across $\mathcal{S}$.

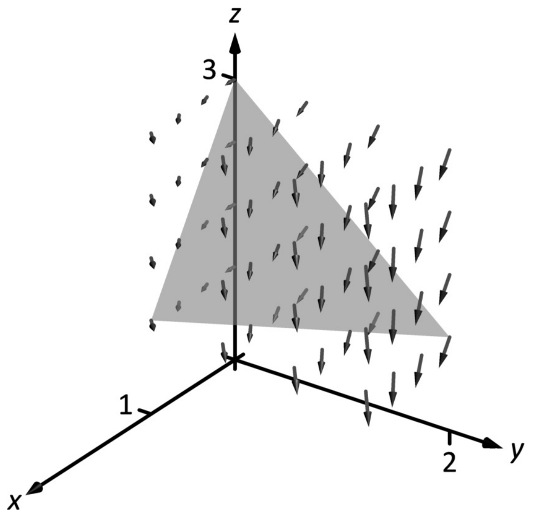


Figure 18.27: The surface and vector field used in Example 18.22.

Solution

Using our work from the previous example, we have $\hat{\boldsymbol{n}} = \vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v = \langle 4-4u, 2-2u, 2-2u \rangle$. We also need $\vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}(u,v)\big) = \big\langle 1, u, -v(2-2u) \big\rangle$.

Thus the flux of $\vec{\boldsymbol{F}}$ across $\mathcal{S}$ is:

$$\begin{aligned}
\text{flux} &= \iint_{\mathcal{S}} \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}\ dS \\
&= \iint_R \big\langle 1, u, -v(2-2u) \big\rangle \cdot \langle 4-4u, 2-2u, 2-2u \rangle\ dA \\
&= \int_0^1 \int_0^1 \big(-4u^2v - 2u^2 + 8uv - 2u - 4v + 4 \big)\ du\ dv \\
&= \frac{5}{3}.
\end{aligned}$$

To make full use of this numeric answer, we need to know the direction in which the field is passing across $\mathcal{S}$. The graph in Figure 18.27 helps, but we need a method that is not dependent on a graph.

Pick a point (u,v) in the interior of R, which is the region in the uv-plane that we get by projecting $\mathcal{S}$ onto that plane, and consider $\hat{\boldsymbol{n}}(u,v)$. For instance, choose $(u,v) = (1/2, 1/2)$ and look at $\hat{\boldsymbol{n}}\big(\vec{\boldsymbol{r}}(1/2,1/2)\big) = \langle 2,1,1 \rangle / \sqrt{6}$. This vector has positive x-, y- and z-components. Generally speaking, one has some idea of what the surface $\mathcal{S}$ looks like, as that surface is for some reason important. In our case, we know $\mathcal{S}$ is a plane with z-intercept of $z = 3$. Knowing $\hat{\boldsymbol{n}}$ and the flux measurement of positive $5/3$, we know that the field must be passing from behind $\mathcal{S}$, i.e., the side the origin is on, to the front of $\mathcal{S}$.

Example 18.23

Let $\mathcal{S}_1$ be the unit disk in the xy-plane, and let $\mathcal{S}_2$ be the paraboloid $z = 1 - x^2 - y^2$, for $z \geq 0$, as graphed in Figure 18.28. Note how these two surfaces each have the unit circle as a boundary.

Let $\vec{\boldsymbol{F}}_1 = \langle 0,0,1 \rangle$ and $\vec{\boldsymbol{F}}_2 = \langle 0,0,z \rangle$. Using normal vectors for each surface that point upward, i.e., with a positive z-component, find the flux of each field across each surface.

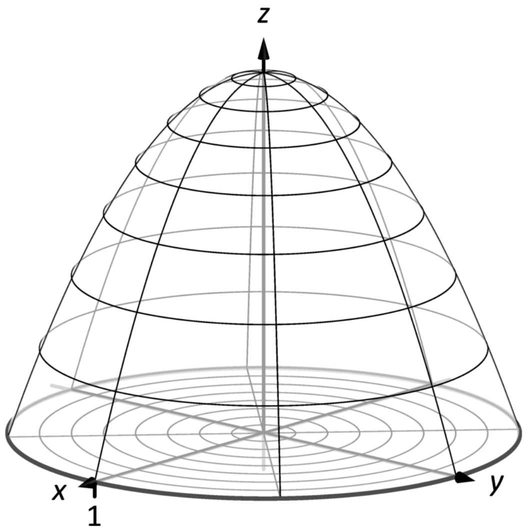


Figure 18.28: The surfaces used in Example 18.23.

Solution

We begin by parameterizing each surface.

The boundary of the unit disk in the xy-plane is the unit circle, which can be described with $\big(\cos(u), \sin(u), 0\big)$, $0 \le u \le 2\pi$. To obtain the interior of the circle as well, we can scale by v, giving

$$\vec{r}_1(u,v) = \big(v\cos(u), v\sin(u), 0\big), \quad 0 \le u \le 2\pi, \quad 0 \le v \le 1.$$

As the boundary of $\mathcal{S}_2$ is also the unit circle, the x- and y-components of $\vec{r}_2$ will be the same as those of $\vec{r}_1$; we just need a different z component. With $z = 1 - x^2 - y^2$, we have

$$\vec{r}_2(u,v) = \Big(v\cos(u), v\sin(u), 1 - v^2\cos^2(u) - v^2\sin^2(u)\Big) = \Big(v\cos(u), v\sin(u), 1 - v^2\Big),$$

where $0 \le u \le 2\pi$ and $0 \le v \le 1$.

We now compute the normal vectors $\hat{n}_1$ and $\hat{n}_2$.

For $\hat{n}_1$: $(\vec{r}_1)_u = \big(-v\sin(u), v\cos(u), 0\big)$, $(\vec{r}_1)_v = \big(\cos(u), \sin(u), 0\big)$, so

$$\hat{n}_1 = (\vec{r}_1)_u \times (\vec{r}_1)_v = (0, 0, -v)\,.$$

As this vector has a negative z-component, we instead use

$$\hat{n}_1 = (\vec{r}_1)_v \times (\vec{r}_1)_u = (0, 0, v)\,.$$

Similarly, $\hat{n}_2$: $(\vec{r}_2)_u = \big(-v\sin(u), v\cos(u), 0\big)$, $(\vec{r}_2)_v = \big(\cos(u), \sin(u), -2v\big)$, so

$$\hat{n}_2 = (\vec{r}_2)_u \times (\vec{r}_2)_v = \Big(-2v^2\cos(u), -2v^2\sin(u), -v\Big).$$

Again, this normal vector has a negative z-component so we use

$$\hat{n}_2 = (\vec{r}_2)_v \times (\vec{r}_2)_u = \Big(2v^2\cos(u), 2v^2\sin(u), v\Big).$$

We are now set to compute flux. Over field $\vec{F}_1 = (0,0,1)$:

$$\begin{aligned}\text{flux across } \mathcal{S}_1 &= \iint_{\mathcal{S}_1} \vec{F}_1 \cdot \hat{n}_1 \; dS \\ &= \iint_R (0,0,1)\cdot(0,0,v)\; dA \\ &= \int_0^1 \int_0^{2\pi} (v)\; du\; dv \\ &= \pi.\end{aligned}$$

$$\begin{aligned}\text{flux across } \mathcal{S}_2 &= \iint_{\mathcal{S}_2} \vec{F}_1 \cdot \hat{n}_2 \; dS \\ &= \iint_R (0,0,1)\cdot\big(2v^2\cos(u), 2v^2\sin(u), v\big)\; dA \\ &= \int_0^1 \int_0^{2\pi} (v)\; du\; dv \\ &= \pi.\end{aligned}$$

These two results are equal and positive. Each are positive because both normal vectors are pointing in the positive z-directions, as does $\vec{F}_1$. As the field passes through each surface in the direction of their normal vectors, the flux is measured as positive.

We can also intuitively understand why the results are equal. Consider $\vec{F}_1$ to represent the flow of air, and let each surface represent a filter. Since $\vec{F}_1$ is constant, and moving straight up, it makes sense that all air passing through $\mathcal{S}_1$ also passes through $\mathcal{S}_2$, and vice–versa.

If we treated the surfaces as creating one piecewise smooth surface $\mathcal{S}$, we would find the total flux across $\mathcal{S}$ by finding the flux across each piece, being sure that each normal vector pointed to the outside of the closed surface. Above, $\hat{n}_1$ does not point outside the surface, though $\hat{n}_2$ does. We would instead want to use $-\hat{n}_1$ in our computation. We would then find that the flux across $\mathcal{S}_1$ is $-\pi$, and hence the total flux across $\mathcal{S}$ is $-\pi+\pi=0$. As 0 is a special number, we should wonder if this answer has special significance. It does, which is briefly discussed following this example and will be more fully developed in the next section.

We now compute the flux across each surface with $\vec{F}_2 = (0,0,z)$. First, the flux across $\mathcal{S}_1$:

$$\iint_{\mathcal{S}_1} \vec{F}_2 \cdot \hat{n}_1 \; dS.$$

Over $\mathcal{S}_1$, $\vec{F}_2 = \vec{F}_2\big(\vec{r}_1(u,v)\big) = (0,0,0)$. Therefore,

$$\begin{aligned}\text{flux across } \mathcal{S}_1 &= \iint_R (0,0,0)\cdot(0,0,v)\; dA \\ &= \int_0^1 \int_0^{2\pi} (0)\; du\; dv \\ &= 0.\end{aligned}$$

Then, the flux across $\mathcal{S}_2$:

$$\iint_{\mathcal{S}_2} \vec{\boldsymbol{F}}_2 \cdot \hat{\boldsymbol{n}}_2 \, dS.$$

Over $\mathcal{S}_2$, $\vec{\boldsymbol{F}}_2 = \vec{\boldsymbol{F}}_2\big(\vec{\boldsymbol{r}}_2(u,v)\big) = \big(0,0,1-v^2\big)$. Therefore,

$$\begin{aligned}
\text{flux across } \mathcal{S}_2 &= \iint_R \big(0,0,1-v^2\big)\cdot\big(2v^2\cos(u),2v^2\sin(u),v\big)\, dA \\
&= \int_0^1\int_0^{2\pi} (v-v^3)\, du\, dv \\
&= \frac{\pi}{2}.
\end{aligned}$$

This time the measurements of flux differ. Over $\mathcal{S}_1$, the field $\vec{\boldsymbol{F}}_2$ is just $\vec{\boldsymbol{0}}$, hence there is no flux. Over $\mathcal{S}_2$, the flux is again positive as $\vec{\boldsymbol{F}}_2$ points in the positive z-direction over $\mathcal{S}_2$, as does $\hat{\boldsymbol{n}}_2$.

In the previous example, the surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ form a closed surface that is piecewise smooth. That the measurement of flux across each surface was the same for some fields (and not for others) is reminiscent of a result from Section 18.4, where we measured flux across curves. The quick answer to why the flux was the same when considering $\vec{\boldsymbol{F}}_1$ is that $\operatorname{div}\vec{\boldsymbol{F}}_1 = 0$. In the next section, we'll see the second part of the Divergence Theorem which will more fully explain this occurrence. We will also explore Stokes' theorem, the spatial analogue to Green's theorem.

18.7 The divergence theorem revisited and Stokes' theorem

18.7.1 The divergence theorem

Theorem 18.6 gives the divergence theorem in the plane, which states that the flux of a vector field across a closed *curve* equals the sum of the divergences over the region enclosed by the curve. Recall that the flux was measured via a line integral, and the sum of the divergences was measured through a double integral.

We now consider the three-dimensional version of the divergence theorem. It states, in words, that the flux across a closed *surface* equals the sum of the divergences over the domain enclosed by the surface. Since we are in space (versus the plane), we measure flux via a surface integral, and the sums of divergences will be measured through a triple integral.

Theorem 18.8 (The divergence theorem (in space))
Let D be a closed domain in space whose boundary is an orientable, piecewise smooth surface $\mathcal{S}$ with outer unit normal vector $\hat{\boldsymbol{n}}$, and let $\vec{\boldsymbol{F}}$ be a vector field whose components are differentiable on D. Then

$$\iiint_D \operatorname{div}\vec{\boldsymbol{F}}\, dV = \iint_{\mathcal{S}} \vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}}\, dS.$$

Note that the term outer unit normal vector used in Theorem 18.8 means $\hat{\boldsymbol{n}}$ points to the outside of $\mathcal{S}$.

Proof It suffices to prove the theorem for rectangular prisms; the Riemann sum nature of the triple integral then guarantees the theorem for arbitrary regions. So, let

$$D = \big\{(x,y,z)\,\big|\,a_1 \le x \le a_2, b_1 \le y \le b_2, c_1 \le z \le c_2\big\}$$

and let S be the boundary surface of R. Then

$$S = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$$

where A_1 and A_2 are those sides perpendicular to the x-axis, A_3 and A_4 are those sides perpendicular to the y-axis and A_5 and A_6 are those sides perpendicular to the z-axis. Furthermore, we let $\vec{\boldsymbol{F}} = M\,\hat{\boldsymbol{\imath}} + N\,\hat{\boldsymbol{\jmath}} + P\,\hat{\boldsymbol{k}}$, where M, N, $P : \mathbb{R}^3 \mapsto \mathbb{R}$.

Then, we have

$$\begin{aligned}
\iiint_D \vec{\nabla}\cdot\vec{\boldsymbol{F}}\,dV &= \iiint_D \left(\frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} + \frac{\partial P}{\partial z}\right) dx\,dy\,dz \\
&= \iiint_D \frac{\partial M}{\partial x}\,dx\,dy\,dz + \iiint_D \frac{\partial N}{\partial y}\,dx\,dy\,dz + \iiint_D \frac{\partial P}{\partial z}\,dx\,dy\,dz \\
&= \int_{c_1}^{c_2}\int_{b_1}^{b_2} \big(M(a_2,y,z) - M(a_1,y,z)\big)\,dy\,dz + \int_{c_1}^{c_2}\int_{a_1}^{a_2} \big(N(x,b_2,z) - N(x,b_1,z)\big)\,dx\,dz \\
&+ \int_{b_1}^{b_2}\int_{a_1}^{a_2} \big(P(x,y,c_2) - P(x,y,c_1)\big)\,dx\,dy \\
&= \iint_{A_2} M\,dy\,dz - \iint_{A_1} M\,dy\,dz + \iint_{A_4} N\,dx\,dz - \iint_{A_3} N\,dx\,dz \\
&+ \iint_{A_6} P\,dx\,dy - \iint_{A_5} P\,dx\,dy
\end{aligned}$$

We turn now to examine $\hat{\boldsymbol{n}}$:

$$\begin{aligned}
\hat{\boldsymbol{n}} &= (-1,0,0) \quad \text{on} \quad A_1, \\
\hat{\boldsymbol{n}} &= (1,0,0) \quad \text{on} \quad A_2, \\
\hat{\boldsymbol{n}} &= (0,-1,0) \quad \text{on} \quad A_3, \\
\hat{\boldsymbol{n}} &= (0,1,0) \quad \text{on} \quad A_4, \\
\hat{\boldsymbol{n}} &= (0,0,-1) \quad \text{on} \quad A_5, \\
\hat{\boldsymbol{n}} &= (0,0,1) \quad \text{on} \quad A_6.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= -M \quad \text{on} \quad A_1, \\
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= M \quad \text{on} \quad A_2, \\
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= -N \quad \text{on} \quad A_3, \\
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= N \quad \text{on} \quad A_4, \\
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= -P \quad \text{on} \quad A_5, \\
\vec{\boldsymbol{F}}\cdot\hat{\boldsymbol{n}} &= P \quad \text{on} \quad A_6.
\end{aligned}$$

Besides, we also have that

$$\begin{aligned} dS &= dy\,dz \quad \text{on} \quad A_1 \text{ and } A_2, \\ dS &= dx\,dz \quad \text{on} \quad A_3 \text{ and } A_4, \\ dS &= dx\,dy \quad \text{on} \quad A_5 \text{ and } A_6, \end{aligned}$$

where dS is the area element. This is true because each side is perfectly flat, and constant with respect to one coordinate. Consequently

$$\iint_{A_2} \vec{F} \cdot \hat{n}\, dS = \iint_{A_2} M\, dy\, dz$$

and in general

$$\sum_{i=1}^{6} \iint_{A_i} \vec{F} \cdot \hat{n}\, dS = \iint_{A_2} M\, dy\, dz - \iint_{A_1} M\, dy\, dz + \iint_{A_4} N\, dx\, dz - \iint_{A_3} N\, dx\, dz + \iint_{A_6} P\, dx\, dy - \iint_{A_5} P\, dx\, dy. \quad \square$$

Hence, the result

$$\iiint_D \operatorname{div} \vec{F}\, dV = \iint_{\mathcal{S}} \vec{F} \cdot \hat{n}\, dS.$$

Example 18.24

Let D be the domain in space bounded by the planes $z = 0$ and $z = 2x$, along with the cylinder $x = 1 - y^2$, as graphed in Figure 18.29, let $\mathcal{S}$ be the boundary of D, and let $\vec{F} = \left(x + y, y^2, 2z\right)$.

Verify the divergence theorem by finding the total outward flux of $\vec{F}$ across $\mathcal{S}$, and show this is equal to $\iiint_D \operatorname{div} \vec{F}\, dV$.

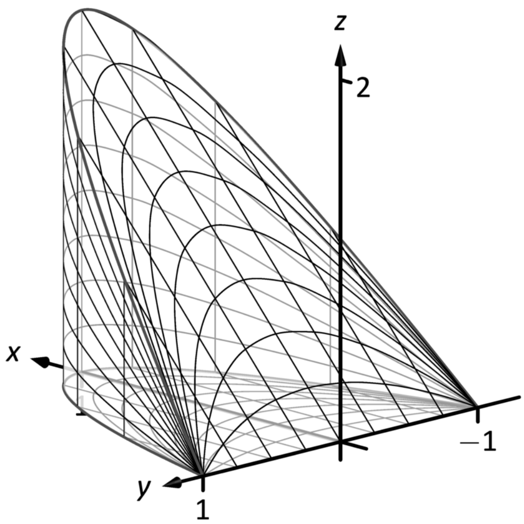


Figure 18.29: The surfaces used in Example 18.24.

Solution

The surface $\mathcal{S}$ is piecewise smooth, comprising surfaces $\mathcal{S}_1$, which is part of the plane $z = 2x$, surface $\mathcal{S}_2$, which is part of the cylinder $x = 1 - y^2$, and surface $\mathcal{S}_3$, which is part of the plane $z = 0$. To find the total outward flux across $\mathcal{S}$, we need to compute the outward flux across each of these three surfaces.

We leave it to the reader to confirm that surfaces $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ can be parameterized by $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$ and $\vec{\boldsymbol{r}}_3$ respectively as

$$\begin{aligned}
\vec{\boldsymbol{r}}_1(u,v) &= \Big(v\big(1-u^2\big), u, 2v\big(1-u^2\big)\Big),\\
\vec{\boldsymbol{r}}_2(u,v) &= \Big(1-u^2, u, 2v\big(1-u^2\big)\Big),\\
\vec{\boldsymbol{r}}_3(u,v) &= \Big(v\big(1-u^2\big), u, 0\Big),
\end{aligned}$$

where $-1 \le u \le 1$ and $0 \le v \le 1$ for all three functions.

We compute a unit normal vector $\hat{\boldsymbol{n}}$ for each as $\frac{\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v}{\|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|}$, though recall that as we are integrating $\vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}\ dS$, we actually only use $\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v$. Finally, in previous flux computations, it did not matter which direction $\hat{\boldsymbol{n}}$ pointed as long as we made note of its direction. When using the divergence theorem, we need $\hat{\boldsymbol{n}}$ to point to the outside of the closed surface, so in practice this means we will either use $\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v$ or $\vec{\boldsymbol{r}}_v \times \vec{\boldsymbol{r}}_u$, depending on which points outside of the closed surface $\mathcal{S}$.

We leave it to the reader to confirm the following cross products and integrations are correct.

For $\mathcal{S}_1$, we need to use $\vec{(r_1)}_v \times \vec{(r_1)}_u = \Big(2\big(u^2-1\big), 0, 1-u^2\Big)$. Note the z-component is non-negative as $-1 \le u \le 1$, therefore this vector always points up, meaning to the outside, of $\mathcal{S}$. The flux across $\mathcal{S}_1$ is:

$$\begin{aligned}
\iint_{\mathcal{S}_1} \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}_1\ dS &= \int_0^1 \int_{-1}^1 \vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}_1(u,v)\big) \cdot \big(\vec{(r_1)}_v \times \vec{(r_1)}_u\big)\ du\ dv\\
&= \int_0^1 \int_{-1}^1 \Big(v\big(1-u^2\big)+u, u^2, 4v\big(1-u^2\big)\Big) \cdot \Big(2\big(u^2-1\big), 0, 1-u^2\Big)\ du\ dv\\
&= \int_0^1 \int_{-1}^1 \Big(2u^4v + 2u^3 - 4u^2v - 2u + 2v\Big)\ du\ dv\\
&= \frac{16}{15}.
\end{aligned}$$

For $\mathcal{S}_2$, we use $\vec{(r_2)}_u \times \vec{(r_2)}_v = \Big(2\big(1-u^2\big), 4u\big(1-u^2\big), 0\Big)$. Note the x-component is always non-negative as $-1 \le u \le 1$, meaning this vector points outside $\mathcal{S}$. The flux across $\mathcal{S}_2$ is:

$$\begin{aligned}
\iint_{\mathcal{S}_2} \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{n}}_2\ dS &= \int_0^1 \int_{-1}^1 \vec{\boldsymbol{F}}\big(\vec{\boldsymbol{r}}_2(u,v)\big) \cdot \big(\vec{(r_2)}_u \times \vec{(r_2)}_v\big)\ du\ dv\\
&= \int_0^1 \int_{-1}^1 \Big(1-u^2+u, u^2, 4v\big(1-u^2\big)\Big) \cdot \Big(2\big(1-u^2\big), 4u\big(1-u^2\big), 0\Big)\ du\ dv
\end{aligned}$$

$$= \int_0^1 \int_{-1}^1 \left(4u^5 - 2u^4 - 2u^3 + 4u^2 - 2u - 2\right)\, du\, dv$$
$$= \frac{32}{15}.$$

For $\mathcal{S}_3$, we use $(r_3)_u \times \vec{(}r_3)_v = \left(0, 0, u^2 - 1\right)$. Note that the z-component is never positive as $-1 \le u \le 1$, meaning this vector points down, outside of $\mathcal{S}$. The flux across $\mathcal{S}_3$ is:

$$\iint_{\mathcal{S}_3} \vec{F} \cdot \hat{n}_3\, dS = \int_0^1 \int_{-1}^1 \vec{F}\big(\vec{r}_3(u,v)\big) \cdot \big((r_3)_u \times \vec{(}r_3)_v\big)\, du\, dv$$
$$= \int_0^1 \int_{-1}^1 \left(v\left(1-u^2\right) + u, u^2, 0\right) \cdot \left(0, 0, u^2 - 1\right)\, du\, dv$$
$$= \int_0^1 \int_{-1}^1 0\, du\, dv$$
$$= 0.$$

Thus the total outward flux, measured by surface integrals across all three component surfaces of $\mathcal{S}$, is $16/15 + 32/15 + 0 = 48/15 = 16/5 = 3.2$. We now find the total outward flux by integrating $\operatorname{div} \vec{F}$ over D.

Following the steps outlined in Section 17.5, we see the bounds of x, y and z can be set as (thinking "surface to surface, curve to curve, point to point"):

$$0 \le z \le 2x, \quad 0 \le x \le 1 - y^2, \quad -1 \le y \le 1.$$

With $\operatorname{div} \vec{F} = 1 + 2y + 2 = 2y + 3$, we find the total outward flux of $\vec{F}$ over $\mathcal{S}$ as:

$$\iiint_D \operatorname{div} \vec{F}\, dV = \int_{-1}^1 \int_0^{1-y^2} \int_0^{2x} \left(2y + 3\right)\, dz\, dx\, dy = \frac{16}{5},$$

the same result we obtained previously.

In Example 18.24 we see that the total outward flux of a vector field across a closed surface can be found two different ways because of the divergence theorem. One computation took far less work to obtain. In that particular case, since $\mathcal{S}$ was comprised of three separate surfaces, it was far simpler to compute one triple integral than three surface integrals (each of which required partial derivatives and a cross product). In practice, if outward flux needs to be measured, one would choose only one method. We will use both methods in this section simply to reinforce the truth of the divergence theorem.

We practice again in the following example.

Example 18.25

Let $\mathcal{S}$ be the surface formed by the paraboloid $z = 1 - x^2 - y^2$, $z \ge 0$, and the unit disk centred at the origin in the xy-plane, graphed in Figure 18.30, and let $\vec{F} = (0, 0, z)$. This surface and vector field were used in Example 18.23.

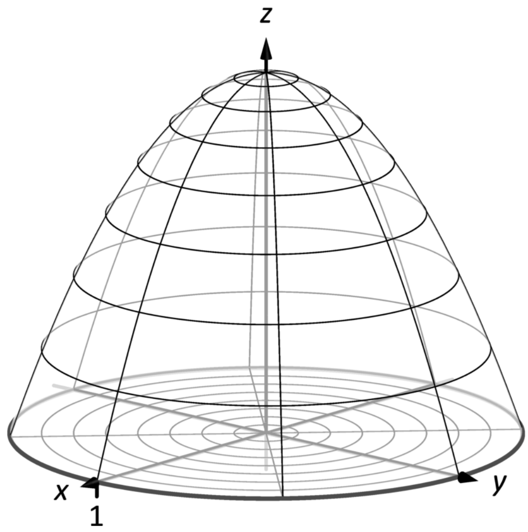


Figure 18.30: The surfaces used in Example 18.25.

Verify the divergence theorem; find the total outward flux across $\mathcal{S}$ and evaluate the triple integral of $\operatorname{div}\vec{F}$, showing that these two quantities are equal.

Solution

We find the flux across $\mathcal{S}$ first. As $\mathcal{S}$ is piecewise smooth, we decompose it into smooth components $\mathcal{S}_1$, the disk, and $\mathcal{S}_2$, the paraboloid, and find the flux across each.

In Example 18.23, we found the flux across $\mathcal{S}_1$ is 0. We also found that the flux across $\mathcal{S}_2$ is $\pi/2$. In that example, the normal vector had a positive z-component hence was an outer normal. Thus the total outward flux is $0+\pi/2=\pi/2$.

We now compute $\iiint_D \operatorname{div}\vec{F}\, dV$. We can describe D as the domain bounded by:

$$0 \le z \le 1-x^2-y^2, \quad -\sqrt{1-x^2} \le y \le \sqrt{1-x^2}, \quad -1 \le x \le 1.$$

This description of D is not very easy to integrate. With polar, we can do better. Let R represent the unit disk, which can be described in polar simply as r, where $0 \le r \le 1$ and $0 \le \theta \le 2\pi$. With $x = r\cos(\theta)$ and $y = r\sin(\theta)$, the surface $\mathcal{S}_2$ becomes

$$z = 1-x^2-y^2 = 1-\big(r\cos(\theta)\big)^2-\big(r\sin(\theta)\big)^2 = 1-r^2.$$

Thus D can be described as the domain bounded by:

$$0 \le z \le 1-r^2, \quad 0 \le r \le 1, \quad 0 \le \theta \le 2\pi.$$

With $\operatorname{div}\vec{F} = 1$, we can integrate, recalling that $dV = r\, dz\, dr\, d\theta$:

$$\iiint_D \operatorname{div}\vec{F}\, dV = \int_0^{2\pi}\int_0^1\int_0^{1-r^2} r\, dz\, dr\, d\theta = \frac{\pi}{2},$$

which matches our flux computation above.

Our interest in the divergence theorem is twofold. First, its truth alone is interesting: to study the behavior of a vector field across a closed surface, one can examine properties of that field within the surface. Secondly, it offers an alternative way of computing flux. When there are multiple methods of computing a desired quantity, one has power to select the easiest computation as illustrated next.

Example 18.26

Let $\mathcal{S}$ be the cube bounded by the planes $x = \pm 1$, $y = \pm 1$, $z = \pm 1$, and let $\vec{\boldsymbol{F}} = \left(x^2y, 2yz, x^2z^3\right)$. Compute the outward flux of $\vec{\boldsymbol{F}}$ over $\mathcal{S}$.

Solution

We compute $\operatorname{div}\vec{\boldsymbol{F}} = 2xy + 2z + 3x^2z^2$. By the divergence theorem, the outward flux is the triple integral over the domain D enclosed by $\mathcal{S}$:

$$\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} (2xy + 2z + 3x^2z^2)\, dz\, dy\, dx = \frac{8}{3}.$$

The direct flux computation requires six surface integrals, one for each face of the cube. The divergence theorem offers a much more simple computation.

18.7.2 Stokes' theorem

Just as the spatial divergence theorem of this section is an extension of the planar Divergence Theorem, Stokes' theorem is the spatial extension of Green's theorem. Recall that Green's theorem states that the circulation of a vector field around a closed curve in the plane is equal to the sum of the curl of the field over the region enclosed by the curve. Stokes' Theorem effectively makes the same statement: given a closed curve that lies on a surface $\mathcal{S}$, the circulation of a vector field around that curve is the same as the sum of the curl of the field across the enclosed surface. We use quotes around the curl of the field to signify that this statement is not quite correct, as we do not sum $\operatorname{curl}\vec{\boldsymbol{F}}$, but $\left(\operatorname{curl}\vec{\boldsymbol{F}}\right)\cdot\hat{\boldsymbol{n}}$, where $\hat{\boldsymbol{n}}$ is a unit vector normal to $\mathcal{S}$. That is, we sum the portion of $\operatorname{curl}\vec{\boldsymbol{F}}$ that is orthogonal to $\mathcal{S}$ at a point.

Green's theorem dictated that the curve was to be traversed counterclockwise when measuring circulation. Stokes' Theorem will follow a right hand rule: when the thumb of one's right hand points in the direction of $\hat{\boldsymbol{n}}$, the path C will be traversed in the direction of the curling fingers of the hand.

Theorem 18.9 (Stokes' theorem)

Let $\mathcal{S}$ be a piecewise smooth, orientable surface whose boundary is a piecewise smooth curve C, let $\hat{\boldsymbol{n}}$ be a unit vector normal to $\mathcal{S}$, let C be traversed with respect to $\hat{\boldsymbol{n}}$ according to the right hand rule, and let the components of $\vec{\boldsymbol{F}}$ have continuous first partial derivatives over $\mathcal{S}$. Then

$$\oint_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \iint_{\mathcal{S}} \left(\operatorname{curl}\vec{\boldsymbol{F}}\right)\cdot\hat{\boldsymbol{n}}\, dS.$$

In general, the best approach to evaluating the surface integral in Stokes' theorem is to parametrize the surface $\mathcal{S}$ with a function $\vec{\boldsymbol{r}}(u,v)$. We can find a unit normal vector $\hat{\boldsymbol{n}}$ as

$$\hat{\boldsymbol{n}} = \frac{\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v}{\|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|}.$$

Since $dS = \|\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\|\; dA$, the surface integral in practice is evaluated as

$$\iint_{\mathcal{S}} \left(\operatorname{curl}\vec{\boldsymbol{F}}\right)\cdot\left(\vec{\boldsymbol{r}}_u \times \vec{\boldsymbol{r}}_v\right)\; dA,$$

where $\vec{r}_u \times \vec{r}_v$ may be replaced by $\vec{r}_v \times \vec{r}_u$ to properly match the direction of this vector with the orientation of the parameterization of C.

Example 18.27

Considering the planar surface $f(x,y) = 7-2x-2y$, let C be the curve in space that lies on this surface above the circle of radius 1 and centred at $(1,1)$ in the xy-plane, let $\mathcal{S}$ be the planar region enclosed by C, as illustrated in Figure 18.31, and let $\vec{F} = \left(x+y, 2y, y^2\right)$. Verify Stoke's theorem by showing $\oint_C \vec{F}\cdot\, d\vec{r} = \iint_{\mathcal{S}} \left(\operatorname{curl}\vec{F}\right)\cdot \hat{n}\; dS$.

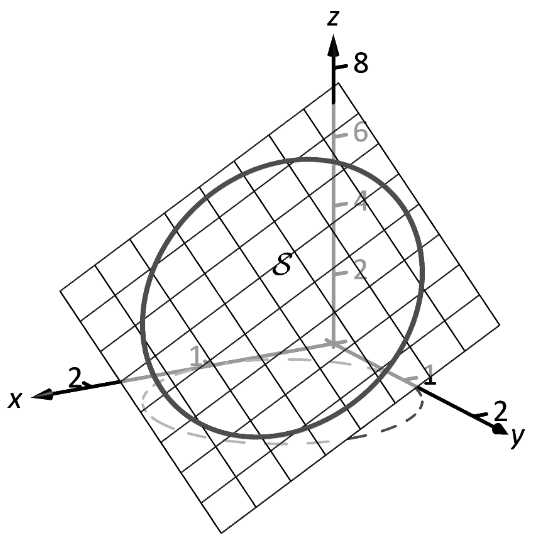


Figure 18.31: As given in Example 18.27, the surface $\mathcal{S}$ is the portion of the plane bounded by the curve.

Solution

We begin by parameterizing C and then find the circulation. A unit circle centred at $(1,1)$ can be parametrized with $x = \cos(t)+1$, $y = \sin(t)+1$ on $0 \le t \le 2\pi$; to put this curve on the surface $f(x,y)$, make the z-component equal $z = 7-2\left(\cos(t)+1\right)-2\left(\sin(t)+1\right) = 3-2\cos(t)-2\sin(t)$. All together, we parametrize C with $\vec{r}(t) = \left(\cos(t)+1, \sin(t)+1, 3-2\cos(t)-2\sin(t)\right)$.

The circulation of $\vec{F}$ around C is

$$\oint_C \vec{F}\cdot d\vec{r} = \int_0^{2\pi} \vec{F}\left(\vec{r}(t)\right)\cdot \vec{r}\,'(t)\; dt$$

which simplifies to

$$\begin{aligned}\oint_C \vec{F}\cdot d\vec{r} &= \int_0^{2\pi} \left(\cos t+\sin t+2, 2\sin t+2, (\sin t+1)^2\right)\cdot(-\sin t, \cos t, -2\sin t+2\cos t)\; dt\\ &= \int_0^{2\pi} \left(2\sin^3(t)-2\cos(t)\sin^2(t)+3\sin^2(t)-3\cos(t)\sin(t)\right)\; dt\\ &= 3\pi.\end{aligned}$$

We now parametrize $\mathcal{S}$. We reuse the letter ‘r’ for our surface as this is our custom. Based on the parametrization of C above, we describe $\mathcal{S}$ with
$\vec{r}(u,v) = \left(v\cos(u)+1, v\sin(u)+1, 3-2v\cos(u)-2v\sin(u)\right)$, where $0 \le u \le 2\pi$ and $0 \le v \le 1$.

We leave it to the reader to confirm that $\vec{r}_u \times \vec{r}_v = (2v, 2v, v)$. As $0 \le v \le 1$, this vector always has a non-negative z-component, which the right–hand rule requires given the orientation of C used above. We also leave it to the reader to confirm $\text{curl}\,\vec{F} = (2y, 0, -1)$.

The surface integral of Stokes' theorem is thus

$$\begin{aligned}\iint_{\mathcal{S}} (\text{curl}\,\vec{F}) \cdot \hat{n}\; dS &= \iint_{\mathcal{S}} \big(\text{curl}\,\vec{F}\big) \cdot (\vec{r}_u \times \vec{r}_v)\; dA \\ &= \int_0^1 \int_0^{2\pi} \big(2v\sin(u) + 2, 0, -1\big) \cdot (2v, 2v, v)\; du\, dv \\ &= 3\pi,\end{aligned}$$

which matches our previous result.

One of the interesting results of Stokes' Theorem is that if two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ share the same boundary, then $\iint_{\mathcal{S}_1} (\text{curl}\,\vec{F}) \cdot \hat{n}\; dS = \iint_{\mathcal{S}_2} (\text{curl}\,\vec{F}) \cdot \hat{n}\; dS$. That is, the value of these two surface integrals is somehow independent of the interior of the surface. We demonstrate this principle in the next example.

Example 18.28

Let C be the curve given in Example 18.27 and note that it lies on the surface $z = 6 - x^2 - y^2$. Let $\mathcal{S}$ be the region of this surface bounded by C, and let $\vec{F} = \big(x + y, 2y, y^2\big)$ as in the previous example. Compute $\iint_{\mathcal{S}} \big(\text{curl}\,\vec{F}\big) \cdot \hat{n}\; dS$ to show it equals the result found in the previous example.

Solution

We begin by demonstrating that C lies on the surface $z = 6 - x^2 - y^2$. We can parametrize the x- and y-components of C with $x = \cos(t) + 1$, $y = \sin(t) + 1$ as before. Lifting these components to the surface f gives the z-component as

$$z = 6 - x^2 - y^2 = 6 - \big(\cos(t) + 1\big)^2 - \big(\sin(t) + 1\big)^2 = 3 - 2\cos(t) - 2\sin(t),$$

which is the same z-component as found in Example 18.27. Thus the curve C lies on the surface $z = 6 - x^2 - y^2$, as illustrated in Figure 18.32.

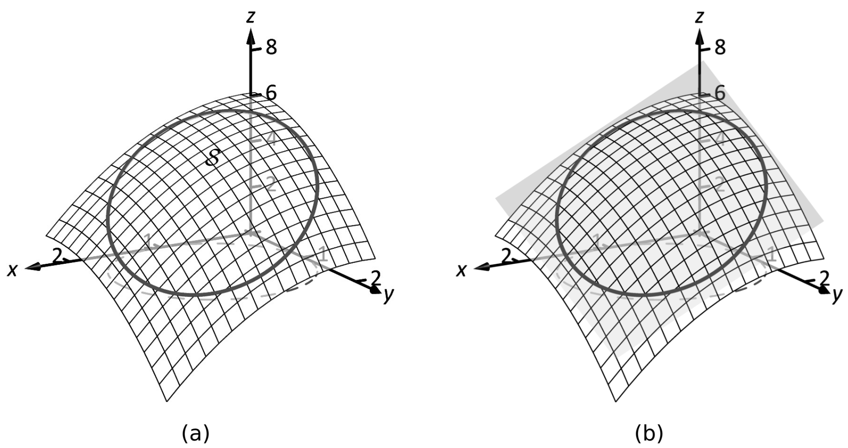


Figure 18.32: As given in Example 18.28, the surface $\mathcal{S}$ is the portion of the plane bounded by the curve.

Since C and $\vec{F}$ are the same as in the previous example, we already know that $\oint_C \vec{F} \cdot d\vec{r} = 3\pi$. We confirm that this is also the value of $\iint_{\mathcal{S}} \big(\text{curl}\,\vec{F}\big) \cdot \hat{n}\; dS$.

We parametrize $\mathcal{S}$ with

$$\vec{r}(u,v)=\Big(v\cos(u)+1, v\sin(u)+1, 6-(v\cos(u)+1)^2-(v\sin(u)+1)^2\Big),$$

where $0 \le u \le 2\pi$ and $0 \le v \le 1$, and leave it to the reader to confirm that

$$\vec{r}_u \times \vec{r}_v = \Big(2v\big(v\cos(u)+1\big), 2v\big(v\sin(u)+1\big), v\Big),$$

which also conforms to the right–hand rule with regard to the orientation of C. With $\operatorname{curl}\vec{F} = (2y, 0, -1)$ as before, we have

$$\begin{aligned}
&\iint_{\mathcal{S}} (\operatorname{curl}\vec{F})\cdot\hat{n}\, dS \\
&= \int_0^1\int_0^{2\pi} \big(2v\sin(u)+2, 0, -1\big)\cdot\Big(2v\big(v\cos(u)+1\big), 2v\big(v\sin(u)+1\big), v\Big)\, du\, dv \\
&= 3\pi.
\end{aligned}$$

Even though the surfaces used in this example and in Example 18.27 are very different, because they share the same boundary, Stokes' theorem guarantees they have equal sum of curls across their respective surfaces.

18.7.3 A common thread of calculus

We have threefold interest in each of the major theorems of this chapter: the fundamental theorem of line integrals, Green's, Stokes' and the divergence theorems. First, we find the beauty of their truth interesting. Second, each provides two methods of computing a desired quantity, sometimes offering a simpler method of computation.

There is yet one more reason of interest in the major theorems of this chapter. These important theorems also all share an important principle with the fundamental theorem of calculus, introduced in Chapter 12.

Revisit this fundamental theorem, adopting the notation used heavily in this chapter. Let I be the interval $[a,b]$ and let $y = F(x)$ be differentiable on I, with $F'(x) = f(x)$. The fundamental theorem of calculus states that

$$\int_I f(x)\, dx = F(b) - F(a).$$

That is, the sum of the rates of change of a function F over an interval I can also be calculated with a certain sum of F itself on the boundary of I (in this case, at the points $x = a$ and $x = b$).

Each of the named theorems above can be expressed in similar terms. Consider the fundamental theorem of line integrals: given a function $z = f(x,y)$, the gradient $\vec{\nabla}f$ is a type of rate of change of f. Given a curve C with initial and terminal points A and B, respectively, this fundamental theorem states that

$$\int_C \vec{\nabla}f\, ds = f(B) - f(A),$$

where again the sum of a rate of change of f along a curve C can also be evaluated by a certain sum of f at the boundary of C (i.e., the points A and B).

Green's theorem is essentially a special case of Stokes' theorem, so we consider just Stokes' theorem here. Recalling that the curl of a vector field $\vec{F}$ is a measure of a rate of change of $\vec{F}$, Stokes' theorem states that over a surface $\mathcal{S}$ bounded by a closed curve C,

$$\iint_S \left(\operatorname{curl} \vec{\boldsymbol{F}}\right) \cdot \hat{\boldsymbol{n}}\ dS = \oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}},$$

i.e., the sum of a rate of change of $\vec{\boldsymbol{F}}$ can be calculated with a certain sum of $\vec{\boldsymbol{F}}$ itself over the boundary of $\mathcal{S}$. In this case, the latter sum is also an infinite sum, requiring an integral.

Finally, the divergence theorems state that the sum of divergences of a vector field (another measure of a rate of change of $\vec{\boldsymbol{F}}$) over a region can also be computed with a certain sum of $\vec{\boldsymbol{F}}$ over the boundary of that region. When the region is planar, the latter sum of $\vec{\boldsymbol{F}}$ is an integral; when the region is spatial, the latter sum of $\vec{\boldsymbol{F}}$ is a double integral.

The common thread among these theorems: the sum of a rate of change of a function over a region can be computed as another sum of the function itself on the boundary of the region. While very general, this is a very powerful and important statement.

18.8 Exercises

Line integrals over a scalar field

Assignment 18.1 — Evaluate the line integral of the scalar functions below along the given curve.

(a) $\displaystyle\int_C x^2\, ds$ along the intersection of $x-y+z=0$ and $x+y+2z=0$ from $(0,0,0)$ to $(3,1,-2)$

(b) $\displaystyle\int_C y\, ds$ from $x=3$ to $x=24$ along the curve $C: y=2\sqrt{x}$

(c) $\displaystyle\int_C (x+y)\, ds$ along the right loop of $r^2 = 2\cos(2\theta)$

(d) $\displaystyle\int_C \frac{ds}{x^2+y^2+z^2}$ along the first bend of $x(t)=8\cos(t),\ y(t)=8\sin(t),\ z(t)=t$

(e) $\displaystyle\int_C \sqrt{2y^2+z^2}\, ds$ along the curve

$$C: \begin{cases} x^2+y^2+z^2=4 \\ y=x \end{cases}$$

(f) $\displaystyle\int_C e^z\, ds$ along the curve $x(t)=e^t\cos(t),\ y(t)=e^t\sin(t),\ z(t)=t$ from $t=0$ to $t=2\pi$

(g) $\displaystyle\int_C \sqrt{1+4x^2z^2}\, ds$ along the curve

$$C: \begin{cases} x^2+z^2=1 \\ y=x^2 \end{cases}$$

(h) $\displaystyle\int_C x\, ds$ along the curve

$$C: \begin{cases} x^2+y^2=a^2 \\ z=x \end{cases}$$

in the first octant

(i) $\displaystyle\int_C z\, ds$ along the curve

$$C: \begin{cases} x^2+y^2+z^2=1 \\ x+y=1 \end{cases} \qquad \text{with } z \geq 0$$

Vector fields

Assignment 18.2 — Which vector field belongs to which graph in Figure 18.33?

(a) $\vec{\boldsymbol{F}}_1 = \dfrac{\vec{\boldsymbol{r}}}{\|\vec{\boldsymbol{r}}\|}$

(b) $\vec{\boldsymbol{F}}_2 = \vec{\boldsymbol{r}}$

(c) $\vec{\boldsymbol{F}}_3 = y\hat{\boldsymbol{\imath}} - x\hat{\boldsymbol{\jmath}}$

(d) $\vec{\boldsymbol{F}}_4 = x\hat{\boldsymbol{\jmath}}$

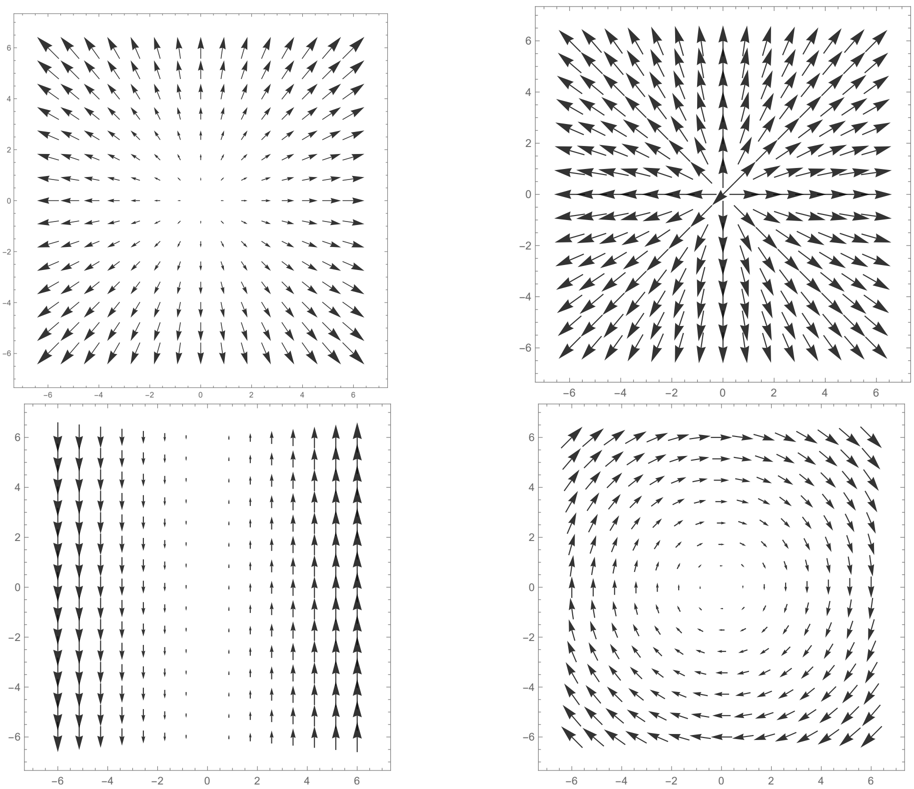


Figure 18.33: Vector fields from Exercise 18.2

Assignment 18.3 — Consider a scalar function $f(x,y,z) = x^2yz^3$ and a vector field $\vec{\boldsymbol{F}} = (xz, -y^2, 2x^2y)$. Find $\vec{\nabla}f$, $\vec{\nabla}\cdot\vec{\boldsymbol{F}}$ and $\vec{\nabla}\times\vec{\boldsymbol{F}}$.

Assignment 18.4 — Consider a scalar function $f(x,y,z) = xy + yz + xz$ and a vector field $\vec{\boldsymbol{F}} = (x^2y, -y^2z, z^2x)$. Find $\vec{\nabla}f$, $\vec{\nabla}\cdot\vec{\boldsymbol{F}}$, $\vec{\nabla}\times\vec{\boldsymbol{F}}$ and $(\vec{\nabla}f)\times\vec{\boldsymbol{F}}$ in the point $(3,-1,2)$.

Assignment 18.5 — Identify all points where the direction of the given vector field does not change.

(a) $\vec{\boldsymbol{F}} = (xy^2, xyz, z-2x)$

(b) $\vec{\boldsymbol{F}} = (xy^3, xyz, z-x^2)$

Assignment 18.6 — Consider the vector field $\vec{\boldsymbol{F}} = \left(\frac{3x}{z}, 2x, 7yz\right)$ and the curve C described by $\vec{\boldsymbol{r}}(t) = (\cos^2(t), \sin(t), -\cos(t))$ with $0 \leq t < 2\pi$. At which point(s) on C does the curl of $\vec{\boldsymbol{F}}$ have a minimal/maximal length.

Assignment 18.7 — The coulomb potential V in vacuum at a point $P(x,y,z)$ cause by a point charge q placed at the origin is given by

$$V(x,y,z)=\frac{q}{4\pi\epsilon_0\sqrt{x^2+y^2+z^2}},$$

with $\epsilon_0\in\mathbb{R}$ a constant. Find the corresponding electric field given by $\vec{\boldsymbol{E}}=-\vec{\nabla}V$.

Line Integrals over vector fields

Assignment 18.8 — Evaluate the line integral over the given vector field along the given curve.

(a) $\vec{\boldsymbol{F}}=\big(\cos(x),y\big)$ along $y=\sin(x)$ from $(0,0)$ to $(\pi,0)$

(b) $\vec{\boldsymbol{F}}=(xy,y-x)$ from $(0,0)$ to $(1,1)$ along the curves $C_1: y=x$, $C_2: y=x^2$ and $C_3: y^2=x$

(c) $\vec{\boldsymbol{F}}=(2xy,x^2)$ from $(0,0)$ to $(1,2)$ along the curves $C_1: y=2x$, $C_2: y=2x^2$, $C_3: y^2=4x$ and $C_4: y=2x^3$

(d) $\vec{\boldsymbol{F}}=\left(\frac{1}{\sqrt{xy}},-\frac{1}{\sqrt{xy}}\right)$ along the curves $C: y=1-x$ from $x=0$ to $x=1$

(e) $\vec{\boldsymbol{F}}=(x^2y,xy^2)$ along the curve

$$C:\begin{cases} x=\frac{t}{2}\\ y=\sqrt{2t}\end{cases}$$

from $(0,0)$ to $(1,2)$

(f) $\vec{\boldsymbol{F}}=\left(\sqrt{b^2-y^2},\sqrt{a^2-x^2}\right)$ along the curve

$$C:\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

with $x\geq 0, y\geq 0$ travelled counterclockwise.

(g) $\vec{\boldsymbol{F}}=(-3y,2x)$ along the curve $C: abca$ with $a(1,2)$, $b(3,1)$ and $c(3,2)$

(h) $\vec{\boldsymbol{F}}=\left(3x^2+2xy,x^2+y^2\right)$ from $(1,1)$ to $(2,2)$ along the curve $C_1: y=x$ and along the path given by $C_2: y=1$ and $C_3: x=2$. Also find the contour integral using a counterclockwise orientation.

(i) $\vec{\boldsymbol{F}}=(y^2,z^2,x^2)$ along the curve

$$C:\begin{cases} y=1\\ x^2+y^2+z^2=5\end{cases}\qquad \text{met } x\geq 0$$

travelled in the direction of increasing z-values

Assignment 18.9 — Do the following sets of points represent an connected or simply connected region or are they not connected?

(a) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x > 0,\ y \geq 0\right\}$

(b) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x = 0,\ y \geq 0\right\}$

(c) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x \neq 0,\ y > 0\right\}$

(d) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 > 1\right\}$

(e) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 > 1\right\}$

(f) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 + z^2 > 1\right\}$

Assignment 18.10 — Verify that the given vector field is conservative and, if possible, find a potential function.

(a) $\vec{F} = \left(xy, \frac{1}{2}x^2 - y^2\right)$

(b) $\vec{F} = (y, x, -2z)$

(c) $\vec{F} = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$

(d) $\vec{F} = \left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2}\right)$

(e) $\vec{F} = \left(2xy - z^2, 2yz + x^2, -2zx + y^2\right)$

(f) $\vec{F} = e^{x^2+y^2+z^2}\,(xz, yz, xy)$

(g) $\vec{F} = \left(xy - \sin(z), \frac{1}{2}x^2 - \frac{e^y}{z}, \frac{e^y}{z^2} - x\cos(z)\right)$

Green's theorem and the divergence theorem

Assignment 18.11 — Verify Green's theorem (Theorem 18.5) for the contour integral

$$\oint_C \left(\left(\sin(x) + 3y^2\right)\, dx + \left(2x - e^{-y^2}\right)\, dy\right)$$

where C is the boundary of the area $x^2 + y^2 \leq a^2,\ y \geq 0$, using a counterclockwise orientation.

Assignment 18.12 — Find the contour integrals below and verify Green's theorem (Theorem 18.5).

(a) $\oint_C \left(y^2\, dx - xy\, dy\right)$

C is the boundary of the region enclosed by $x = 0$, $y = 1$ and $x^2 + y^2 = 2x$

(b) $\oint_C \left(\left(2xy - x^2\right)\, dx + \left(x + y^2\right)\, dy\right)$

C is the boundary of the region enclosed by $x = y^2$ and $y = x^2$

(c) $\oint_C \left(\ln\left(1 + y^{2/3}\right)\, dx + x^2\, dy\right)$

C is the boundary of the region enclosed by $y = 0$, $x = 2$ and $y = x^3$

(d) $\oint_C \left(\left(x^2 - xy\right)\, dx + \left(xy - y^2\right)\, dy\right)$

C is the boundary of the triangle with vertices $(0, 0)$, $(1, 1)$ and $(2, 0)$, using a clockwise orientation

(e) $\oint_C \left(\left(x\sin\left(y^2\right) - y^2\right)\, dx + \left(x^2 y\cos\left(y^2\right) + 3x\right)\, dy\right)$

C is the boundary of the trapezium with vertices $(0,-2)$, $(1,-1)$, $(1,1)$ and $(0,2)$, using a counterclockwise orientation

Assignment 18.13 — Using Green's theorem (Theorem 18.5) find the area of the plane region enclosed by the curve $\vec{\boldsymbol{r}} = \left(a\cos^3(t), b\sin^3(t)\right)$ with $0 \le t \le 2\pi$.

Assignment 18.14 — In each case, verify the divergence theorem in the plane (Theorem 18.6).

(a) $\vec{\boldsymbol{F}} = \left(ay^2, bx^2\right)$, R is the region given by $0 \le x \le 1, 0 \le y \le x$.

(b) $\vec{\boldsymbol{F}} = \left(\sin(x)\cos(y), \cos(x)\sin(y)\right)$, R is the region given by $0 \le x \le \pi/2, 0 \le y \le x$.

(c) $\vec{\boldsymbol{F}} = (y, -x)$, R is the region enclosed by $x^2 + y^2 \le 1$.

Parameterized surfaces and surface area

Assignment 18.15 — Describe and sketch, if possible, the following surfaces with given vector function.

(a) $\vec{\boldsymbol{r}}(u,v) = (u+v, 3-v, 1+4u+5v)$

(b) $\vec{\boldsymbol{r}}(u,v) = \left(2\sin(u), 3\cos(u), v\right)$

(c) $\vec{\boldsymbol{r}}(s,t) = \left(s, t, t^2 - s^2\right)$

(d) $\vec{\boldsymbol{r}}(s,t) = \left(s\sin(2t), s^2, s\cos(2t)\right)$

Assignment 18.16 — Find a parameter representation, $\vec{\boldsymbol{r}}(u,v)$, for the following regions.

(a) the plane through $(1,2,-3)$ and parallel with the vectors $(1,1,-1)$ and $(1,-1,1)$

(b) the bottom half of the ellipsoid $2x^2 + 4y^2 + z^2 = 1$

(c) the part of the sphere centered at the origin and of radius 4 that lies between the planes $z = -2$ and $z = 2$

Surface integrals

Assignment 18.17 — Evaluate the surface integral $\iint_{\mathcal{S}} z \, dS$ where $\mathcal{S}$ is the region $z^2 = 1 + x^2 + y^2$ between $z = 1$ and $z = \sqrt{5}$.

Assignment 18.18 — Evaluate the surface integral $\iint_{\mathcal{S}} x \, dS$ where $\mathcal{S}$ is the part of $z = x^2/2$ inside $x^2 + y^2 = 1$ in the first octant.

Assignment 18.19 — Find the area of the regions below.

(a) the part of $x^2 + z^2 = a^2$ that lies inside $y^2 + z^2 = a^2$

(b) the part of $2x + 4y + z = 0$ inside $x^2 + y^2 = 1$

(c) the part of $z = \sqrt{x^2 + y^2}$ that lies above $z = 2$

(d) the part of $x^2 + y^2 + z^2 = a^2$ that lies above the interior of the circle $r = a\cos(\theta)$

Assignment 18.20 — Find the flux of the given vector field through the given surface.

(a) $\vec{\boldsymbol{F}} = (x, y, z)$, $\quad x^2 + y^2 \leq a^2$, with $-h \leq z \leq h$ where $\hat{\boldsymbol{n}}$ is pointing outwards

(b) $\vec{\boldsymbol{F}} = (z, 0, x^2)$, $\quad$ the part of $z = x^2 + y^2$that lies above the square R with $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$ where $\hat{\boldsymbol{n}}$ is pointing outwards

(c) $\vec{\boldsymbol{F}} = (y, 0, z)$, $\quad 0 \leq z \leq 1 - \sqrt{x^2 + y^2}$ where $\hat{\boldsymbol{n}}$ is pointing outwards

(d) $\vec{\boldsymbol{F}} = (x, x, 1)$, $\quad$ the part of $z = x^2 - y^2$ that lies inside $x^2 + y^2 = a^2$ where $\hat{\boldsymbol{n}}$ is pointing outwards

(e) $\vec{\boldsymbol{F}} = (y^3, z^2, x)$, $\quad$ the part of $z = 4 - x^2 - y^2$ that lies above $z = 2x + 1$ where $\hat{\boldsymbol{n}}$ is pointing outwards

(f) $\vec{\boldsymbol{F}} = (x, y, z^2)$, $\quad \vec{\boldsymbol{r}}(u, v) = \big(u\cos(v), u\sin(v), u\big)$ met $0 \leq u \leq 2, 0 \leq v \leq \pi$ where $\hat{\boldsymbol{n}}$ is pointing outwards

The divergence theorem revisited and Stokes' theorem

Assignment 18.21 — Use the divergence theorem (Theorem 18.8) to calculate the flux of the given vector field through the surface $x^2 + y^2 + z^2 = a^2$ $(a > 0)$ where $\hat{\boldsymbol{n}}$ is pointed outwards.

(a) $\vec{\boldsymbol{F}} = (x, -2y, 4z)$

(b) $\vec{\boldsymbol{F}} = \big(x^2 + y^2, y^2 - z^2, z\big)$

(c) $\vec{\boldsymbol{F}} = \big(x^3, 3yz^2, 3y^2 z + x^2\big)$

Assignment 18.22 — In each case, verify the divergence theorem (Theorem 18.8).

(a) $\vec{\boldsymbol{F}} = \big(3x, y^3, -2z^2\big)$, $\quad$ area bounded by $x^2 + y^2 = 9$, $z = 0$ and $z = 5$

(b) $\vec{\boldsymbol{F}} = (x, 2y, 3z)$, $\quad$ volume described by $0 \leq x \leq 1, 0 \leq y \leq x, 0 \leq z \leq x + y$

(c) $\vec{\boldsymbol{F}} = \big(x^3, y^3, z^3\big)$, $\quad$ volume described by $x^2 + y^2 + z^2 \leq 4$

Assignment 18.23 — Verify Stokes' theorem (Theorem 18.9).

(a) $\vec{\boldsymbol{F}} = (xy^2, -x^2 y, xyz)$, $\quad$ with $\mathcal{S}$ the part of $z = 1 - x^2 - y^2$ above the xy-plane where $\hat{\boldsymbol{n}}$ is pointed outwards.

(b) $\vec{\boldsymbol{F}} = (y, z, -x)$, $\quad$ with $\mathcal{S}$ the part of $z = 2x + 5y$ inside $x^2 + y^2 = 1$ where $\hat{\boldsymbol{n}}$ is pointed outwards

Assignment 18.24 — Calculate the requested closed curve line integral and verify Stokes' theorem (Theorem 18.9).

(a) $\displaystyle\oint_C (xy\ dx + yz\ dy + xz\ dz)$

C is the edge of the triangle with vertices $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, oriented clockwise as seen from the point $(1, 1, 1)$

(b) $\displaystyle\oint_C \Big(3y\ dx - 2xz\ dy + \big(x^2 - y^2\big)\ dz\Big)$

C is the cross section of $x^2 + y^2 + z^2 = a^2$ and $z = 0$, counterclockwise oriented viewed from the positive z-axis

(c) $\oint_C (y\ dx + z\ dy + x\ dz)$

C is the cross section of $x^2 + y^2 + z^2 = a^2$ and $x + y + z = 0$, where $\hat{\boldsymbol{n}}$ is orientated inside.

PART IV

APPENDICES

A

Proofing Techniques

In this appendix we gather information to help you read, understand and construct definitions, theorems and proofs.

A.1 Definitions

A definition is a term conceived by humans and used as a shortcut for a complicated idea. For example, we say that an integer is even as a shortcut to say that if we divide this number by two, we get a remainder of zero. With a precise definition we can answer certain questions unambiguously. For example, have you ever wondered if zero was an even number? Now the answer should be clear as we have a precise definition of what we mean by the term even.

A single term can have multiple definitions. For example, they could say that the integer n is even if there is another integer k such that $n = 2k$. We call this an equivalent definition, because it categorizes even numbers in the same way as our first definition. Definitions are like two-way streets — we can use a definition to replace something rather complicated with its definition (if it fits) and we can replace a definition with its more complicated description. A definition is usually written as some form of implication, such as “If something-nice-happens, then party.’ However, this also means that “If party, then something-nice-happens,” even though this may not be formally stated. This is what we mean when we say a definition is a two-way street — it’s actually two implications going in opposite “directions”. Anyone (including you) can come up with a definition, as long as it’s unambiguous, but the real test of a definition’s usefulness is whether or not it’s useful for describing interesting or common situations.

A.2 Theorems

Advanced mathematics is about understanding theorems. Reading them, understanding them, applying them, proving them. Each theorem is a shortcut — we prove something in general and then when we encounter a specific case that falls under the theorem, we can immediately say that we know something else about the situation by applying the theorem. In many cases, this new information can be obtained with much less effort than when we would not have the theorem at our disposal. The first step to arrive at an understanding of a theorem is realising that the statement of any theorem can be rewritten with statements of the form “If something-happens, then something-else-happens.” The “something-happens” part is the **hypothesis** or **assumption** (*hypothese*) and the “something-else-happening” is the **conclusion** or **decision** (*conclusie*). To understand a theorem, it helps to rewrite its statements using this construction. To apply a theorem, we verify that in a certain case “something-happens” and immediately conclude that “something-else-happens”. To prove a theorem, we must argue based on the assumption that the hypothesis is true and via the process of logic obtain that the conclusion must then be true as well.

A.3 Logic

In proofs and theorems we often use statements that are either true (T) or false (F). For example, ‘2 is an odd number’ is a false statement, while ‘2 is an even number’ is a true statement. With such simple statements we can form combined statements that are true or false according to the simple statements of which they are composed and in which way (and versus or).

Let p and q be two statements, then the negation of the statement p ($\neg p$) is true if p is false and vice versa. We can represent this schematically in the following truth table:

p	$\neg p$
T	F
F	T

In an analogous way, we can construct truth tables for the statement ‘p and q’, i.e.

p	q	$p \wedge q$
T	T	T
T	F	T
F	T	F
F	F	F

and for the statement ‘$p \vee q$’:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

From these truth tables it immediately follows that $p \wedge q \Longleftrightarrow (\neg p \vee \neg q)$ and that $p \vee q \Longleftrightarrow (\neg p \wedge \neg q)$.

The statements p and q also allow us to write down implications and equivalences. For example, the implication ‘if p, then q’, ‘p implies q’ or ‘p is sufficient for q’, symbolically represented by $p \Rightarrow q$, is true if p and q are true. In the implication $p \Rightarrow q$ we call p the assumption or hypothesis and q the conclusion. The corresponding truth table is given by

p	q	$p \Rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

Based on these truth tables, we can verify that the statement $p \Rightarrow q$ is logically equivalent to 'not p or q'. The inverse of this implication is $q \Rightarrow p$, but this is not logically equivalent to $p \Rightarrow q$. Thus, we cannot prove the latter implication by proving the inverse implication, nor can we say that $q \Rightarrow p$ is true after we have proven that $p \Rightarrow q$ is true. On the other hand, the contraposition of the implication $p \Rightarrow q$, i.e. the implication $\neg q \Rightarrow \neg p$, is logically equivalent to the given implication. Proving an implication $p \Rightarrow q$ can therefore be done by proving its contraposition.

The truth table for the equivalence 'p if and only if q', 'p equivalent to q' or 'p is needed and sufficient for q' is given by

p	q	$p \Leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

Since the contraposition of $q \Rightarrow p$, being $\neg p \Rightarrow \neg q$, is equivalent to $q \Rightarrow p$, we can prove the equivalence $p \Leftrightarrow q$ by proving that $p \Rightarrow q$ and $\neg p \Rightarrow \neg q$ hold.

A.4 Methods of proof

"I don't know how to start!" is often the lament of the novice proof builder. Here are a few pieces of advice.

1. As mentioned above, you rewrite the statement of the theorem in an "if-then" form. This simplifies identifying the hypothesis and conclusion, referred to in the following paragraphs.
2. Ask yourself what kind of statement you are trying to prove. This is always a part of your conclusion. Are you being asked to conclude that two numbers are equal, that a function is differentiable, or that a set is a subset of another? You cannot apply other techniques if you do not know what type of conclusion you have.
3. Write down reformulations of your hypotheses. Interpret and translate each definition correctly.
4. Write down your hypothesis at the top of a piece of paper and your conclusion at the bottom. See if you can formulate a statement that precedes and implies the conclusion. Work downwards from your hypothesis and upwards from your conclusion, and see if you can equate them in the middle. When you're done, neatly rewrite the proof, from hypothesis to conclusion, with verifiable implications with each subsequent statement.
5. As you work through your proof, think about the types of objects your symbols represent. For example, suppose A is a set and $f(x)$ is a real-valued function. Then the expression $A+f$ might not make sense if we have not defined what it means to 'add a set' to a function, so that we can stop at that point and adjust it accordingly. On the other hand, we can understand $2f$ as the function whose rule is described by $(2f)(x) = 2f(x)$."Think about your objects" means always checking that your objects and operations are compatible.

A.4.1 Proof by construction

Conclusions of proofs come in different types. Often a theorem will simply claim that something exists. The best, but not the only way to show that something exists is to actually construct it. Such a proof is called **constructive** (*opbouwend*). What you need to realize about constructive proofs is that the proof itself will contain a procedure that can be used computationally to construct the desired object. If the procedure is not too cumbersome, the proof itself is just as useful as the statement of the theorem.

A.4.2 Equivalences

When a statement uses the expression "if and only if" (or the abbreviation "iff"), it is a shorter way of saying that two if-then statements are true. So if a theorem says "P if and only if Q," then it is true that "if P, then Q" and it is also true that "if Q, then P". Statements such as "I wear bright yellow knee-high rubber boots if and only if it rains." This means that I never forget to wear my yellow boots when it rains and I am never seen in such crazy boots unless it rains. You never have one without the other. I have my boots on and it is raining or I am not wearing my boots and it is dry. The result of proving such statements is like a 2-for-1 sale, we get to do two proofs. Suppose P and conclude Q, then start again and assume Q and conclude P. For this reason, "if and only if" is sometimes abbreviated with $\iff$. Proofs indicate which implication is proved by prefixing each with $\Rightarrow$ (sufficient condition) or $\Leftarrow$ (necessary condition). A carefully written proof will remind the reader which statement is being used as the hypothesis, a faster version will let the reader infer it from the direction of the arrow. Tradition dictates that we do the "easy" half first, but that's difficult for a student to know until you're done with both halves! If you rewrite your proofs (a good habit), you can choose to put the easy half first. These types of theorems are called **equivalences** or **characterizations** (*equivalenties* of *gelijkwaardigheden*) and are one of the most pleasing results in mathematics. They say that two objects, or two situations, really are the same. You do not have one without the other. The more different P and Q seem to be, the more pleasant it is to discover that they are really equivalent. And if P describes a very mysterious solution or involves a difficult calculation, while Q is transparent or involves simple calculations, then we have found a great shortcut for a better understanding or faster calculation. Remember that every statement is basically a shortcut in one form or another. You will also find that if proving $P \Rightarrow Q$ is very easy, then proving $Q \Rightarrow P$ is probably proportionally more difficult. Sometimes the two halves are about equally difficult. And on rare occasions, you can string together a whole host of other equivalencies to form the one you want and not even have to do two halves. In this case, one half's argument is just the other half's argument, but reversed. One last remark about equivalences. If you see a statement of a theorem that says that two things are "equivalent", you should first translate it into an "if and only if" statement.

A.4.3 Negation

When we construct the contrapositive of a theorem, we must negate the two statements in the implication. And if we construct a proof by **contradiction** (*contradictie* of *tegenstelling*) we have to negate the conclusion of the theorem. One way to proceed is to simultaneously negate the hypothesis and conclusion of an implication (but remember that this is not guaranteed to be a true statement). We thus often need to negate statements and in some situations this can be difficult.

If a statement says that a set is empty, then its negation is the statement that the set is not empty. That is obvious. Suppose a statement says "something-happens" for all i, or every i, or some random i. Then the negation is that "something-doesn't-happening" for at least one value of i. If a statement says that at least one "thing" exist, then the negation is the statement that there is no "thing". If a statement says that a "thing" is unique, then the negation is that there are zero or more than one of the "thing.' We will not cover all possibilities, but we would like to point out that logical quantifiers such as "there exists" or "for every" should be treated with caution when negating statements. Studying proofs that use contradiction is a good first step in understanding the range of possibilities.

A.4.4 Contrapositive

The **contrapositive** (*contrapositief*) of an implication $P \Rightarrow Q$ is the implication $\text{not}(Q) \Rightarrow \text{not}(P)$, where "not " means the logical negation. An implication is true if and only if its contrapositive is true. In symbols, $(P \Rightarrow Q) \iff (\text{not}(Q) \Rightarrow \text{not}(P))$ is a statement. Such statements about logic, which are always true, are called **tautologies** (*tautologieën*).

For example, it's a statement like "If a vehicle is a fire truck, it has big tires and a siren". The contrapositive is "if a vehicle doesn't have big tires or it doesn't have a siren, then it's not a fire truck". Notice how the "and" became an "or" when we negated the conclusion of the original theorem. It will often happen that it is easier to construct a proof of the contrapositive than of the original implication. If you have trouble formulating a proof of an implication, see if the contrapositive is easier to prove. The trick is to accurately construct the negation of complex statements.

A.4.5 Converse

The **converse** (*omgekeerde*) of the implication $P \Rightarrow Q$ is the implication $Q \Rightarrow P$. There is no guarantee that the truth of these two statements is related. In particular, if an implication is proven to be a theorem, do not try to use its converse as if it were a theorem as well. Sometimes the converse is true (and we have an equivalence). But more likely the converse is false, especially if it was not included in the statement of the original theorem. For example, we have the statement "if a vehicle is a fire engine, it has big tires and a siren." The converse is false. The statement "if a vehicle has big tires and a siren, it's a fire truck" is false. A police vehicle for use on a public beach could also have large tires and a siren, but it is not equipped to fight fires.

A.4.6 Contradiction

Another method of proof is known as **proof by contradiction** (*bewijs d.m.v. contradictie*) and can be a powerful (and satisfying) method. Simply put, suppose you want to prove the implication "if A, then B". As usual, we assume A to be true, but we also assume B to be false. If our original implication is true, then these two assumptions should lead us to a logical contradiction. In practice, we assume that the negation of B is true (see proof technique contrapositive). So we argue, based on the assumptions A and $\text{not}(B)$, and we look for an incorrect conclusion such as $1 = 6$ or a set that is simultaneously empty and non-empty or a matrix that is both non-singular and singular. You have to be careful with formulating proofs that look like proofs by contradiction, but in reality are not. This happens when you assume A and $\text{not}(B)$ and go on to give a "normal" and direct proof that B is true by only using the assumption that A is true. Your last step then is to claim that B is true and then call upon the assumption that $\text{not}(B)$ is true, obtaining the desired contradiction. Instead, you could have avoided the overhead of a proof by contradiction and worked with the direct proof. Here is a simple example of a proof by contradiction. There are direct proofs that are just as easy, but this will prove the point.

Statement: If a and b are odd integers, then their product, ab, is odd.

Proof: To begin a proof by contradiction, assume the hypothesis that a and b are odd. Assume also the negation of the conclusion, in this case being that ab is even. Then there are integers j, k, ℓ such that $a = 2j+1$, $b = 2k+1$, $ab = 2\ell$. Then

$$\begin{aligned} 0 &= ab - ab \\ &= (2j+1)(2k+1) - (2\ell) \\ &= 4jk + 2j + 2k - 2\ell + 1 \end{aligned}$$

$$= 2\,(2jk + j + k - \ell) + 1\,.$$

Note that we used both our hypothesis and the negation of the conclusion in the second line. Now divide the integer on each side of this set of equalities by 2. At the left hand side we will get a remainder of 0, while at the right hand side the remainder will be 1. Both remainders cannot be equal, so this is our desired contradiction. The conclusion (that ab is odd) is thus true.

A.4.7 Uniqueness

A theorem will sometimes claim that an object, with a certain desirable property, is unique. In other words, there is only one such object. To prove this, a standard technique is to assume that there are two such objects and then analyze the consequences. The end result can be a contradiction or the conclusion that the two supposedly different objects are really equal.

A.4.8 Proving identities

Many theorems have conclusions that say that two objects are equal. Perhaps one object is difficult to calculate or understand while the other is easy to calculate or understand. This would make an interesting theorem. Whether the result is interesting or not, we follow the same method to formulate a proof. Sometimes we have to use specialized notions of equality, but in other cases we can string together a list of equalities. The wrong way to prove an identity is to start by writing it down and then search until it becomes an obvious identity. The first mistake you make is writing the statement you want to prove down as if you already believe it to be true. But more dangerous is the possibility that some of your operations are not reversible. Here's an example. Let's prove that $3 = -3$.

$$\begin{aligned}
3 &= -3 && \text{(This is a bad start.)}\\
3^2 &= (-3)^2 && \text{(Square both sides.)}\\
9 &= 9 && \\
0 &= 0 && \text{(Subtract 9 from both sides.)}
\end{aligned}$$

So since $0 = 0$ is a true statement, it follows that $3 = -3$ is a true statement? No. Of course we don't really expect a valid proof for $3 = -3$, but this attempt should illustrate the dangers of this (incorrect) method. What you have just seen are proofs of the following form. To prove that $A = D$ we write

$$\begin{aligned}
A &= B && \text{(Theorem, Definition or Hypothesis that justifies } A = B.\text{)}\\
&= C && \text{(Theorem, Definition or Hypothesis that justifies } B = C.\text{)}\\
&= D && \text{(Theorem, Definition or Hypothesis that justifies } C = D.\text{)}
\end{aligned}$$

In your drafts, where you explore possible methods for proving a theorem, you can derive a variety of expressions by sometimes making connections with different bits and pieces, while sometimes leaving some parts out. Once you see a direct line, rewrite your proof and mimic this style carefully.

A.4.9 Induction

Induction or **mathematical induction** (*wiskundige inductie*) is a method for proving statements that are indexed by integers. In other words, suppose you have a statement to prove that actually contains multiple statements, one for $n = 1$, another for $n = 2$, a third for $n = 3$, etc. If there is enough similarity between the statements, you can use a script to prove them all at once. Consider for example the statement

Statement $1+2+3+\cdots+n = \frac{n(n+1)}{2}$ for $n \geq 1$.

This is an abbreviation for the statements $1 = \frac{1(1+1)}{2}$, $1+2 = \frac{2(2+1)}{2}$, $1+2+3 = \frac{3(3+1)}{2}$, $1+2+3+4 = \frac{4(4+1)}{2}$, and so on. You can do the calculations in each of these statements and check if all four are true. We may not be surprised that the fifth statement is also true (go ahead and check). However, do we think that the statement is also true for $n = 872$? Or $n = 1,234,529$?

To see that these questions are not so ridiculous, take a look at the following example. The statement "$n^2 - n + 41$ is a prime number" is true for integers $1 \leq n \leq 40$ (verify a few). However, if we check $n = 41$, we find $41^2 - 41 + 41 = 41^2$, which is not a prime number. So how do we prove infinitely many statements at once? More formally, let's denote our statements as $P(n)$. If we can then prove the following two claims

1. $P(1)$ is true.
2. If $P(k)$ is true, then $P(k+1)$ is true.

It then follows that $P(n)$ is true for all $n \geq 1$. To understand this, consider the process of climbing an infinitely long ladder with equally spaced steps. Faced with such a ladder, suppose I tell you that you are able to step on the first step, and if you are on a certain step then you are able to go to the next step. It follows that you can climb the ladder as far as you want. The first formal statement above is related to entering the first step and the second formal statement is related to the assumption that if you stand on one step, you can always reach the next step. In practice, determining that $P(1)$ is true is called the "base case" and is simple in most cases. Establishing that $P(k) \Rightarrow P(k+1)$ is referred to as the "induction step," or in this course (and elsewhere) we will generally refer to the assumption of $P(k)$ as the "induction hypothesis". This is perhaps the most mysterious part of a proof by induction, because it seems like you are assuming ($P(k)$), which you are trying to prove ($P(n)$). Sometimes it's even worse because, as you become more familiar with induction, we often do not bother using a different letter (k) for the index (n) in the induction step. Note that the second formal statement never says that $P(k)$ is true. It just says what logically could follow if $P(k)$ is true. We can make statements like "If I lived on the moon, I could jump over a 12 meter high beam". This may be a true statement, but it does not say we live on the moon, and we indeed may never live there. Enough generalities. Let's work out an example and prove the above statement about sums of integers. Formally, our statement is $P(n):\ 1+2+3+\cdots+n = \frac{n(n+1)}{2}$.

Proof: Base case. $P(1)$ is the statement $1 = \frac{1(1+1)}{2}$, which simplifies to the true statement $1 = 1$.

Induction step: We assume that $P(k)$ is true and will try to prove $P(k+1)$. Given what we want to achieve, it is natural to start by examining the sum of the first $k+1$ integers.

$$\begin{aligned}
1+2+3+\cdots+(k+1) &= (1+2+3+\cdots+k)+(k+1) \\
&= \frac{k(k+1)}{2} + (k+1) && \text{(Induction hypothesis.)} \\
&= \frac{k^2+k}{2} + \frac{2k+2}{2} \\
&= \frac{k^2+3k+2}{2} \\
&= \frac{(k+1)(k+2)}{2} \\
&= \frac{(k+1)((k+1)+1)}{2}
\end{aligned}$$

We then recognize the two ends of this chain of equalities as $P(k+1)$. So, by mathematical induction, the theorem is true for all n.

How do you recognize when to use induction? The first clue is a statement that consists of many statements, one for each integer. The second clue would be that you start a more standard proof and find yourself using a lot of words like “and so on” (as above) or a lot of dots to establish patterns that you believe will last forever. However, there are many minor cases where induction may be justified, but we make no effort to use this proof technique.

Induction is important enough and used often enough that it comes in different variations. The base case sometimes starts with $n = 0$ or possibly with an integer greater than 1. Some formulate the induction step as $P(k-1) \Rightarrow P(k)$. There is also a “strong form” of induction where we assume $P(1)$, $P(2)$, $P(3)$, $\ldots P(k)$ all as hypotheses to prove the conclusion $P(k+1)$.

B

Calculus in Mathematica and Wolfram|Alpha

The methods given in this course can be used to solve mathematical problems with pen and paper, but in practice this only works for relatively simple problems that require little computation, such as those covered in the board exercise sessions. Today, however, we usually leave the repetitive work to computers, which are specially designed to perform a huge number of calculations in a very short time.

Here we use Wolfram Mathematica (or Mathematica for short) and Wolfram|Alpha. Mathematica is a so-called computer algebra system capable of performing symbolic mathematical calculations on the computer. Wolfram|Alpha is freely accessible online at the following url: `https://www.wolframalpha.com/` and uses the same syntax as Mathematica, but is limited in available computation time. You can use both Mathematica and and Wolfram|Alpha to solve complex problems, but also to support and check your calculations when solving problems with pen and paper. First we give an introduction to using Mathematica, then we discuss in detail how Mathematica can be used to (help) solve calculus problems. For more extensive documentation, we refer you to the Mathematica Documentation Centre (you can find it under Help > Documentation).

B.1 Mathematica Notebooks

Mathematica uses so-called Notebooks, denoted by the .nb extension, an interactive document containing both formatted text and code. This document is structured in what are known as **cells**, indicated by straight brackets on the right side of the notebook. To create a new cell, move your cursor below/above one or between two existing cell(s) and start typing. Each cell has a particular style, which defines its properties and formatting. This style can be changed by right-clicking on the cell and then under Styles selecting the desired style. We give a brief overview of the most commonly used cell styles.

B.1.1 Input cell

When we create a new cell in the Mathematica notebook, the default cell style is **Input**. In these cells we can enter mathematical operations, which can be performed with Shift + Enter . If the computation time gets too high (which often indicates a bug in the code), the evaluation can be aborted by pressing keystrokeAlt + keystroke. or in the menu bar via Evaluation > Abort Evaluation. The result of the evaluated Input cell is written out to a linked Output cell, which appears below it.

```
In[31]:= 1+1

Out[31]= 2
```

B.1.2 Text and layout

Text cells contain formatted text and are used to provide explanations for the notebook. In addition, there are numerous cell types that help structure the notebook. Cell types such as **Title**, **Chapter**, **Section**. . . are arranged according to a clear hierarchy, where cell types higher up the ladder (e.g. Title), include all subordinate cell types (e.g. Section, Text and Input). This group of cells is indicated by a straight bracket on the right side of the document and can be closed with a double-click so that only the cell with the highest rank is shown.

B.2 Mathematica for dummies

This section goes over the basics of the Mathematica language. Feel free to modify the following examples or experiment on your own! If something is unclear, you can consult the Documentation Centre via the Help menu in the taskbar at the top of the document or via a right-click at the level of a specific command.

B.2.1 Operations, evaluations and lists

B.2.1.1 Mathematical and realtional operations

$+$, $-$, $*$, /, $\wedge$	Basic mathematical operators: sum, difference, multiplication, division, exponentiation
$==, !=, >, >=, <, =<$	relational operators: (un)equality, bigger than (or equal to), less than (or equal to)
()	brackets to group operations

Below we illustrate the use of some of these operators.

```
In[32]:= (10-5)*2^3

Out[32]= 40

In[33]:= 1 < 2 ≤ 3

Out[33]= True
```

```
In[34]:= 5/2 == 10/3

Out[34]= False
```

B.2.1.2 Evaluation of expressions

expr	the exact result is shown in an Output cell below after executing the expression *expr*
expr//N or N[*expr*]	the numerical (approximate) result is given after executing the expression expr
N[*expr*,*n*]	the numerical (approximate) result is given after executing the expression expr; With a precision of *n* significant numbers in the Output cell
expr;	the result is not displayed after executing the expression *expr*

B.2.1.3 Assignements

x = *value*	direct assignement (*x* will from now on always be replaced by *value*)
x=*y*=*value*	Assume *x* and *y* both equal to *value*
x=. or Clear[x]	deletes the value assigned to *x*
expr /. *x* -> *value*	replace all *x*'s in the expression *expr* by value

It should be noted that *value* need not be a scalar value, but can equally be a symbolic expression. In the last expression, the order to assign is indicated by /. and *x* -> *value* is the rule that determines what should be replaced (the so-called. **replacement rule**).

We illustrate all this in the example below.

Example B.1

Save the expression $xy^2 - 2x(1+y^{-1})$ in the variable `expr`:

```
In[35]:= expr =  x*y^2-2*x(1+1/y)
```

Out[35]= $-2x\left(1+\frac{1}{y}\right) + xy^2$

Exact and numerical evaluation of `expr`, voor $x = y = 1/3$, by direct assignment:

```
In[36]:= x = y = 1/3;
         expr

          71
Out[36]= - ──
          27

In[37]:= expr // N

Out[37]= -2.62963

In[38]:= N[expr,2]

Out[38]= -2.6
```

Evaluation of `expr`, for $x = 3, y = 2$ and $x = uv, y = u^2$, by exchange:

```
In[39]:= Clear[x, y];
         expr/.{x→3, y→2}
         expr/.{x→u*v, y→u^2}
         expr

Out[39]= 3

                  1
Out[40]= -2( 1+ ── )uv + u^5 v
                 u^2

                 1
Out[41]= -2x( 1+ ─ ) + xy^2
                 y
```

B.2.1.4 Lijsten

Often we want to work with several objects (values, functions ...) at the same time. This can be done by using lists:

List[e1, e2, ...] of {e1, e2, . . . }   ordered sequence of elements e1, e2 . . .

Note that these elements can also be Lists themselves. Lists of lists are called **nested lists**. Lists can be used to pass multiple arguments to functions (see below), but can also be considered vectors.We can select elements from a list. For example, consider the vector v, then

$v[[k]]$ element k in v

Example B.2

Create the vector $\vec{\mathbf{v}}_1 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$ and select the first element.

```
In[42]:= v1 = {1, 2, 3};
         First[v1]

Out[42]= 1
```

B.2.2 Functions

Functions with arguments x, y, etc. are called as follows:

f[x,y,. . .]

and entered:

f[x_,y_,. . .] := *expr*,

where, *expr* is an expression with the variables x, y . . . .

Implicit functions can also be defined, but this requires specifying the dependent variables in *expr* (see example).

Piecewise functions are implemented as follows:

f [*x*_, *y*_, ...] := Piecewise[*expr1*, *cond1*, *expr2*, *cond2*, ..., Indeterminate]

where *expr1, expr2,. . .* are expressions with variables *x, y,. . .* and *cond1, cond2,. . .* the conditions they must satisfy. Indeterminate indicates that in the regions where the variables do not satisfy any of the conditions, f is undefined.

In addition to self-defined functions, Mathematica has numerous built-in functions:

Log[x]	$\ln(x)$
Log[x,b]	$\log_b(x)$
Exp[x]	e^x
Sin[x], Cos[x], Sec[x], Csc[x],Tan[x], Cot[x]	trigonometric functions
ArcSin[x], ArcCos[x], ArcSec[x], ArcCsc[x]	inverse trigonometric functions
ArcTan[x], ArcCot[x]	
Sinh[x], Cosh[x], Sech[x], Csch[x], Tanh[x], Coth[x]	hyperbolic functions
ArcSinh[x], ArcCosh[x], ArcSech[x], ArcCsch[x]	inverse hyperbolic functions
ArcTanh[x], ArcCoth[x]	

Note that Mathematica can work both symbolically and numerically. It always works with exact values, unless it is explicitly stated that it should work numerically. The latter can be done with `//N` or `N[expr]`, or by using a decimal.

Finally, we note that there are numerous other functions available in Mathematica, such as the previously cited **`Piecewise`** or the function `Manipulate`, with which we can generate interactive ouput.

Manipulate[*expr*, $\{k, k_0, k_1\}$]	gives an interactive result of expr in which we can adjust the value of k through a controller.

The values we can get k to assume lie between k_0 and k_1.

Example B.3

Calculate $e^{\ln(x)} = x$:

```
In[43]:=  Exp[Log[x]]

Out[43]=  x
```

Exact and numerical evaluation of $f(x) = \sin(x)^2/5$ for $x = 5$:

```
In[44]:= f[x_] := 2/10*Sin[x]^2
         f[5]
         f[5] // N

           Sin[5]^2
Out[44]= --------
              5

Out[45]= 0.183907
```

Once more the function: $f(x) = \sin(x)^2/5$, but numerically defined:

```
In[46]:= f[x_] := 0.2*Sin[x]^2
         f[5]

Out[46]= 0.183907
```

Implicitly defined function $\sin(y) + y^3 = 6 - x^3$:

```
In[47]:= fImpl[x_] := Sin[y[x]] + y[x]^3 == 6 -x^3
```

Remark that x is the only independent variable.

```
In[48]:= fImpl[x]
         fImpl[2]

Out[48]= Sin[y[2]] + y[x]^3 == 6-x^3

         Sin[y[2]] + y[2]^3 == 6-2^3
```

A piecewise defined function:

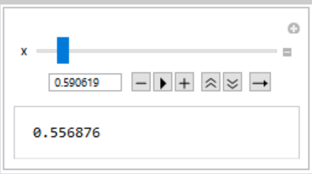
```
In[49]:= fpw[x_] := Piecewise[{{x+1, x<0},{-x^2+1, x>0}},Indeterminate]
         fpw[-.5]
         fpw[0]
         fpw[.5]

Out[49]= 0.5
         Indeterminate
         0.75
```

Interactive solution of Sin[x] for $x \in [0, 2\pi]$:

```
In[50]:= Manipulate[Sin[x], {x, 0, 2*Pi}]
```

Out[50]=



B.2.3 Solving (un)equality's

Solve[*expr*, *vars*]	tries to find solutions of the equation, inequality, or system of equations/inequalities in *expr* as a function of the variable *vars*
NSolve[*expr*, *vars*]	tries to find a numerical approximation of the solutions of the equation, inequality, or system of equations/inequalities in *expr* to the variables in *vars*
Simplify[expr]	ries to find a simplified form of the expression in expr
Apart[expr]	splits the expression in expr into partial fractions

The results of **`Solve`** and **`NSolve`** are given as lists of replacement rules.

Example B.4

Find the exact and approximate values of the zero points of $x^2 - 2$:

```
In[51]:= ZerosExact = Solve[x^2 - 2 == 0, x]
         ZerosNumeriek = NSolve[x^2 - 2 == 0, x]

Out[51]= {{x→-√2},{x→√2}}

         {{x→-1.41421,{x→1.41421}}
```

Retrieve the replacement rule of the first zero point from the list of solutions :

```
In[52]:= nulpuntenExact[[1, 1]]

Out[52]= x→-√2
```

The output of an unsolvable equation or system is an empty list.

```
In[53]:= Solve[{x^2-2 == 0, 2x-5 == 0}, x]

Out[53]= {}
```

Split the fraction $\frac{x^2-2}{x+4}$ in partial fractions:

```
In[54]:= Apart[(x^2 - 2)/(x + 4)]

                     14
Out[54]= -4 + x + -----
                   4 + x
```

B.2.4 Visualisatie

Finally, we go over some of the functions that serve to create plots.

Plot[$f[x]$,{x, a, b}]	Create a plot of the function f over the interval $[a, b]$
Plot[{$f_1[x]$,$f_2[x]$,...},{x, a, b}]	Creat a plot of the functions f_1 , f_2, ... over the interval $[a, b]$.
Plot3D[$g[x, y]$, {x, a, b}, {y, c, d}]	Create a 3D plot of g over the range $[a, b] \times [c, d]$.
ListPlot[{{x_1, y_1},{x_2, y_2},...}]	Create a plot for the points (x_i, y_i).
ContourPlot[$g[x, y]$, {x, a, b}, {y, c, d}]	plot the contour(s) of g over the range $[a, b] \times [c, d]$ (g can also be implicitly defined)
RegionPlot[*cond*, {x, a, b}, {y, c, d}]	plot the subarea of $[a, b] \times [c, d]$ where the conditions in *cond* are met
ParametricPlot[{$k[u], l[u]$}, {u, u_{min}, u_{max}}]	plot the parametric equations $x = k(u)$ and $y = l(u)$ for $u \in [u_{min}, u_{max}]$

There are numerous options for customizing plot formatting. Below we give an overview of the most important ones:

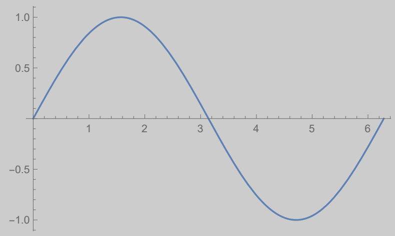
PlotLabel -> "title"	gives a title to the plot
AxesStyle -> Arrowheads[*s*]	places arrows on the axes of the plot (pointing in an increasing sense). *s* is a number that determines the size of the arrows.
AxesLabel -> {"x","y"}	labels the axes
PlotRange -> {*ymin, ymax*}	specifies the y-range of the plot
ImageSize -> grootte	specifies the size of the plot
PlotStyles -> {*stijl1*, *stijl2*, ...}	plots the first function in *stijl1*, the second in *stijl2* ...
PlotLegend -> {*name1*, *name2*, ...}	creates a legend with the names of the plotted function

Example B.5

Plot $\sin(x)$ the default layout:

```
In[55]:= Plot[Sin[x],{x,0,2*Pi}]
```

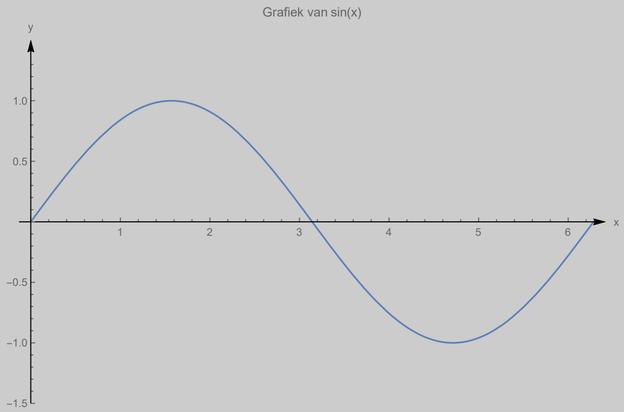
Out[55]=



Plot $\sin(x)$ with custom layout:

```
In[56]:= Plot[Sin[x],{x,0,2*Pi},
         PlotRange → {-1.5, 1.5},
         PlotLabel → "Grafiek van sin(x)",
         AxesStyle → Arrowheads[0.02],
         AxesLabel → "x", "y",
         ImageSize → Large ]
```

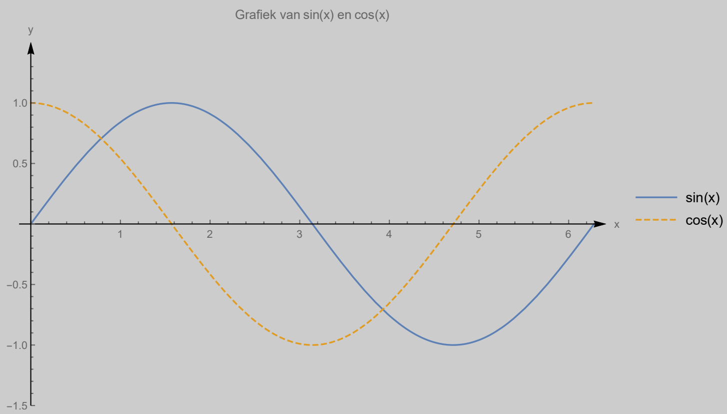
Out[56]=



Plot sin(x) en $cos(x)$ with custom layout, where $cos(x)$ is drawn as a dotted line:

```
In[57]:= Plot[{Sin[x], Cos[x]}, {x, 0, 2*Pi},
         PlotRange → {-1.5, 1.5},
         PlotLabel → "Graph of sin(x) and cos(x)",
         AxesStyle → Arrowheads[0.02],
         AxesLabel → {"x", "y"},
         ImageSize → Large,
         PlotStyle → {Line, Dashed},
         PlotLegends → {"sin(x)", "cos(x)"}]
```

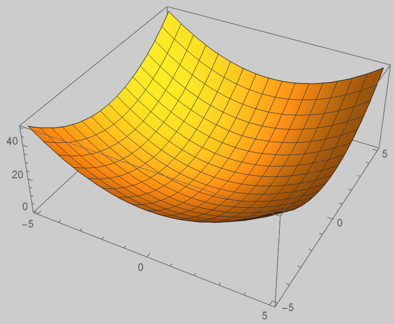
Out[57]=



Consider the function $g(x, y) = x^2 + y^2$ over $[-5, 5] \times [-5, 5]$. Create a 3D and contour plot:

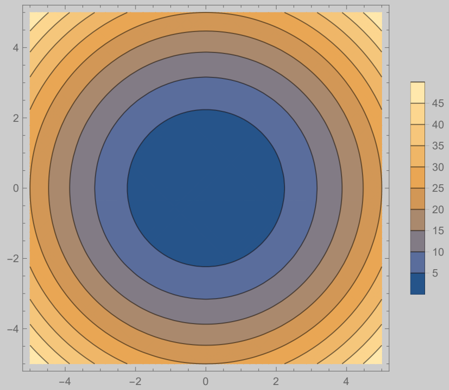
```
In[58]:= Plot3D[x^2 + y^2, {x, -5, 5}, {y, -5, 5}]
```

Out[58]=



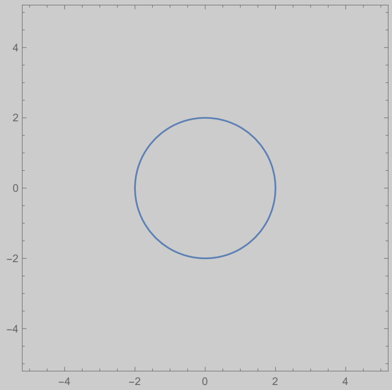

```
In[59]:= ContourPlot[x^2 + y^2, {x, -5, 5}, {y, -5, 5}]
```

Out[59]=

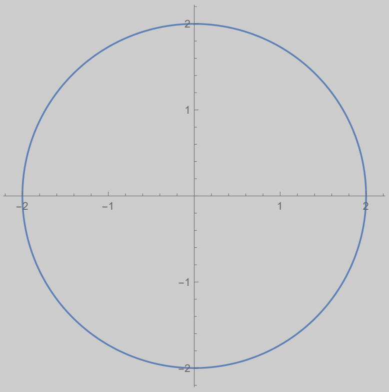


Plot the circle of radius 2, using the implicit equation and the parametric equation:

In[60]:= `ContourPlot[x^2 + y^2 == 4, {x, -5, 5}, {y, -5, 5}]`

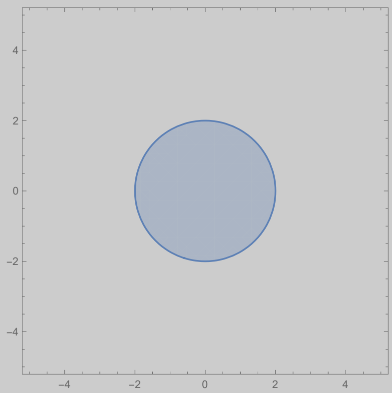
Out[60]=


In[61]:= `ParametricPlot[{2*Cos[u],2*Sin[u]}, {u, 0, 2*Pi}]`

Out[61]=


Plot the subarea of $[-5,5] \times [-5,5]$, delimited by the circle of radius 2:

In[62]:= `RegionPlot[x^2 + y^2 ≤ 4, {x, -5, 5}, {y, -5, 5}]`

Out[62]=


B.3 Calculus specific intstructions

B.3.1 Limits

In Mathematica we can calculate limits with the function **`Limit`**.

Limit[$f[x]$, x -> c]	determine the limit of f in c
Limit[$f[x]$, x -> c, Direction -> "FromAbove"]	determine the right limit of f in c
Limit[$f[x]$, x -> c, Direction -> "FromBelow"]	determine the right limit of f in c

Example B.6

Determine the total, right and left limit of $1/x$ in 0:

```
In[63]:= Limit[ 1/x, x → 0]

Out[63]= Indeterminate

In[64]:= Limit[ 1/x, x → 0, Direction → "FromAbove"]

Out[64]= ∞

In[65]:= Limit[ 1/x, x → 0, Direction → "FromBelow"]

Out[65]= -∞
```

Determine the limit of $1/x$ for $x \to +\infty$:

```
In[66]:= Limit[ 1/x, x → Infinity]

Out[66]= 0
```

B.3.2 Derivatives

Derivatives of functions with one variable can be calculated using **`Derivative`**:

Derivative$[n][f][x]$	n-th order derivative of the function $f(x)$ towards x

Usually we will be using the short notation of **`Derivative`**:

$f'[x]$	first order derivative of $f(x)$
$f''[x]$	second order derivative of $f(x)$
D$[expr, x]$	first order partial derivative of *expr* to x
D$[expr, \{x, n\}]$	n-th order partial derivative of *expr* to x

The difference between these notations is subtle and often they can be used interchangeably. However, we recommend to use $f'[x]$ when possible. Remark that $f(x)$ can be implicitly defined as well.

Example B.7

Determine the first and second derivatives of $x^2 + x$:

```
In[67]:= f[x_] := x^2 + x;
         f'[x]
         f''[x]

Out[67]= 1 + 2x
         2
```

We can keep the derivative as a function:

```
In[68]:= df[x_] := f'[x]
         df[x]

Out[68]= 1 + 2 x
```

Determine the derivative of the implicit function $y^3 + y\sin = 6 - x^3$:

```
In[69]:= fImpl[x_] := Sin[y[x]] + y[x]^3 == 6 - x^3
         fImpl'[x]

Out[69]= Cos[y[x]] y'[x] + 3 y[x]^2 y'[x] == -3 x^2
```

Try to rewrite the outcome in explicit form:

```
In[70]:= Solve[fImpl'[x], y'[x]]

Out[70]= {{y'[x]→-3x^2/(Cos[y[x]]+3y[x]^2)}}
```

With **Derivative** we can also determine derivatives of functions of multiple variables:

Derivative$[n_1, n_2, \ldots][f][x_1, x_2, \ldots]$	derivative of the function $f(x_1, x_2, \ldots)$ that is derived n_i times with respect to the variable x_i.

The short notation $f'[x_1, x_2, \ldots]$ can no longer be used for functions of multiple variables since it does not indicate to which variable(s) it should be derived. **D** can be used:

D$[expr, \{x_1, n_1\}, \{x_2, n_2\}, \ldots]$	derivative of *expr*, that is n_i derived with respect to the variable x_i,

where *expr* may again be explicitly or implicitly defined. For functions with multiple variables, we can also compute the gradient:

Grad$[f, \{x_1, x_2, \ldots\}]$	gradient of f

Example B.8

Determine $f_x(x,y)$ and $f_{xy}(x,y)$ if $f(x,y) = x^2 + xy + y^2$.

```
In[71]:= D[x^2 + x y + y^2, {x, 1}]

Out[71]= 2x + y

In[72]:= D[x^2 + x y + y^2, {x, 1},{y, 1}]

Out[72]= 1
```

Determine the gradient of $f(x,y) = x^2 + xy + y^2$ in the point $(1,2)$.

```
In[73]:= gGrad = Grad[x^2 + x y + y^2, {x, y}]

Out[73]= {2x+y, x+2y}

In[74]:= gGrad /. {x→1, y→2}

Out[74]= {4, 5}
```

B.3.3 Integrals

Mathematica has two functions for integrating functions, being **Integrate** and **NIntegrate**. The former calculates an integral analytically, while the latter provides a numerical approximation.

$\text{Integrate}[f, x]$	determines the indefinite integral $\int f(x)dx$,
$\text{Integrate}[f, \{x, x_{min}, x_{max}\}]$	determines the definite integral $s \int_{x_{min}}^{x_{max}} f(x)dx$,
$\text{NIntegrate}[f, \{x, x_{min}, x_{max}\}]$	determines the numerical approximation of $\int_{x_{min}}^{x_{max}} f(x)dx$,

Example B.9

Determine the following integral

$$\int \left(4x - x^2\right) dx.$$

```
In[75]:= Integrate[4x-x^2,x]

Out[75]= 2x^2 - x^3/3
```

Determine

$$\int_0^{-\infty} e^x dx.$$

```
In[76]:= Integrate[Exp[x],{x,0,-Infinity}]

Out[76]= -1
```

Determine the numerical value of the definite integral

$$\int_0^1 \frac{\sin(x)}{x} dx.$$

```
In[77]:= NIntegrate[Sin[x]/x,{x,0,1}]

Out[77]= 0.946083
```

B.3.4 Series

We use following Mathematica functions to determine the Taylor series expansion of a function:

Series[f, $\{x, x_0, n\}$]	gives the taylor series expansion of $f(x)$ around x_0 with terms up to and including the n-the order (+ the error term of the $n+1$th order)
Normal[s]	Returns round the Taylor development s to the nth order

Example B.10

Determine the Taylor series expansion of $\ln(x)$ in the area of $x = 1$ up to the 2—the order term.

```
In[78]:= s = Series[Log[x], {x, 1, 2}]

Out[78]= (x-1) - 1/2 (x-1)^2 +O[x-1]^3

In[79]:= Normal[s]

Out[79]= -1 - 1/2 (-1+x)^2 +x
```

C

Python Tutorial

This tutorial is available as a Jupyter notebook (extension .ipynb). Note that this tutorial is made for **Python 3.x** and is not compatible with older Python versions. Make sure you are working with an appropriate version!

C.1 What are Python and Jupyter Notebooks?

Python is a programming language used to write computer programs. These programs consist of lines of code that can be interpreted and executed by a computer one by one. Like all (programming) languages, Python has its own vocabulary and grammar. This is called the **syntax**, i.e. the rules that the code must follow in order for computers to be able to read it.

Here we use Python in a **Jupyter Notebook**, available through a web browser (Jupyter.org). Jupyter Notebooks are structured in so-called cells. We distinguish two types of cells:

1. **Markdown cells** contain text with titles, explanations, assignments,. . . . This text is entered as code, which is converted to formatted text when executed.
2. **Input cells** contain Python code that can be executed. In front of such cells you see `In [x]:`, where x keeps track of the number of cells executed. When the cell is executed, the output(s) appear below the input cell. These are then preceded by `Out [x]:`.

Two modes are possible for a cell:

1. **Command Mode** (left bar is blue): in this mode, operations can be done on the entire cell (e.g. changing cell type, cutting and pasting cells, inserting cells,...).
2. **Edit Mode** (left bar is green): in this mode, the text/code in the cell can be edited.

Press Enter to switch to edit mode and press Esc to switch back to command mode. Use Ctrl + Enter to enter a cell, and Shift + Enter to select the cell below after entering. An overview of commands for both modes is listed under Help > Keyboard Shortcuts.

Question 1 What should you enter to create a new Input cell under an existing cell? And in what mode should you enter this? Enter your answer below and convert this cell back to formatted text.

> Answer: . . .

Question 2 Create an Input cell below with the command

```
print("Hello, world!")
```

and enter this cell.

C.2 Python for dummies

C.2.1 Objects

Python is a so-called object-oriented programming language, which means that we perform operations on objects. The operations covered here are mainly mathematical ones. Objects are data stored in computer memory. Each object has a particular class, which determines which operations can be performed on it. The object class can be checked with the command `type`.

```
>>> type(1)
int
>>> type(1.5)
float
```

The examples above are both numbers, yet belong to different object classes! `1` is an integer, whereas `1.5` is a is a so-called floating point number. Note that Python distinguishes these two classes from each other by the presence or absence of a decimal sign.

```
>>> type(1.)
float
```

Besides number objects, other types exist as well. A first example is a so-called `list`. This datatype stores a set of objects together in an ordered sequence, indicated by square brackets `[]`. A second example is the data type `string`, which is used to store text. Strings are indicated by single (`'`) or double (`"`) quotation marks. Text not between quotation marks is interpreted as a command.

```
>>> [1,1.5,2]
[1,1.5,2]

>>> type([1,1.5,2])
list

>>> type("This is a string.")
str
```

```
>>> type('This is a string as well.')
str

>>> This is not a string and will thus result in an error.
        File "<ipython-input-5-ec99100478b7>", line 1
          This is not a string and will thus result in an error.
                           ^
     SyntaxError: invalid syntax
```

C.2.2 Variables

Often we want to store a certain object (temporarily), in order to use it further. To do this, we use variables. A variable is a name we assign to a particular object in computer memory. This is done as follows: `variable_name = object`.

```
>>> a=1
    a
1
```

In the cell above, a value of 1 is assigned to the variable `a` on the first line. This operation gives no output. When we call `a` again, we get a value of 1 as output. Keep in mind the order! When assigning, the name of the variable must always be to the left and that of the object to the right of "=". Otherwise we get an error message.

```
>>> 1 = a
        File "<ipython-input-14-7596acd8e627>", line 1
         1 = a
              ^
      SyntaxError: can't assign to literal
```

Variables can also be overwritten, which means that we assign a new object to the name of the variable.

```
>>> a = 2
    a
2
```

Moreover, to determine the new object in the variable, reference can be made to the current object linked to the variable.

```
>>> a = a+1
    a
3
```

We can also delete a variable.

```
>>> del a
    a
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-22-ef9d13752aff> in <module>()
----> 1 del a
2 a

NameError: name 'a' is not defined
```

After closing the Python notebook, all variables are deleted from the memory. So, when restarting the notebook, we will have to enter all the (necessary) input cells again before we can continue working.

C.2.3 Mathematical operations

When we want to perform mathematical operations on number objects we can use following symbols:

$+$, $-$	addition and subtraction
$*$, $/$	multiplication and division
$**$	exponentiation
$()$	grouping of elements

The classical order of operations applies for parenthesis, power-law, multiplication/division, addition/-subtraction.

```
>>> a = 1
    a + 2
3
>>> b = 2
    c = a+b
    c
3
>>> b**(a+c)
16
```

C.2.4 Logical operations

The following symbols allow logical operations:

$<$, $>$	greater and less than
$<=$, $>=$	greater and less than or equal to
$==$, $!=$	equal to or not equal to

The result of such an operation is an object of the class `bool`, the Boolean numbers. These can take on only two values, viz. **`True`** (1) and **`False`** (0). Boolean numbers are widely used to control programs with so-called Control statements (see Section C.2.7).

```
>>> a == 1
True
>>> 1 == a
True
>>> d = a>b
    print(d)
    type(d)
False
bool
```

C.2.5 List operations

The object type `list` comes with a number of class specific operations. The most important is the so-called indexing, which allows us to retrieve an element from the list based on its rank. This is done using brackets.

```
>>> l = [1,2,3,4,5]
    l[0]
1
>>> l[2]
3
```

Note that indexing in Python always starts at 0!

Moreover, the function `sum` can be used to sum the elements in a list.

```
>>> sum(l)
15
```

C.2.6 Python functions

In addition to the (symbolic) operators, there are numerous built-in Python functions to perform operations with. These are called using:

`functiename(arguments).`

If the function takes multiple (n) arguments, it can be called in two ways:

1. `functionname(argument_1, ... ,argument_n),`
2. `functionname(argument_1 = argument_1, ... ,argument_n = argument_n),`

where `argument_i` is the name of argument i. We can also define functions ourselves.

```
    def functionname(argument_1, argument_2, ..., argument_n):
        (operations with arguments)
        return outputs
```

The operations to be performed by the function may span several lines of code. To indicate which expressions belong to the function, expressions that belong together are aligned in the same way using (*tabs*).

It is often necessary to include comments in our code. This way, we make the code not only readable for others, but also for ourselves when we look back after a while. In Python, we can use # at the beginning of a line to indicate that it should not be interpreted as code. When comments span multiple lines, they are between """.

It is important to note that functions work with local memory, which is separate from the notebook's global memory. The variables in local memory are included as arguments or defined in the function itself. Once is executed, everything following `return` is returned to the notebook, after which the local function memory is closed.

This is made clear by the following example.

```
>>> a = 2       # a is defined in the notebook memory

        def f(x,y):
                """
                This function calculates the sum of the squares of x and y.
```

```
                Inputs: x, y
                """
                z = x**2+y**2 # local variable z is the sum of the squares of x and y
                a = 10        # variable a is defined in local function memory
                return z      # just the value of z (and thus not the variable)
                              # is returned to the notebook memory

# we can now call the function in two ways for e.g. x=2 and y=3
>>> f(2,3)
13

>>> f(x=2,y=3)
13


>>> x   # x was only defined in local memory
        # and thus does not exist in the Notebook memory.
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-34-401b30e3b8b5> in <module>()
----> 1 x

NameError: name 'x' is not defined


>>> z   # z was only defined in local memory
# and therefore does not exist in notebook memory
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-35-a8a78d0ff555> in <module>()
----> 1 z

NameError: name 'z' is not defined


>>> a   # a was already defined in Notebook memory,
        # but has thus remained unchanged after the function evaluation
2
```

Finally, we mention that while defining a function, we can already pass so-called *default* values with the arguments. This makes these arguments optional, since the function has a default value for them.

```
#example:
def f(x,y=2):
        return x*y

>>> f(2)
4

>>> f(2,3)
6
```

Question 3 Write a function to find the weighted sum of two numbers. Where p is the weight of the first number and $(1-p)$ is the weight of the second. If no value p is given, the average of the two numbers should be returned.

C.2.7 Control statements

Lastly, we introduce two more so-called **control statements** (the **while** and **for** loop) and a **decision statement** (**if**), which are frequently used in programming. These statements define the so-called **flow of the program**, i.e. the way code is interpreted. By default this is done sequentially, line by line, but control statements can ensure that certain parts of the code are repeated or skipped.

C.2.7.1 While-loop

First we have the **while-loop**. This performs a number of operations as long as a certain condition is met and thus yields True as the result. The Python syntax is as follows:

```
        while condition:
              operations
```

The expressions within the while loop are aligned in the same way as functions. The control flow is illustrated in Figure C.1.

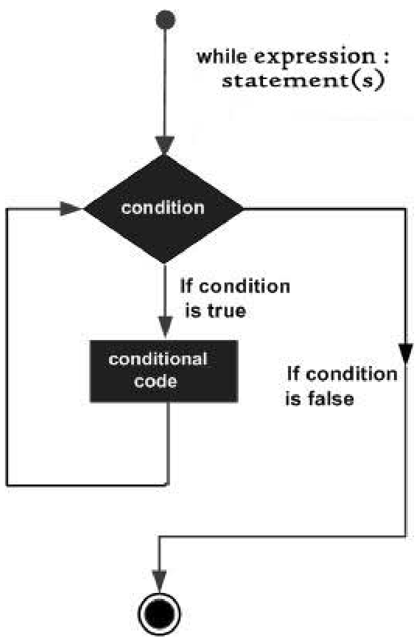


Figure C.1: The control flow of a while loop

Below we provide a simple example.

```
#Example
>>> x = 10       # assign an initial value to x
    print(x)     # prints the initial value
    while x > 1: # as long as x is bigger than 1:
        x=x/2    #      - divide x by 2
        print(x) #      - print the new value of x

 10
 5.0
 2.5
```

```
1.25
0.625
```

Note that if the condition is always met, the program will be stuck in a while loop and must be aborted manually. This can be done via **Kernel > Interrupt**.

C.2.7.2 For-loop

A second control statement is the **for-loop**. This repeats, like the while loop, a number of operations. However, the number of iterations here is predefined by the so-called **iterator**. This iterator iterates through the elements of a specified set of data. In Python, we write a for loop as:

```
for element in iterator:
    operations
```

The expressions within the for-loop are also grouped together. The control flow is illustrated in Figure C.2.

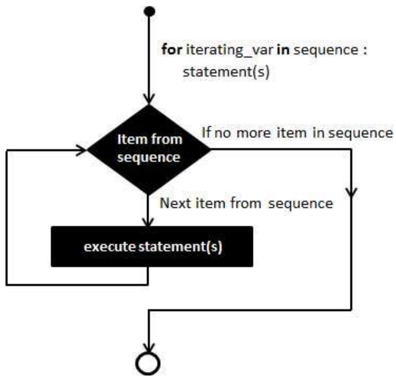


Figure C.2: Controlflow of a for-loop.

```
#example
>>> for value in [1,2,3,4]:
        print(value)

 1
 2
 3
 4
```

A for-loop can be used to quickly create lists.

```
new_list = [operation with element for element in iterator]
```

```
#example
>>> l_1 = [1,2,3,4]
    l_2 = [value**2 for value in l_1] #squares the values in list l_1
    l_2

[1, 4, 9, 16]
```

C.2.7.3 If, else-control statement

Finally, we discuss the decision structure **if**. This imposes a certain condition that must be met before a sequence of operations may be performed. In Python, we enter this as follows.

```
if condition == True:
    operations
```

The control flow is illustrated in Figure C.3.

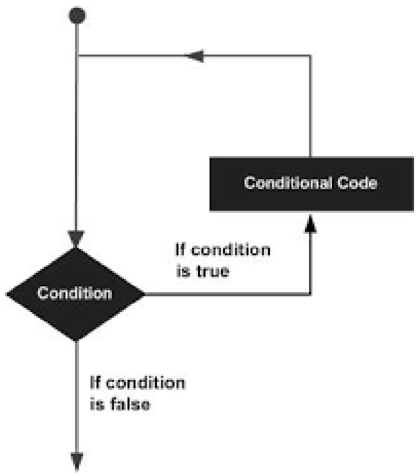


Figure C.3: Control flow of an if-statement.

In order to perform different operations depending on whether a condition is met or not, we can also use **else**. Then, we specify the operations to be performed if the condition is not met.

```
if condition == True:
    operations # are executed if condition was met
else:
    operations # are executed if condition was not met
```

The expressions within if and else are aligned in the same way.

```
#Example
>>> cd = True
    if cd:
        print("The condition was met")
    else:
        print("The condition was not met")

    The condition was met
```

C.3 Packages

In addition to the basic syntax, Python has numerous additional functionalities. These are collected in so-called **packages**, which can be loaded via:

```
import package as packagename
```

This makes that all the definitions in the package are stored in the variable `packagename`. Next, when we want to use a function from the imported package, we do so as follows:

```
packagename.functionname()
```

We can also import the function straight from the package using `from`:

```
from package import functiename
```

after which we can call the function directly:

```
functiename()
```

The input cell below loads a number of packets.

```
import numpy as np
import matplotlib.pyplot as plt
from ipywidgets import interact, fixed
```

C.3.1 Numpy

Numpy is the package that provides the foundation for scientific programming in Python. It contains numerous mathematical functions and numbers, such as:

- trigonometric functions: `np.sin(), np.cos() ...`
- exponential and logarithmic functions: `np.exp(), np.log()`
- the number π : `np.pi`

C.3.2 Sympy

C.3.3 Other Packages

In addition to Numpy, a number of other packages can be loaded:

- `matplotlib.pyplot` to make figures
- `ipywidgets` to make these figures interactive

D

Answers to the exercises

Chapter 2

Assignment 2.1 —

(a) $\forall\, x \in \mathbb{Z},\, \exists\, y \in \mathbb{Z}:\ x \geq y$

(b) $\exists\, y \in \mathbb{Z},\, \forall\, x \in \mathbb{Z}:\ x \geq y$

(c) $\exists\, x \in \mathbb{Z},\, \forall\, y \in \mathbb{Z}:\ x + y \neq 0$

(d) $\forall\, y \in \mathbb{Z},\, \exists\, x \in \mathbb{Z}:\ x + y \neq 0$

Assignment 2.2 —

(a) The set A contains the natural numbers between 2 and 6. $A = \{3, 4, 5\}$

(b) The set B contains the positive fractions that are solutions of $2x^2 + x - 6 = 0$. $B = \left\{\frac{3}{2}\right\}$

(c) The set C contains the positive integers that are solutions of $x^2 - 5 = 0$. $C = \varnothing$

Assignment 2.3 —

(a) $A = \left\{a \;\middle|\; \frac{a}{2} \in \mathbb{N} \;\wedge\; a > 100\right\}$

(b) $A = \left\{(a, b) \;\middle|\; a, b \in \mathbb{Z} \;\wedge\; \frac{a}{2} \in \mathbb{Z} \;\wedge\; \frac{b+1}{2} \in \mathbb{Z}\right\}$

(c) $A = \left\{a \;\middle|\; a \in \mathbb{Z}_0 \;\wedge\; \frac{a}{3} \in \mathbb{Z}_0\right\}$

(d) $A = \left\{a \;\middle|\; a \in \mathbb{Q}^+ \;\wedge\; \sqrt{a} > 3\right\}$

(e) $A = \left\{a \;\middle|\; a \in \mathbb{R} \;\wedge\; \frac{a-2}{6} \in \mathbb{Z}\right\}$

Assignment 2.4 —

(a) $\{1,3,5,7,9,11,...\} = \{x \in \mathbb{N} | \, x \text{ is an odd number}\}$

(b) $\{x | \, x \text{ is a rose}\} \subset \{x | \, x \text{ is a flower}\}$

(c) $\{1,3,5,7,9\} \not\ni 2$

(d) $\{1\} \subset \{1,3,5,7,9\}$

(e) $\{1,3\} \not\subset \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$ also $\{1,3\} \notin \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(f) $\{1,3\} \subset \{1,3,5,7,9\}$

(g) $\{1,3,5,7,9\} \neq \emptyset$

(h) $\{1\} \in \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(i) $\{1,3,\{5,7,9\}\} \not\ni 5$

Assignment 2.5 —

(a) $1 \in A$: True

(b) $\{1\} \in A$: True

(c) $\{1\} \subseteq A$: True

(d) $\{\{1\}\} \subseteq A$: True

(e) $2 \in A$: False

(f) $\{\{2\}\} \subseteq A$: True

(g) $\{\{2\}\} \subset A$: True

(h) $\{2\} \subseteq A$: False

Assignment 2.6 —

(a) $(A \cup B) \cap C = \{3,5\}$

(b) $A \cup (B \cap C) = \{1,2,3,4,5\} = A$

(c) $\overline{C} \cup \overline{D} = U$

(d) $\overline{C \cap D} = \overline{C} \cup \overline{D} = U$ (rules of De Morgan)

(e) $(A \cup B) \backslash C = \{1,2,4\}$

(f) $A \cup (B \backslash C) = \{1,2,3,4,5\} = A$

(g) $(B \backslash C) \backslash D = \{1\}$

(h) $B \backslash (C \backslash D) = \{1,2,4\}$

(i) $(A \cup B) \backslash (C \cap D) = \{1,2,3,4,5\} = A$

Assignment 2.7 —

(a) $A \cap (B \backslash A) = \emptyset$

(b) $(A \backslash B) \cup (A \cap B) = A$

(c) $\overline{A} \cup \overline{B} \cup \left(A \cap B \cap \overline{C}\right) = \overline{A \cap B \cap C}$

(d) $(A \cap B) \cup \left(A \cap B \cap \overline{C} \cap D\right) \cup \left(\overline{A} \cap B\right) = B$

(e) $(A \cap B) \cup \left(B \cap \left((C \cap D) \cup \left(C \cap \overline{D}\right)\right)\right) = (A \cap B) \cup (B \cap C)$

Assignment 2.8 — The answer is left as an exercise to the reader.

Assignment 2.9 — The answer is left as an exercise to the reader.

Assignment 2.10 — The numbering of the areas is done from left to right.

(a) $A \cup (B \cap C)$

(b) $((B \cap C)\backslash A) \cup ((A \cap C)\backslash B)$

(c) $A\backslash(B \cup C)$ of $(A\backslash B) \cap (A\backslash C)$

(d) $(A \cap C)\backslash B$

(e) $((A \cap C)\backslash B) \cup (B \cap D)$

(f) $(D\backslash(A \cup C)) \cup ((C \cap A)\backslash B)$

Assignment 2.11 —

(a) b is an upper bound of the set W: $\forall w \in W : w \leq b$

(b) b is not an upper bound of the set W: $\exists w \in W : b < w$

(c) W is a set bounded from above: $\exists b \in \mathbb{R} | \forall w \in W : w \leq b$

(d) W is a set not bounded from above: $\forall b \in \mathbb{R} : \exists w \in W : b < w$

Assignment 2.12 —

(a) inf = $1/2$, sup = 1

(b) inf = $1/3$, sup = 2

(c) inf = 0, sup = 1

Assignment 2.13 —

(a) inf $A = 1$, sup $A = 7$, min $A = 1$, max A does not exist, boundary points $A : 1, 2, 5, 7$, internal points $A :]2, 5[\cup]5, 7[$

(b) inf $A = -3$, sup A does not exist, min $A = -3$, max A does not exist, boundary points $A : -3, 0, 4, 7$, internal points $A :]0, 4[\cup]7, +\infty[$

(c) inf A does not exist, sup $A = 9$, min A does not exist, max $A = 9$, boundary points $A : -2, 2, 3, 4, 5, 9$, internal points $A :]-\infty, -2[\cup]-2, 2[\cup]3, 4[\cup]5, 9[$

Assignment 2.14 —

(a) rational

(b) rational

(c) irrational

(d) irrational

(e) rational

(f) rational

Assignment 2.15 —

(a) $\sum_{j=1}^{99} x^j$

(b) $\sum_{j=1}^{25} \sqrt{2j+1}$

(c) $\prod_{j=1}^{13} \frac{j^2}{a+j}$

Assignment 2.16 —

(a) 15

(b) 14

(c) 65

(d) $-\frac{7}{12}$

(e) 165

(f) -2368450

(g) -482295

(h) 165

Assignment 2.17 —

(a) $\frac{n(3n-1)}{2}$

(b) $\frac{3n-7}{2n}$

(c) $\frac{n}{2}(6n^2-15n+11)$

(d) $\frac{(n+1)(2n-11)+24}{6n^2}$

(e) $(n!)^3$

(f) $\frac{n+1}{2n}$

Assignment 2.18 —

(a) $2a^m b^{2n} c^{3p}$

(b) $1+\sqrt{x}$

(c) $6-2\sqrt{2}-2\sqrt{3}+2\sqrt{2}\sqrt{3}$

(d) $\frac{19\sqrt{a}}{2b}$

(e) $\sqrt[6]{\frac{x+1}{x-1}}$

(f) $a+\frac{b}{2}$

(g) a^2+b^2

(h) $\frac{x^2-1}{x}$

(i) $x^{\frac{5}{2}}$

(j) $\frac{81}{256}a^{\frac{7}{4}}b^{-\frac{9}{2}}$

Assignment 2.19 —

	$z+w$	zw	z^2	z^{-1}	$\frac{z}{w}$	$\frac{w}{z}$	$\bar{z}$	$z\bar{z}$	$(\bar{z})^2$
(a)	$2+7i$	$8i$	$-5+12i$	$\frac{2-3i}{13}$	$\frac{3-2i}{4}$	$\frac{12+8i}{13}$	$2-3i$	13	$-5-12i$
(b)	1	$1-i$	$2i$	$\frac{1-i}{2}$	$-1+i$	$\frac{-1-i}{2}$	$1-i$	2	$-2i$
(c)	$5+2i$	$41+11i$	$-16-30i$	$\frac{3+5i}{34}$	$\frac{-29-31i}{53}$	$\frac{-29+31i}{34}$	$3+5i$	34	$-16+30i$
(d)	$2\sqrt{2}$	4	$-4i$	$\frac{\sqrt{2}+\sqrt{2}i}{4}$	$-i$	i	$\sqrt{2}+\sqrt{2}i$	4	$4i$
(e)	$-2\sqrt{3}i$	-4	$-2-2\sqrt{3}i$	$\frac{1+\sqrt{3}i}{4}$	$\frac{1+\sqrt{3}i}{2}$	$\frac{1-\sqrt{3}i}{2}$	$1+\sqrt{3}i$	4	$-2+2\sqrt{3}i$
(f)	$-\sqrt{2}$	1	$-i$	$\frac{-\sqrt{2}-\sqrt{2}i}{2}$	$-i$	i	$\frac{-\sqrt{2}-\sqrt{2}i}{2}$	1	i

Assignment 2.20 —

(a) $19-4i$

(b) $-4+11i$

(c) $-3+4i$

(d) $3+4i$

(e) 61

(f) $\frac{5-2i}{29}$

(g) $\frac{5-i}{13}$

(h) $\frac{-2}{5}$

Chapter 3

Assignment 3.1 —

(a) no function

(b) function

(c) no function

(d) no function

Assignment 3.2 —

(a) $f(x) = x^3 - 1$ and $g(x) = \dfrac{x+1}{x-1}$

a) $(f+g)(x) = \dfrac{x^4 - x^3 + 2}{x-1}$ with $\operatorname{dom}(f+g) = \mathbb{R} \setminus \{1\}$

b) $(f-g)(x) = \dfrac{x(x^3 - x^2 - 2)}{x-1}$ and $\operatorname{dom}(f-g) = \mathbb{R} \setminus \{1\}$

c) $(fg)(x) = (x^2 + x + 1)(x+1)$ with $\operatorname{dom}(fg) = \mathbb{R} \setminus \{1\}$

d) $\left(\dfrac{f}{g}\right)(x) = \dfrac{(x^3-1)(x-1)}{(x+1)}$ met $\operatorname{dom}\left(\dfrac{f}{g}\right) = \mathbb{R} \setminus \{-1, 1\}$

(b) $f(x) = \dfrac{x}{2}$ and $g(x) = \dfrac{2}{x}$

a) $(f+g)(x) = \dfrac{x^2+4}{2x}$ with $\operatorname{dom}(f+g) = \mathbb{R}_0$

b) $(f-g)(x) = \dfrac{x^2-4}{2x}$ with $\operatorname{dom}(f-g) = \mathbb{R}_0$

c) $(fg)(x) = 1$ with $\operatorname{dom}(fg) = \mathbb{R}_0$

d) $\left(\dfrac{f}{g}\right)(x) = \dfrac{x^2}{4}$ with $\operatorname{dom}\left(\dfrac{f}{g}\right) = \mathbb{R}_0$

(c) $f(x) = x$ and $g(x) = \sqrt{x+1}$

a) $(f+g)(x) = x + \sqrt{x+1}$ with $\operatorname{dom}(f+g) = [-1, +\infty[$

b) $(f-g)(x) = x - \sqrt{x+1}$ met $\operatorname{dom}(f-g) = [-1, +\infty[$

c) $(fg)(x) = x\sqrt{x+1}$ with $\operatorname{dom}(fg) = [-1, +\infty[$

d) $\left(\dfrac{f}{g}\right)(x) = \dfrac{x}{\sqrt{x+1}}$ with $\operatorname{dom}\left(\dfrac{f}{g}\right) =]-1, +\infty[$

Assignment 3.3 —

(a) $f \circ g(0) = 2$

(b) $g(f(0)) = 22$

(c) $f \circ f(-5) = 5$

(d) $g(g(2)) = -2$

Assignment 3.4 —

(a) $(g \circ f)(x) = \dfrac{1}{5x-2}$ with $\operatorname{dom}(g \circ f) = \mathbb{R} \setminus \left\{\dfrac{2}{5}\right\}$

$(f \circ g)(x) = \dfrac{5}{x-2}$ with $\operatorname{dom}(f \circ g) = \mathbb{R} \setminus \{2\}$

(b) $(g \circ f)(x) = |x|$ with $\operatorname{dom}(g \circ f) = \mathbb{R}$

$(f \circ g)(x) = x$ with $\operatorname{dom}(f \circ g) = \mathbb{R}^+$

(c) $(g \circ f)(x) = \sqrt[3]{1-x^3}$ with $\operatorname{dom}(g \circ f) = \mathbb{R}$

$(f \circ g)(x) = 1-x$ with $\operatorname{dom}(f \circ g) = \mathbb{R}$

(d) $(g \circ f)(x) = |x^2-5|$ with $\operatorname{dom}(g \circ f) = \mathbb{R}$

$(f \circ g)(x) = x^2-2x-3$ with $\operatorname{dom}(f \circ g) = \mathbb{R}$

(e) $(g \circ f)(x) = \sqrt{4-|x|}$ with $\operatorname{dom}(g \circ f) = [-4,4]$

$(f \circ g)(x) = \sqrt{4-x}$ with $\operatorname{dom}(f \circ g) =]-\infty,4]$

Assignment 3.5 — $(g \circ g)(x) = x$ with $\operatorname{dom}(g \circ g) = \mathbb{R}\setminus\{-1\}$.

Assignment 3.6 —

(a) $f+g$: Neither odd, nor even

(b) fg: odd

(c) f/g: odd

(d) g/f: odd

(e) f^2: even

(f) g^2: even

(g) $f \circ g$: even

(h) $g \circ f$: even

(i) $f \circ f$: even

(j) $g \circ g$: odd

Assignment 3.7 — Consider the graphs below.

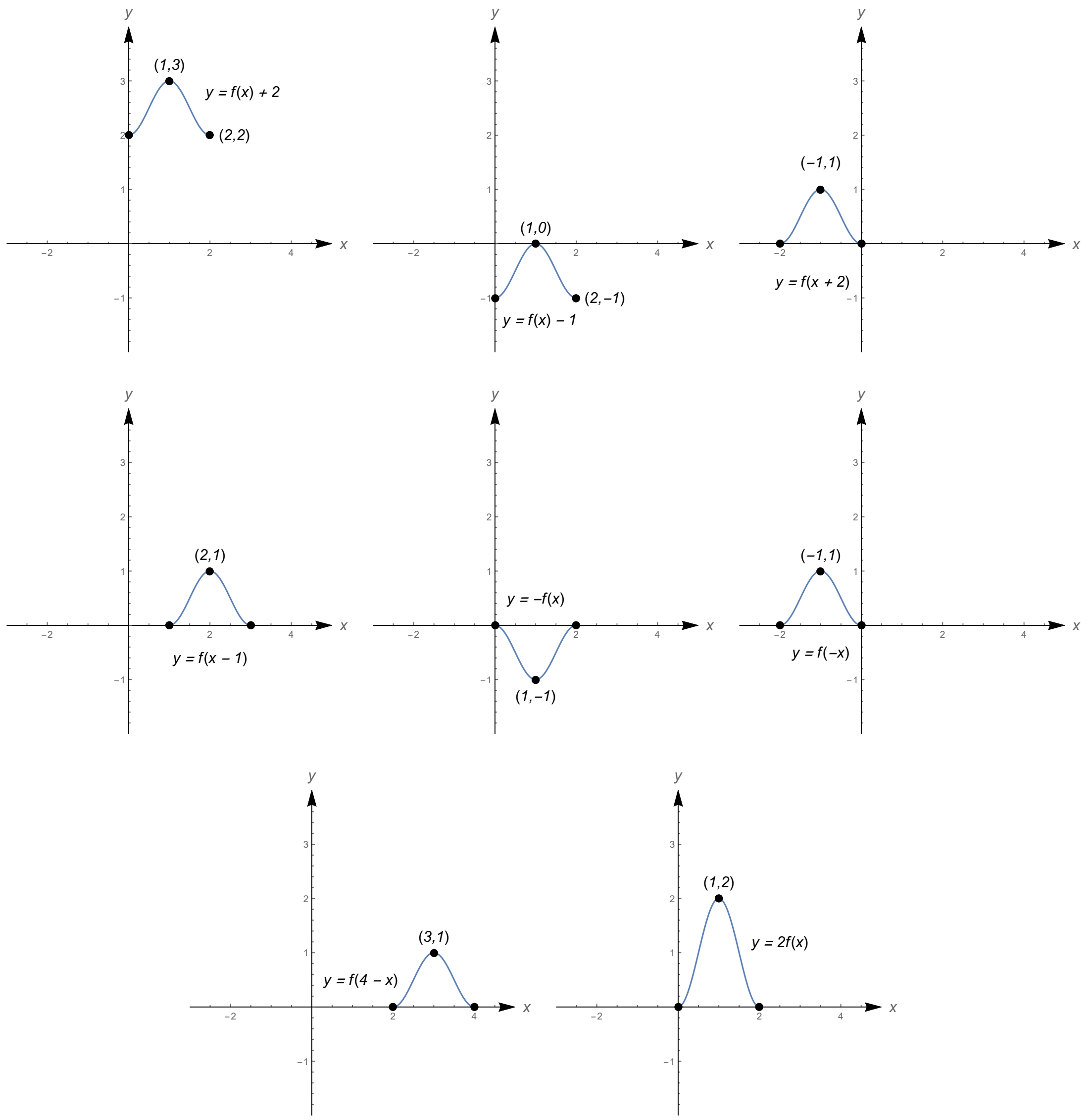


Figure 3.27: Graphs from the transformations from Exercise 3.7 (part 1).

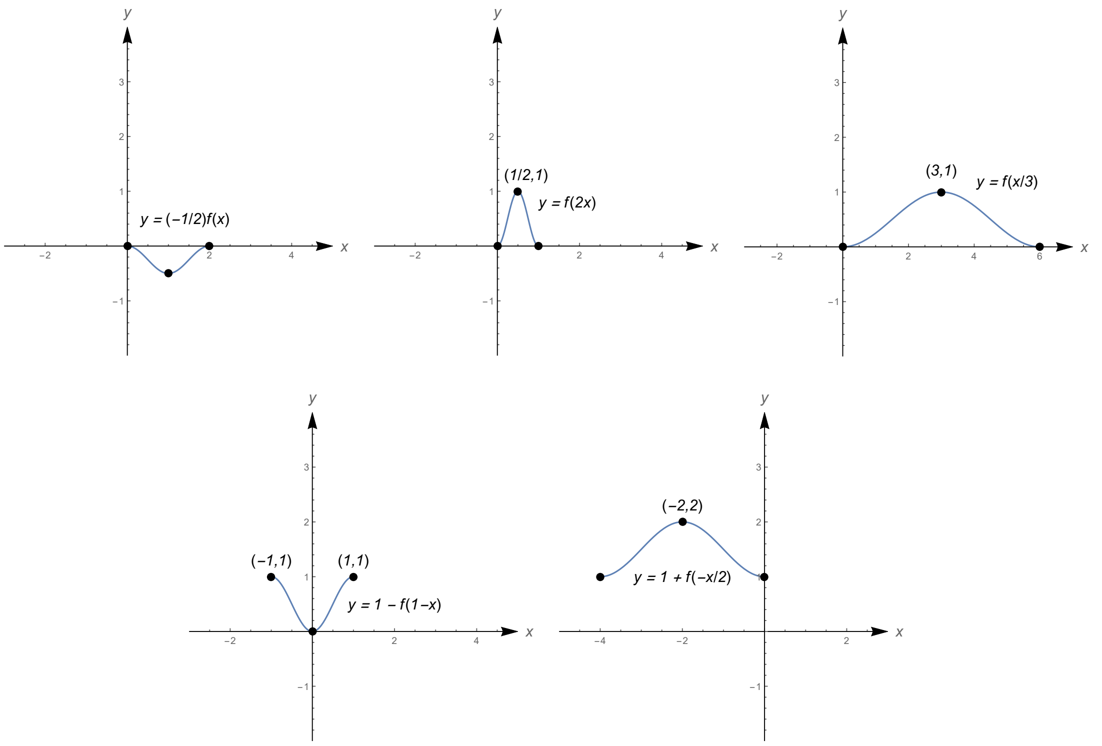


Figure 3.28: Graphs for the transformations in Exercise 3.7 (part 2).

Assignment 3.8 — Consider the graphs below.

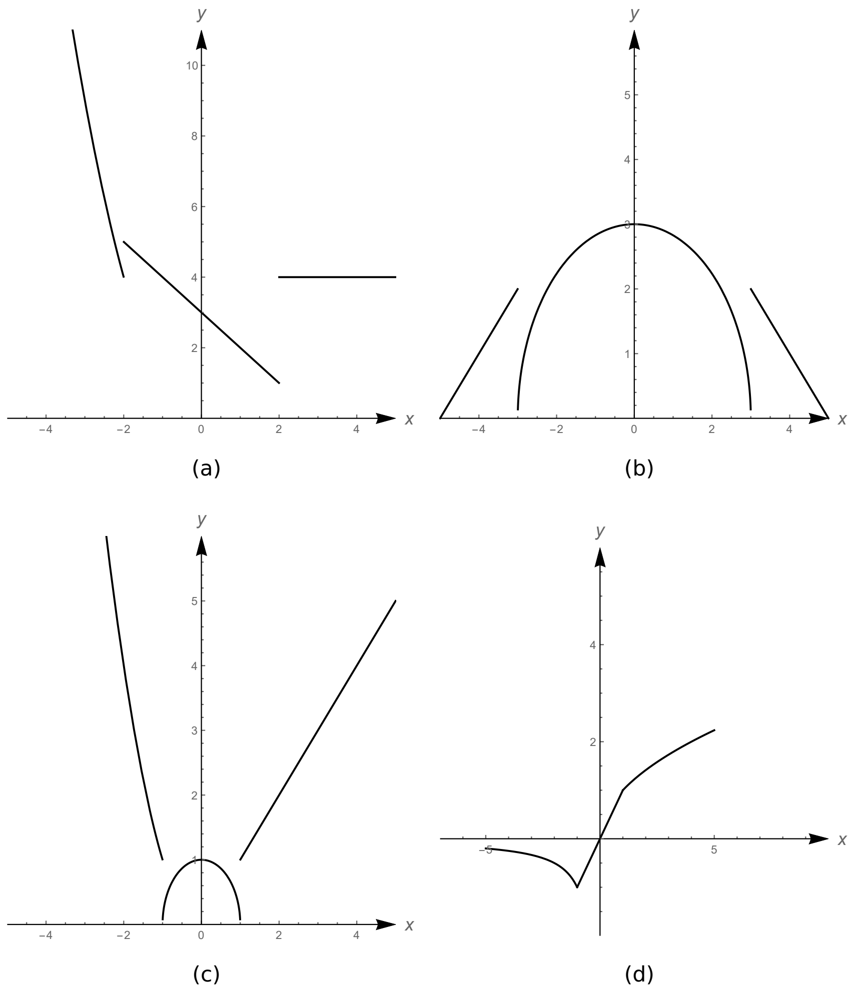


Figure 3.29: Graphs of the piecewise functions in Exercise 3.8.

Assignment 3.9 — $f(x) = \begin{cases} |x+1|-1, & \text{if} \quad x < 2, \\ 2, & \text{if} \quad x \geq 2. \end{cases}$ or: $f(x) = \begin{cases} -(x+2), & \text{if} \quad x < -1, \\ x, & \text{if} \quad -1 \leq x \leq 2, \\ 2, & \text{if} \quad x > 2. \end{cases}$

Assignment 3.10 —

(a) $x = -\frac{13}{8} \vee x = \frac{53}{8}$

(b) $x = -\frac{3}{10}$

(c) $x = \frac{-4}{3} \vee x = 2$

(d) $x = -1 \vee x = 2$

(e) No solution

(f) No solution

(g) $x \in \left]\frac{5}{3}, 3\right[$

(h) $x \in [0, 4]$

(i) $x = -5$

(j) $x = -\frac{49}{18} \vee x = \frac{17}{6}$

(k) $x \geq 2$

(l) $x > 1$

(m) $x < -3 \vee x > 1$

(n) $x < 2$

(o) $2 < x < 9$

(p) $x < -10 \vee -4 < x < 0 \vee x > 6$

(q) $\frac{3}{4} < x < 1 \vee x > 1$

(r) $x \in]-1, +\infty[$

(s) No solution

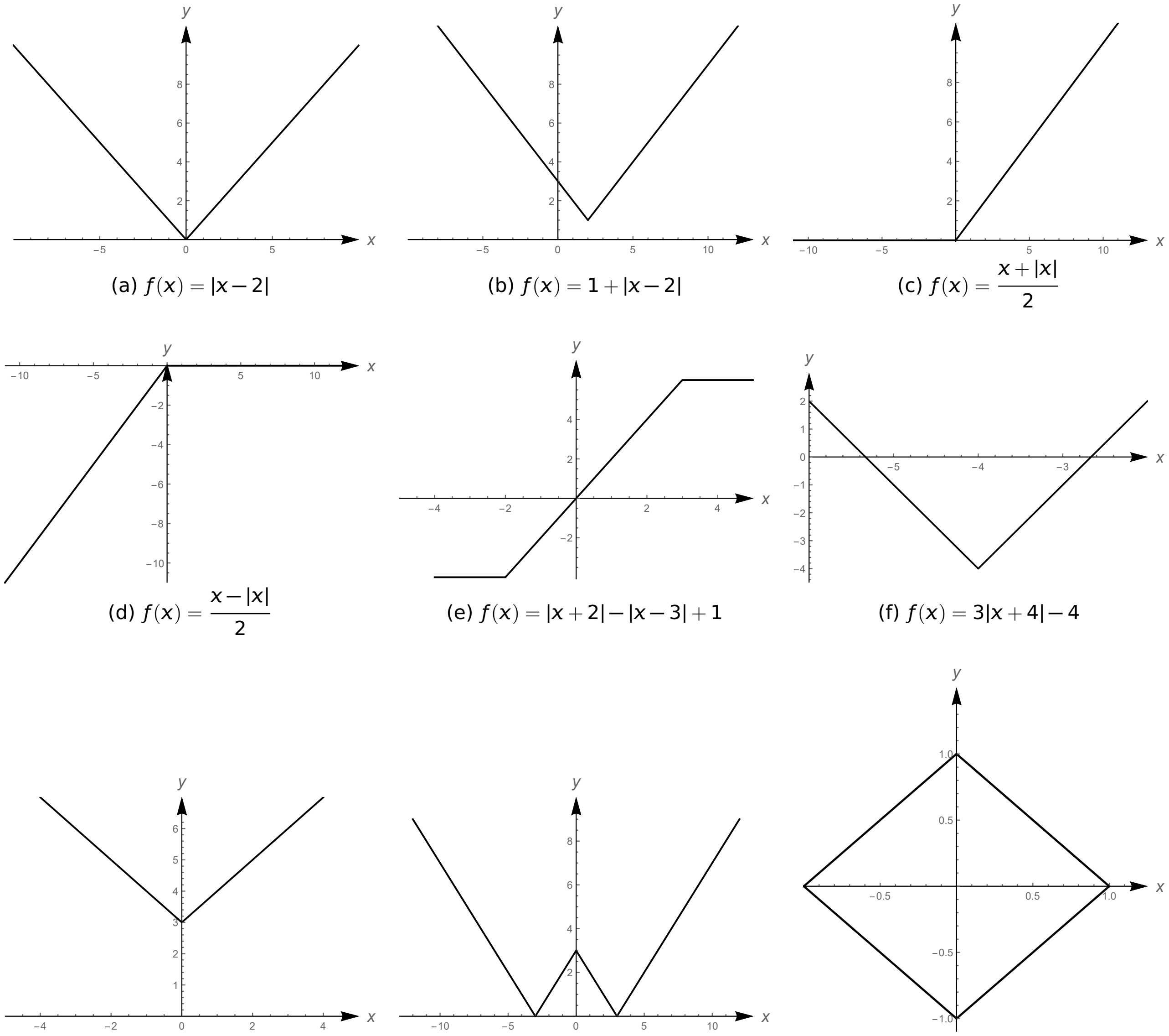
(t) $x = 1 \vee x = 2 - \sqrt{3} \vee x = -2 - \sqrt{7}$

(u) $-1 \leq x \leq \frac{3}{5} \vee x \geq 3$

(v) $x \in \mathbb{R}$

(w) $x \geq 1$

Assignment 3.11 — Consider the graphs below.



(a) $f(x) = |x-2|$ (b) $f(x) = 1+|x-2|$ (c) $f(x) = \dfrac{x+|x|}{2}$

(d) $f(x) = \dfrac{x-|x|}{2}$ (e) $f(x) = |x+2|-|x-3|+1$ (f) $f(x) = 3|x+4|-4$

(g) $f(x) = ||x|+3|$ (h) $f(x) = ||x|-3|$ (i) $|x|+|y| = 1$

Figure 3.31: Graphs of the absolute value functions in Exercise 3.11.

Assignment 3.12 —

(a) $f^{-1}(x) = \dfrac{x-8}{2}$ Function

(b) $f^{-1}(x) = -\dfrac{5}{3}x + \dfrac{1}{3}$ Function

(c) $f^{-1}(x) = \dfrac{6-5x}{x-1}$ Function

(d) $f^{-1}(x) = \sqrt[3]{x-1}$ Function

(e) $f^{-1}(x) = -\dfrac{1}{2} \pm \sqrt{x+\dfrac{1}{4}} \quad \wedge \quad x \geq -\dfrac{1}{4}$ No Function

(f) $f^{-1}(x) = \pm\sqrt{1-x^2} \quad \wedge \quad x \in [0,1]$ No Function

(g) $f^{-1}(x) = \dfrac{1}{9}(x+4)^2+1 \quad \wedge \quad x \geq -4$ Function

(h) $f^{-1}(x) = 3 \pm \sqrt{x+4} \quad \wedge \quad x \geq -4$ No Function

Assignment 3.13 —

	dom f	codom f	im f	inj	surj	bij	periodic	even/odd	mon.	bounded
(a)	$[-2,2]$	$\mathbb{R}^+$	$[0,2]$	no	no	no	no	even	no	upper/under
(b)	$\mathbb{R}$	$\mathbb{R}^+$	$[2,+\infty[$	no	no	no	no	even	no	under
(c)	$\mathbb{R}\backslash\{-1,1\}$	$\mathbb{R}$	$\mathbb{R}\backslash[0,1[$	no	no	no	no	even	no	no
(d)	$\mathbb{R}$	$\mathbb{R}^+$	$]0,1]$	no	no	no	no	even	no	upper/under
(e)	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	yes	yes	yes	no	no	yes	no
(f)	$\mathbb{R}$	$\mathbb{R}^+$	$[1,+\infty[$	no	no	no	no	even	no	under
(g)	$]-\infty,1]$	$\mathbb{R}^+$	$\mathbb{R}^+$	yes	yes	yes	no	no	yes	under
(h)	$\mathbb{R}$	$\mathbb{R}^+$	$\mathbb{R}^+$	no	yes	no	no	even	no	under
(i)	$\mathbb{R}\backslash\{-1\}$	$\mathbb{R}\backslash\{1\}$	$\mathbb{R}\backslash\{1\}$	yes	yes	yes	no	no	yes	no
(j)	$\mathbb{R}_0$	$\mathbb{R}_0$	$\mathbb{R}_0$	yes	yes	yes	no	oneven	no	no
(k)	$[-2,+\infty[$	$\mathbb{R}$	$]-\infty,3]$	yes	no	no	no	no	yes	upper
(l)	$\mathbb{R}$	$\mathbb{R}^+$	$\mathbb{R}^+$	no	yes	no	no	no	no	under

	lok. max/min
(a)	$(0,2),(\pm 2,0)$
(b)	$(0,2)$
(c)	$(0,1)$
(d)	$(0,1)$
(e)	None
(f)	$(0,1)$
(g)	$(1,0)$
(h)	$(0,0)$
(i)	None
(j)	None
(k)	$(-2,3)$
(l)	$(-2,0)$

Chapter 4

Assignment 4.1 —

(a) All points located in the half plane under the line $y = x - 1$, including the points on the line.

(b) All points located in the area between the lines $y = -x - 4$, $y = x - 4$ and $y = 2 - x$, which includes the origin.

(c) All points located in the region understood between the line $y = x+2$ and the parabola $y = x^2$, including points on the parabola.

(d) All points located within the circle with center $(-1,0)$ and radius 3.

(e) All points located within the circle with center $(1,0)$ and radius 1 ánd within the circle with center $(0,1)$ and radius 1.

(f) All points located outside the circle with center $(2,-1)$ and radius 3 ánd to the right of the line $x+y=1$.

Assignment 4.2 —

(a) $2(x-2)\left(x^2+2x+4\right)(x+2)\left(x^2-2x+4\right)$

(b) $(2x+1)^3$

(c) $\left(x^2+1\right)(2x+3)$

(d) $(x-2)^3(x-1)$

(e) $(x-2)(x+2)\left(x^2+x+2\right)$

(f) $(2x-1)(x+1)(x-2)$

(g) $(2a-5b)^3$

(h) $(a-1-b-2c)(a-1+b+2c)$

(i) $(x-y)(x^2-3xy+4y^2)$

(j) $9(x+1)^2(x^2+4x+1)$

Assignment 4.3 —

(a) $16x^4-8x^2+1 = (2x+1)^2(2x-1)^2 = (4x^2-1)^2$

(b) $x^4-1 = (x^2+1)(x+1)(x-1)$

(c) $x^5-x^4-16x+16 = (x-2)(x+2)(x^2+4)(x-1)$

(d) $x^5+x^3+8x^2+8 = (x^2+1)(x+2)(x^2-2x+4)$

(e) $x^4+6x^3+9x^2 = x^2(x+3)^2$

(f) $x^6-3x^4+3x^2-1 = (x-1)^3(x+1)^3$

(g) $x^9-4x^7-x^6+4x^4 = x^4(x-2)(x-1)(x+2)(x^2+x+1)$

Assignment 4.4 —

(a) $x^3-3x^2+20 = (x+2)(x^2-5x+10) = (x+2)\left(x-\dfrac{5+\sqrt{15}i}{2}\right)\left(x-\dfrac{5-\sqrt{15}i}{2}\right)$
real zeros: $x=-2$, complex zeros: $x = \dfrac{5 \pm \sqrt{15}i}{2}$

(b) $2x^3-4x^2-10x+12 = 2(x-3)(x-1)(x+2)$
real zeros: $x=3$, $x=1$ and $x=-2$

(c) $x^6-16x^3+64 = (x^3-8)^2 = (x-2)^2(x^2+2x+4)^2 = (x-2)^2\left(x+1-\sqrt{3}i\right)^2\left(x+1+\sqrt{3}i\right)^2$
real zeros: $x=2$ (2x), complex zeros: $x = -1 \pm \sqrt{3}i$ (2x)

(d) $8x^4-20x^3+18x^2-7x+1 = (x-1)(2x-1)^3$
real zeros: $x = \dfrac{1}{2}$ (3x) and $x=1$

(e) $x^3-16x^2+48x+72=(x-6)\left(x-5-\sqrt{37}\right)\left(x-5+\sqrt{37}\right)$
real zeros: $x=6$, $x=5+\sqrt{37}$ and $x=5-\sqrt{37}$

(f) $4x^3-14x^2+8x+8=4(x-2)^2\left(x+\dfrac{1}{2}\right)$
real zeros: $x=2$ (2x) and $x=-\dfrac{1}{2}$

(g) $x^5+6x^4+x^3-26x^2-32=(x^3+6x^2-32)(x^2+1)=(x-2)(x+4)^2(x-i)(x+i)$
real zeros: $x=2$ and $x=-4$ (2x), complex zeros: $x=\pm i$

(h) $-2x^6-10x^5-16x^4-8x^3=-2x^3(x+1)(x+2)^2$
real zeros: $x=0$ (3x), $x=-1$ and $x=-2$ (2x)

Assignment 4.5 —

(a) $a=-3, b=-8$

(b) $a=3, b=30$

Assignment 4.6 —

(a) $\left]-\infty,\dfrac{1}{2}\right[\cup\,]4,5[$

(b) $\{-2\}\cup[1,3]$

(c) $]-\infty,-1[\,\cup\,]-1,0[\,\cup\,]2,+\infty[$

(d) $]-\infty,-2[\,\cup\,]-\sqrt{2},\sqrt{2}[$

(e) $]-\infty,-\sqrt{3}[\,\cup\,]\sqrt{3},+\infty[$

Assignment 4.7 —

(a) $-4x^4+x^3+x^2+x+1$

(b) $x+\dfrac{2x-1}{x^2-2}$

(c) $1-\dfrac{5x+3}{x^2+5x+1}$

(d) $x-2+\dfrac{x+6}{x^2+2x+3}$

(e) $2x+9+\dfrac{44x-68}{x^2-6x+7}$

Assignment 4.8 —

(a) $\varnothing$

(b) $]-1,0[\,\cup\,]1,+\infty[$

(c) $[0,+\infty[$

(d) $\left]-3,-\dfrac{1}{3}\right[\cup\,]2,3[$

(e) $[-3,0[\,\cup\,]0,4[\,\cup\,[5,+\infty[$

(f) $\left]-1,-\dfrac{1}{2}\right]\cup\,]1,+\infty[$

Assignment 4.9 —

(a) $\operatorname{dom} f$: $\mathbb{R}$, VA: none, HA: $y=0$

(b) $\operatorname{dom} f$: $\mathbb{R}\backslash\{-1,0,1\}$, VA: $x=-1$, $x=0$, and $x=1$, HA: $y=0$

(c) $\operatorname{dom} f$: $\mathbb{R}\backslash\{-1,0\}$, VA: $x=-1$ and $x=0$, HA: $y=0$

(d) $\operatorname{dom} f\colon \mathbb{R}\setminus\left\{\dfrac{-1\pm\sqrt{5}}{2}\right\}$, VA: $x=\dfrac{-1\pm\sqrt{5}}{2}$, HA: none

(e) $\operatorname{dom} f\colon \mathbb{R}\setminus\{-3,2\}$, VA: $x=2$, HA: $y=1$

(f) $\operatorname{dom} f\colon \mathbb{R}\setminus\{-1\}$, VA: $x=-1$, HA: none

Assignment 4.10 —

(a) No solution

(b) $x=10+4\sqrt{5}$

(c) $x=2$

(d) $x>0$

(e) $x=8$

(f) $\dfrac{1}{10}\left(\sqrt{5}-5\right)\leq x\leq 0$

Assignment 4.11 — $v=\dfrac{\sqrt{3}c}{2}$

Assignment 4.12 —

(a)
- $\operatorname{dom} f=[-4,+\infty[$,
- zero(s): $(17/2,0)$,
- intersection(s) x-axis: $(17/2,0)$, intersection(s) y-axis: $(0,5-2\sqrt{2})$,
- asymptotes: none.

(b)
- $\operatorname{dom} f=]-3,0]\cup]3,+\infty[$,
- zero(s): $(0,0)$,
- intersection(s) x-axis: $(0,0)$, intersection(s) y-axis: $(0,0)$,
- asymptotes: $x=\pm 3$, $y=0$.

(c)
- $\operatorname{dom} f=\mathbb{R}\setminus\{-2\}$,
- zero(s): $(0,0)$,
- intersection(s) x-axis: $(0,0)$, intersection(s) y-axis: $(0,0)$,
- asymptotes: $x=-2$, $y=5$.

(d)
- $\operatorname{dom} f=\mathbb{R}$,
- zero(s): $(0,0)$, $(7,0)$
- intersection(s) x-axis: $(0,0)$, $(7,0)$, intersection(s) y-axis: none,
- asymptotes: none.

(e)
- $\operatorname{dom} f=]-\infty,-2[\cup]2,+\infty[$,
- zero(s): none,
- intersection(s) x-axis: none, intersection(s) y-axis: none,
- asymptotes: $x=\pm 2$, $y=0$.

(f)
- $\operatorname{dom} f=[3,17[\cup]17,+\infty[$,
- zero(s): $(3,0)$,
- intersection(s) x-axis: $(3,0)$, intersection(s) y-axis: none,

- asymptotes: $x = 17$.

(g)
- $\operatorname{dom} f =]-\infty, -1/2] \cup]2, +\infty[$,
- zero(s): $(-1/2, 0)$,
- intersection(s) x-axis: $(-1/2, 0)$, intersection(s) y-axis: none,
- asymptotes: $x = 2$

Assignment 4.13 —

(a) Graph (c)

(b) Graph (a)

(c) Graph (b)

(d) Graph (d)

Assignment 4.14 —

(a) $\operatorname{dom} f_t =]0, 2t[$, no zeros, asymptotes: $x = 0$ en $x = 2t$

(b) This sketch is left as an exercise

(c) Two graphs for different values of t do not have an intersection point.

Assignment 4.15 —

(a) $(x-2)^2 = y+1$

- top V: $(2, -1)$,
- focal point F: $(2, -3/4)$,
- axis of symmetry: $x = 2$,
- directrix d: $y = -5/4$,
- intersection(s) x-axis: $(1, 0)$ en $(3, 0)$, intersection(s) y-axis: $(0, 3)$.

(b) $(x-1)^2 = y+1$

- top V: $(1, -1)$,
- focal point F: $(1, -3/4)$,
- axis of symmetry: $x = 1$,
- directrix d: $y = -5/4$,
- intersection(s) x-axis: $(0, 0)$ en $(2, 0)$, intersection(s) y-axis: $(0, 0)$.

(c) $(y+1)^2 = -2(x-1/2)$

(d) top V: $(1/2, -1)$,
- focal point F: $(0, -1)$,
- axis of symmetry: $y = -1$,
- directrix d: $x = 1$,
- intersection(s) x-axis: $(0, 0)$, intersection(s) y-axis: $(0, 0)$ en $(0, -2)$.

(e) $(x+1/2)^2 = -(y-1/4)$

- top V: $(-1/2, 1/4)$,

- focal point F: $(-1/2, 0)$,
- axis of symmetry: $x = -1/2$,
- directrix d: $y = 1/2$,
- intersection(s) x-axis: $(0,0)$ en $(-1,0)$, intersection(s) y-axis: $(0,0)$.

(f) $\left(x - \frac{5}{6}\right)^2 = -\frac{1}{3}\left(y - \frac{73}{12}\right)$

- top V: $\left(\frac{5}{6}, \frac{73}{12}\right)$,
- focal point F: $\left(\frac{5}{6}, 6\right)$,
- axis of symmetry: $x = 5/6$,
- directrix d: $y = 37/6$,
- intersection(s) x-axis: $\left(\frac{5 \pm \sqrt{73}}{6}, 0\right)$, intersection(s) y-axis: $(0,4)$.

Assignment 4.16 —

(a) $y = x^2 - 3$

(b) $y = (x-4)^2$

(c) $y = (x-3)^2 + 3$

(d) $y = (x-4)^2 - 2$

Assignment 4.17 —

(a) $x^2 + \frac{y^2}{2/3} = 1$: ellipse

(b) $y^2 = -\frac{3}{2}x$: parabola

(c) $-\frac{x^2}{3} + \frac{y^2}{3/2} = 1$: hyperbole

(d) $x^2 = \frac{4}{3}y$: parabola

(e) $\frac{x^2}{3/2} + \frac{y^2}{3/4} = 1$: ellipse

(f) $x^2 = -\frac{3}{2}y$: parabola

(g) $y^2 = 3x$: parabola

(h) $\frac{x^2}{3} - \frac{y^2}{3} = 1$: hyperbole

(i) $\frac{x^2}{3/2} - y^2 = 1$: hyperbole

Assignment 4.18 —

(a) $\frac{x^2}{5} + \frac{y^2}{9} = 1$

(b) $\frac{(x-2)^2}{16} + \frac{(y-1)^2}{12} = 1$

(c) $\frac{x^2}{12} + \frac{y^2}{4} = 1$

(d) $(x-2)^2 = -4(y-4)$

(e) $x^2 = -4y$

(f) $y^2 = -8x$

(g) $-\frac{x^2}{17/8} + \frac{y^2}{17} = 1$

(h) $-\frac{x^2}{3} + y^2 = 1$

(i) $\frac{x^2}{25/2} - \frac{(y-1)^2}{25/2} = 1$

Assignment 4.19 —

(a) $(x+1)^2+\left(y-\frac{3}{2}\right)^2=\frac{13}{4}$: circle

(b) $3\left(x+\frac{1}{3}\right)^2+3\left(y+\frac{7}{6}\right)^2=\frac{89}{12}$: circle

(c) $-\frac{(x+1/2)^2}{5/12}+\frac{(y-1)^2}{5/8}=1$: hyperbola

(d) $\frac{(x-1)^2}{2}+\frac{(y-2)^2}{4/3}=1$: ellipse

(e) $(y-1)^2=-\frac{3}{2}x$: parabola

(f) $\left(x-\frac{1}{4}\right)^2=\frac{1}{2}\left(y-\frac{55}{8}\right)$: parabola

(g) $\frac{x^2}{2/3}+\frac{(y-1)^2}{2/7}=1$: ellipse

(h) $-\frac{(x-1)^2}{1/2}+y^2=1$: hyperbola

(i) $-x^2+\frac{(y+2)^2}{4}=1$: hyperbola

(j) $\frac{(x-1)^2}{4}+\frac{(y+1)^2}{9}=1$: ellipse

Assignment 4.20 — $\frac{x^2}{225}-\frac{7y^2}{3600}=1$, with $x \geq 15$

Assignment 4.21 — $\frac{x^2}{14400}-\frac{13y^2}{921600}=1$, <ith $x \geq 120$

Assignment 4.22 —

(a) $x \in \left[\frac{1}{2},1\right] \cup \left[2,\frac{5}{2}\right]$

(b) $x \in \left]-\infty,-2\right[\cup \left]\frac{4}{3},+\infty\right[$

(c) $x \in \mathbb{R}$

(d) $x \in \left]-\infty,1\right[\cup \left]\frac{7}{3},+\infty\right[$

(e) $x \in \left]-\infty,\frac{5-\sqrt{73}}{6}\right] \cup \left[\frac{5+\sqrt{73}}{6},+\infty\right[$

(f) $x \in \left]-3\sqrt{2},-\sqrt{11}\right] \cup \left[-\sqrt{7},0\right[\cup \left]0,\sqrt{7}\right] \cup \left[\sqrt{11},3\sqrt{2}\right[$

(g) $x \in \left[-2-\sqrt{7},\sqrt{7}-2\right] \cup [1,3]$

(h) $x \in [-6,-3] \cup [-2,+\infty[$

(i) $x \in \left]-\infty,1\right[\cup \left]2,\frac{3+\sqrt{17}}{2}\right[$

(j) $x \in \left]-1-\sqrt{2},-1\right[\cup \left]0,1\right[\cup \left]1,+\infty\right[$

Chapter 5

Assignment 5.1 —

(a) $\frac{-3}{2}$

(b) $\frac{3}{2}$

(c) 27

(d) $-2x$

(e) $1+\log_x(2)$

(f) 2

(g) -2

(h) $\ln\left(x^2(x-2)^5\right)$

Assignment 5.2 — Self demonstration.

Assignment 5.3 —

(a) $x = -\dfrac{1}{3}$

(b) $x = 2$

(c) $x = \dfrac{1}{2}$

(d) $x = 81$

(e) $x = -\log_4\left(\dfrac{512}{5}\right)$

(f) $x = 1$ of $x = \frac{1}{e}$

(g) $x = \dfrac{15}{22}$

(h) $x = 3$ of $x = 3 + \dfrac{\ln(2)}{\ln(5)}$

(i) $\varnothing$

(j) $x = -2$

(k) $x = 1 \vee x = 2$

(l) $x = 0 \vee x = 1$

(m) $x = \dfrac{1}{81} \vee x = 729$

(n) $x = 4$

(o) $x = 16$

(p) $x = \dfrac{1}{3}$

(q) $x = -1$

(r) $x = -2 \vee x = 1$

(s) $x = \log_2(7)$

(t) $x = 3$

(u) $x = 6$

Assignment 5.4 —

(a) $\sqrt{2} < x \leq 2$

(b) $x > -1$

(c) $0 < x < 2 \vee 8 < x < 10$

(d) $1 < x < \dfrac{9}{5}$

(e) $x > e$

(f) $-\sqrt{3} < x < \sqrt{3}$

Assignment 5.5 —

(a) $x = 1, y = 2$

(b) $x = -1, y = 1/20$

(c) $x = 5, y = 20$

Assignment 5.6 —

(a)
- $\text{dom} f =]-2,-1[\cup]1,+\infty[$,
- intersections(s) x-axis: $x = \dfrac{1 \pm \sqrt{13}}{2}$,
- intersections(s) y-axis: none.

(b)
- $\text{dom} f =]5,+\infty[$,
- intersections(s) x-axis: $x \approx 5$,
- intersections(s) y-axis: none.

(c)
- $\text{dom} f =]13,+\infty[$,
- intersections(s) x-axis: $x = 20$,
- intersections(s) y-axis: none.

(d)
- $\text{dom} f = \varnothing$,
- intersections(s) x-axis: none,
- intersections(s) y-axis: none.

(e)
- $\text{dom} f =]-1,1[$,
- intersections(s) x-axis: $x = 0$,
- intersections(s) y-axis: $y = 0$.

(f)
- $\text{dom} f =]0,1[$,
- intersections(s) x-axis: $x = 0.2$,
- intersections(s)) y-axis: none.

Assignment 5.7 — The culture will contain 1012 cells after another 12 hours.

Assignment 5.8 — It takes 182 years to reduce radioactivity by 10%.

Assignment 5.9 —

(a) After one week there will be approximately 390 million bacteria.

(b) After 125 hours there will be 5 million bacteria.

Assignment 5.10 — After 5 minutes the thermometer will indicate 22.35°C.

Assignment 5.11 — It takes 92.88 minutes for the object to cool to 0°C.

Assignment 5.12 —

(a) Prove yourself.

(b) $\left(f^{-1} \circ f\right)(x) = x \quad \Leftrightarrow \quad x \in \text{dom} f \quad \Rightarrow \quad x \in \mathbb{R}_0^+ \setminus \{10\}$

$\left(f \circ f^{-1}\right)(x) = x \quad \Leftrightarrow \quad x \in \text{dom} f^{-1} \quad \Rightarrow \quad x \in \mathbb{R} \setminus \{-1\}$

(c) $\text{im} f = \text{dom} f^{-1} \quad \Rightarrow \quad \text{im} f = \mathbb{R} \setminus \{-1\}$

(d) Prove yourself.

Assignment 5.13 —

(a) $-\dfrac{\sqrt{2}}{2}$

(b) 1

(c) $\dfrac{\sqrt{3}}{2}$

(d) $\dfrac{\sqrt{2}}{2}\left(\dfrac{\sqrt{3}}{2} + \dfrac{1}{2}\right)$

(e) $\dfrac{\sqrt{2}}{2}\left(\dfrac{\sqrt{3}}{2} - \dfrac{1}{2}\right)$

(f) $\dfrac{\sqrt{2}}{2}\left(\dfrac{\sqrt{3}}{2} - \dfrac{1}{2}\right)$

(g) $2 + \sqrt{3}$

(h) $-\sqrt{2}\left(1 - \sqrt{3}\right)$

(i) $\dfrac{-2\sqrt{3}}{3}$

(j) $\sqrt{3}$

Assignment 5.14 —

(a) $\cos(\theta)=-\frac{4}{5}, \quad \csc(\theta)=\frac{5}{3}, \quad \sec(\theta)=-\frac{5}{4}, \quad \tan(\theta)=-\frac{3}{4}, \quad \cot(\theta)=-\frac{4}{3}$

(b) $\cos(\theta)=\frac{\sqrt{5}}{5}, \quad \sin(\theta)=\frac{2\sqrt{5}}{5}, \quad \csc(\theta)=\frac{\sqrt{5}}{2}, \quad \sec(\theta)=\sqrt{5}, \quad \cot(\theta)=\frac{1}{2}$

(c) $\cos(\theta)=\frac{1}{3}, \quad \sin(\theta)=-\frac{2\sqrt{2}}{3}, \quad \csc(\theta)=-\frac{3\sqrt{2}}{4}, \quad \tan(\theta)=-2\sqrt{2}, \quad \cot(\theta)=-\frac{\sqrt{2}}{4}$

(d) $\sin(\theta)=\frac{12}{13}, \quad \csc(\theta)=\frac{13}{12}, \quad \sec(\theta)=-\frac{13}{5}, \quad \tan(\theta)=-\frac{12}{5}, \quad \cot(\theta)=-\frac{5}{12}$

(e) $\cos(\theta)=-\frac{\sqrt{3}}{2}, \quad \sin(\theta)=-\frac{1}{2}, \quad \sec(\theta)=-\frac{2\sqrt{3}}{3}, \quad \tan(\theta)=\frac{\sqrt{3}}{3}, \quad \cot(\theta)=\sqrt{3}$

(f) $\cos(\theta)=-\frac{2\sqrt{5}}{5}, \quad \sin(\theta)=-\frac{\sqrt{5}}{5}, \quad \csc(\theta)=-\sqrt{5}, \quad \sec(\theta)=-\frac{\sqrt{5}}{2}, \quad \cot(\theta)=2$

Assignment 5.15 — Prove yourself.

Assignment 5.16 —

(a) $f(x)=3\sin\left(\frac{x}{2\pi}\right)$

- Period: $4\pi^2$
- Amplitude: 3
- Phase shift: none
- Vertical shift: none
- $\operatorname{dom} f=\mathbb{R}$
- $\operatorname{im} f=[-3,3]$
- zeros: $x=2\pi^2 k,\ k\in\mathbb{Z}$

(b) $f(x)=\frac{2}{3}\sin\left(\frac{2}{3}\left(x-\frac{\pi}{4}\right)\right)-11$

- Period: 3π
- Amplitude: $\frac{2}{3}$
- Phase shift: $\frac{\pi}{4}$
- Vertical shift: -11
- $\operatorname{dom} f=\mathbb{R}$
- $\operatorname{im} f=\left[-\frac{35}{3},-\frac{31}{3}\right]$
- zeros: none

(c) $f(x)=\sin\left(10\pi\left(x+\frac{1}{2}\right)\right)+3$

- Period: $\frac{1}{5}$
- Amplitude: 1
- Phase shift: -$\frac{1}{2}$
- Vertical shift: 3
- $\operatorname{dom} f=\mathbb{R}$
- $\operatorname{im} f=[2,4]$
- zeros: none

(d) $f(x)=2\sin(3x-2)+1$

- Period: $\frac{2\pi}{3}$
- Amplitude: 2
- Phase shift: $\frac{2}{3}$
- Vertical shift: 1
- $\operatorname{dom} f=\mathbb{R}$
- $\operatorname{im} f=[-1,3]$
- zeros: $x=\frac{2}{3}-\frac{\pi}{18}(12k+5),\ k\in\mathbb{Z}$,
  $x=\frac{2}{3}-\frac{\pi}{18}(12k+1),\ k\in\mathbb{Z}$

Assignment 5.17 —

(a) $x = \frac{\pi}{8} + \frac{k\pi}{2} \quad \vee \quad x = \frac{\pi}{20} + \frac{k\pi}{5} \qquad (k \in \mathbb{Z})$

(b) $x = -\frac{\pi}{2} + k\pi \quad \vee \quad x = \frac{3\pi}{2} + k2\pi \qquad (k \in \mathbb{Z})$

(c) $x = \frac{\pi}{2} + k\pi \quad \vee \quad x = \arctan\left(\frac{1}{3}\right) + k\pi \qquad (k \in \mathbb{Z})$

(d) $x = \arcsin\left(\frac{1}{3}\right) + k2\pi \quad \vee \quad x = \pi - \arcsin\left(\frac{1}{3}\right) + k2\pi \qquad (k \in \mathbb{Z})$

(e) $x = -\frac{\pi}{3} + k2\pi \quad \vee \quad x = \frac{\pi}{3} + k2\pi \qquad (k \in \mathbb{Z})$

(f) $x = k\pi \quad \vee \quad x = \arctan\left(\pm\sqrt{\frac{3}{5}}\right) + k\pi \qquad (k \in \mathbb{Z})$

(g) $x = \pm\frac{\pi}{4} + k\pi \quad \vee \quad x = \pm\frac{\pi}{3} + k2\pi \qquad (k \in \mathbb{Z})$

(h) $x = -\frac{5\pi}{6} + k\pi \quad \vee \quad x = -\frac{\pi}{6} + k\pi \qquad (k \in \mathbb{Z})$

(i) $x = \pm\frac{\pi}{4} + k\pi \quad \vee \quad x = \pm\frac{4\pi}{3} + k\pi \qquad (k \in \mathbb{Z})$

(j) $x = \frac{3\pi}{4} + k\pi \quad \vee \quad x = \frac{\pi}{4} + k\pi \qquad (k \in \mathbb{Z})$

Assignment 5.18 —

(a) $-\frac{\pi}{4} + k\frac{\pi}{2} < x < 0,1608\ldots + k\frac{\pi}{2} \qquad (k \in \mathbb{Z})$

(b) $\frac{3\pi}{4} + k\pi < x < \frac{13\pi}{12} + k\pi \qquad (k \in \mathbb{Z})$

(c) $-\frac{14\pi}{45} + k\frac{2\pi}{3} < x < -\frac{4\pi}{45} + k\frac{2\pi}{3} \qquad (k \in \mathbb{Z})$

(d) $\frac{\pi}{6} + k\frac{\pi}{2} < x < \frac{7\pi}{12} + k\frac{\pi}{2} \qquad (k \in \mathbb{Z})$

(e) $\frac{5\pi}{12} + k\pi < x < \frac{\pi}{2} + k\pi \quad \vee \quad \frac{\pi}{2} + k\pi < x < \frac{7\pi}{12} + k\pi \qquad (k \in \mathbb{Z})$

(f) $-\frac{\pi}{6} + 2k\pi < x < \arcsin\left(\frac{-1+\sqrt{3}}{2}\right) + 2k\pi$

$\vee \quad \pi - \arcsin\left(\frac{-1+\sqrt{3}}{2}\right) + 2k\pi < x < \frac{7\pi}{6} + 2k\pi$

$\vee \quad \frac{\pi}{6} + 2k\pi < x < \frac{5\pi}{6} + 2k\pi \qquad (k \in \mathbb{Z})$

Assignment 5.19 —

(a) $\mathbb{R}$

(b) $]-\infty,-1] \cup [1,+\infty[$

(c) $\left[-2,-\sqrt{2}\right] \cup \left[\sqrt{2},2\right]$

(d) $\left[-\frac{1}{2},\frac{1}{2}\right]$

(e) $\left[\frac{\sqrt{2}}{4},\frac{1}{2}\right]$

Assignment 5.20 —

(a) $-\frac{\pi}{4}$

(b) $\frac{\pi}{6}$

(c) $\frac{\pi}{6}$

(d) $\frac{1}{2}$

(e) -1

(f) $\frac{\sqrt{5}}{3}$

(g) $\frac{3}{4}$

(h) $\frac{56\sqrt{2}}{17}$

(i) 1

(j) $\frac{\sqrt{2}}{2}$

(k) -1

(l) $\frac{17}{25}$

(m) $\frac{3\pi}{4}$

Assignment 5.21 —

(a) $\tan(\arccos x) = \dfrac{\sqrt{1-x^2}}{x}$, with domain = $[-1,0[\cup]0,1]$

(b) $\cot(\arcsin x) = \dfrac{\sqrt{1-x^2}}{x}$, with domain = $[-1,0[\cup]0,1]$

(c) $\sin(2\arctan x) = \dfrac{2x}{x^2+1}$, with domain = $\mathbb{R}$

(d) $\cos(2\operatorname{arccot} x) = \dfrac{x^2-1}{x^2+1}$, with domain = $\mathbb{R}$

(e) $\cos(2\arcsin x) = 1-2x^2$, with domain = $[-1,1]$

Assignment 5.22 — Prove yourself

Assignment 5.23 — Prove yourself

Assignment 5.24 — Prove yourself

Assignment 5.25 — Use the definition of the hyperbolic cosine (5.29) and the hyperbolic sine (5.30).

Assignment 5.26 — $\dfrac{\cosh(\ln(x))+\sinh(\ln(x))}{\cosh(\ln(x))-\sinh(\ln(x))} = x^2$

Assignment 5.27 — Prove yourself

Assignment 5.28 — Prove yourself

Assignment 5.29 — The distance between the suspension points is $160\ln(2)$ m.

Assignment 5.30 —

(a) $a=0,\ b=0,\ c=9 \quad \Rightarrow b^2-4ac=0 \quad \Rightarrow$ parabole

$9\left(y+\frac{2}{3}\right)^2=0$ (falls apart into 2 parallel lines after square root).

(b) $a=3,\ b=0,\ c=-1 \quad \Rightarrow b^2-4ac=12>0 \quad \Rightarrow$ hyperbole.

$$\frac{(x-2)^2}{5/3}-\frac{(y+1)^2}{5}=1$$

Chapter 6

Assignment 6.1 —

(a) $\vec{\boldsymbol{a}}=(3,-2) \quad \Rightarrow \quad \|\vec{\boldsymbol{a}}\|=\sqrt{13}$ and $\theta=\arctan\left(-\frac{2}{3}\right)$

(b) $\vec{\boldsymbol{b}}=\left(\pi,-\frac{4}{3}\right) \quad \Rightarrow \quad \left\|\vec{\boldsymbol{b}}\right\|=\frac{4}{3}$ and $\theta=\pi$

(c) $\vec{\boldsymbol{c}}=(0,2) \quad \Rightarrow \quad \|\vec{\boldsymbol{c}}\|=2$ and $\theta=\frac{\pi}{2}$

(d) $\vec{\boldsymbol{d}}=(1,-1) \quad \Rightarrow \quad \left\|\vec{\boldsymbol{d}}\right\|=\sqrt{2}$ and $\theta=-\frac{\pi}{4}$

Assignment 6.2 —

(a) $\vec{\boldsymbol{a}}+\vec{\boldsymbol{x}}+\vec{\boldsymbol{b}}=\vec{\boldsymbol{o}} \quad \Leftrightarrow \quad \vec{\boldsymbol{x}}=-\vec{\boldsymbol{a}}-\vec{\boldsymbol{b}} \quad \Rightarrow \quad \vec{\boldsymbol{x}}=(-4,-6)$

(b) $\vec{\boldsymbol{a}}-\vec{\boldsymbol{b}}=2\vec{\boldsymbol{b}}+\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}} \quad \Leftrightarrow \quad \vec{\boldsymbol{x}}=2\vec{\boldsymbol{a}}-3\vec{\boldsymbol{b}} \quad \Rightarrow \quad \vec{\boldsymbol{x}}=(3,-8)$

(c) $3(\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}})=\vec{\boldsymbol{x}}-\vec{\boldsymbol{b}} \quad \Leftrightarrow \quad \vec{\boldsymbol{x}}=\frac{3\vec{\boldsymbol{a}}-\vec{\boldsymbol{b}}}{2} \quad \Rightarrow \quad \vec{\boldsymbol{x}}=(4,1)$

(d) $2(\vec{\boldsymbol{x}}-\vec{\boldsymbol{a}})=3\big(\vec{\boldsymbol{x}}-\vec{\boldsymbol{b}}\big) \quad \Leftrightarrow \quad \vec{\boldsymbol{x}}=-2\vec{\boldsymbol{a}}+3\vec{\boldsymbol{b}} \quad \Rightarrow \quad \vec{\boldsymbol{x}}=(-3,8)$

Assignment 6.3 — Write $\overrightarrow{AX}$ as $\overrightarrow{AM}+\overrightarrow{MX}$ and $\overrightarrow{BX}$ als $\overrightarrow{BM}+\overrightarrow{MX}$.
M at the middle of $[AB]$: $\overrightarrow{AM}=\overrightarrow{MB}=-\overrightarrow{BM}$ and $\left(\overrightarrow{AM}\right)^2=\left(\overrightarrow{BM}\right)^2$

(a) $\left(\overrightarrow{AX}\right)^2+\left(\overrightarrow{BX}\right)^2=\left(\overrightarrow{AM}\right)^2+2\overrightarrow{AM}\cdot\overrightarrow{MX}+\left(\overrightarrow{MX}\right)^2+\left(\overrightarrow{BM}\right)^2+2\overrightarrow{BM}\cdot\overrightarrow{MX}+\left(\overrightarrow{MX}\right)^2$
$=2\left(\overrightarrow{BM}\right)^2+2\left(\overrightarrow{MX}\right)^2+2\overrightarrow{BM}\cdot\overrightarrow{MX}-2\overrightarrow{BM}\cdot\overrightarrow{MX}=2\left(\overrightarrow{MX}\right)^2+2\left(\overrightarrow{MB}\right)^2$

(b) $\left(\overrightarrow{AX}\right)^2-\left(\overrightarrow{BX}\right)^2=\left(\overrightarrow{AM}\right)^2+2\overrightarrow{AM}\cdot\overrightarrow{MX}+\left(\overrightarrow{MX}\right)^2+\left(\overrightarrow{BM}\right)^2+2\overrightarrow{BM}\cdot\overrightarrow{MX}+\left(\overrightarrow{MX}\right)^2$
$=2\left(\overrightarrow{AM}-\overrightarrow{BM}\right)\cdot\overrightarrow{MX}=2\left(\overrightarrow{AM}+\overrightarrow{MB}\right)\cdot\overrightarrow{MX}=2\overrightarrow{AB}\cdot\overrightarrow{MX}$

Assignment 6.4 — $d(A,B) = d(A,C) = 5$

Assignment 6.5 — $d(A,B) = d(A,C) = d(B,C) = 2$

Assignment 6.6 — The wind is coming from the southwest and its speed is 502 km/hr.

Assignment 6.7 —

(a) $\overrightarrow{AB} = 3\hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}}$

(b) $\overrightarrow{BA} = -3\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}}$

(c) $\overrightarrow{AC} = 2\hat{\boldsymbol{\imath}} - 5\hat{\boldsymbol{\jmath}}$

(d) $\overrightarrow{AB} - \overrightarrow{BC} = 4\hat{\boldsymbol{\imath}} + \hat{\boldsymbol{\jmath}}$

(e) $\overrightarrow{AC} - 2\overrightarrow{AB} + 3\overrightarrow{CD} = -7\hat{\boldsymbol{\imath}} + 20\hat{\boldsymbol{\jmath}}$

(f) $\dfrac{\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{AD}}{3} = 2\hat{\boldsymbol{\imath}} - \dfrac{5}{3}\hat{\boldsymbol{\jmath}}$

Assignment 6.8 — $x = -4$

Assignment 6.9 — $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} = 4$ and $\cos\left(\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}}\right) = \dfrac{4}{5}$

Assignment 6.10 — $\left(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}\right) \vec{\boldsymbol{c}} = (a_1b_1c_1 + a_2b_2c_1, a_1b_1c_2 + a_2b_2c_2)$
$\vec{\boldsymbol{a}} \left(\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{c}}\right) = (a_1b_1c_1 + a_1b_2c_2, a_2b_1c_1 + a_2b_2c_2)$

Assignment 6.11 — Write out the left-hand member using the properties/definitions below.

- $(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}})^2 = \|\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}\|^2 = (\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}) \cdot (\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}})$
- $\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{y}} = \vec{\boldsymbol{y}} \cdot \vec{\boldsymbol{x}} = \|\vec{\boldsymbol{x}}\|\|\vec{\boldsymbol{y}}\| \cos(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}})$

Assignment 6.12 — Use the definition of the dot product: $\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{y}} = \vec{\boldsymbol{y}} \cdot \vec{\boldsymbol{x}} = \|\vec{\boldsymbol{x}}\|\|\vec{\boldsymbol{y}}\| \cos(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}})$.

Assignment 6.13 — Assume equality for equivalence, square both members, and apply the same properties as in Exercise 6.11.

Assignment 6.14 — Vertices of a square:

- $\overrightarrow{AB} \perp \overrightarrow{BC} \quad \Leftrightarrow \quad \overrightarrow{AB} \cdot \overrightarrow{BC} = 0$
- $d(A,B) = d(B,C) = \sqrt{17}$

The fourth vertex is $(-2,-2)$.

Assignment 6.15 —

(a) $\vec{\boldsymbol{v}}(-1,2,3) \perp \vec{\boldsymbol{w}}(1,2,h) \Leftrightarrow h = -1$

(b) $\vec{\boldsymbol{a}}(\sqrt{3},h,8) \perp \vec{\boldsymbol{b}}(h,-4,2) \Leftrightarrow h = \dfrac{-16}{\sqrt{3}-4}$

Assignment 6.16 —

(a) Prove yourself

(b) Prove yourself

Assignment 6.17 — Use the following properties/definitions.

- $(\vec{\boldsymbol{x}}+\vec{\boldsymbol{y}})^2 = \|\vec{\boldsymbol{x}}+\vec{\boldsymbol{y}}\|^2 = (\vec{\boldsymbol{x}}+\vec{\boldsymbol{y}})\cdot(\vec{\boldsymbol{x}}+\vec{\boldsymbol{y}})$
- $\vec{\boldsymbol{x}}\cdot\vec{\boldsymbol{y}} = \vec{\boldsymbol{y}}\cdot\vec{\boldsymbol{x}} = \|\vec{\boldsymbol{x}}\|\|\vec{\boldsymbol{y}}\|\cos(\vec{\boldsymbol{x}},\vec{\boldsymbol{y}})$

Assignment 6.18 —

(a) $\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}} = (5,13,7)$

(b) $\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}} = (3,-2,1)$

Assignment 6.19 — $\vec{\boldsymbol{u}}\times(\vec{\boldsymbol{v}}\times\vec{\boldsymbol{w}}) = \vec{\boldsymbol{u}}\times(3\hat{\boldsymbol{\imath}}+2\hat{\boldsymbol{\jmath}}+2\hat{\boldsymbol{k}}) = -2\hat{\boldsymbol{\imath}}+7\hat{\boldsymbol{\jmath}}-4\hat{\boldsymbol{k}}$

$\rightarrow \vec{\boldsymbol{u}}\times(\vec{\boldsymbol{v}}\times\vec{\boldsymbol{w}})$ lies in the plane of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$

$(\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}})\times\vec{\boldsymbol{w}} = (9\hat{\boldsymbol{\imath}}+6\hat{\boldsymbol{\jmath}}+-7\hat{\boldsymbol{k}})\times\vec{\boldsymbol{w}} = \hat{\boldsymbol{\imath}}+9\hat{\boldsymbol{\jmath}}+9\hat{\boldsymbol{k}}$

$\rightarrow (\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}})\times\vec{\boldsymbol{w}}$ lies in the plane of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$

Assignment 6.20 — Calculate the cross product of $\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}} = \vec{\boldsymbol{0}}$ with $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

Assignment 6.21 —

(a) $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\imath}}-\hat{\boldsymbol{\jmath}}$ en $\vec{\boldsymbol{v}} = \hat{\boldsymbol{\jmath}}+2\hat{\boldsymbol{k}}$

a) $\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}} = \hat{\boldsymbol{\imath}}+2\hat{\boldsymbol{k}}$, $\vec{\boldsymbol{u}}-\vec{\boldsymbol{v}} = \hat{\boldsymbol{\imath}}-2\hat{\boldsymbol{\jmath}}-2\hat{\boldsymbol{k}}$, $2\vec{\boldsymbol{u}}-3\vec{\boldsymbol{v}} = 2\hat{\boldsymbol{\imath}}-5\hat{\boldsymbol{\jmath}}-6\hat{\boldsymbol{k}}$

b) $\|\vec{\boldsymbol{u}}\| = \sqrt{2}$, $\|\vec{\boldsymbol{v}}\| = \sqrt{5}$

c) $\hat{\boldsymbol{u}} = \dfrac{1}{\sqrt{2}}(\hat{\boldsymbol{\imath}}-\hat{\boldsymbol{\jmath}})$, $\hat{\boldsymbol{v}} = \dfrac{1}{\sqrt{5}}\left(\hat{\boldsymbol{\jmath}}+2\hat{\boldsymbol{k}}\right)$

d) $\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = -1$

e) $\theta = \arccos\left(\dfrac{-1}{\sqrt{10}}\right)$

(b) $\vec{\boldsymbol{u}} = 3\hat{\boldsymbol{\imath}}+4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}$ en $\vec{\boldsymbol{v}} = 3\hat{\boldsymbol{\imath}}-4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}$

a) $\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}} = 6\hat{\boldsymbol{\imath}}-10\hat{\boldsymbol{k}}$, $\vec{\boldsymbol{u}}-\vec{\boldsymbol{v}} = 8\hat{\boldsymbol{\jmath}}$, $2\vec{\boldsymbol{u}}-3\vec{\boldsymbol{v}} = -3\hat{\boldsymbol{\imath}}+20\hat{\boldsymbol{\jmath}}+5\hat{\boldsymbol{k}}$

b) $\|\vec{\boldsymbol{u}}\| = 5\sqrt{2}$, $\|\vec{\boldsymbol{v}}\| = 5\sqrt{2}$

c) $\hat{\boldsymbol{u}} = \dfrac{1}{5\sqrt{2}}\left(3\hat{\boldsymbol{\imath}}+4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}\right)$, $\hat{\boldsymbol{v}} = \dfrac{1}{5\sqrt{2}}\left(3\hat{\boldsymbol{\imath}}-4\hat{\boldsymbol{\jmath}}-5\hat{\boldsymbol{k}}\right)$

d) $\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = 18$

e) $\theta = \arccos\left(\dfrac{9}{25}\right)$

Chapter 7

Assignment 7.1 — Line ℓ through A and B: $\dfrac{x-1}{1}=\dfrac{y-2}{1}=\dfrac{z-3}{1}$.

$C(3,4,5)\in\ell:\quad 3-1=4-2=5-3 \quad\rightarrow$ ok

Assignment 7.2 — $\ell:\begin{cases} x=8/7-t\\ y=4/7+10t\\ z=7t\end{cases}$

Assignment 7.3 —

(a)
- Cartesian equation: $\dfrac{x}{1}=\dfrac{y}{2}=\dfrac{z}{3}$
- Vector equation: $\vec{\boldsymbol{l}}(t)=(0,0,0)+t(1,2,3)$
- Parameter equations: $\ell:\begin{cases} x=t\\ y=2t\\ z=3t\end{cases}$

(b)
- Cartesian equation: $\begin{cases} \dfrac{x-3}{-2}=\dfrac{z-1}{-1}\\ y=4\end{cases}$
- Vector equation: $\vec{\boldsymbol{l}}(t)=(3,4,1)+t(-2,0,-1)$
- Parameter equations: $\ell:\begin{cases} x=3-2t\\ y=4\\ z=1-t\end{cases}$

(c)
- Cartesian equation: $\dfrac{x-1}{3/2}=\dfrac{y-2}{2}=\dfrac{z}{1/2}$
- Vector equation: $\vec{\boldsymbol{l}}(t)=(1,2,0)+t(3/2,2,1/2)$
- Parameter equations: $\ell:\begin{cases} x=1+3/2t\\ y=2+2t\\ z=1/2t\end{cases}$

(d)
- Cartesian equation: $\dfrac{x-1}{2}=\dfrac{y-2}{-3}=\dfrac{z-3}{-4}$
- Vector equation: $\vec{\boldsymbol{l}}(t)=(1,2,3)+t(2,-3,-4)$
- Parameter equations: $\ell:\begin{cases} x=1+2t\\ y=2-3t\\ z=3-4t\end{cases}$

(e) • Cartesian equation: $\dfrac{x+1}{2} = \dfrac{y}{-1} = \dfrac{z-1}{7}$

• Vector equation: $\vec{\boldsymbol{l}}(t) = (-1,0,1) + t(2,-1,7)$

• Parameter equations: $\ell : \begin{cases} x = -1+2t \\ y = -t \\ z = 1+7t \end{cases}$

(f) • Cartesian equation: $\dfrac{x}{7} = \dfrac{y}{-6} = \dfrac{z}{-5}$

• Vector equation: $\vec{\boldsymbol{l}}(t) = (0,0,0) + t(7,-6,-5)$

• Parameter equations: $\begin{cases} x = 7t \\ y = -6t \\ z = -5t \end{cases}$

(g) • Cartesian equation: $\dfrac{x-2}{1} = \dfrac{y+1}{-1} = \dfrac{z+1}{-1}$

• Vector equation: $\vec{\boldsymbol{l}}(t) = (2,-1,-1) + t(1,-1,-1)$

• Parameter equations: $\begin{cases} x = 2+t \\ y = -1-t \\ z = -1-t \end{cases}$

Assignment 7.4 —

(a) $\vec{\boldsymbol{d}}_1 = (3,4,2)$, $\vec{\boldsymbol{d}}_2 = (6,8,4) \quad \Rightarrow \quad \vec{\boldsymbol{d}}_2 = 2\vec{\boldsymbol{d}}_1 \quad \Rightarrow \quad \ell_1 \parallel \ell_2$

(b) ℓ_1 and ℓ_2 are skew.

(c) $\vec{\boldsymbol{d}}_1 = (8,-7,-5)$, $\vec{\boldsymbol{d}}_2 = (8,-7,-5) \quad \Rightarrow \quad \vec{\boldsymbol{d}}_1 = \vec{\boldsymbol{d}}_2 \quad \Rightarrow \quad \ell_1 \parallel \ell_2$

(d) ℓ_1 and ℓ_2 are skew.

Assignment 7.5 —

(a) $3x + y - 4z + 1 = 0$

(b) $-2x + z + 1 = 0$

(c) $x - 2y + 3z + 12 = 0$

(d) $x + 6y - 4z + 4 = 0$

(e) $7x - 3y - 5z + 18 = 0$

(f) $x - 5y - 3z + 7 = 0$

(g) $5x - 9y + 5z + 15 = 0$

Assignment 7.6 —

(a) The planes p_1 and p_2 are perpendicular if and only if $\vec{\boldsymbol{n}}_{p_1} \cdot \vec{\boldsymbol{n}}_{p_2} = 0$.

(b) The planes p_1 and p_3 are parallel if and only if $\vec{\boldsymbol{n}}_{p_1}$ and $\vec{\boldsymbol{n}}_{p_3}$ are parallel.

(c) The planes p_2 and p_3 are perpendicular to each other.

Assignment 7.7 — The angle between the planes p_1 and p_2 is $\arccos\left(\frac{-4}{\sqrt{91}}\right)$.

Assignment 7.8 —

(a) $4x^2+9y^2+4z^2=36$

(b) $x=4(y^2+z^2)$

(c) $x^2+y^2 \pm 2z\sqrt{x^2+y^2}+2z^2=1$

Assignment 7.9 —

(a) $3x^2-2y^2+z^2+3=0$: 2-leaf hyperboloid.

(b) $-4x^2+2z^2-3=0$: hyperbolic cylinder parallel to the y-axis.

(c) $3x^2-y^2=z^2$: cone with top in the origin and located around the x-axis.

(d) $(x-1)^2+(y-2)^2=(z-4)^2$: cone with top in $(1,2,4)$.

(e) $2x^2+3z^2=1$: elliptical cylinder parallel to the y-axis.

(f) $y^2+2z^2=x$: elliptical paraboloid.

(g) $x^2+4y^2+9z^2=36$: ellipsoid.

(h) $\frac{25}{9}x^2-25y^2+z^2=25$: 1-leaf hyperboloid.

Assignment 7.10 — $\frac{x-2}{3}=\frac{y+1}{2}=\frac{z-5}{-2}$

Assignment 7.11 —

(a) The line ℓ lies in the plane p_1.

(b) The line ℓ is parallel to the plane p_2.

(c) The line ℓ intersects the plane p_3, but not perpendicularly in $(6,-25/2,3/2)$.

(d) The line ℓ intersects the plane p perpendicularly in $(1,1,1)$.

(e) The line ℓ is parallel to the plane p.

(f) The line ℓ lies in the plane p.

(g) The line ℓ intersects with the plane p, but not perpendicularly in $(3,2,-1)$.

Chapter 8

Assignment 8.1 — Prove with the (ε, δ)-definition.

Assignment 8.2 —

(a) Choose $\delta = \varepsilon$.

(b) Choose $\delta = \dfrac{\varepsilon}{10}$.

(c) Choose $\delta = \varepsilon$.

Assignment 8.3 — Prove similar to sum.
Hint:

$$\begin{aligned}
|f(x)g(x) - L_1L_2| &= |f(x)g(x) - L_1g(x) + L_1g(x) - L_1L_2| \\
&= |(f(x) - L_1)g(x) + L_1(g(x) - L_2)| \\
&\leq |(f(x) - L_1)g(x)| + |L_1(g(x) - L_2)| \\
&= |g(x)||f(x) - L_1| + |L_1||g(x) - L_2|
\end{aligned}$$

Make each term in the last line smaller than $\varepsilon/2$ by choosing x close enough to c.

Assignment 8.4 —

(a) 1

(b) 2

(c) 0

(d) $-\dfrac{1}{16}$

(e) -2

(f) $\dfrac{7}{12}$

(g) $\dfrac{7}{3}$

(h) -6

(i) 0

(j) $\dfrac{1}{4}$

(k) 8

(l) $\dfrac{8}{3}$

(m) 4

(n) $x \underset{>}{\to} 2 : +\infty \;\wedge\; x \underset{<}{\to} 2 : -\infty$

(o) $\dfrac{1}{3}$

(p) 3

Assignment 8.5 — $\lim\limits_{x\to 0}\left(x^2 \sin\left(\dfrac{1}{x}\right)\right) = 0$

Use the fact that $-1 \leq \sin\left(\dfrac{1}{x}\right) \leq 1$, multiply both parts with x^2 and take the limit for $x \to 0$.

Assignment 8.6 —

(a) $\lim\limits_{x\to a} f(x) = 0$

(b) $-3 \leq \lim\limits_{x\to a} f(x) \leq 3$

Assignment 8.7 —

(a) 1

(b) $+\infty$

(c) 1

(d) 2

(e) $-\infty$

(f) $+\infty$

(g) 2

(h) 0

(i) -1

(j) 0

(k) 1

Assignment 8.8 —

(a) -2

(b) 2

(c) π^2

(d) 1

Assignment 8.9 —

(a) $\lim\limits_{\substack{x \to 0 \\ <}} f(x^2 - x^4) = A$

(b) $\lim\limits_{\substack{x \to 0 \\ >}} f(x^2 - x^4) = A$

Assignment 8.10 — $f(x)$ is discontinuous in $x = 1, x = 2, x = 3, x = 4$ and $x = 5$. f is left continuous in $x = 4$ and right continuous in $x = 2$ and $x = 5$. $f(x)$ can not be redefined in $x = 1$ such that the function becomes continuous because $\lim\limits_{x\to 1} f(x) = +\infty$ does not exist ($+\infty \notin \mathbb{R}$).

Assignment 8.11 —

(a) $a = -1$

(b) $a = -4, \quad b = -2$

(c) $a = \dfrac{5}{3}, \quad b = \dfrac{2}{3}$

(d) $a = 2$

Assignment 8.12 —

(a) There is a problem with continuity in $x = -3$. We determine that $\lim\limits_{\substack{x \to -3 \\ <}} f(x) = +\infty$ and $\lim\limits_{\substack{x \to -3 \\ >}} f(x) = -\infty$. Therefore we consider an essential discontinuity in $x = -3$. However, $f(-3)$ is not defined, so it can't be an essential discontinuity, we can talk about a continuity in $\mathbb{R} \setminus \{-3\}$. The discontinuity arrives at $x = -3$, but this point is not a part of the domain.

(b) There is a problem with continuity in $x = -2$. We determine that $\lim\limits_{\substack{x \to -2 \\ <}} f(x) = \lim\limits_{\substack{x \to -2 \\ >}} f(x) = 1 \neq f(-2)$. $f(-2)$ is not defined. This is a removable discontinuity.

Assignment 8.13 —

(a) yes, between $0,2$ en $0,25$

(b) yes, slightely smaller than -8

(c) no

(d) yes, between $\dfrac{\pi}{2}$ and $\dfrac{3\pi}{4}$

(e) yes, one between -4 en -3, one between -3 and 1 and one between 1 and 4

Assignment 8.14 — Use the consequence of the intermediate value theorem: if f is continuous in $[a, b]$ and $f(a)f(b) < 0$, then there exists at least one point c in $]a, b[$ such that $f(c) = 0$.

Assignment 8.15 — $c = 3$

Assignment 8.16 —

(a) $+\infty : 0 \quad \wedge \quad -\infty : 0$

(b) $+\infty : 3 \quad \wedge \quad -\infty : 1$

(c) 2

(d) $+\infty : -1 \quad \wedge \quad -\infty : -1$

(e) $+\infty : 0 \quad \wedge \quad -\infty : +\infty$

(f) $+\infty : -\frac{5}{4} \quad \wedge \quad -\infty : -\infty$

(g) $-\frac{3}{5}$

(h) $+\infty : \frac{2}{\sqrt{3}} \quad \wedge \quad -\infty : -\frac{2}{\sqrt{3}}$

(i) $-\frac{2}{3}$

(j) $+\infty : +\infty \quad \wedge \quad -\infty : 2$

(k) $-\frac{\sqrt{2}}{4}$

(l) $+\infty : -1 \quad \wedge \quad -\infty : 0$

Assignment 8.17 —

	VA	HA	SA
(a)	$x = \frac{1}{2}$	$y = -\frac{1}{2}$	none
(b)	$x = 3$	none	$y = x + 2$
(c)	none	$y = 3$	none
(d)	none	none	$y = 2x$
(e)	$x = 0$	none	$y = x$
(f)	$x = -2$ $x = 0$ $x = 2$	none	$y = \frac{1}{3}x$
(g)	none	none	$y = x - 3\ (+\infty)$ $y = -x - 3\ (-\infty)$
(h)	none	none	$y = x + 1\ (+\infty)$ $y = -x - 1\ (-\infty)$
(i)	none	$y = 0\ (+\infty)$	$y = 2x\ (-\infty)$
(j)	$x = -2$ $x = 3$	none	$y = x + \frac{3}{2}\ (+\infty)$ $y = -x - \frac{3}{2}\ (-\infty)$

Assignment 8.18 —

(a) 0

(b) $-\infty$

(c) 0

(d) 3

(e) 1

(f) $\frac{2}{5}$

(g) $\frac{1}{2}$

(h) $\frac{49}{4}$

(i) 8

(j) $\frac{1}{2}$

Assignment 8.19 —

(a) $\sqrt{e}$
(b) e^3
(c) -1
(d) a
(e) e^a
(f) e^{-2}
(g) e^{-1}

Assignment 8.20 —

(a) $f(x)$ is continuous over $\mathbb{R}$.

(b) $f(x)$ is continuous over $]-\infty,-3[$, in $]-3,3[$ and in $]3,+\infty[$.

(c) $f(x)$ is continuous over $\mathbb{R}$.

(d) $f(x)$ is continuous over $]-\infty,-1]$ and in $[1,+\infty[$.

(e) $f(x)$ is continuous over $\mathbb{R}$.

(f) $f(x)$ is continuous over $\mathbb{R}_0^-$ and in $\mathbb{R}_0^+$.

(g) $f(x)$ is continuous for $|x| > 2$ and for $|x| < 2$ and in $x=-2$. $f(x)$ is left continuously in $x=2$, but not right continuously.

(h) $f(x)$ is continuous in $\mathbb{R}_0$. $f(x)$ is right continuous in $x=0$.

Assignment 8.21 —

(a) perforation in $x=-3$

(b) 2 vertical asymptotes in $x=-3$ and $x=2$

(c) 1 vertical asymptote in $x=-1$ and 1 perforation in $x=3$

Chapter 9

Assignment 9.1 —

(a) $f(x)=|x|$ is continuous in $x=0$, but not differentiable because the left and right derivative are not equal.

(b) $f(x) = |x^2-1|$ is continuous in $x=1$, but not differentiable because the left and right derivative are not equal.

(c) $f(x) = |\sin(x)|$ is continuous in $x=0$, but not differentiable because the left and right derivative are not equal.

(d) $f(x)=\sqrt{x}$ is right-continuous in $x=0$, but not differentiable because the derivative does not exist at $x=0$. The function is differentiable over $\mathbb{R}_0^+$.

(e) $f(x)=\sqrt{1-x^2}$ is right-continuous in $x=-1$, but not differentiable because the derivative does not exist at $x=-1$.

Assignment 9.2 — f is differentiable over the intervals $[-1,0]$, $[0,1]$ and $[1,2]$ but it is not differentiable over $[-1,1]$ and $[-1,2]$.

Assignment 9.3 —

(a) tangent: $y=2$, normal: $x=0$

(b) tangent: $y=2x+1$, normal: $y=-\dfrac{1}{2}x-\dfrac{3}{2}$

(c) tangent: $y=\dfrac{\sqrt{2}}{2}x+\dfrac{\sqrt{2}}{2}\left(1-\dfrac{\pi}{4}\right)$, normal: $y=-\sqrt{2}\,x+\dfrac{\sqrt{2}}{2}\left(1+\dfrac{\pi}{2}\right)$

(d) tangent: $y=2x+1-\dfrac{\pi}{2}$, normal: $y=-\dfrac{1}{2}x+1+\dfrac{\pi}{8}$

(e) tangent: $y=2x+2$, normal: $y=-\dfrac{1}{2}x+2$

(f) tangent: $y=4$, normal: $x=2$

Assignment 9.4 —

(a) $0^+ : \; y=x \quad \wedge \quad 0^- : \; y=-x$

(b) $0^+ : \; x=0 \quad \wedge \quad 0^- : \; x=0$

(c) $y=0$

(d) $y=0$

Assignment 9.5 — $y=6x-9$ and $y=-2x-1$

Assignment 9.6 — $N'(t)=N(t)\,\ln(2)=N(0)\,2^t\,\ln(2)$

Assignment 9.7 —

(a) $f^{(n)}(x)=\dfrac{(-1)^n\,n!}{x^{n+1}}$

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

(c) $f^{(n)}(x)=\dfrac{(-1)^n\,(2n)!}{n!\,2^{2n}\,x^{n+\frac{1}{2}}}$

(d) $f^{(n)}(x)=na^{n-1}e^{ax}+xa^n e^{ax}$

Assignment 9.8 —

(a) $f'(x)=5x^4-16x^3+6x-6$

(b) $f'(x)=\sqrt{x^2-1}+\dfrac{x^2}{\sqrt{x^2-1}}$

(c) $f'(x)=-\dfrac{7}{(2x-1)^2}$

(d) $f'(x)=\dfrac{1}{(x+1)^2}\sqrt{\dfrac{x+1}{x-1}}$

(e) $f'(x)=-\dfrac{2(14x+3)}{(7x^2+3x-6)^3}$

(f) $f'(x)=\dfrac{4}{3\sqrt[3]{2x+3}}$

(g) $f'(x)=\dfrac{4x+3}{3\sqrt[3]{(x+1)^2}}$

(h) $f'(x)=\dfrac{1+6x}{(1+2x+6x^2)^2}$

(i) $f'(x)=\dfrac{2(\sqrt{x-2}+\sqrt{3})^3}{9\sqrt{x-2}}$

Assignment 9.9 —

(a) $f'(x) = \sin^2(x) + x\sin(2x)$

(b) $f'(x) = x^2(3\cos(x) - x\sin(x))$

(c) $f'(x) = x^4\left(e^x(x+5) + \left(x\tan(x)+5\right)\sec(x)\right)$

(d) $f'(x) = \dfrac{-2e^{-2x}}{x^3}(x+1)$

(e) $f'(x) = \dfrac{-e^x}{e^{2x}-1}$

(f) $f'(x) = -\dfrac{e^{-\arcsin(x)}}{\sqrt{1-x^2}}$

(g) $f'(x) = \sec^2(e^x)e^x$

(h) $f'(x) = \dfrac{x}{\ln(10)\left(x^2-9\right)}$

(i) $f'(x) = \dfrac{4(2x+1)}{\ln(2)\left(x^2+x+2\right)}$

(j) $f'(x) = \dfrac{2}{\ln(3)\left(4x^2-1\right)}$

(k) $f'(x) = \dfrac{\pm 1}{(x-1)\sqrt{1-2x}}$

(l) $f'(x) = \dfrac{1}{2\sqrt{x}\left(1+\cos\left(\sqrt{x}\right)\right)}$

(m) $f'(x) = -\dfrac{1}{\sqrt{a^2-(x-b)^2}}$

(n) $f'(x) = \dfrac{x}{\sqrt{1-x^4}\sqrt{\arcsin(x^2)}}$

(o) $f'(x) = \sqrt{\dfrac{a-x}{a+x}}$

Assignment 9.10 —

(a) $f'(x) = \dfrac{1}{\cos(x)}$

(b) $f'(x) = \dfrac{1}{\cos^3(x)}$

(c) $f'(x) = \dfrac{\sqrt{x^2-a^2}}{x^2}$

(d) $f'(x) = \left(1-x^2\right)^{\frac{3}{2}}$

(e) $f'(x) = \dfrac{1}{x\sqrt{x+1}}$

(f) $f'(x) = \dfrac{x}{(ax+b)^2}$

(g) $f'(x) = 8x^2\sqrt{a^2-x^2}$

(h) $f'(x) = \left(3-2x-x^2\right)^{\frac{3}{2}}$

Assignment 9.11 —

(a) $f'(x) = \tanh^2(x)$

(b) $f'(x) = \tanh(x) + x\left(1-\tanh^2(x)\right)$

(c) $f'(x) = 3^{\cosh(x)}\sinh(x)\ln 3$

(d) $f'(x) = \tanh(x) - \coth(x)$

(e) $f'(x) = 1$

(f) $f'(x) = \dfrac{x^2}{(x\cosh(x)-\sinh(x))^2}$

(g) $f'(x) = \dfrac{x\cosh(\cosh(x))\sinh(x) - 2\sinh(\cosh(x))}{x^3}$

(h) $f'(x) = -\dfrac{1}{x\sqrt{1-x^2}}$

(i) $f'(x) = \left(\dfrac{x}{1-x^2}+1\right)e^{\operatorname{artanh}(x)}$

(j) $f'(x) = \dfrac{2x}{\sqrt{x^4+1}}$

(k) $f'(x) = \dfrac{1}{\sqrt{1+x^2}}$

(l) $f'(x) = -\dfrac{\sqrt{2}}{|1+x|\sqrt{1+x^2}}$

Assignment 9.12 —

(a) $\dfrac{d}{dx}\left(f\left(\dfrac{2}{x}\right)\right)^3 = \dfrac{-6}{x^2}f'\left(\dfrac{2}{x}\right)\left(f\left(\dfrac{2}{x}\right)\right)^2$

(b) $\dfrac{d}{dx}f\big(2f(3f(x))\big) = 6f'\big(2f(3f(x))\big)f'(3f(x))f'(x)$

Assignment 9.13 —

(a) $y' = -\dfrac{2e^2x}{2^y\ln(2)}$

(b) $y' = -\dfrac{x}{y^2}$

(c) $y' = \dfrac{1-y}{2+x}$

(d) $y' = \dfrac{-3x^2y-y^5}{x^3+5xy^4}$

(e) $y' = \dfrac{x}{4(1-y)}$

(f) $y' = -\dfrac{3x^2+2xy}{x^2+4y}$

(g) $y' = \dfrac{1-\cos(x)}{1+\sin(y)}$

(h) $y' = -\dfrac{2x+y}{2y+x}$

Assignment 9.14 —

(a) $y = x$

(b) $y = -x+2$

(c) $y = -x+1$

(d) $y = -x/e^2+2/e$

Assignment 9.15 — $y' = \dfrac{-x}{4y}, \quad y'' = -\dfrac{1}{4y^3}$

Assignment 9.16 —

(a) $y' = \dfrac{x^x\big(x+(x+1)\ln(x)\big)}{(x+1)^2}$

(b) $y' = x^{\sin(x)+2}\left(\cos(x)\ln(x)+\dfrac{\sin(x)+2}{x}\right)$

(c) $y' = (\sin(x))^{\ln(x)}\left(\dfrac{\ln(\sin(x))}{x}+\ln(x)\cot(x)\right)$

(d) $y' = -\frac{2\ln(x)}{x}\left(\frac{1}{x}\right)^{\ln(x)}$

(e) $y' = (\cos x)^x(\ln(\cos(x)) - x\tan(x)) - x^{\cos(x)}\left(-\sin(x)\ln(x) + \frac{1}{x}\cos(x)\right)$

(f) $y' = \frac{x^{\ln(x)}(\sin(x))^x}{x^x\ln(x)}\left(\frac{2\ln(x)}{x} + \ln(\sin(x)) + x\cot(x) - \ln(x) - 1 - \frac{1}{x\ln(x)}\right)$

(g) $y' = x^{x^2+1}\left(2\ln(x)+1\right)$

(h) $y' = x^{x^x+x}\left(\ln^2(x) + \ln(x) + \frac{1}{x}\right)$

Assignment 9.17 — $(f^{-1})'(x) = 1 + f^{-1}(x)$

Assignment 9.18 —

(a) $(f^{-1})'(x) = \frac{1}{6y^2} = \frac{1}{6(f^{-1}(x))^2}$

(b) $(f^{-1})'(-2) = \frac{1}{f'(-1)} = \frac{2}{5}$

(c) $(f^{-1})'(2) = \frac{1}{f'(1)} = \frac{1}{4}$

(d) $(f^{-1})'(10) = \frac{1}{f'(2)} = \frac{1}{13}$

(e) $(f^{-1})'(\sqrt{3}/2) = \frac{1}{f'(\pi/6)} = 1$

(f) $(f^{-1})'(6) = \frac{1}{f'(0)} = \frac{1}{18}$

Assignment 9.19 —

(a) $+\infty$

(b) $\frac{3}{5}$

(c) 0

(d) $\frac{1}{2}$

(e) 1

(f) 0

(g) $-\frac{1}{2}$

(h) 1

(i) e^3

(j) 1

(k) 1

(l) $e^{-\frac{3}{2}}$

(m) $\frac{1}{\sqrt{e}}$

(n) 1

(o) $-\frac{1}{\pi}$

(p) 1

(q) $\frac{1}{e^2}$

(r) 1

(s) 0

(t) 1

(u) 3

(v) -1

(w) -5

(x) -8

(y) 0

(z) e^3

Assignment 9.20 — $0.01\,\mathrm{cm}$

Assignment 9.21 — $1005\,\mathrm{cm}^3$

Chapter 10

Assignment 10.1 — $f'(x) = x^{(x^2)} x (2\ln(x)+1) = 0 \quad \Leftrightarrow \quad x = e^{-1/2}$

$f''(x) = x^{(x^2)} x^2 (2\ln(x)+1)^2 + x^{(x^2)}(2\ln(x)+3) \quad \Rightarrow f''\left(e^{-1/2}\right) = 2\left(e^{-1/2}\right)^{e^{-1}} > 0$

$\Rightarrow$ $x = e^{-1/2}$ is a minimum

Assignment 10.2 — $f'(v) = \dfrac{\sqrt{2/\pi}\, m^{3/2}\, v\, e^{-mv^2/(2kT)}\left(2kT - mv^2\right)}{(kT)^{5/2}}$. $f(v)$ is at a maximum if $v = \sqrt{\dfrac{2kT}{m}}$.

Assignment 10.3 —

(a) $c = 0$

(b) $c = 3/\sqrt{2}$

(c) The mean value theorem is not applicable.

Assignment 10.4 —

(a) $c = -1/2$

(b) Rolle's theorem is not applicable.

(c) Rolle's theorem is not applicable.

Assignment 10.5 —

(a) The function with parameter $k > 0$ has a horizontal asymptote $y = 0$ for $x \to \pm\infty$.
Derivatives: $f'(x) = e^{-kx^2}\left(1 - 2x^2 k\right)$ en $f''(x) = -2kxe^{-kx^2}\left(3 - 2x^2 k\right)$

(b) $k = \dfrac{1}{2}$

(c) Summary for: $f(x) = xe^{-\frac{x^2}{2}}$

x	$-\infty$		$-\sqrt{3}$		-1		0		1		$\sqrt{3}$		$+\infty$
$f'(x)$	$-$	$-$	$-$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	$-$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	(0)	$\searrow$	$f(-\sqrt{3})$	$\searrow$	$f(-1)$	$\nearrow$	0	$\nearrow$	$f(1)$	$\searrow$	$f(\sqrt{3})$	$\searrow$	(0)
	H.A.	$\cap$	inf.p.	$\cup$	min.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

Assignment 10.6 —

(a) Asymptotes: $y = x - 2$ and $y = -x + 2$

Derivatives: $f'(x) = \frac{x-2}{\sqrt{x^2-4x+3}}$ and $f''(x) = -\frac{1}{(x^2-4x+3)^{\frac{3}{2}}}$

x	$-\infty$		1		3		$+\infty$
$f'(x)$	$-$	$-$	|	///	|	$+$	$+$
$f''(x)$	$-$	$-$	|	///	|	$-$	$-$
$f(x)$	$(+\infty)$	$\searrow$	0	///	0	$\nearrow$	$(+\infty)$
	S.A.	$\cap$				$\cap$	S.A.

(b) Asymptotes: $x = -\sqrt{3}$, $x = \sqrt{3}$, $y = -1$ and $y = 1$

Derivatives: $f'(x) = -\frac{7x+3}{(x^2-3)^{\frac{3}{2}}}$ and $f''(x) = \frac{14x^2+9x+21}{(x^2-3)^{\frac{5}{2}}}$

x	$-\infty$		-7		$-\sqrt{3}$		$\sqrt{3}$		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	|	///	|	$-$	$-$
$f''(x)$	$+$	$+$	$+$	$+$	|	///	|	$+$	$+$
$f(x)$	(-1)	$\nearrow$	0	$\nearrow$	|	///	|	$\searrow$	(1)
	H.A.	$\cup$	$\cup$	$\cup$	V.A.		V.A.	$\cup$	H.A.

(c) Asymptotes: $y = 0$

Derivatives: $f'(x) = -xe^{-\frac{x^2}{2}}$ and $f''(x) = \left(x^2-1\right)e^{-\frac{x^2}{2}}$

x	$-\infty$		-1		0		1		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	(0)	$\nearrow$	$\nearrow$	$\nearrow$	1	$\searrow$	$\searrow$	$\searrow$	(0)
	H.A.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

(d) Asymptotes: $x = 0$ and $y = 0$

Derivatives: $f'(x) = -\frac{e^{-x}(x+3)}{x^4}$ and $f''(x) = \frac{e^{-x}\left(x^2+6x+12\right)}{x^5}$

x	$-\infty$		-3		0		$+\infty$
$f'(x)$	$+$	$+$	0	$-$	|	$-$	$-$
$f''(x)$	$-$	$-$	$-$	$-$	|	$+$	$+$
$f(x)$	$(-\infty)$	$\nearrow$	$-\frac{e^3}{27}$	$\searrow$	|	$\searrow$	(0)
		$\cap$	max.	$\cap$	V.A.	$\cup$	H.A.

(e) Asymptotes: $x = 0$ and $y = \ln(2)x$

Derivatives: $f'(x) = \dfrac{2^x \ln(2)}{2^x - 1}$ and $f''(x) = -\dfrac{2^x \ln^2(2)}{(2^x - 1)^2}$

x		0		1		$+\infty$
$f'(x)$	///	$\vert$	$+$	$+$	$+$	$+$
$f''(x)$	///	$\vert$	$-$	$-$	$-$	$-$
$f(x)$	///	$\vert$	$\nearrow$	0	$\nearrow$	$(+\infty)$
		V.A.	$\cap$	$\cap$	$\cap$	S.A.

(f) Asymptotes: $x = 0$ and $y = 0$

Derivatives: $f'(x) = \dfrac{1 - 2\ln(x)}{x^3}$ and $f''(x) = \dfrac{6\ln(x) - 5}{x^4}$

x		0		1		$\sqrt{e}$		$e^{\frac{5}{6}}$		$+\infty$
$f'(x)$	///	$\vert$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	///	$\vert$	$-$	$-$	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	///	$\vert$	$\nearrow$	$\nearrow$	$\nearrow$	$\dfrac{1}{2e}$	$\searrow$	$\searrow$	$\searrow$	(0)
		V.A.	$\cap$	$\cap$	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

(g) Asymptotes: $y = \dfrac{x}{2}$ and $y = -\dfrac{x}{2}$

Derivatives: $f'(x) = \dfrac{e^x - e^{-x}}{2(e^x + e^{-x})}$ and $f''(x) = \dfrac{2}{(e^x + e^{-x})^2}$

x	$-\infty$		0		$+\infty$
$f'(x)$	$-$	$-$	0	$+$	$+$
$f''(x)$	$+$	$+$	$+$	$+$	$+$
$f(x)$	$(+\infty)$	$\searrow$	$\dfrac{\ln(2)}{2}$	$\nearrow$	$(+\infty)$
	S.A.	$\cup$	min.	$\cup$	S.A.

(h) Asymptotes: $x = e$ and $y = -2$

Derivatives: $f'(x) = \dfrac{2}{(\ln(x) - 1)^2\, x}$ and $f''(x) = -\dfrac{2(\ln(x) + 1)}{(\ln(x) - 1)^3\, x^2}$

x		0		e^{-1}		e		$+\infty$
$f'(x)$	///	$\vert$	$+$	$+$	$+$	$\vert$	$+$	$+$
$f''(x)$	///	$\vert$	$-$	0	$+$	$\vert$	$-$	$-$
$f(x)$	///	(-2)	$\nearrow$	$\nearrow$	$\nearrow$	$\vert$	$\nearrow$	(-2)
			$\cap$	inf.p.	$\cup$	V.A.	$\cap$	H.A.

(i) Asymptotes: $x = \dfrac{\pi}{2} + k\pi,\ k \in \mathbb{Z}$

Derivatives: $f'(x) = -\tan(x)$ and $f''(x) = -\dfrac{1}{\cos^2(x)}$

x	$\cdots$		$-\frac{\pi}{2}$		0		$\frac{\pi}{2}$		$\frac{3\pi}{2}$		2π		$\frac{5\pi}{2}$		$\cdots$
$f'(x)$	$\cdots$	$///$	$\vert$	$+$	0	$-$	$\vert$	$///$	$\vert$	$+$	0	$-$	$\vert$	$///$	$\cdots$
$f''(x)$	$\cdots$	$///$	$\vert$	$-$	$-$	$-$	$\vert$	$///$	$\vert$	$-$	$-$	$-$	$\vert$	$///$	$\cdots$
$f(x)$	$\cdots$	$///$	$\vert$	$\nearrow$	max.	$\searrow$	$\vert$	$///$	$\vert$	$\nearrow$	max.	$\searrow$	$\vert$	$///$	$\cdots$
			V.A.	$\cap$	$\cap$	$\cap$	V.A.		V.A.	$\cap$	$\cap$	$\cap$	V.A.		

There are an infinite number of maxima at $x = 2k\pi$, with $k \in \mathbb{Z}$.

(j) Asymptotes: $y = \dfrac{\pi}{2}$

Derivatives: $f'(x) = \dfrac{1}{x(1+\ln^2(x))}$ and $f''(x) = -\dfrac{(1+\ln(x))^2}{x^2(1+\ln^2(x))^2}$

x		0		e^{-1}		1		$+\infty$
$f'(x)$	$///$	$\vert$	$+$	$+$	$+$	$+$	$+$	$+$
$f''(x)$	$///$	$\vert$	$-$	0	$-$	$-$	$-$	$-$
$f(x)$	$///$	$\left(-\frac{\pi}{2}\right)$	$\nearrow$	$\nearrow$	$\nearrow$	0	$\nearrow$	$\left(\frac{\pi}{2}\right)$
			$\cap$	$\cap$	$\cap$	$\cap$	$\cap$	H.A.

(k) Asymptotes: $y = 0$

Derivatives: $f'(x) = -\dfrac{1}{1+x^2}$ and $f''(x) = \dfrac{2x}{\left(1+x^2\right)^2}$

x	$-\infty$		0		$+\infty$
$f'(x)$	$-$	$-$	$\vert$	$-$	$-$
$f''(x)$	$-$	$-$	$\vert$	$+$	$+$
$f(x)$	(0)	$\searrow$	$\left(-\frac{\pi}{2} \middle\vert \frac{\pi}{2}\right)$	$\searrow$	(0)
	H.A.	$\cap$	V.A.	$\cup$	H.A.

(l) Asymptotes: none

Derivatives: $f'(x) = x^x(\ln(x)+1)$ and $f''(x) = x^x\left((\ln(x)+1)^2 + \dfrac{1}{x}\right)$

x		0		e^{-1}		$+\infty$
$f'(x)$	$///$	$\vert$	$-$	0	$+$	$+$
$f''(x)$	$///$	$\vert$	$+$	$+$	$+$	$+$
$f(x)$	$///$	(1)	$\searrow$	$e^{-e^{-1}}$	$\nearrow$	$(+\infty)$
			$\cup$	min.	$\cup$	

(m) Asymptotes: $y = 0$

Derivatives: $f'(x) = (x^2)^x(\ln(x^2)+2)$ and $f''(x) = (x^2)^x\left(\left(\ln(x^2)+2\right)^2+\dfrac{2}{x}\right)$

x	$-\infty$		$-0,8$		$-\frac{1}{e}$		0		$\frac{1}{e}$		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	0	$-$	$\vert$	$-$	0	$+$	$+$
$f''(x)$	$+$	$+$	0	$-$	$-$	$-$	$\vert$	$+$	$+$	$+$	$+$
$f(x)$	(0)	$\nearrow$	$\nearrow$	$\nearrow$	$e^{2e^{-1}}$	$\searrow$	(1)	$\searrow$	$e^{-2e^{-1}}$	$\nearrow$	$(+\infty)$
	H.A.	$\cup$	inf.p.	$\cap$	max.	$\cap$		$\cup$	min.	$\cup$	$\cup$

(n) Asymptotes: none

Derivatives: $f'(x) = 1-2\cos(x)$ and $f''(x) = 2\sin(x)$

x	$\cdots$	0		$\frac{\pi}{3}$		π		$\frac{5\pi}{3}$		2π		$\frac{7\pi}{3}$		3π	$\cdots$
$f'(x)$	$\cdots$	$-$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$\cdots$
$f''(x)$	$\cdots$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$+$	0	$\cdots$
$f(x)$	$(-\infty)$	0	$\searrow$	$f\left(\frac{\pi}{3}\right)$	$\nearrow$	$f(\pi)$	$\nearrow$	$f\left(\frac{5\pi}{3}\right)$	$\searrow$	$f(2\pi)$	$\searrow$	$f\left(\frac{7\pi}{3}\right)$	$\nearrow$	$f(3\pi)$	$(+\infty)$
	$\cdots$	inf.p.	$\cup$	min.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	min.	$\cup$	inf.p.	$\cdots$

(o) Asymptotes: y=0

Derivatives: $f'(x) = e^{-x}(-\sin(x)+\cos(x))$ and $f''(x) = -2e^{-x}\cos(x)$

x	0		$\frac{\pi}{4}$		$\frac{\pi}{2}$		π		$\frac{5\pi}{4}$		$\frac{3\pi}{2}$		2π		$\frac{9\pi}{4}$		$\frac{5\pi}{2}$	$\cdots$
$f'(x)$	$+$	$+$	0	$-$	$-$	$-$	$-$	$-$	0	$+$	$+$	$+$	$+$	$+$	0	$-$	$-$	$\cdots$
$f''(x)$	$-$	$-$	$-$	$-$	0	$+$	$+$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$	$-$	0	$\cdots$
$f(x)$	0	$\nearrow$	$f\left(\frac{\pi}{4}\right)$	$\searrow$	$f\left(\frac{\pi}{2}\right)$	$\searrow$	0	$\searrow$	$f\left(\frac{5\pi}{4}\right)$	$\nearrow$	$f\left(\frac{3\pi}{2}\right)$	$\nearrow$	0	$\nearrow$	$f\left(\frac{9\pi}{4}\right)$	$\searrow$	$f\left(\frac{5\pi}{2}\right)$	(0)
		$\cap$	max.	$\cap$	inf.p.	$\cup$		$\cup$	min.	$\cup$	inf.p.	$\cap$		$\cap$	max.	$\cap$	inf.p.	$\cdots$

(p) Asymptotes: none

Derivatives: $f'(x) = 1+\cos(x)$ and $f''(x) = -\sin(x)$

x	$\cdots$	$\cdots$	-2π		$-\pi$		0		π		2π		3π	$\cdots$	$\cdots$
$f'(x)$	$\cdots$	$\cdots$	$+$	$+$	0	$+$	$+$	$+$	0	$+$	$+$	$+$	0	$\cdots$	$\cdots$
$f''(x)$	$\cdots$	$+$	0	$-$	0	$+$	0	$-$	0	$+$	0	$-$	0	$+$	$\cdots$
$f(x)$	$(-\infty)$	$\nearrow$	$f(-2\pi)$	$\nearrow$	$f(-\pi)$	$\nearrow$	$f(0)$	$\nearrow$	$f(\pi)$	$\nearrow$	$f(2\pi)$	$\nearrow$	$f(3\pi)$	$\nearrow$	$(+\infty)$
	$\cdots$	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	$\cdots$

(q) $f(x) = \dfrac{|1+x|-1}{x} = \begin{cases} 1, & \text{if } x \geq -1, \\ -\dfrac{x+2}{x} = -1-\dfrac{2}{x}, & \text{if } x < -1. \end{cases}$

Asymptotes: $y=-1$ (for $x \to -\infty$)

Derivatives: $f'(x) = \begin{cases} 0, & \text{if } x \geq -1, \\ \dfrac{2}{x^2}, & \text{if } x < -1. \end{cases}$ and $f''(x) = \begin{cases} 0, & \text{if } x \geq -1, \\ \dfrac{-4}{x^3}, & \text{if } x < -1. \end{cases}$

x	$-\infty$		-2		-1		0		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	2	0	0	0	0
$f''(x)$	$+$	$+$	$+$	$+$	4	0	0	0	0
$f(x)$	(-1)	$\nearrow$	0	$\nearrow$	1	1	(1)	1	1
	H.A.	$\cup$		$\cup$		$-$		$-$	

(r) $f(x) = \left|2 - \sqrt{2x+4}\right| \quad \Rightarrow \quad \text{dom} f(x) = [-2, +\infty[, \quad f(x) = 0 \quad \Leftrightarrow \quad x = 0$

$$\Rightarrow f(x) = \begin{cases} 2 - \sqrt{2x+4}, & \text{if } -2 \le x < 0, \\ -2 + \sqrt{2x+4}, & \text{if } x > 0. \end{cases}$$

- $f(x) = 2 - \sqrt{2x+4}$ with $-2 \le x < 0$

  Asymptotes: none

  Derivatives: $f'(x) = -\dfrac{1}{\sqrt{2x+4}}$ and $f''(x) = \dfrac{1}{\sqrt{(2x+4)^3}}$

- $f(x) = -2 + \sqrt{2x+4}$ with $x > 0$

  Asymptotes: none

  Derivatives: $f'(x) = \dfrac{1}{\sqrt{2x+4}}$ and $f''(x) = -\dfrac{1}{\sqrt{(2x+4)^3}}$

x	-2		0			$+\infty$
$f'(x)$	$\vert$	$-$	$\left(-\frac{1}{2} \middle\vert \frac{1}{2}\right)$	$+$	$+$	$+$
$f''(x)$	$\vert$	$+$	$\left(\frac{1}{8} \middle\vert -\frac{1}{8}\right)$	$-$	$-$	$-$
$f(x)$	2	$\searrow$	0	$\nearrow$	$\nearrow$	$(+\infty)$
		$\cup$	inf.p.	$\cap$	$\cap$	$\cap$

(s) $f(x) = \dfrac{x^2}{x|x|+1} \quad \Rightarrow \quad f(x) = 0 \quad \Leftrightarrow \quad x = 0$

$$\Rightarrow f(x) = \begin{cases} \dfrac{x^2}{-x^2+1}, & \text{if } x < 0, \\ \dfrac{x^2}{x^2+1}, & \text{if } x > 0. \end{cases}$$

- $f(x) = \dfrac{x^2}{-x^2+1}$ with $x < 0$

  Asymptotes: $x = -1$ and $y = -1$

  Derivatives: $f'(x) = \dfrac{2x}{(-x^2+1)^2} \quad (x \neq -1)$ and $f''(x) = \dfrac{2(3x^2+1)}{(-x^2+1)^3} \quad (x \neq -1)$

- $f(x) = \dfrac{x^2}{x^2+1}$ with $x > 0$

  Asymptotes: $y = 1$

  Derivatives: $f'(x) = \dfrac{2x}{(x^2+1)^2}$ and $f''(x) = \dfrac{2(-3x^2+1)}{(x^2+1)^3}$

x	$-\infty$		-1		0		$\frac{\sqrt{3}}{3}$		$+\infty$
$f'(x)$	$-$	$-$	|	$-$	0	$+$	$+$	$+$	$+$
$f''(x)$	$-$	$-$	|	$+$	$+$	$+$	0	$-$	$-$
$f(x)$	(-1)	$\searrow$	|	$\searrow$	0	$\nearrow$	$f\left(\frac{\sqrt{3}}{3}\right)$	$\nearrow$	(1)
		$\cap$	V.A	$\cup$	min.	$\cup$	inf.p.	$\cap$	

(t) $f(x) = \left|(x-2)^2 - 4\right| \quad \Rightarrow \quad f(x) = 0 \quad \Leftrightarrow \quad x = 0 \quad \vee \quad x = 4$

$$\Rightarrow f(x) = \begin{cases} (x-2)^2 - 4, & \text{if } x < 0, \\ -(x-2)^2 + 4, & \text{if } 0 < x < 4, \\ (x-2)^2 - 4, & \text{if } x > 4. \end{cases}$$

Asymptotes: none

Derivatives: $f'(x) = 2(x-2)$ and $f''(x) = 2$ if $x < 0$ or $x > 4$

$f'(x) = -2(x-2)$ and $f''(x) = -2$ if $0 < x < 4$

x	$-\infty$		0		2		4		$+\infty$
$f'(x)$	$-$	$-$	$(-4\,\vert\,4)$	$+$	0	$-$	$(-4\,\vert\,4)$	$+$	$+$
$f''(x)$	$+$	$+$	$+$	$-$	$-$	$-$	$+$	$+$	$+$
$f(x)$	$(+\infty)$	$\searrow$	0	$\nearrow$	4	$\searrow$	0	$\nearrow$	$(+\infty)$
		$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	

(u) Asymptotes: none

Derivatives: $f'(x) = \cosh(x) - 1$ and $f''(x) = \sinh(x)$

x	$-\infty$		0		$+\infty$
$f'(x)$	$+$	$+$	0	$+$	$+$
$f''(x)$	$-$	$-$	0	$+$	$+$
$f(x)$	$(-\infty)$	$\nearrow$	0	$\nearrow$	$(+\infty)$
		$\cap$	inf.p.	$\cup$	

(v) Asymptotes: $y = -\frac{1}{2}$

Derivatives: $f'(x) = e^{2x}$ and $f''(x) = 2e^{2x}$

x	$-\infty$		0		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	$+$
$f''(x)$	$+$	$+$	$+$	$+$	$+$
$f(x)$	$\left(-\frac{1}{2}\right)$	$\nearrow$	0	$\nearrow$	$(+\infty)$
		$\cup$	$\cup$	$\cup$	

(w) Asymptotes: $x = 0$, $y = x+1$ and $y = x-1$

Derivatives: $f'(x) = 1 - \dfrac{1}{\sinh^2(x)}$ and $f''(x) = \dfrac{2\cosh(x)}{\sinh^3(x)}$

x	$-\infty$		$-\ln(1+\sqrt{2})$		0		$\ln(1+\sqrt{2})$		$+\infty$
$f'(x)$	$+$	$+$	0	$-$	|	$-$	0	$+$	$+$
$f''(x)$	$-$	$-$	$-$	$-$	|	$+$	$+$	$+$	$+$
$f(x)$	$(-\infty)$	$\nearrow$	$-\ln(1+\sqrt{2})-\sqrt{2}$	$\searrow$	|	$\searrow$	$\ln(1+\sqrt{2})+\sqrt{2}$	$\nearrow$	$(+\infty)$
	S.A.	$\cap$	max.	$\cap$	V.A.	$\cup$	min.	$\cup$	S.A.

(x) Asymptotes: none

Derivatives: $f'(x) = \dfrac{1}{2\sqrt{x-2}\sqrt{x-3}}$ and $f''(x) = \dfrac{5-2x}{4(x-2)^{\frac{3}{2}}(x-3)^{\frac{3}{2}}}$

x		3		$+\infty$
$f'(x)$	///	|	$+$	$+$
$f''(x)$	///	|	$-$	$-$
$f(x)$	///	0	$\nearrow$	$(+\infty)$
			$\cap$	

(y) Asymptotes: $x = -4$, $x = 4$ and $y = 0$

Derivatives: $f'(x) = -\dfrac{4}{x^2-16}$ and $f''(x) = \dfrac{8x}{(x^2-16)^2}$

x	$-\infty$		-4		4		$+\infty$
$f'(x)$	$-$	$-$	|	///	|	$-$	$-$
$f''(x)$	$-$	$-$	|	///	|	$+$	$+$
$f(x)$	(0)	$\searrow$	|	///	|	$\searrow$	(0)
	H.A.	$\cap$	V.A.		V.A.	$\cup$	H.A.

Assignment 10.7 — Graph (c) is the graph of f, (d) is the graph of f', (b) is the graph of f'' and (a) is that of the function g.

Assignment 10.8 — Graph (c) is the graph of $f(x) = \dfrac{x}{1-x^2}$, graph (b) is the graph of $g(x) = \dfrac{x^3}{1-x^4}$, graph (d) is the graph of $h(x) = \dfrac{x^3-x}{\sqrt{1+x^6}}$ graph (a) is the graph of $k(x) = \dfrac{x^3}{\sqrt{|x^4-1|}}$.

Assignment 10.9 —

(a)
- local maxima/minima: none,
- infliction points: $x = 1/2$,
- concave over $]-\infty, 1/2[$, convex over $]1/2, +\infty[$.

(b)
- local maxima/minima: global max. in $(0,1)$,
- infliction points: $x = \pm\sqrt{3}/3$,

- concave over $]-\sqrt{3}/3, \sqrt{3}/3[$, convex over $]-\infty, -\sqrt{3}/3[\cup]\sqrt{3}/3, +\infty[$.

(c)
- local maxima/minima: global max. in $(\pi/4, \sqrt{2})$, global min. in $(-3\pi/4, -\sqrt{2})$,
- infliction points: $x = -\pi/4, 3\pi/4$,
- concave over $]-\pi/4, 3\pi/4[$, convex over $]-\pi, -\pi/4[\cup]3\pi/4, \pi[$.

(d)
- local maxima/minima: global min. in $(e^{-1/2}, -e/2)$,
- infliction points: $x = 1/e^{3/2}$,
- concave over $]0, 1/e^{3/2}[$, convex over $]1/e^{3/2}, +\infty[$.

Assignment 10.10 —

(a) Critical points: $x = 0$

Infliction points: $x = \pm\sqrt{\dfrac{b^2}{3}}$

(b) Critical points: $x = \dfrac{n\pi/2 - b}{a}$, with $n \in \mathbb{Z}$ odd

Infliction points: $x = \dfrac{n\pi - b}{a}$, with $n \in \mathbb{Z}$

Chapter 11

Assignment 11.1 —

(a) $(3, 0)$

(b) $(-3, 0)$

(c) $\left(2, \dfrac{2\pi}{3}\right)$

(d) $\left(2, \dfrac{7\pi}{3}\right)$

(e) $(-3, \pi)$

(f) $\left(2, \dfrac{\pi}{3}\right)$

(g) $(-3, 2\pi)$

(h) $\left(-2, -\dfrac{\pi}{3}\right)$

(i) $\left(2, \dfrac{13\pi}{3}\right)$

Assignment 11.2 —

(a) $(1, 1)$

(b) $(1, 0)$

(c) $(0, 0)$

(d) $(-1, -1)$

(e) $\left(-\sqrt{3}, 3\right)$

Assignment 11.3 —

(a) $y = x$ → line

(b) $y = \frac{2}{5}x + \frac{7}{5}$ → line

(c) $(x-1)^2 + y^2 = 1$ → circle

(d) $x^2 + (y+2)^2 = 4$ → circle

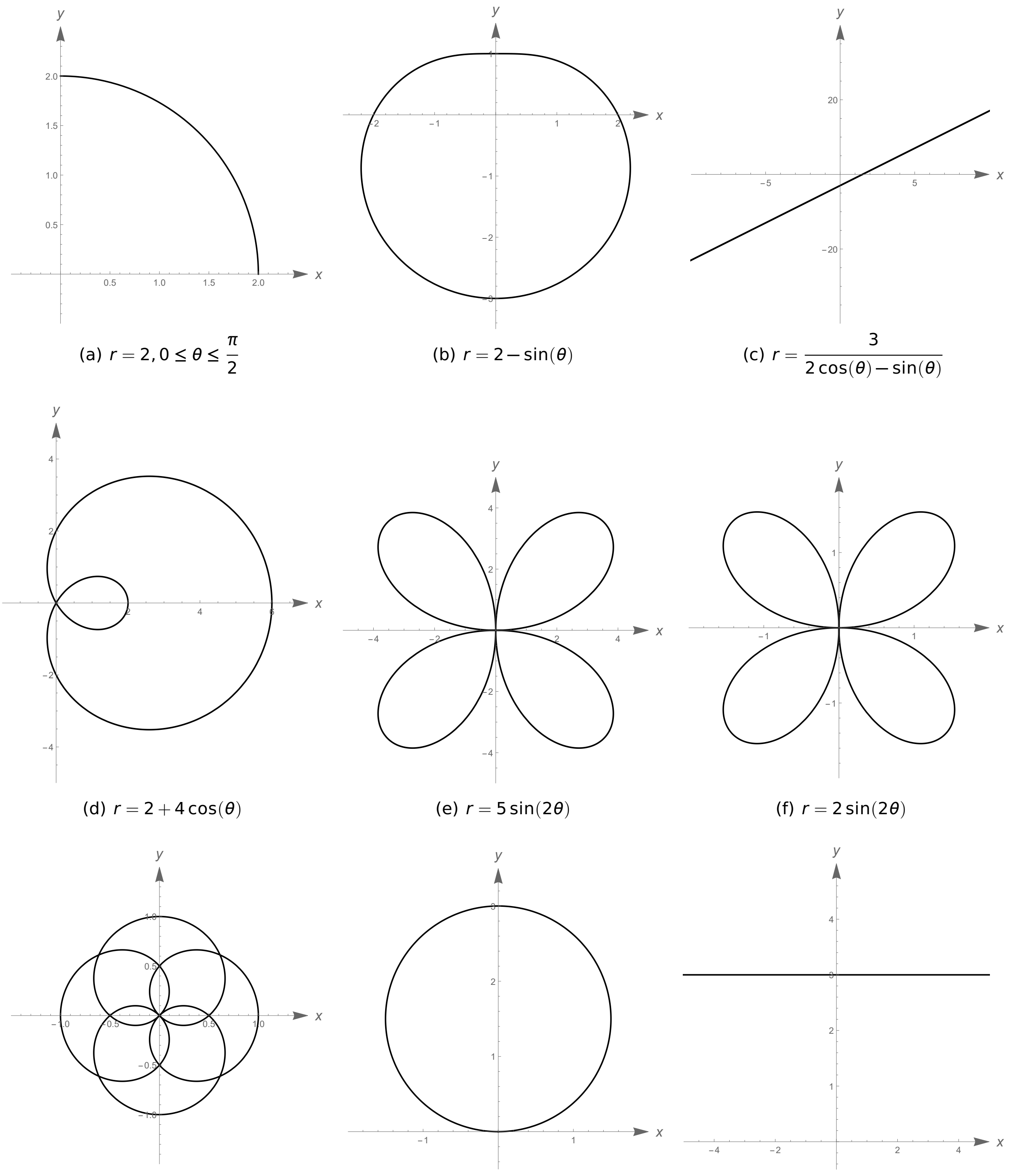
(e) $\left(x - \frac{1}{2}\right)^2 + \left(y - \frac{1}{2}\right)^2 = \frac{1}{2}$ → circle

(f) $x^2 + 4y^2 = 4$ → ellipse

(g) $y^2 = 1 + 2x$ → parabola

(h) $-3x^2 + 9\left(y + \frac{2}{3}\right)^2 = 1$ → hyperbola

Assignment 11.4 — Consider the graphs below.



(a) $r = 2, 0 \leq \theta \leq \dfrac{\pi}{2}$

(b) $r = 2 - \sin(\theta)$

(c) $r = \dfrac{3}{2\cos(\theta) - \sin(\theta)}$

(d) $r = 2 + 4\cos(\theta)$

(e) $r = 5\sin(2\theta)$

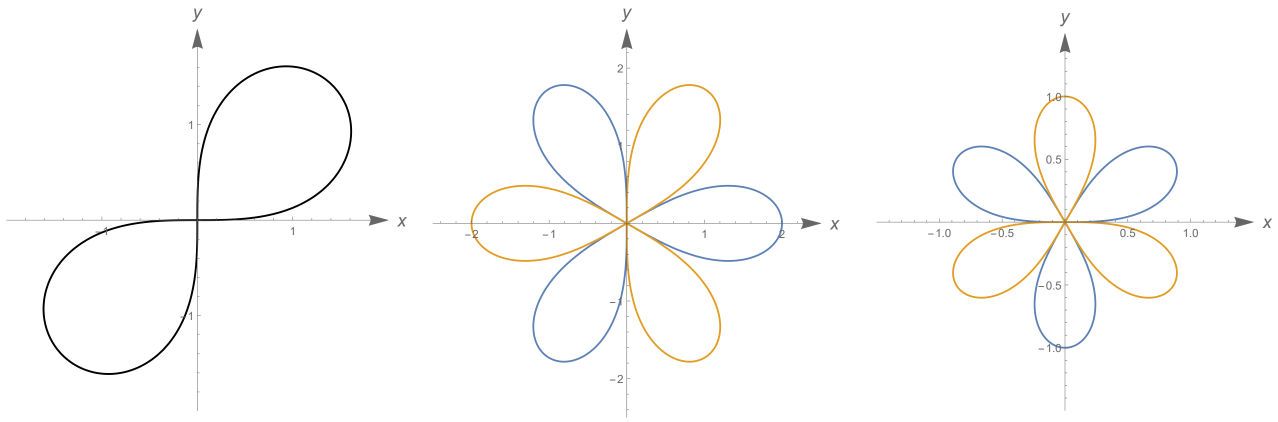
(f) $r = 2\sin(2\theta)$

(g) $r = \cos(2\theta/3), 0 \leq \theta \leq 6\pi$

(h) $r = 3\sin(\theta), 0 \leq \theta \leq \pi$

(i) $r = 3\csc(\theta), 0 < \theta < \pi$

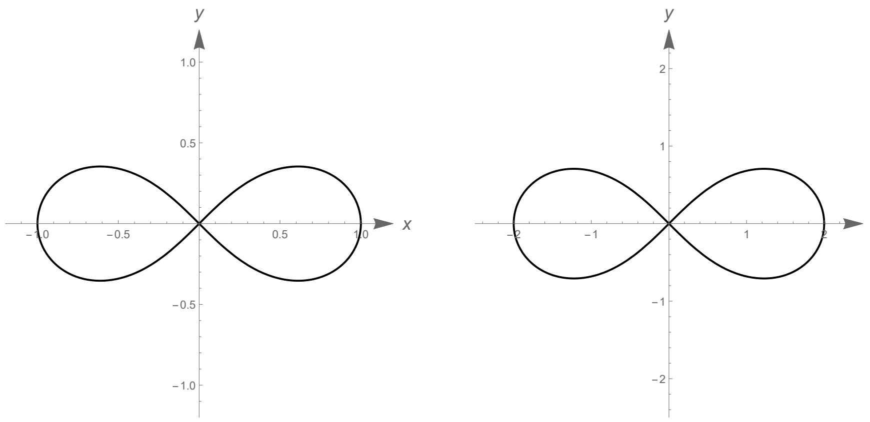
Figure 11.16: Graphs of the polar curves in Ex. 11.4 (deel 1).



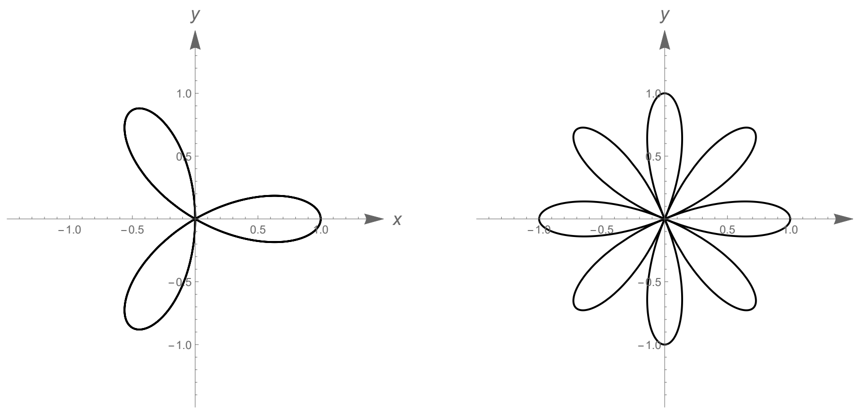
(j) $r^2 = 4\sin(2\theta)$

(k) $r^2 = 4\cos(3\theta)$

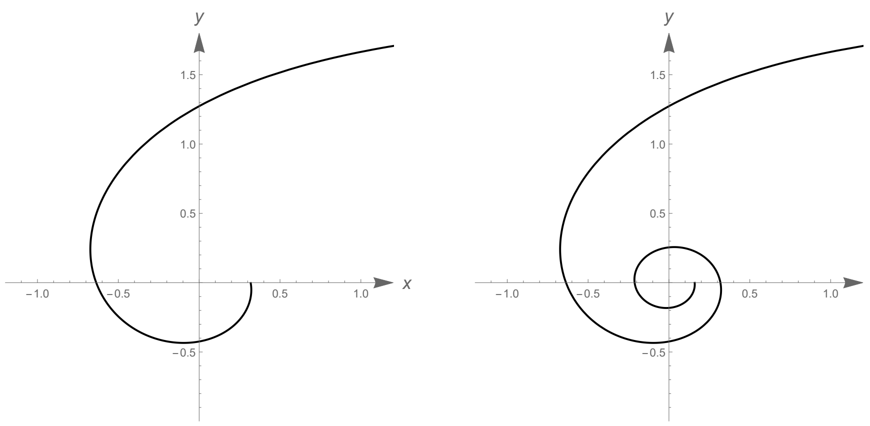
(l) $r^2 = \sin(3\theta)$



(m) $r = a\sqrt{\cos(2\theta)}$, $\quad a > 0$. Links $a = 1$, $\quad$ right $a = 2$



(n) $r = a\cos(n\theta)$, $\quad a > 0$. If n is odd, the curve has n 'leaves'. If n is even, the curve has $2n$ 'leaves'. The curve was plotted for $n = 3$ (left) and $n = 4$ (right).



(o) $r = a/\theta$. The curve was plotted for $a = 2$ with two different intervals for θ. Left: $0 \le \theta \le 2\pi$, $\quad$ right: $0 \le \theta \le 4\pi$.

Figure 11.17: Graphs of the polar curves in Exercise 11.4 (part 2).

Assignment 11.5 —

(a) The origin and $\left(\frac{3}{2}, \pm\frac{\pi}{3}\right)$

(b) $\left(4, \pm\frac{2\pi}{3}\right)$

(c) The origin, $\left(\frac{1}{2}, \frac{\pi}{6}\right)$ and $\left(\frac{1}{2}, \frac{5\pi}{6}\right)$

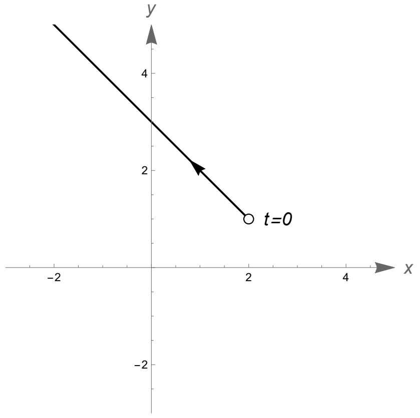
(d) The origin and $\left(\frac{\sqrt{3}}{2}, \frac{\pi}{3}\right)$

(e) $\left(1, \pm\frac{\pi}{6}\right)$ and $\left(1, \pm\frac{5\pi}{6}\right)$

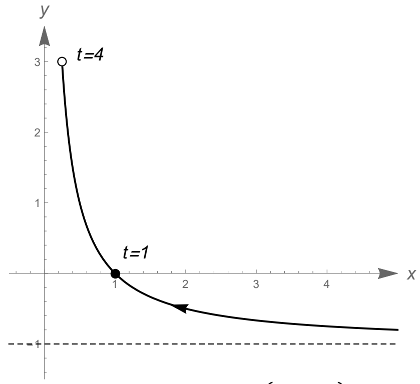
(f) The origin, $\left(\frac{\sqrt{2}}{2}, \frac{\pi}{12}\right)$, $\left(-\frac{\sqrt{2}}{2}, \frac{5\pi}{12}\right)$ and $\left(\frac{\sqrt{2}}{2}, \frac{3\pi}{4}\right)$

(g) The origin, $\left(1+\frac{\sqrt{2}}{2}, \frac{3\pi}{4}\right)$ and $\left(1-\frac{\sqrt{2}}{2}, \frac{7\pi}{4}\right)$

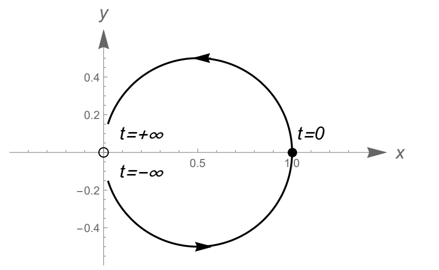
Assignment 11.6 — Consider the graphs below.



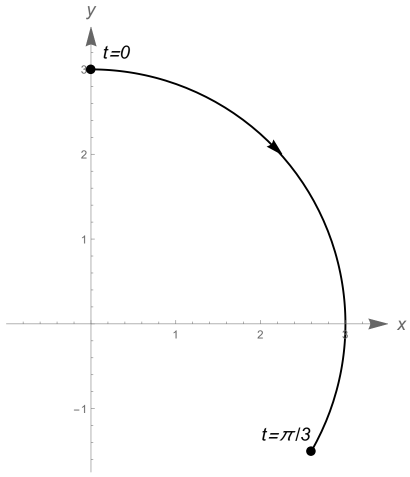
(a) $y = 3 - x \qquad (-\infty < x < 2)$



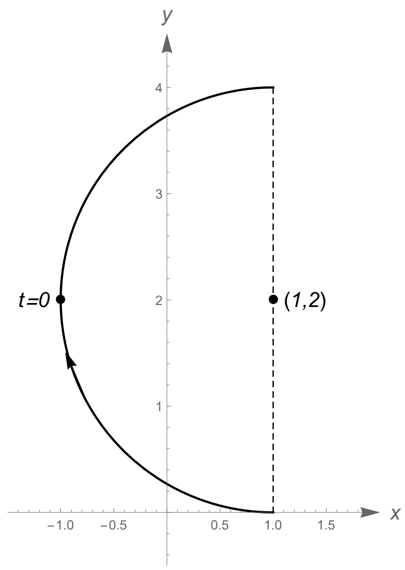
(b) $y = \dfrac{1}{x} - 1 \qquad \left(x > \dfrac{1}{4}\right)$



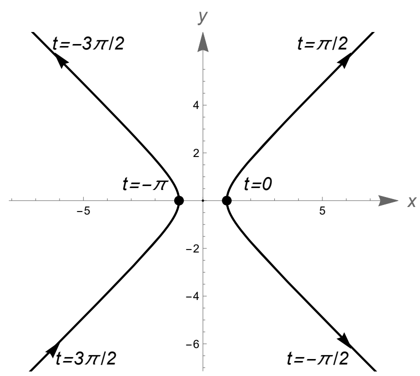
(c) $\left(x - \dfrac{1}{2}\right)^2 + y^2 = \dfrac{1}{4}$



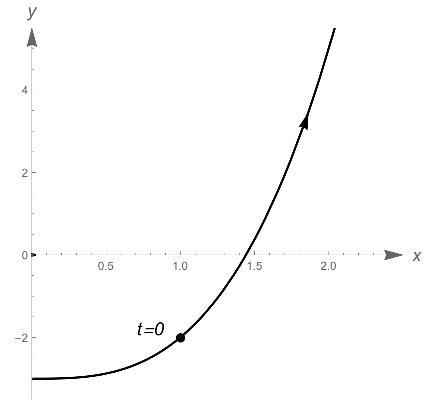
(d) $x^2 + y^2 = 9$ (part of the circle indicated on the figure)



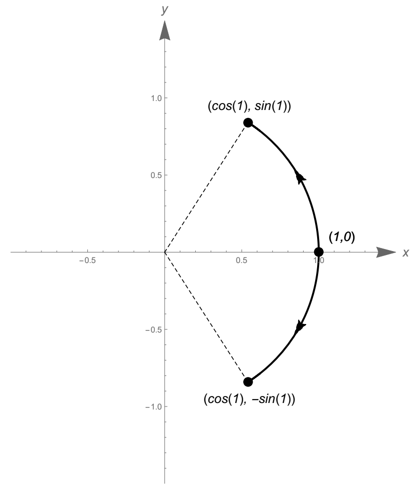
(e) $(x-1)^2 + (y-2)^2 = 4 \qquad (x \le 1)$



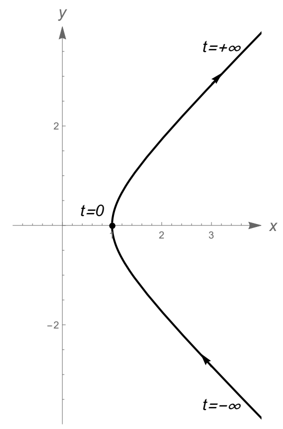
(f) $x^2 - y^2 = 1$



(g) $y = x^3 - 3$



(h) $x^2 + y^2 = 1.$



(i) $x^2 - y^2 = 1 \qquad (x \ge 1)$

Figure 11.18: Graphs of the parametric curves in Exercise 11.6.

Assignment 11.7 —

(a) $\begin{cases} x = t \\ y = -\sqrt{t-1} \end{cases} \quad (t > 1)$ (b) $\begin{cases} x = 6\cos^3(t) \\ y = 6\sin^3(t) \end{cases} \quad (0 < t < 2\pi)$

Assignment 11.8 — $\vec{\boldsymbol{r}}(t) = \left(\dfrac{t+4}{2}, t, \left(\dfrac{3}{2}t+7\right)\right)$ with $0 \le t \le 2$

Assignment 11.9 —
parameter $t = x$: $\vec{\boldsymbol{r}}(t) = \left(t, 1-t, 1-2t+2t^2\right)$ with $-\infty < t < +\infty$
parameter $t = y$: $\vec{\boldsymbol{r}}(t) = \left(1-t, t, 1-2t+2t^2\right)$ with $-\infty < t < +\infty$
parameter $t = z$: We have to solve the system of equations $\begin{cases} x+y=1 \\ x^2+y^2=t \end{cases}$ for x and y. There are two possible solutions each corresponding to one half of the parabola starting at the lowest point $(1/2, 1/2, 1/2)$ because there are two points on the parabola at every $z > 1/2$. The entire parabola can not be parameterized by using z as a parameter.

Assignment 11.10 —

(a) $x^2 + y^2 = 9$ and $z = x + y$

A possible parameterization is $\vec{\boldsymbol{r}}(t) = (3\cos(t), 3\sin(t), 3(\cos(t)+\sin(t)))$.

(b) $z = \sqrt{1-x^2-y^2}$ and $x + y = 1$

A possible parameterization is $\vec{\boldsymbol{r}}(t) = \left(t, 1-t, \sqrt{2(t-t^2)}\right)$.

(c) $z = x^2 + y^2$ and $2x - 4y - z - 1 = 0$

A possible parameterization is $\vec{\boldsymbol{r}}(t) = (1+2\cos(t), -2(1-\sin(t)), 9+4\cos(t)-8\sin(t))$.

Assignment 11.11 — Consider the graphs below.

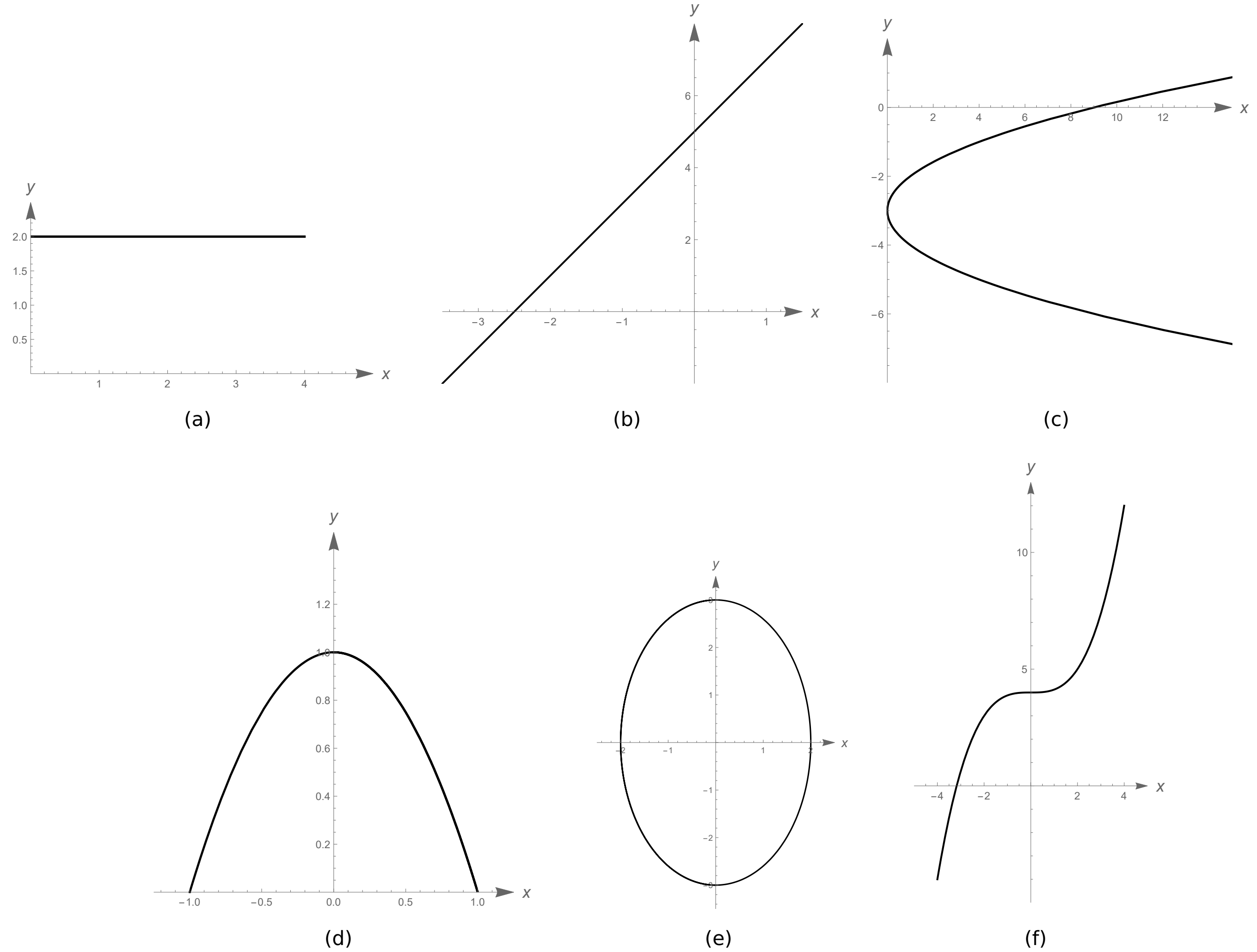


Figure 11.19: Grafieken van de parameterkrommen in Oefening 11.11.

Assignment 11.12 —

(a) Circle with radius 1, counterclockwise, 1 loop.

(b) Circle with radius 1, counterclockwise, 6 loops.

(c) Circle of radius 1, clockwise, infinite number of loops.

(d) Arc of a circle of radius 1, from angle −1 radians to 1 radians, 2 loops.

Assignment 11.13 — Consider the figure below for the graphs of the parametric equations.

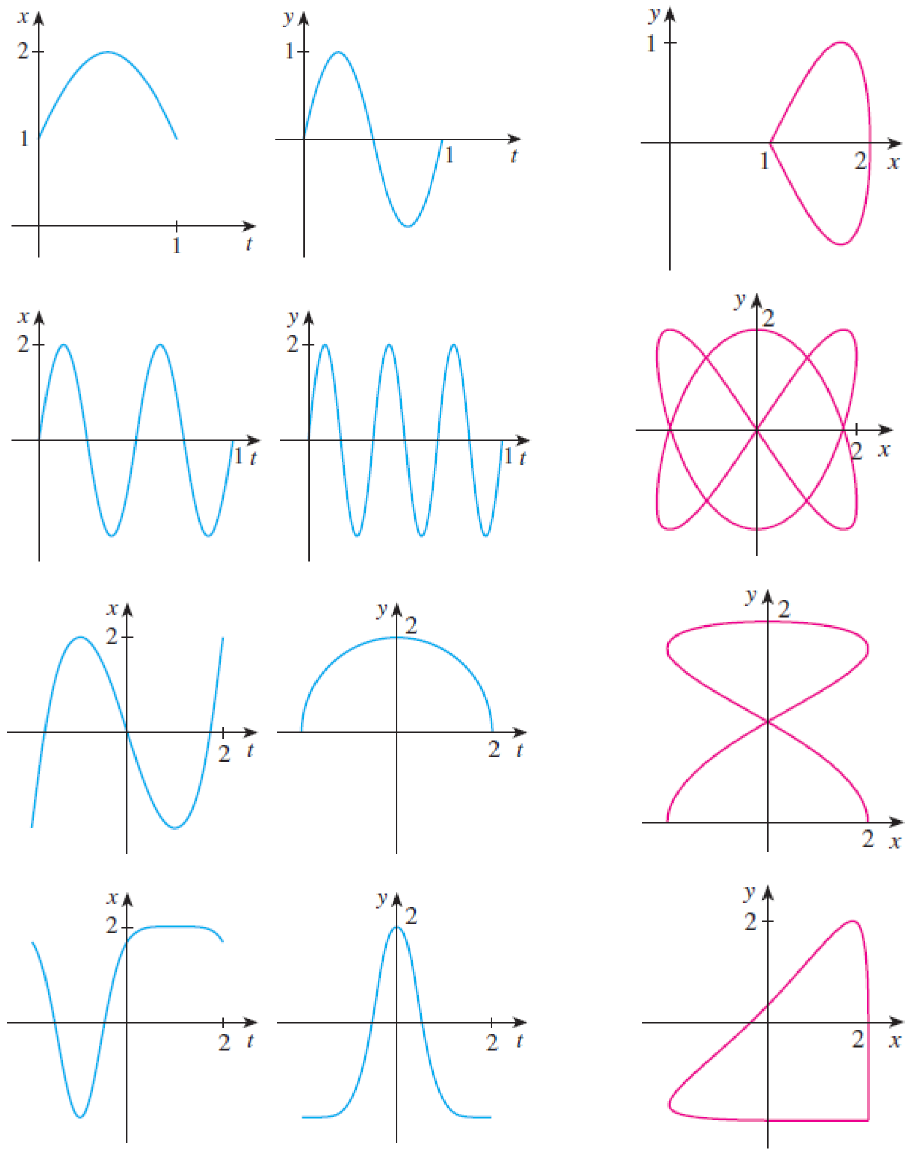


Figure 11.22: Graphs of the parametric equations from Exercise 11.13.

Assignment 11.14 —

(a) $C_t = \big((a-b)\cos(t), (a-b)\sin(t)\big)$

(b) $x = (a-b)\cos(t) + b\cos(\theta_t - t)$

(c) $y = (a-b)\sin(t) - b\sin(\theta_t - t)$

(d) $\theta_t = \dfrac{at}{b}$

Assignment 11.15 —

(a) $y' = \dfrac{\sqrt{2}}{6+\sqrt{2}}$

(b) $y' = 5\sqrt{3}$

(c) $y' = -\dfrac{3}{4}$

(d) $y' = -\dfrac{3}{2}$

Assignment 11.16 —

(a) Tangent: $x = \dfrac{3\sqrt{3}}{4}$, normal: $y = \dfrac{3}{4}$

(b) Tangent: $y = x + 1$, normal: $y = -x - 1$

(c) Tangent: $y = 3x + 2$, normal: $y = -\dfrac{1}{3}x + 2$

(d) Tangent: $y = 1$, normal: $x = \dfrac{\sqrt{2}}{2}$

(e) Tangent: $y = -\dfrac{x}{10} + e^{\pi/20}$, normal: $y = 10x + e^{\pi/20}$

Assignment 11.17 —

(a) Horizontal tangent at $t = 0$, this is in $(0,0)$. Vertical tangent at $t = 1$, in $(-2,5)$.

(b) Horizontal tangent at $t = n\pi$, this is in $(0, -(-1)^n n\pi)(n \in \mathbb{Z})$. Vertical tangent at $t = \left(n + \dfrac{1}{2}\right)\pi$, this is in $(1,1)$ and $(-1,-1)$.

(c) Horizontal tangents at $t = 0$ and $t = 2^{1/3}$, this is in $(0,0)$ en $(2^{1/3}, 2^{2/3})$. Vertical tangent at $t = 2^{-1/3}$, this is in $(2^{2/3}, 2^{1/3})$. The curve approximates $(0,0)$ vertically if $t \to \pm\infty$.

(d) There are no horizontal tangents. Vertical tangent at $t = 0$, this is in $(1,0)$.

(e) Horizontal tangents at $t = k\pi$ and $t = \pm \arctan\left(\sqrt{2}\right) + k\pi$. Vertical tangents at $t = \dfrac{\pi}{2} + k\pi$ and $t = \pm \arctan\left(\dfrac{\sqrt{2}}{2}\right) + k\pi$.

(f) Horizontal tangent at $\left(\dfrac{3}{2}, \pm\dfrac{\pi}{3}\right)$. Vertical tangents at $(2,0)$ and $\left(\dfrac{1}{2}, \pm\dfrac{2\pi}{3}\right)$.

(g) Horizontal tangents at $\left(\dfrac{1}{\sqrt{2}}, \pm\dfrac{\pi}{6}\right)$ and $\left(\dfrac{1}{\sqrt{2}}, \pm\dfrac{5\pi}{6}\right)$. Vertical tangents in $(1,0)$ and $(1,\pi)$.

(h) Horizontal tangents at $\left(4, -\dfrac{\pi}{2}\right)$, $\left(1, \dfrac{\pi}{6}\right)$ and $\left(1, \dfrac{5\pi}{6}\right)$. Vertical tangents at $\left(3, -\dfrac{\pi}{6}\right)$ and $\left(3, -\dfrac{5\pi}{6}\right)$.

Assignment 11.18 —

(a) $t = 2$

(b) $t = 0$

(c) $t = 2k\pi, \quad k \in \mathbb{Z}$

(d) $t = 0$

(e) $t = 1$

(f) $t = k\pi$ or $t = \dfrac{\pi}{2} + k\pi, \quad k \in \mathbb{Z}$

Assignment 11.19 — See the graphs below.

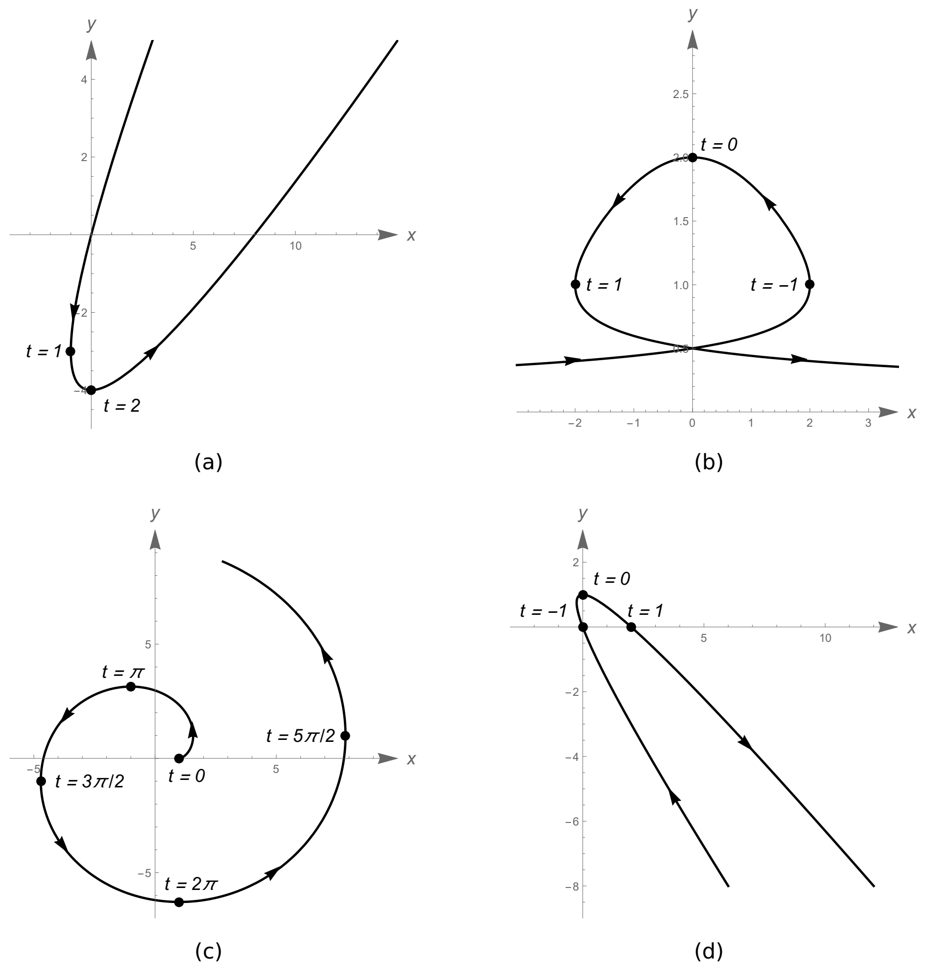


Figure 11.24: Graphs of the curves in Exercise 11.19.

Chapter 12

Assignments 12.1 — The area is 2.

Assignments 12.2 —

(a) $S_l(8) = \dfrac{7}{4}; \quad S_u(8) = \dfrac{9}{4}$

(b) $S_l(5) = 0,3153168; \quad S_u(5) = 0,4539462$

(c) $S_l(6) = 1,43; \quad S_u(6) = 2,48$

(d) $S_l(4) = -\pi; \quad S_u(4) = \pi$

(e) $S_l(6) = 10; \quad S_u(6) = 28$

(f) $S_l(4) = \dfrac{496}{315}; \quad S_u(4) = \dfrac{4352}{105}$

Assignments 12.3 —

(a) $\int_0^1 \sqrt{x}\,dx$ (b) $\int_0^1 \sqrt{x}\,dx$ (c) $\int_0^2 \ln(1+x)\,dx$

Assignments 12.4 — $\bar{f}$ is the mean of f.

Assignments 12.5 —

(a) $F'(x) = \dfrac{3x^2+1}{x^3+x}$

(b) $F'(x) = -3x^{11}$

(c) $F'(t) = \dfrac{\cos(t)}{1+t^2}$

(d) $F'(t) = -\dfrac{\sin(t)}{t}$

(e) $F'(x) = 2x\displaystyle\int_0^{x^2} \frac{\sin(u)}{u}\,du + 2x\sin(x^2)$

(f) $F'(\theta) = -\dfrac{1}{\sin(\theta)} - \dfrac{1}{\cos(\theta)}$

(g) $F'(x) = 3\displaystyle\int_4^{x^2} e^{-\sqrt{t}}\,dt + 6x^2 e^{-|x|}$

(h) $F'(x) = 2x^3 + 3x - 2$

(i) $F'(x) = e^x \sin(e^x) - \dfrac{\sin\big(\ln(x)\big)}{x}$

Assignments 12.6 —

(a) $I = e^x - \arcsin(x) + C$

Split the integral to obtain two basic integrals.

(b) $I = \dfrac{1}{4}\ln\Big(4x^2+4x+3\Big) + C$

Substitute $4x^2+4x+3$ by t.

(c) $I = \dfrac{1}{5\cos^5(x)} + C$

Substitute $\cos(x)$ by t.

(d) $I = \sin(x) - \dfrac{2}{3}\sin^3(x) + \dfrac{1}{5}\sin^5(x) + C$

Rewrite the integrand as $\Big(\cos^2(x)\Big)^2\cos(x)$ and then use the Pythagorean identity $\cos^2(x) = 1-\sin^2(x)$. Assume $t = \sin(x)$, expand the product and split the integral.

(e) $I = -\ln|\sin(x)+\cos(x)| + C$

Use $t = \sin(x)+\cos(x)$.

(f) $I = \dfrac{1}{2}\arcsin\big(2\tan(x)\big) + C$

Use $t = 2\tan(x)$.

(g) $I = -\dfrac{1}{1+\tan(x)} + C$

Rewrite the denominator by using $(\cos(x)+\sin(x))^2 = \cos^2(x)\big(1+\tan(x)\big)^2$ and then use $t = 1+\tan(x)$.

(h) $I = x\ln\left(x+\sqrt{x^2+5}\right) - \sqrt{x^2+5} + C$

Use integration by parts with $u(x) = \ln\left(x+\sqrt{x^2+5}\right)$ and $dv(x) = dx$. Then, use $t^2 = x^2+5$.

(i) $I = \dfrac{3}{5}\ln|x-2| + \dfrac{7}{5}\ln|x+3| + C$

Split the integrand into partial fractions.

(j) $I = 4\ln|x-2| - \dfrac{1}{x-2} - 4\ln|x-3| - \dfrac{4}{x-3} + C$

There are two ways to start. Either you first take the square of the integrand (square of the numerator over square of the denominator) and split the expression into partial fractions, or you first split the $\dfrac{x-1}{x^2-5x+6}$ into partial fractions and then square the result.

(k) $I = x - \ln(x^2+2x+2) + \arctan(x+1) + C$

First, perform Euclidean division. Rewrite the numerator of the integral as $2x+2-1$ and split it into two integrals. One integral can be found using $x^2+2x+2 = t$. The denominator of the second integral should be written as a perfect square to obtain $\dfrac{1}{1+u^2}$ as integrand.

(l) $I = \dfrac{x-1}{\sqrt{1+x^2}} + C$

The denominator contains the square root of the sum of two squares, therefore use $x = \tan(t)$or $x = \sinh(t)$.

(m) $I = \arcsin\left(\dfrac{x-2}{2}\right) + C$

Write the denominator as the square root of the difference of two squares by completing to a perfect square. Then work toward the integral with integrand $\dfrac{1}{\sqrt{1-u^2}}$ or use $x-2 = 2\sin(t)$.

(n) $I = -\dfrac{e^{2x}\cos(4x)}{5} + \dfrac{e^{2x}\sin(4x)}{10} + C$

By applying integration by parts twice with $u(x) = e^{2x}$ (choice) you arrive at the original integral I. Solve the resulting integral equation for I.

(o) $I = \dfrac{1}{16}\left(x - \dfrac{\sin(4x)}{4} - \dfrac{\sin^3(2x)}{3}\right) + C$

Use the angle doubling formula for $\cos(2x)$:

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \qquad \text{en} \qquad \cos^2(x) = \frac{1+\cos(2x)}{2}.$$

Then expand the powers in the integrand. For integrands with an even power of $\cos(x)$ the doubling formula can be used again for $\cos(2x)$. For integrands with odd powers of $\cos(x)$ you can use the Pythagorean identity.
Remark: If you use a different trigonometric formula, you will arrive at a different formula for I.

(p) $I = \dfrac{1}{3}\ln\left|\dfrac{2\sin(x)+1}{\sin(x)-1}\right| + C$

Rewrite $\cos^2(x)$ by using the Pythagorean identity $1-\sin^2(x)$ and perform the substitution $\sin(x) = t$. Afterwards, split the integrand in partial fractions.

(q) $I = \dfrac{1}{3}\tan^3(x) - \cot(x) + 2\tan(x) + C$

Write $\sin(x)$ as $\tan(x)\cos(x)$ and perform the substitution $\tan(x) = u$ together with the trigonometric formula $1 + \tan^2(x) = \sec^2(x)$.

(r) $I = \ln\left|\dfrac{e^x - 1}{e^x + 1}\right| + C$

Use the definition $\sinh(x) = \dfrac{e^x - e^{-x}}{2}$ and multiply numerator and denominator of the integrand by e^x. Use $e^x = t$ and split in partial fractions.

(s) $I = \dfrac{\operatorname{sech}^2(x)}{2} + \ln\big(\cosh(x)\big) + C$

Rewrite $\tanh^3(x)$ as $\sinh^3(x)/\cosh^3(x)$ and then $\sinh^3(x)$ as $(\cosh^2(x) - 1)\sinh(x)$. Afterwards use $\cosh(x) = t$.

Assignments 12.7 —

(a) $I = \dfrac{1}{2}\big(\sin(x)\cosh(x) - \sinh(x)\cos(x)\big) + C$

Use integration by parts twice respectively with $u(x) = \sin(x)$ or $\cos(x)$ and $dv(x) = \sinh(x)\,dx$ or $\cosh(x)\,dx$. We obtain an equation in I that can be solved for I.

(b) $I = 3x - 7\arctan(x) + C$

Rewrite the numerator as $3\left(x^2 + 1\right) - 7$ and split the integral into two integrals or perform Euclidean division. We then obtain two basic integrals.

(c) $I = \dfrac{x^2}{2} + \dfrac{4}{3}\ln|x - 2| - \dfrac{2}{3}\ln|x^2 + 2x + 4| + \dfrac{4\sqrt{3}}{3}\arctan\left(\dfrac{x+1}{\sqrt{3}}\right) + C$

Perform a Euclidean division and split $\dfrac{8x}{x^3 - 8}$ in partial fractions. Then, the integral $\displaystyle\int \dfrac{x-2}{x^2 + 2x + 4}$ d has to be found. Rewrite the numerator as $x + 1 - 3$ and split the integral into two integrals. To find integral $\displaystyle\int \dfrac{1}{x^2 + 2x + 4}\,dx$ rewrite the denominator as a perfect square.

(d) $I = \dfrac{\sqrt{x^2 - 1}\left(2x^2 + 1\right)}{3x^3} + C$

let $x = \sec(t)$ and then rewrite $\dfrac{1}{\cos^2(t)} - 1$ as $\tan^2(t)$. Next, write $\cos^3(t)$ as $(1 - \sin^2(t))\cos(t)$ and perform the substitution $u = \sin(t)$.

(e) $I = -x - 2\sqrt{x} + \dfrac{1}{1 - 2\sqrt{x}} + C$

Let $t^2 = x$ and then perform a Euclidean division.

(f) $I = -\cot(x) + \ln\big|1 + \cot(x)\big| + C$

Let $t = \cot(x)$.

(g) $I = \dfrac{x\,e^{2x}}{2} - \dfrac{e^{2x}}{4} + C$

Use integration by parts with $u(x) = x$ and $dv(x) = e^{2x}\,dx$.

(h) $I = \frac{5}{2} \ln\left|x^2 + \sqrt{x^4+1}\right| + C$

Let $x^2 = \tan(t)$ or let $x^2 = t$.

(i) $I = \frac{\sin(2x)}{4} + C$

Use Simpsons formula $2\sin(\alpha)\sin(\beta) = \cos(\alpha-\beta) - \cos(\alpha+\beta)$.

(j) $I = -2\sqrt{5-x} + 4\sqrt[4]{5-x} - 4\ln|\sqrt[4]{5-x}+1| + C$

Let $t^4 = 5-x$ and then perform a Euclidean division.

Assignments 12.8 —

(a) $I = \dfrac{x^3 \ln\left(\sqrt{1-x}\right)}{3} - \dfrac{x^3}{18} - \dfrac{x^2}{12} - \dfrac{x}{6} - \dfrac{\ln|1-x|}{6} + C$

Use integration by parts with $u(x) = \ln(\sqrt{1-x})$ and $dv(x) = x^2\,dx$ and then perform a Euclidean division.

(b) $I = x - 2\ln|2x+3| + C$

Rewrite the numerator as $2x+3-4$ ans split the integral in two basic integrals.

(c) $I = \dfrac{\sin^2(2x)}{4} + C$

Use $t = \sin(2x)$.

(d) $I = -\ln(1+e^x) + x + C$

Use $t = e^x + 1$ and split the integrand in partial fractions.

(e) $I = \dfrac{2\sqrt{3}}{3}\arctan\left(\dfrac{2x+1}{\sqrt{3}}\right) + C$

Rewrite the denominator as $1+t^2$ with $t = \dfrac{2x+1}{\sqrt{3}}$.

(f) $I = -\dfrac{1}{x^2+x+1} + \dfrac{2}{3}\dfrac{2x+1}{x^2+x+1} + \dfrac{8\sqrt{3}}{9}\arctan\left(\dfrac{2x+1}{\sqrt{3}}\right) + C$

Rewrite the numerator as $2x+1+2$ and split the integrals in two integrals. The first integral $\displaystyle\int \frac{2x+1}{(x^2+x+1)^2}\,dx$ can be found using $t = x^2+x+1$. In the second integral $\displaystyle 2\int \frac{dx}{(x^2+x+1)^2}$ the denominator should be rewritten as a perfect square. Then let $u = \dfrac{2x+1}{\sqrt{3}}$. This integral can be calculated using $u = \tan(\theta)$.

(g) $I = 2a\arctan\left(\sqrt{\dfrac{a+x}{a-x}}\right) - (a-x)\sqrt{\dfrac{a+x}{a-x}} + C$

Let $t^2 = \dfrac{a+x}{a-x}$. Rewrite the obtained numerator as t^2+1-1 and split the integral in two integrals, then use $t = \tan(\theta)$.

(h) $I = \frac{4}{7}(x-1)^{7/4} - \frac{2}{3}(x-1)^{3/2} - \frac{4}{5}(x-1)^{5/4} + x - 1 + C$

Use $t^4 = x-1$.

(i) $I = x\arctan(\sqrt{x}) - \sqrt{x} + \arctan(\sqrt{x}) + C$

Use integration by parts with $u(x) = \arctan(\sqrt{x})$ and $dv(x) = dx$. Then, let $t^2 = x$ and rewrite the numerator of the integrand as $t^2 + 1 - 1$.

(j) $I = \dfrac{2}{15}(1+x^3)^{5/2} - \dfrac{2}{9}(1+x^3)^{3/2} + C$

Let $1 + x^3 = t^2$ and rewrite x^5 as $x^3.x^2$.

(k) $I = \dfrac{\tan^4(x)}{4} + \dfrac{3}{2}\tan^2(x) + 3\ln|\tan(x)| - \dfrac{1}{2\tan^2(x)} + C$

Use $t = \tan(x)$ and write $\sin(x)$ and $\cos(x)$ as a function of t.

(l) $I = -\dfrac{1}{5\tan^5(x)} - \dfrac{2}{3\tan^3(x)} - \dfrac{1}{\tan(x)} + C = -\dfrac{\cot^5(x)}{5} - \dfrac{2\cot^3(x)}{3} - \cot(x) + C$

Use $t = \tan(x)$ or $t = \cot(x)$.

(m) $I = \ln\left(\sqrt{1+e^x} - 1\right) - \ln\left(\sqrt{1+e^x} + 1\right) + C$

Let $t^2 = 1 + e^x$.

(n) $I = \dfrac{2^x\left(\sinh(x) - \ln 2\cosh(x)\right)}{1 - \ln^2(2)} + C$

Use integration by parts twice, respectively with $u(x) = 2^x$ and $dv(x) = \cosh(x)\,dx$ or $\sinh(x)\,dx$. We then obtain an equation in I.

(o) $I = \dfrac{1}{32}(12x - 8\sin(2x) + \sin(4x)) + C$

Use the formula $\sin^2(x) = \dfrac{1-\cos(2x)}{2}$, then $\cos^2(2x) = \dfrac{1+\cos(4x)}{2}$.

(p) $I = \ln\left|1 + \tan\left(\frac{x}{2}\right)\right| + C$

Rewrite the denominator first by using $\cos(x) = 1 - 2\sin^2\left(\frac{x}{2}\right)$ en $\sin(x) = 2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right)$. Then, multiply both numerator and denominator with $\sec^2\left(\frac{x}{2}\right)$ and perform the substitution $u = \tan\left(\frac{x}{2}\right) + 1$.

Assignments 12.9 —

(a) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(b) $I = \dfrac{1}{2} \quad \Rightarrow \quad I$ is convergent.

(c) $I = 9 \quad \Rightarrow \quad I$ is convergent.

(d) $I = 2\left(1 + \sqrt{2}\right) \quad \Rightarrow \quad I$ is convergent.

(e) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(f) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(g) $I = 2e^2 \quad \Rightarrow \quad I$ is convergent.

(h) We write the integral as a sum:

$$I = \underbrace{\int_0^1 \frac{x^2}{x^5+1}\,dx}_{I_1} + \underbrace{\int_1^{+\infty} \frac{x^2}{x^5+1}\,dx}_{I_2}.$$

We conclude that I_1 converges, since this is a proper integral. The integral $\int_1^{+\infty} x^{-3}\, dx$ is a convergent upper limit I_2, thus I_2 converges as well. We conclude that I is the sum of convergent integrals and thus converges.

(i) We write the integral as a sum:

$$I = \underbrace{\int_0^1 \frac{dx}{1+\sqrt{x}}}_{I_1} + \underbrace{\int_1^{+\infty} \frac{dx}{1+\sqrt{x}}}_{I_2}.$$

We conclude that I_1 converges, since this is a proper integral. The integral $\int_1^{+\infty} \frac{dx}{2\sqrt{x}}$ is a divergent lower bound of I_2, Consequently, I_2 diverges as well. We conclude that I is a sum of a convergent and a divergent integral and therefore diverges.

(j) We write the integral as a sum:

$$I = \underbrace{\int_0^1 \frac{dx}{\sqrt{x}+x^2}}_{I_1} + \underbrace{\int_1^{+\infty} \frac{dx}{\sqrt{x}+x^2}}_{I_2}.$$

Both integrals are improper. The integral $\int_0^1 \frac{dx}{\sqrt{x}}$ is a convergent upper bound of I_1, therefore I_1 converges as well. I_2 will converge as well, because $\int_1^{+\infty} \frac{dx}{x^2}$ is a convergent upper bound of I_2. We conclude that I is a sum of convergent integrals and thus converges.

(k) The integral diverges as the integral $\int_{-1}^{1} \frac{e^{-1}}{x+1}\, dx$ is a divergent lower bound of I.

(l) The integral diverges as the integral $\int_2^{+\infty} \frac{dx}{\sqrt{x}}$ is a divergent lower bound of I.

(m) The integral converges as the integral $\int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{x}}\, dx$ is a convergent upper bound of I.

(n) The integral converges as the integral $\int_1^{+\infty} \frac{1}{x^2}\, dx$ is a convergent upper bound of I.

Assignments 12.10 — $\int_{\omega_0}^{+\infty} g(\omega)\, d\omega = \lim_{b\to+\infty} \int_{\omega_0}^{b} g(\omega)\, d\omega = \lim_{b\to+\infty} \frac{1}{\pi} \arctan[T(\omega-\omega_0)]\Big|_{\omega_0}^{b} = \frac{1}{2}$

Assignments 12.11 — The recursion formula can be proved by partial integration where $u(x) = \sin^{n-1}(x)$ en $dv(x) = \sin(x)\,dx$.

$\int_0^{\frac{\pi}{2}} \sin^9(x)\,dx = \frac{128}{315}$.

Assignments 12.12 — For the indefinite integrals, we obtain the following expressions

$$\begin{aligned} \int \sin^n(x)\,dx &= \frac{1}{n}\cos(x)\sin^{n-1}(x) + \frac{n-1}{n}\int \sin^{n-2}(x)\,dx \\ \int \cos^n(x)\,dx &= \frac{1}{n}\cos^{n-1}(x)\sin(x) + \frac{n-1}{n}\int \cos^{n-2}(x)\,dx \end{aligned}$$

(a) $I_n = \int_0^{\pi} \sin^n(x)\,dx = \frac{n-1}{n} I_{n-2}$

(b) $I_n = \int_0^{\pi} \cos^n(x)\,dx = \frac{n-1}{n} I_{n-2}$

(c) $I_n = \int_0^{2\pi} \sin^n(x)\,dx = \begin{cases} \frac{n-1}{n} I_{n-2}, & n \text{ even}, \\ 0, & n \text{ odd} \end{cases}$

(d) $I_n = \int_0^{2\pi} \cos^n(x)\,dx = \begin{cases} \frac{n-1}{n} I_{n-2}, & n \text{ even}, \\ 0, & n \text{ odd} \end{cases}$

Assignments 12.13 —

(a) $I = -1 + 4\ln\frac{3}{4}$

(b) $I = \frac{\pi}{3} - \sqrt{3}$

Assignments 12.14 —

(a) We write the given integral as a sum of two integrals:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t}\,dt = \underbrace{\int_0^{a} t^{x-1} e^{-t}\,dt}_{I_1} + \underbrace{\int_a^{+\infty} t^{x-1} e^{-t}\,dt}_{I_2}.$$

We know that $t^{x-1} e^{-t} \leq t^{x-1}$, $\forall\, t \in [0, +\infty[$. By using paragraph 12.5.3 we conclude that $\int_0^{a} t^{x-1}\,dt$ converges if $x > 0$, with $0 < a < +\infty$.

From this it follows that $\int_0^{a} t^{x-1}\,dt$ is a convergent integral larger than I_1. Thus I_1 converges as well.

We then seek an upper bound for I_2. The integral $\displaystyle\int_a^{+\infty} e^{-t}\,dt = e^{-a}$ converges. It would be interesting should the following apply:

$$t^{x-1}\,e^{-t} \overset{?}{\leq} e^{-t} \quad\Leftrightarrow\quad t^{x-1} \overset{?}{\leq} 1.$$

However, this is not always true. In general, we can say that $\displaystyle\int_a^{+\infty} e^{-ct}\,dt$ converges if $c > 0$. It is easy to verify that

$$\lim_{t\to+\infty} t^{x-1}\,e^{-kt} = 0, \qquad \text{met} \quad k > 0.$$

Thus, there exists an a for which it holds that:

$$t^{x-1}\,e^{-kt} \leq 1, \qquad \forall\, t \in [a, +\infty[.$$

From this it follows that

$$t^{x-1}\,e^{-kt}\,e^{-ct} \leq e^{-ct}, \qquad \forall\, t \in [a, +\infty[.$$

We choose $k + c = 1$, such that we may say

$$\int_a^{+\infty} t^{x-1}\,e^{-t}\,dt \leq \int_a^{+\infty} e^{-ct}\,dt.$$

For the integral I_2 there thus exists a convergent integral that is larger, by which I_2 converges as well. $\Gamma(x)$ is a sum of convergent integrals and will consequently converge.

(b) Perform integration by parts on $\Gamma(x+1)$ with $u(x) = t^x$ en $dv(x) = e^{-t}\,dt$.

(c) This can be shown by means of induction.

(d) Basic step: ($n = 0$): $\displaystyle\Gamma(1) = \int_0^{+\infty} t^0\,e^{-t}\,dt = 1 = 1!$

(e) Induction hypothesis ($n = k-1$): $\Gamma(k) = (k-1)!$

(f) Prove the property for $n = k$ using the induction hypothesis:

$$\Gamma(k+1) = k\Gamma(k) \overset{I.H.}{=} k(k-1)! = k!$$

(g)

- $\displaystyle\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} t^{-1/2}\,e^{-t}\,dt \overset{t=x^2}{=} \int_0^{+\infty} x^{-1}e^{-x^2}2x\,dx = \sqrt{\pi}$
- $\displaystyle\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$

Assignments 12.15 — Choose $c = 4\pi\left(\dfrac{m}{2\pi kT}\right)^{3/2}$ and $a = \dfrac{m}{2kT}$, then

$$\overline{v} = c\int_0^{+\infty} v^3\,e^{-av^2}\,dv = \frac{c}{a}\int_0^{+\infty} v\,e^{-av^2}\,dv = \sqrt{\frac{8kT}{\pi m}}.$$

Chapter 13

Assignment 13.1 —

(a) $A = \int_{-2}^{4} \left(\frac{y}{2} + 2 - \frac{y^2}{4}\right) dy = 9$ or $A = 4\int_{0}^{1} \sqrt{x}\, dx + \int_{1}^{4} \left(2\sqrt{x} - 2x + 4\right) dx = 9$

(b) $A = 2\int_{0}^{2} (4 - y^2)\, dy = \frac{32}{3}$ or $A = 2\int_{0}^{4} \sqrt{4-x}\, dx = \frac{32}{3}$

(c) $A = 2\int_{3}^{5} \sqrt{25 - x^2}\, dx = -12 + \frac{25\pi}{2} - 25\arcsin\left(\frac{3}{5}\right) = 25\arccos\left(\frac{3}{5}\right) - 12$

or $A = 2\int_{0}^{4} \left(\sqrt{25 - y^2} - 3\right) dy = 25\arccos\left(\frac{3}{5}\right) - 12$

(d) $A = \int_{0}^{1} \left((4-x) - \left(4x - x^2\right)\right) dx + \int_{1}^{4} \left(\left(4x - x^2\right) - (4-x)\right) dx = \frac{19}{3}$

This exercise is not suited to use y as variable of integration.

(e) $A = \int_{0}^{4} \left(6x - x^2 - \left(x^2 - 2x\right)\right) dx = \int_{0}^{4} \left(8x - 2x^2\right) dx = \frac{64}{3}$

This exercise is not suited to use y as variable of integration.

(f) $A = 2\int_{0}^{3} \sqrt{x}\, dx + 2\int_{3}^{2\sqrt{3}} \sqrt{12 - x^2}\, dx = 2\left(\frac{\sqrt{3}}{2} + \pi\right)$

or $A = 2\int_{0}^{\sqrt{3}} \left(\sqrt{12 - y^2} - y^2\right) dy = 2\left(\frac{\sqrt{3}}{2} + \pi\right)$

(g) $A = 2\int_{\pi/2}^{3\pi/2} \cos^2(x)\, dx = \pi$ of $A = 4\int_{0}^{1} \arccos\left(\sqrt{y}\right) dy = \pi$

(h) $A = \int_{1}^{e} \left(\ln(2x) - \ln(x)\right) dx = (e-1)\ln(2)$

or $A = \int_{0}^{\ln(2)} \left(e^y - 1\right) dy + \int_{\ln(2)}^{1} \left(e^y - \frac{e^y}{2}\right) dy + \int_{1}^{\ln(2e)} \left(e - \frac{e^y}{2}\right) dy = (e-1)\ln(2)$

(i) $A = \int_{0}^{1} \left(\cosh(x) - \sinh(x)\right) dx = 1 - \frac{1}{e}$

or $\quad A = \displaystyle\int_0^1 \operatorname{arsinh}(y)\ dy + \int_1^{\sinh(1)} \big(\operatorname{arsinh}(y) - \operatorname{arcosh}(y)\big)\ dy + \int_{\sinh(1)}^{\cosh(1)} \big(1 - \operatorname{arcosh}(y)\big)\ dy =$
$1 - \dfrac{1}{e}$

(j) $A = \displaystyle\int_{-\pi/2-1}^{-\pi/2} \left(x + \frac{\pi}{2} + 1\right) dx + \int_{-\pi/2}^{0} \big(-\sin(x)\big)\, dx = \frac{3}{2}$

or $\quad A = \displaystyle\int_{-1}^{0} \left(1 + y + \frac{\pi}{2} + \arcsin(y)\right) dy = \frac{3}{2}$

Assignment 13.2 —

(a) $A = 2\displaystyle\int_0^3 \left(t - \frac{t^3}{9}\right) 2t\, dt = 4\int_0^3 \left(t^2 - \frac{t^4}{9}\right) dt = \frac{72}{5}$

(b) $A = 4\displaystyle\int_{\pi/2}^{0} |4\sin(\theta)|\big(-\sin(\theta)\big)\, d\theta = 16\int_0^{\pi/2} \sin^2(\theta)\, d\theta = 4\pi$

(c) $A = 4\displaystyle\int_{\pi/2}^{0} |3\sin(2t)|\, 2\big(-\sin(t)\big)\, dt = 48\int_0^{\pi/2} \sin^2(t)\cos(t)\, dt = 16$

(d) $A = \displaystyle\int_0^{\pi} \big(a(1 + \cos(\theta))\big)^2\, d\theta = a^2\int_0^{\pi} \big(1 + 2\cos(\theta) + \cos^2(\theta)\big)\, d\theta = \frac{3\pi a^2}{2}$

(e) $A = \displaystyle\int_0^{\pi/4} \big(4\cos(2\theta)\big)^2\, d\theta = 16\int_0^{\pi/4} \cos^2(2\theta)\, d\theta = 2\pi$

(f) $A = 4\displaystyle\int_{\pi/2}^{0} c\sin^3(t)\, 3c\cos^2(t)\big(-\sin(t)\big)\, dt = 12c^2\int_0^{\pi/2} \sin^4(t)\cos^2(t)\, dt = \frac{3c^2\pi}{8}$

(g) $A = \dfrac{1}{2}\displaystyle\int_{2\pi}^{4\pi} a^2\theta^2\, d\theta - \frac{1}{2}\int_0^{2\pi} a^2\theta^2\, d\theta = 8a^2\pi^3$

(h) $A = \dfrac{3}{2}\displaystyle\int_{\pi/18}^{5\pi/18} \left(\big(2a\sin(3\theta)\big)^2 - a^2\right) d\theta = 3a^2\int_{\pi/18}^{5\pi/18} (4\sin^2(3\theta) - 1)\, d\theta = \frac{(3\sqrt{3} + 2\pi)a^2}{6}$

(i) $A = \dfrac{1}{2}\displaystyle\int_{-\pi/2}^{\pi/6} \left(\frac{1 + \sin(\theta)}{\sqrt{3}}\right)^2 d\theta + \frac{1}{2}\int_{\pi/6}^{\pi/2} \cos^2(\theta)\, d\theta$
$= \dfrac{1}{6}\displaystyle\int_{-\pi/2}^{\pi/6} \big(1 + 2\sin(\theta) + \sin^2(\theta)\big)\, d\theta + \frac{1}{2}\int_{\pi/6}^{\pi/2} \cos^2(\theta)\, d\theta = \frac{\pi - \sqrt{3}}{4}$

(j) $A = \frac{1}{2}\int_0^{\pi/3} 3a^2 \sin^2(3\theta)\, d\theta = \frac{a^2\pi}{4}$

(k) $A = \frac{1}{2}\int_0^{2\pi} \left(3e^{2\theta}\right)^2 d\theta = \frac{9}{2}\int_0^{2\pi} e^{4\theta}\, d\theta = \frac{9}{8}\left(e^{8\pi}-1\right)$

Assignment 13.3 —

(a) disk method: $V = \pi\int_0^2 8x\, dx = 16\pi$ (most efficient)

shell method: $V = 2\pi\int_0^4 \left(2-\frac{y^2}{8}\right) y\, dy = 16\pi$

(b) disk method: $V = 2\pi\int_0^4 2^2\, dy - 2\pi\int_0^4 \left(\frac{y^2}{8}\right)^2 dy = 32\pi - \frac{\pi}{32}\int_0^4 y^4\, dy = \frac{128\pi}{5}$

shell method: $V = 4\pi\int_0^2 x\sqrt{8x}\, dx = 8\sqrt{2}\pi\int_0^2 x\sqrt{x}\, dx = \frac{128\pi}{5}$ (most efficient)

(c) disk method: $V = \pi\int_0^1 \left(\left(\sqrt{x}\right)^2 - \left(x^2\right)^2\right) dx = \pi\int_0^1 \left(x-x^4\right) dx = \frac{3\pi}{10}$

shell method: $V = 2\pi\int_0^1 y\left(\sqrt{y}-y^2\right) dy = \frac{3\pi}{10}$ (most efficient)

(d) disk method: $V = 2\pi\int_0^4 \left(2-\frac{y^2}{8}\right)^2 dy = 2\pi\int_0^4 \left(\frac{y^4}{64}-\frac{y^2}{2}+4\right) dy = \frac{256\pi}{15}$

shell method: $V = 4\pi\int_0^2 (2-x)\sqrt{8x}\, dx = 8\sqrt{2}\pi\int_0^2 (2-x)\sqrt{x}\, dx = \frac{256\pi}{15}$ (most efficient)

(e) a) disk method: $V = \pi\int_0^3 \left(x^2 - \left(2-\sqrt{-x+4}\right)^2\right) dx + \pi\int_3^4 \left(\left(2+\sqrt{-x+4}\right)^2 - \left(2-\sqrt{-x+4}\right)^2\right) dx$

shell method: $V = 2\pi\int_0^3 y(4y-y^2-y)\, dy = 2\pi\int_0^3 y(3y-y^2)\, dy = \frac{27\pi}{2}$ (most efficient)

b) disk method: $V = \pi\int_0^3 \left(\left(4y-y^2\right)^2 - y^2\right) dy = \pi\int_0^3 \left(y^4-8y^3+15y^2\right) dy = \frac{108\pi}{5}$

(most efficient)

shell method: $V = 2\pi\int_0^3 x\left(x-2+\sqrt{-x+4}\right) dx + 2\pi\int_3^4 x\left(2\sqrt{-x+4}\right) dx$

(f) disk method: $V = \pi \int\limits_0^4 6^2\,dx - \pi \int\limits_0^4 \left(6-(4x-x^2)\right)^2\,dx = 144\pi - \pi \int\limits_0^4 \left(6-4x+x^2\right)^2\,dx = \dfrac{1408\pi}{15}$

shell method: $V = 4\pi \int\limits_0^4 \sqrt{4-y}(6-y)\,dy = \dfrac{1408\pi}{15}$ (most efficient)

Assignment 13.4 —

(a) $V = 2\pi \int\limits_0^3 (3-x)(2x)\,dx = 18\pi$

(b) a) about the x-axis: $V = 2\pi \int\limits_0^1 x^2\left(1-x^2\right)\,dx = \dfrac{4\pi}{15}$

b) about the y-axis: $V = 4\pi \int\limits_0^1 x^2\sqrt{1-x^2}\,dx = \dfrac{\pi^2}{4}$

(c) a) about the y-axis: $V = 2\pi \int\limits_0^{2\pi} (\theta-\sin(\theta))(1-\cos(\theta))(1-\cos(\theta))\,d\theta = 6\pi^3$

b) about $y=2$: $V = \pi \int\limits_0^{2\pi} 2^2\,d\theta - \pi \int\limits_0^{2\pi} \left(2-\left(1-\cos(\theta)\right)\right)^2\left(1-\cos(\theta)\right)\,d\theta$

$$= 8\pi^2 - 2\pi \int\limits_0^{\pi} \sin^2(\theta)\left(1+\cos(\theta)\right)\,d\theta = 7\pi^2$$

(d) $V = \pi \int\limits_{-1}^1 1\,dx - \pi \int\limits_{-1}^1 x^4\,dx = \dfrac{8\pi}{5}$

(e) $V = \pi \int\limits_{-5}^3 \left(\sqrt{25-x^2}\right)^2\,dx - \pi \int\limits_{-5}^3 \left(\dfrac{x+5}{2}\right)^2\,dx = \pi \int\limits_{-5}^3 \left(\dfrac{75}{4} - \dfrac{5x^2}{4} - \dfrac{5x}{2}\right)\,dx = \dfrac{320\pi}{3}$

(f) $V = 2\pi \int\limits_{\pi/2}^0 \left(c\sin^3(t)\right)^2\left(-3c\cos^2(t)\sin(t)\right)\,dt = 2\pi \int\limits_0^{\pi/2} 3c^3\sin^7(t)\cos^2(t)\,dt = \dfrac{32}{105}c^3\pi$

(g) $V = \pi \int\limits_{\pi}^0 16\cos^4(\theta)\sin^2(\theta)\left(-8\cos^2(\theta)\sin(\theta) - 4\cos^2(\theta)\sin(\theta)\right)\,d\theta$

$$= 192\pi \int\limits_0^{\pi} \cos^6(\theta)\sin^3(\theta)\,d\theta = \dfrac{256\pi}{21}$$

Remark:

If a curve K is given by $y = f(x)$, the volume when rotated about the x-axis is given by

$$V = \pi \int_{x_a}^{x_b} \left(y(x)\right)^2 \, dx.$$

This expression can be transformed to polar coordinates by using

$$\begin{cases} x = r(\theta)\cos(\theta) \\ y = r(\theta)\sin(\theta). \end{cases}$$

The resulting volume when rotated about the x-axis is given by

$$V = \pi \int_{\alpha}^{\beta} r^2 \sin^2(\theta)\left(r'\cos(\theta) - r\sin(\theta)\right) d\theta.$$

(h) For simplicity, we consider a circle with center $(a,0)$ on the x-axis and radius a. We rotate this circle about the y-axis (a tangent).

disk method: $V = 2\pi \int_0^a \left(a + \sqrt{a^2 - y^2}\right)^2 dy - 2\pi \int_0^a \left(a - \sqrt{a^2 - y^2}\right)^2 dy = 2\pi \int_0^a 4a\sqrt{a^2 - y^2}\, dy = 2\pi^2 a^3$

shell method: $V = 2\pi \int_0^{2a} 2x\sqrt{a^2 - (x-a)^2}\, dx = 2\pi^2 a^3$ h

(i) $V = 2\pi \int_0^{+\infty} (2a - x)^2 \, dy = 2\pi \int_0^{2a} (2a - x)^2 \underbrace{\frac{(3a - x)x^{1/2}}{(2a - x)^{3/2}}\, dx}_{dy} = 2\pi^2 a^3$

(j) $V = \pi \int_{\pi/2}^{0} y^2\, dx + \pi \int_{2\pi/3}^{\pi/2} y^2\, dx - \pi \int_{2\pi/3}^{\pi} y^2\, dx$

$= \pi \int_{\pi}^{0} a^2 \left(1 + \cos(\theta)\right)^2 \sin^2(\theta)\left(-a\sin(\theta)\left(1 + 2\cos(\theta)\right)\right) d\theta = \frac{8}{3} a^3 \pi$

See remarks exercise 4(c).

Assignment 13.5 —

(a) $L = 4 \int_0^a \sqrt{1 + \frac{a^{2/3} - x^{2/3}}{x^{2/3}}}\, dx = 4a^{1/3} \int_0^a x^{-1/3}\, dx = 6a$

(b) $L = 2 \int_0^{\pi/2} \sqrt{20}\, d\theta = 2\sqrt{5}\pi$

(c) $L = \int_0^{2\pi} \sqrt{a^2 \sin^2(\theta) + a^2\left(1 + \cos(\theta)\right)^2}\, d\theta = \sqrt{2} a \int_0^{2\pi} \sqrt{1 + \cos(\theta)}\, d\theta = 8a$

(d) $L = \int_0^{\pi/3} \frac{1}{\cos(y)}\, dy = \ln(\sqrt{3}+2)$

(e) $L = \int_{1/2}^{2} \sqrt{\frac{1}{\theta^2} + \left(-\frac{1}{\theta^2}\right)^2}\, d\theta = \int_{1/2}^{2} \frac{\sqrt{1+\theta^2}}{\theta^2}\, d\theta = \frac{\sqrt{5}}{2} + \ln\left(\frac{3+\sqrt{5}}{2}\right)$

(f) $L = \int_{-\pi/2}^{\pi/2} \sqrt{a^2 \sin^2\left(\frac{\theta}{2}\right)\cos^{-6}\left(\frac{\theta}{2}\right) + a^2\cos^{-4}\left(\frac{\theta}{2}\right)}\, d\theta = 2a\int_0^{\pi/2} \frac{1}{\cos^3\left(\frac{\theta}{2}\right)}\, d\theta = 2a\left(\sqrt{2} + \ln\left(1+\sqrt{2}\right)\right)$

(g) $L = 2\int_0^{\pi} \sqrt{4\left(\sin(2t)-\sin(t)\right)^2 + 4\left(\cos(t)-\cos(2t)\right)^2}\, dt = 4\sqrt{2}\int_0^{\pi} \sqrt{1-\cos(t)}\, dt = 16$

(h) $L = 2\int_0^{4} \sqrt{1+\left(\frac{3\sqrt{x}}{2}\right)^2}\, dx = \int_0^{4} \sqrt{4+9x}\, dx = \frac{16}{27}\left(10\sqrt{10}-1\right)$

(i) $L = 2\int_0^{3} \sqrt{1+\frac{1}{4}\left(\frac{x-1}{\sqrt{x}}\right)^2}\, dx = \int_0^{3} \left(\sqrt{x}+\frac{1}{\sqrt{x}}\right) dx = 4\sqrt{3}$

(j) $L = \int_1^{3} \sqrt{r'^2+r^2}\, \underbrace{\frac{1}{2}\left(1-\frac{1}{r^2}\right) dr}_{d\theta} \quad \text{met } r'(\theta) = \frac{dr}{d\theta} = \frac{2}{1-\frac{1}{r^2}}$

$\Rightarrow L = \frac{1}{2}\int_1^{3} \frac{\sqrt{r^4+2r^2+1}}{r}\, dr = 2 + \frac{\ln(3)}{2}$

Assignment 13.6 —

(a) $SA = 2\pi\int_0^{\pi} \sin(x)\sqrt{1+\cos^2(x)}\, dx = 2\pi\left(\sqrt{2} - \ln\left(\sqrt{2}-1\right)\right)$

(b) $SA = 4\pi\int_0^{4} \sqrt{\frac{16-x^2}{4}}\sqrt{1+\frac{x^2}{4(16-x^2)}}\, dx = \pi\int_0^{4} \sqrt{64-3x^2}\, dx = \frac{8\pi}{9}\left(9+4\sqrt{3}\pi\right)$

(c) $SA = 2\pi\int_0^{6} \left(3-\frac{y}{2}\right)\sqrt{1+\frac{1}{4}}\, dy = 9\pi\sqrt{5}$

(d) a) about the y-axis: $SA = 2\pi\int_0^{1} y^3\sqrt{1+9y^4}\, dy = \frac{1}{27}(10\sqrt{10}-1)\pi$

b) the line $x=1$: $SA = 2\pi\int_0^{1} (1-y^3)\sqrt{1+9y^4}\, dy$

(e) $SA = 2\pi\int_0^{1} \sqrt{\frac{x^2-x^4}{8}}\sqrt{1+\frac{\left(1-2x^2\right)^2}{8(1-x^2)}}\, dx = \frac{\pi}{4}\int_0^{1} (3x-2x^3)\, dx = \frac{\pi}{4}$

(f) a) about the x-axis:

$$SA = 4\pi \int_0^{\pi/2} \sin^3(t)\, 3\cos(t)\sin(t)\,dt = 12\pi \int_0^{\pi/2} \sin^4(t)\cos(t)\,dt = \frac{12\pi}{5}$$

b) about the line $y = -1$:

$$SA = 4\pi \int_0^{\pi/2} \left(1 + |\sin^3(t)|\right) 3\cos(t)\sin(t)\,dt + 4\pi \int_{-\pi/2}^{0} \left(1 - |\sin^3(t)|\right) 3\cos(t)\sin(t)\,dt$$

$$= 12\pi \int_{-\pi/2}^{\pi/2} \left(1 + \sin^3(t)\right)|\sin(t)|\cos(t)\,dt = 12\pi$$

(g) $$SA = 4\pi \int_0^{\pi/4} a\sqrt{\cos(2\theta)}\sin(\theta)\sqrt{a^2\cos(2\theta) + \frac{a^2\sin^2(2\theta)}{\cos(2\theta)}}\,d\theta = 4a^2\pi \int_0^{\pi/4} \sin(\theta)\,d\theta$$

$$= 2(2 - \sqrt{2})\pi a^2$$

(h) a) about the x-axis:

$$SA = 2\pi \int_0^{2\pi} a\left(1 - \cos(\theta)\right)\sqrt{2}\,a\sqrt{1 - \cos(\theta)}\,d\theta = 2\sqrt{2}\pi a^2 \int_0^{2\pi} \left(1 - \cos(\theta)\right)^{3/2}\,d\theta = \frac{64\pi a^2}{3}$$

b) about the line $x = a\pi$:

$$SA = 2\pi \int_0^{\pi} \left(a\pi - a\left(\theta - \sin(\theta)\right)\right)\sqrt{2}\,a\sqrt{1 - \cos(\theta)}\,d\theta$$

$$= 2\sqrt{2}\pi a^2 \int_0^{\pi} \left(\pi - \left(\theta - \sin(\theta)\right)\right)\sqrt{1 - \cos(\theta)}\,d\theta = 8\pi a^2\left(\pi - \frac{4}{3}\right)$$

(i) $$SA = 2\pi \int_0^{\pi} a(1 + \cos(\theta))\sin(\theta)\, \underbrace{\sqrt{2}a\sqrt{1 + \cos(\theta)}\,d\theta}_{\text{Oef 5(f)}} = 2\sqrt{2}a^2\pi \int_0^{\pi} (1 + \cos(\theta))^{3/2}\sin(\theta)\,d\theta$$

$$= \frac{32\pi a^2}{5}$$

Assignment 13.7 —

(a) $$V_{kir} = \pi \int_0^{5/2} \left(16 - (y-4)^2\right)dy - \pi \int_1^{5/2} \left(1 - (y-2)^2\right)dy = \frac{56}{3}\pi$$

(b) $$SA = 2\pi \int_{5/2}^{3} \sqrt{1 - (y-2)^2}\,\frac{dy}{\sqrt{1 - (y-2)^2}} = 2\pi \int_{5/2}^{3} dy = \pi$$

Chapter 14

Assignment 14.1 —

(a) $\left\{\dfrac{2n^2}{n^2+1}\right\} = \left\{2 - \dfrac{2}{n^2+1}\right\}$ is bounded, positive, increasing and converges to 2.

(b) $\left\{\dfrac{(-1)^n\, n}{e^n}\right\}$ is bounded, alternating, and converges to 0.

(c) $\left\{\dfrac{e^n}{\pi^{n/2}}\right\}$ is bounded from below, positive, increasing and diverging to $+\infty$.

(d) $\left\{n\cos\left(\dfrac{n\pi}{2}\right)\right\} = \{0,-2,0,4,0,-6,\ldots\}$ is divergent.

(e) $\left\{\dfrac{(n!)^2}{(2n)!}\right\}$ is bounded, positive, decreasing and converges to 0.

Assignment 14.2 —

(a) $\lim\limits_{n\to+\infty} \dfrac{e^n - e^{-n}}{e^n + e^{-n}} = \lim\limits_{n\to+\infty} \dfrac{1-e^{-2n}}{1+e^{-2n}} = 1$

(b) $\lim\limits_{n\to+\infty} \left(\dfrac{n-3}{n}\right)^n = \lim\limits_{n\to+\infty} \left(1 + \dfrac{-3}{n}\right)^n = e^{-3}$

(c) $\lim\limits_{n\to+\infty} \left(\dfrac{n-1}{n+1}\right)^n = \lim\limits_{n\to+\infty} \left(\dfrac{n-1}{n}\right)^n \left(\dfrac{n}{n+1}\right)^n = \lim\limits_{n\to+\infty} \left(\dfrac{n-1}{n}\right)^n \left(\dfrac{n+1}{n}\right)^{-n} = e^{-2}$

(d) $\lim\limits_{n\to+\infty} \dfrac{n}{\ln(n+1)} = +\infty$

(e) $\lim\limits_{n\to+\infty} \left(n - \sqrt{n^2-4n}\right) = 2$

(f) $\lim\limits_{n\to+\infty} \dfrac{\pi^n}{1+2^{2n}} = 0 \qquad 0 < a_n < (\pi/4)^n, \quad \pi/4 < 1 \ \Rightarrow\ (\pi/4)^n \to 0$ als $n \to +\infty$

Assignment 14.3 —

(a) $\lim\limits_{n\to+\infty} \arctan\left(\dfrac{2n}{n+1}\right) = \arctan(2) \quad \Rightarrow$ convergence to $\arctan(2)$

(b) The sine function is periodic, so it is divergent

(c) $\lim\limits_{n\to+\infty} \dfrac{2^n}{n!} = 0 \quad \Rightarrow$ convergence to 0

Assignment 14.4 —

(a) The increasing nature and boundedness from above can be proved by induction. An increasing sequence that is bounded from above is convergent.

$$a = \lim_{n\to+\infty} a_{n+1} = \lim_{n\to+\infty} \sqrt{1+2a_n} = \sqrt{1+2a} \quad \Rightarrow \quad a = \sqrt{1+2a} \quad \Rightarrow \quad a = 1+\sqrt{2}$$

(b) The increasing nature and boundedness from above can be proved by induction. An increasing sequence that is bounded from above is convergent.

$$a = \lim_{n\to+\infty} a_{n+1} = \lim_{n\to+\infty} \sqrt{15+2a_n} = \sqrt{15+2a} \quad \Rightarrow \quad a = \sqrt{15+2a} \quad \Rightarrow \quad a = 5$$

Assignment 14.5 —

(a) If $\alpha \neq 0$, the series is divergent. If $\alpha = 0$, the series converges to 0.

(b) The series is divergent. $\lim_{n\to+\infty} a_n = +\infty \neq 0$.

(c) The series is convergent. Compare to the p-series with $p = 2$.

(d) The sequence is convergent. Apply ratio test ($L = 1/4$).

(e) The series is divergent. Compare with harmonic series or apply integral test.

(f) The series is divergent. Apply ratio test.

(g) The sequence is convergent. Apply integral test.

(h) The sequence is convergent. Apply ratio test ($L = 1/4$).

(i) The sequence is convergent. Compare with p-series with $p = 2$.

(j) The series is divergent. Compare with harmonic series.

(k) The series is divergent. Apply ratio test ($L = +\infty$).

(l) The sequence is convergent. Apply ratio test ($L = 0$).

(m) The series is divergent. $\lim_{n\to+\infty} a_n = e^{-2} \neq 0$.

(n) The series is convergent. Limit equation test with p-series with $p = 3/2$.

(o) The sequence is divergent. Apply integral test.

(p) The sequence is convergent. Apply ratio test ($L = 1/3$).

(q) The series is divergent. Compare with harmonic series.

(r) The sequence is convergent. Apply root test

(s) The series is divergent. $\lim_{n\to+\infty} a_n = 1 \neq 0$ or apply limit comparison test.

Assignment 14.6 —

(a) Absolute convergence.

(b) Conditional convergence.

(c) Absolute convergence.

(d) Absolute convergence.

(e) Divergent.

(f) Conditional convergence.

Assignment 14.7 —

(a) Absolute convergence: $-\sqrt{5} < x < \sqrt{5}$. Divergence: $x < -\sqrt{5}$ of $x > \sqrt{5}$.

End points: $x = \sqrt{5} \rightarrow$ divergence $x = -\sqrt{5} \rightarrow$ div.

(b) The convergence radius is equal to 0. The power series is divergent for all values of x.

(c) The convergence interval is equal to $+\infty$. The power sequence is absolute convergent for all x.

(d) Absolute convergence: $-1 < x < 3$. Divergence: $x < -1$ of $x > 3$.

End points: $x = 3 \to$ convergence, $x = -1 \to$ Absolute convergence

(e) Absolute convergence: $-4 < x < 0$. Divergence: $x < -4$ or $x > 0$.

End points: $x = 0 \to$ divergence, $x = -4 \to$ conditional convergence

(f) Absolute convergence: $0 < x < 4$. Divergence: $x < 0$ or $x > 4$.

End points: $x = 4 \to$ convergence, $x = 0 \to$ Absolute convergence

(g) Absolute convergence: $-2 < x < 0$. Divergence: $x < -2$ or $x > 0$.

Bounday points: $x = 0 \to$ divergence, $x = -2 \to$ conditional convergence

(h) Absolute convergence: $2 < x < 4$. Divergence: $x < 2$ or $x > 4$.

End points: $x = 4 \to$ convergence, $x = 2 \to$ Absolute convergence

(i) Absolute convergence: $-3 < x < 7$. Divergence: $x < -3$ or $x > 7$.

End points: $x = 7 \to$ divergence, $x = -3 \to$ conditional convergence

Assignment 14.8 —

(a) $$\frac{1}{(2-x)^2} = \sum_{n=0}^{+\infty} \frac{(n+1)x^n}{2^{n+2}} = \frac{1}{4} + \frac{2x}{2^3} + \frac{3x^2}{2^4} + \frac{4x^3}{2^5} + \frac{5x^4}{2^6} + \ldots \qquad (-2 < x < 2)$$

(b) $$\ln(2-x) = \ln(2) - \sum_{n=1}^{+\infty} \frac{x^n}{n\,2^n} = \ln(2) - \left(\frac{x}{2} + \frac{x^2}{2.2^2} + \frac{x^3}{3.2^3} + \frac{x^4}{4.2^4} + \ldots\right) \qquad (-2 \le x < 2)$$

(c) $$\frac{x^3}{1-2x^2} = \sum_{n=0}^{+\infty} 2^n x^{2n+3} = x^3\left(1 + 2x^2 + 4x^4 + 8x^6 + \ldots\right) \qquad \left(-\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2}\right)$$

(d) $$\ln(x) = \ln(4) - \sum_{n=1}^{+\infty} \frac{(-1)^n (x-4)^n}{n\,4^n}$$
$$= \ln(4) - \left(-\frac{x-4}{4} + \frac{(x-4)^2}{2.4^2} - \frac{(x-4)^3}{3.4^3} + \frac{(x-4)^4}{4.4^4} - \frac{(x-4)^5}{5.4^5} + \ldots\right) \qquad (0 < x \le 8)$$

(e) $$\frac{1-x}{1+x} = 1 + 2\sum_{n=1}^{+\infty} (-x)^n = 1 + 2\left(-x + x^2 - x^3 + x^4 - x^5 + \ldots\right) \qquad (-1 < x < 1)$$

Assignment 14.9 —

(a) $$\cos^2\left(\frac{x}{2}\right) = 1 + \frac{1}{2}\sum_{n=1}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!} \qquad (x \in \mathbb{R})$$

(b) $$\frac{e^{2x^2} - 1}{x^2} = \sum_{n=0}^{+\infty} \frac{2^{n+1}}{(n+1)!} x^{2n} \qquad (x \in \mathbb{R}\backslash\{0\})$$

(c) $$\sinh(x) - \sin(x) = \sum_{n=0}^{+\infty} \frac{x^{2n+1}}{(2n+1)!} - \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = 2\sum_{n=0}^{+\infty} \frac{x^{4n+3}}{(4n+3)!} \qquad (x \in \mathbb{R})$$

(d) $\cosh(x)-\cos(x)=\sum_{n=0}^{+\infty}\frac{x^{2n}}{(2n)!}-\sum_{n=0}^{+\infty}(-1)^n\frac{x^{2n}}{(2n)!}=2\sum_{n=0}^{+\infty}\frac{x^{4n+2}}{(4n+2)!}\qquad(x\in\mathbb{R})$

(e) $x^2\sin\left(\frac{x}{3}\right)=x^2\sum_{n=0}^{+\infty}\frac{(-1)^n x^{2n+1}}{3^{2n+1}(2n+1)!}=\sum_{n=0}^{+\infty}\frac{(-1)^n x^{2n+3}}{3^{2n+1}(2n+1)!}\qquad(x\in\mathbb{R})$

(f) $(1+x)^{\frac{1}{2}}\cos(x)=1+\frac{x}{2}-\frac{5x^2}{8}-\frac{3x^3}{16}+\frac{25x^4}{384}+\frac{13x^5}{768}+\dots\qquad(-1\leq< x\leq 1)$

Assignment 14.10 —

(a) $\sin(x)-\cos(x)=\sqrt{2}\sum_{n=0}^{+\infty}\frac{(-1)^n}{(2n+1)!}\left(x-\frac{\pi}{4}\right)^{2n+1}\qquad(x\in\mathbb{R})$

(b) $x\ln(x)=(x-1)+\sum_{n=2}^{+\infty}\frac{(-1)^n}{n(n-1)}(x-1)^n\qquad(0<x\leq 2)$

(c) $xe^x=-\frac{2}{e^2}+\frac{1}{e^2}\sum_{n=1}^{+\infty}\frac{n-2}{n!}(x+2)^n\qquad(x\in\mathbb{R})$

(d) $\ln(2+x)=\ln 4+\sum_{n=1}^{+\infty}(-1)^{n-1}\frac{(x-2)^n}{n4^n}\qquad(-2<x\leq 6)$

(e) $\cos^2(x)=\frac{1}{2}+\frac{1}{2\sqrt{2}}+\frac{1}{2\sqrt{2}}\sum_{n=1}^{+\infty}(-1)^n\left(\frac{2^{2n-1}}{(2n-1)!}\left(x-\frac{\pi}{8}\right)^{2n-1}+\frac{2^{2n}}{(2n)!}\left(x-\frac{\pi}{8}\right)^{2n}\right)\qquad(x\in\mathbb{R})$

(f) $\frac{1}{x^2}=\frac{1}{4}\sum_{n=0}^{+\infty}\frac{(n+1)(x+2)^n}{2^n}\qquad(-4<x<0)$

(g) $\frac{1}{x}=\sum_{n=1}^{+\infty}(-1)^{n+1}(x-1)^{n-1}\qquad(0<x\leq 2)$

Assignment 14.11 —

(a) $\ln(1+x)=\sum_{n=0}^{+\infty}(-1)^n\frac{x^{n+1}}{n+1}=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots$

(b) $\ln(2)=\sum_{n=0}^{+\infty}\frac{(-1)^n}{n+1}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots$

(c) $\ln\left(\frac{1+x}{1-x}\right)=\sum_{n=0}^{+\infty}\frac{2x^{2n+1}}{2n+1}=2x+\frac{2x^3}{3}+\frac{2x^5}{5}+\cdots$

(d) $\ln(2)=\sum_{n=0}^{+\infty}\frac{2(1/3)^{2n+1}}{2n+1}=\frac{2}{3}+\frac{2}{3}.\frac{1}{27}+\frac{2}{5}.\frac{1}{243}+\cdots$

$\frac{2}{3}+\frac{2}{3}.\frac{1}{27}=\frac{56}{81}>\frac{1}{2}\quad\Rightarrow$ faster convergence.

Assignment 14.12 —

(a) $\int_0^{\sqrt{\pi}} \sin\left(x^2\right)\, dx = \int_0^{\sqrt{\pi}} \left(x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \ldots\right) dx = \frac{\pi^{3/2}}{3} - \frac{\pi^{7/2}}{7.3!} + \frac{\pi^{11/2}}{11.5!} - \frac{\pi^{15/2}}{15.7!} + \ldots$

(b) $\int_0^{\pi^2/4} \cos\left(\sqrt{x}\right)\, dx = \int_0^{\pi^2/4} \left(1 - \frac{x}{2!} + \frac{x^2}{4!} - \frac{x^3}{6!} + \ldots\right) dx = \frac{\pi^2}{4} - \frac{\pi^4}{4^2.2.2!} + \frac{\pi^6}{4^3.3.4!} - \frac{\pi^8}{4^4.4.6!} + \ldots$

Assignment 14.13 — $f(x) = \int_0^x e^{-u^2}\, du = \int_0^x \left(1 - u^2 + \frac{u^4}{2!} - \frac{u^6}{3!} + \ldots\right) du = x - \frac{x^3}{3!} + \frac{x^5}{5.2!} - \frac{x^7}{7.3!} + \ldots$

Chapter 15

Assignment 15.1 — The three functions $\vec{\boldsymbol{r}}_i(t)$ are on the x-axis $(y \geq 0)$. The points on the 3 curves are at distance 1 from the origin. Calculate the norm of the vector functions to verify this. The vector functions $\vec{\boldsymbol{r}}_i(t)$ correspond to the semicircle $y = \sqrt{1-x^2}$, travelled from left to right. To do this, calculate the values at start and end points.

Chapter 15

Assignment 15.2 —

(a) $\left(11, 74, \sin(5)\right)$

(b) $\left(e^3, 0\right)$

(c) $(1, e)$

(d) $(2t, 1, 0)$

Chapter 15

Assignment 15.3 —

(a) $\left(-\frac{1}{t^2}, \frac{5}{(3t+1)^2}, \frac{1}{\cos^2(t)}\right)$

(b) $\left(2t\sin(t) + t^2\cos(t), 6t^2 + 10t\right)$

(c) $2t\sin(t) + \left(t^2+1\right)\cos(t) + 4t + 3$

(d) $\left(-1, -2t + \cos(t), 6t^2 + 10t + 2 - (t-1)\cos(t) - \sin(t)\right)$

Chapter 15

Assignment 15.4 —

(a) $t = 2k\pi, \quad k \in \mathbb{Z}$

(b) $t = 1$

(c) $t = \dfrac{3\pi}{4} + k\pi, \quad k \in \mathbb{Z}$

(d) $t = \pm 1$

Chapter 15

Assignment 15.5 —

(a) $\left(\dfrac{t^4}{4}, \sin(t), e^t(t-1)\right) + \vec{\boldsymbol{c}}$

(b) $\big(\arctan(t), \tan(t)\big) + \vec{\boldsymbol{c}}$

(c) $\left(e^{\sin(t)}, \dfrac{t^2}{4} - \dfrac{t\sin(2t)}{4} - \dfrac{\cos(2t)}{8}, -t\right) + \vec{\boldsymbol{c}}$

(d) $\left(0, \dfrac{2}{3}\right)$

(e) $\left(1 - \dfrac{1}{\sqrt{e}}, \sqrt{e} - 1, e - 1\right)$

Chapter 15

Assignment 15.6 —

(a) $\vec{\boldsymbol{r}}(t) = \left(t^2 - t + 5, \dfrac{3}{2}t^2 - t - \dfrac{5}{2}\right)$

(b) $\vec{\boldsymbol{r}}(t) = \big(1 - \cos(t), \sin(t)\big)$

(c) $\vec{\boldsymbol{r}}(t) = \big(10t, -16t^2 + 50t\big)$

Chapter 15

Assignment 15.7 —

(a) $\vec{\boldsymbol{r}}(t) = (1, t)$, $\quad \vec{\boldsymbol{v}}(t) = (0, 1)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = 1$, $\quad \vec{\boldsymbol{a}}(t) = (0, 0)$

path: $x = 1$ in the xy-plane

(b) $\vec{\boldsymbol{r}}(t) = \left(0, t^2, t\right)$, $\quad \vec{\boldsymbol{v}}(t) = (0, 2t, 1)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = \sqrt{4t^2+1}$, $\quad \vec{\boldsymbol{a}}(t) = (0, 2, 0)$

path: $y = z^2$ in the plane $x = 0$

(c) $\vec{\boldsymbol{r}}(t) = (1, t, t)$, $\quad \vec{\boldsymbol{v}}(t) = (0, 1, 1)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = \sqrt{2}$, $\quad \vec{\boldsymbol{a}}(t) = (0, 0, 0)$

path: the intersection of the planes $x = 1$ and $y = z$

(d) $\vec{\boldsymbol{r}}(t) = \left(t^2, -t^2, 1\right)$, $\quad \vec{\boldsymbol{v}}(t) = (2t, -2t, 0)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = 2\sqrt{2}t$, $\quad \vec{\boldsymbol{a}}(t) = (2, -2, 0)$

path: the ray $\begin{cases} x = -y \geq 0 \\ z = 1 \end{cases}$

(e) $\vec{\boldsymbol{r}}(t) = \big(3\cos(t), 4\cos(t), 5\sin(t)\big)$, $\quad \vec{\boldsymbol{v}}(t) = \big(-3\sin(t), -4\sin(t), 5\cos(t)\big)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = 5$,

$\vec{\boldsymbol{a}}(t) = \big(-3\cos(t), -4\cos(t), -5\sin(t)\big)$

path: circle that is the cross section of the sphere $x^2 + y^2 + z^2 = 25$ and the plane $4x = 3y$

(f) $\vec{\boldsymbol{r}}(t) = \big(3\cos(t), 4\sin(t), t\big)$, $\quad \vec{\boldsymbol{v}}(t) = \big(-3\sin(t), 4\cos(t), 1\big)$, $\quad \|\vec{\boldsymbol{v}}(t)\| = \sqrt{10 + 7\cos^2(t)}$,

$\vec{\boldsymbol{a}}(t) = \big(-3\cos(t), -4\sin(t), 0\big)$

path: a spiral around the elliptical cylinder $\dfrac{x^2}{9} + \dfrac{y^2}{16} = 1$

Assignment 15.8 —

(a) Parameterization C: $\vec{\boldsymbol{r}}(t) = \left(x(t), x^2(t)\right) = x(t)\hat{\boldsymbol{\imath}} + x^2(t)\hat{\boldsymbol{\jmath}}$

(b) $\vec{\boldsymbol{v}}(t) = \dfrac{dx(t)}{dt}\,(\hat{\boldsymbol{\imath}} + 2x\hat{\boldsymbol{\jmath}})$

(c) $\vec{\boldsymbol{a}}(t) = \dfrac{d^2x(t)}{dt^2}\,(\hat{\boldsymbol{\imath}} + 2x\hat{\boldsymbol{\jmath}}) + 2\left(\dfrac{dx}{dt}\right)^2 \hat{\boldsymbol{\jmath}}$

(d) $\left\|\vec{\boldsymbol{v}}(t)\right\| = \left|\dfrac{dx}{dt}\right| \sqrt{1+4x^2} = 5 \;\Rightarrow\; \dfrac{dx}{dt} = \dfrac{5}{\sqrt{1+4x^2}} \;\overset{x=1}{\Rightarrow}\; \dfrac{dx}{dt} = \sqrt{5}$

(e) $\dfrac{d^2x(t)}{dt^2} = -\dfrac{100x}{(1+4x^2)^2} \;\overset{x=1}{\Rightarrow}\; \dfrac{d^2x(t)}{dt^2} = -4$

(f) $\vec{\boldsymbol{v}}(t) \overset{x=1}{=} \sqrt{5}\hat{\boldsymbol{\imath}} + 2\sqrt{5}\hat{\boldsymbol{\jmath}} = \left(\sqrt{5}, 2\sqrt{5}\right)$

(g) $\vec{\boldsymbol{a}}(t) \overset{x=1}{=} -4\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} = (-4, 2)$

Assignment 15.9 — Prove yourself

Assignment 15.10 —

(a) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{20t^2 - 4t + 1}}\,(4t, 2t - 1)$

(b) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1 + 16t^2 + 81t^4}}\left(1, -4t, 9t^2\right)$

(c) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1+t^2+t^4}}\left(1, t, t^2\right)$

(d) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1+\sin^2(t)}}\left(\cos(2t), \sin(2t), -\sin(t)\right)$

(e) $\widehat{\boldsymbol{T}}(t) = \left(-\cos(t), \sin(t)\right) \quad \Rightarrow \quad \widehat{\boldsymbol{T}}\left(\dfrac{\pi}{6}\right) = \left(-\dfrac{\sqrt{3}}{2}, \dfrac{1}{2}\right)$

Assignment 15.11 —

(a) $\widehat{\boldsymbol{N}}(t) = \left(\dfrac{3}{5}, -\dfrac{4}{5}\right)$

(b) $\widehat{\boldsymbol{N}}(t) = \left(-\cos(t), -\sin(t)\right)$

(c) $\widehat{\boldsymbol{N}}(t) = \left(\dfrac{e^{-t}}{\left(e^{2t}+e^{-2t}\right)^2}, \dfrac{e^{t}}{\left(e^{2t}+e^{-2t}\right)^2}\right)$

(d) $\widehat{\boldsymbol{N}}(t) = \left(0, -\sin(t), -\cos(t)\right)$

(e) $\widehat{\boldsymbol{N}}(t) = \left(\sin(t), \cos(t)\right) \quad \Rightarrow \quad \widehat{\boldsymbol{N}}\left(\dfrac{\pi}{6}\right) = \left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right)$

(f) $\widehat{\boldsymbol{N}}(t) = (-\cos(t), -\sin(t), 0)$

Assignment 15.12 —

(a) $L = \displaystyle\int_0^1 t\sqrt{8+9t^2}\, dt = \dfrac{17\sqrt{17}-16\sqrt{2}}{27} \qquad$ (assume $8+9t^2 = u^2$)

(b) $L = \displaystyle\int_0^{2\pi} \sqrt{2e^{2t}+1}\, dt = \sqrt{2e^{4\pi}+1} - \sqrt{3} + \ln\left(\sqrt{2e^{4\pi}+1}-1\right) - 2\pi - \ln\left(\sqrt{3}-1\right)$

(assume $2e^{2t}+1 = u^2$)

(c) $L = \displaystyle\int_{-1}^{0} (-t)\sqrt{9t^2+4}\, dt + \int_0^2 t\sqrt{9t^2+4}\, dt = \dfrac{1}{27}\left(13^{3/2}+40^{3/2}-16\right) \qquad$ (stel $9t^2+4 = u^2$)

Assignment 15.13 —

(a) $\vec{\boldsymbol{r}}(s) = \left(\dfrac{2s}{3}, \dfrac{s}{3}, -\dfrac{2s}{3}\right)$

(b) $\vec{\boldsymbol{r}}(s) = \left(7\cos\left(\dfrac{s}{7}\right), 7\sin\left(\dfrac{s}{7}\right)\right)$

(c) $\vec{\boldsymbol{r}}(s) = \left(3\cos\left(\dfrac{s}{\sqrt{13}}\right), 3\sin\left(\dfrac{s}{\sqrt{13}}\right), \dfrac{2s}{\sqrt{13}}\right)$

(d) $\vec{\boldsymbol{r}}(s) = \left(\dfrac{s+\sqrt{s^2+4}}{2}, \sqrt{2}\ln\left(\dfrac{s+\sqrt{s^2+4}}{2}\right), \dfrac{-2}{s+\sqrt{s^2+4}}\right) \quad$ met $s(t) = e^t - e^{-t}$

Assignment 15.14 — $\vec{r}(s) = \left(a\left(4\arcsin\left(\sqrt{\frac{s}{8a}}\right) - \sin\left(4\arcsin\left(\sqrt{\frac{s}{8a}}\right)\right)\right), a\left(1 - \cos\left(4\arcsin\left(\sqrt{\frac{s}{8a}}\right)\right)\right)\right)$

met $s(\theta) = 8a\sin^2\left(\frac{\theta}{4}\right)$

Assignment 15.15 —

(a) $\kappa(x) = \dfrac{2}{(1+4x^2)^{3/2}} \quad \Rightarrow \kappa(0) = 2 \quad$ and $\quad \kappa(\sqrt{2}) = 2/27$

The radii of curvature at $x = 0$ and $x = \sqrt{2}$ are $1/2$ and $27/2$.

(b) $\kappa(x) = \dfrac{|\cos(x)|}{(1+\sin^2(x))^{3/2}} \quad \Rightarrow \kappa(0) = 1 \quad$ and $\quad \kappa(\pi/2) = 0$

The radii of curvature at $x = 0$ and $x = \pi/2$ are 1 and infinite.

(c) $\kappa(x) = \dfrac{\left|2\dfrac{\tan(x)}{\cos^2(x)}\right|}{(1+\cos^{-4}(x))^{3/2}} \quad \Rightarrow \kappa(\pi/4) = 4/5\sqrt{5}$

The radius of curvature at $x = \pi/4$ is $5\sqrt{5}/4$.

(d) $\kappa(t) = \dfrac{\left\|\left(\frac{4}{t^3}, 0, \frac{4}{t^3}\right)\right\|}{\left\|\left(2, -\frac{1}{t^2}, -2\right)\right\|^3} = \dfrac{4\sqrt{2}t^3}{\left(8t^4+1\right)^{3/2}}$. In $(2,1,-2)$ is $t = 1$. $\quad \Rightarrow \kappa(1) = \dfrac{4\sqrt{2}}{27}$

The radius of curvature at $t = 1$ is $\dfrac{27}{4\sqrt{2}}$.

(e) $\kappa(t) = \dfrac{\left\|(-2, 6t, -6t^2)\right\|}{\left\|(3t^2, 2t, 1)\right\|^3} = \dfrac{\sqrt{4+36t^2+36t^4}}{(9t^4+4t^2+1)^{3/2}} \quad \Rightarrow \kappa(1) = \dfrac{2\sqrt{19}}{14^{3/2}}$

The radius of curvature at $t = 1$ is $\dfrac{14^{3/2}}{2\sqrt{19}}$.

(f) The given curve has a vertical tangent ($y' = +\infty$) in $(\pm 2, 0)$, which makes $\kappa(x) = 0$. The radius of curvature at $x = 2$ equals the distance between the origin and $(2,0)$, being 2.

(g) $\kappa(t) = \dfrac{|-18(t^2+1)|}{(3t^2+3)^3} \quad \Rightarrow \kappa(1) = 1/6$

The radius of curvature at $t = 1$ is 6.

(h) $\kappa(t) = \dfrac{\left|9\sin(t)\sin(3t) + 3\cos(t)\cos(3t)\right|}{\left(\sin^2(t) + 9\cos^2(t)\right)^{3/2}} \quad \Rightarrow \kappa(0) = 1/9$

The radius of curvature at $t = 0$ is 9.

Assignment 15.16 —

(a) $\kappa(x) = \dfrac{\left|2(x^2+1)^4\left(4x^2(x^2+1)^{-1} - 1\right)\right|}{\left((x^2+1)^4 + 4x^2\right)^{3/2}}$.

The radius of curvature is $\dfrac{\left((x^2+1)^4+4x^2\right)^{3/2}}{\left|2(x^2+1)^4\left(4x^2(x^2+1)^{-1}-1\right)\right|}$.

(b) $\kappa(x)=1$. The radius of curvature is 1.

(c) $\kappa(t)=\dfrac{\left\|\left(-\cos(t),\dfrac{\sin(t)}{\sqrt{2}},\dfrac{-\sin(t)}{\sqrt{2}}\right)\right\|}{\left\|\left(-\sqrt{2}\sin(t),-\cos(t),\cos(t)\right)\right\|}=\dfrac{\sqrt{2}}{2}$. The radius of curvature is $\sqrt{2}$.

(d) $\kappa(x)=\dfrac{e^x}{(1+e^{2x})^{3/2}}$. The radius of curvature is $\dfrac{(1+e^{2x})^{3/2}}{e^x}$.

(e) $\kappa(\theta)=\dfrac{3a^2(1-\cos(\theta))}{\left(2a^2(1-\cos(\theta))\right)^{3/2}}=\dfrac{3}{2\sqrt{2ar}}$. The radius of curvature is $\dfrac{2\sqrt{2ar}}{3}$.

(f) $\kappa(t)=\dfrac{2}{\left(3\sin^2(t)+1\right)^{3/2}}$.

The radius of curvature is $\dfrac{\left(3\sin^2(t)+1\right)^{3/2}}{2}$.

(g) $\kappa(t)=\dfrac{\left|-\frac{1}{\sin^2(t)}\right|}{(1+\cot^2(t))^{3/2}}=\sin(t)$. The radius of curvature is $\csc(t)$.

Assignment 15.17 — Prove yourself

Assignment 15.18 — $\kappa(\theta)=\dfrac{\left|r^2\cos(\theta)-1\right|}{\left(2r^2(1-\cos(\theta))\right)^{3/2}}=\dfrac{1}{r2^{3/2}(1-\cos(\theta))^{1/2}}\quad\Rightarrow\kappa(\pi/2)=\dfrac{1}{2\sqrt{2}r}$

The radius of curvature at $\theta=\dfrac{\pi}{2}$ is $2\sqrt{2}r$.

Assignment 15.19 —

(a) $\vec{\boldsymbol{r}}(y)=\left(\sqrt{a^2-y^2},y\right)$ with $0\le y\le a$

(b) $\vec{\boldsymbol{r}}(\phi)=\left(a\sin(\phi),-a\cos(\phi)\right)$ with $\dfrac{\pi}{2}\le\phi\le\pi$

(c) $\vec{\boldsymbol{r}}(s)=\left(a\sin\left(\dfrac{s}{a}\right),a\cos\left(\dfrac{s}{a}\right)\right)$ with $0\le s\le\dfrac{a\pi}{2}$

Chapter 16

Assignment 16.1 —

(a) $\operatorname{dom} f=\left\{(x,y)|\ x^2+y^2<4\right\}$ and $\operatorname{im} f=\mathbb{R}_0^+$

(b) $\mathrm{dom} f = \left\{(x,y) | \, x \in \mathbb{R}_0^+ \ \wedge \ y \in \mathbb{R} \backslash \{k\pi, k \in \mathbb{Z}\}\right\}$ and $\mathrm{im} f = \mathbb{R}$

(c) $\mathrm{dom} f = \left\{(x,y) | \, x \in \mathbb{R}^+ \backslash \{0, 1/2\} \ \wedge \ y \in [-1,1]\right\}$ and $\mathrm{im} f = \mathbb{R}_0$

(d) $\mathrm{dom} f = \left\{(x,y) | \, x^2 + y^2 \leq 1\right\}$ and $\mathrm{im} f = [0,1]$

(e) $\mathrm{dom} f = \left\{(x,y) | \, x \in \mathbb{R} \ \wedge \ y \in \mathbb{R}\right\}$ and $\mathrm{im} f = [-1,1]$

(f) $\mathrm{dom} f = \left\{(x,y) | \, x \in \mathbb{R} \ \wedge \ y \in \mathbb{R}\right\}$ and $\mathrm{im} f = \mathbb{R}$

(g) $\mathrm{dom} f = \left\{(x,y) | \, x \neq y\right\}$ and $\mathrm{im} f = \mathbb{R}_0$

(h) $\mathrm{dom} f = \left\{(x,y) | \, x^2 + y^2 > 3; \wedge \, x^2 + y^2 \neq 4\right\}$ and $\mathrm{im} f = \mathbb{R}_0$

(i) $\mathrm{dom} f = \left\{(x,y) | \, x^2 + 2y^2 \leq 1\right\}$ and $\mathrm{im} f = \left[\frac{\pi}{2}, \pi\right]$

Assignment 16.2 —

(a) The level curves corresponding to $c \neq 0$ are parabolas with their vertex at the origin and the y-axis as axis of symmetry. The level curves corresponding to $c = 0$ is al ine (the y-axis).

(b) The level curves corresponding to $c \neq 0$ are circles with the center at the y-axis. The level curves corresponding to $c = 0$ are lines (the x-axis).

(c) The level curves corresponding to $c \neq 0$ are bell-shaped curves with the y-axis as the axis of symmetry. The level curve for $c = 0$ is a curve with the y-axis as its vertical asymptote.

Assignment 16.3 —

(a) The circles in figure (a) all have the origin as their center. The circle with radius 5 corresponds to $C = 0$, the origin to $C = 5$. The relationship between the radius r of the circles and the constant C is $r + C = 5$. The circles are given by $x^2 + y^2 = (5 - C)^2$. Therefore, we conclude that $C = 5 - \sqrt{x^2 + y^2}$ $(C \leq 5)$. The level curves are derived from the surface $z = f(x,y) = 5 - \sqrt{x^2 + y^2}$. This is a semicircular cone with top in $(0,0,5)$.

(b) The lines in figure (b) are not equidistant. We see that $C = 10$ corresponds to $y = 5$ and $C = 0$ with $y = -5$. The relationship between C and y is $C - y = 5$. Therefore we conclude that $C = y + 5$. The level curves are derived from the surface $z = f(x,y) = \sqrt{y + 5}$. This is a parabolic cylinder parallel to the x-axis.

Assignment 16.4 —

$$f_x(0,0) = \lim_{h \to 0} \frac{1}{h}\left(h^3 \sin\left(\frac{1}{h^2}\right)\right) = 0$$

$$f_y(0,0) = \lim_{k \to 0} \frac{1}{k}\left(k \sin\left(\frac{1}{k^2}\right)\right) \text{ does not exist}$$

$$f_x(x,y) = 3x^2 \sin\left(\frac{1}{x^2 + y^2}\right) - \frac{(x^3 + y)2x}{\left(x^2 + y^2\right)^2} \cos\left(\frac{1}{x^2 + y^2}\right)$$

$\lim_{(x,y) \to (0,0)} f_x(x,y)$ does not exist, thus $f_x(x,y)$ is not continuous in $(0,0)$.

Assignment 16.5 —

	f_x	in a	f_y	in a	f_z	in a
(a)	$\frac{1}{1+(x+y+z)^2}$	$\frac{1}{2}$	$\frac{1}{1+(x+y+z)^2}$	$\frac{1}{2}$	$\frac{1}{1+(x+y+z)^2}$	$\frac{1}{2}$
(b)	$2x-2y+6z+4$	4	$6y-2x+7z-3$	-3	$12z+6x+7y$	0
(c)	$\frac{y}{2\sqrt{xy+z^2}}$	$\frac{\sqrt{2}}{4}$	$\frac{x}{2\sqrt{xy+z^2}}$	$\frac{\sqrt{2}}{4}$	$\frac{z}{\sqrt{xy+z^2}}$	$\frac{\sqrt{2}}{2}$
(d)	$(1+xy)\,e^{xy+z}$	1	$x^2\,e^{xy+z}$	1	$x\,e^{xy+z}$	1
(e)	$e^{x+y^2+z^3}$	1	$2y\,e^{x+y^2+z^3}$	0	$3z^2\,e^{x+y^2+z^3}$	0
(f)	$\sin(y)$	0	$x\cos(y)+\ln(z)$	-1	$\frac{y}{z}$	π
(g)	$\frac{(xy)^z z}{x}+z^{xy}\,y\ln(z)$	1	$\frac{(xy)^z z}{y}+z^{xy}\,x\ln(z)$	1	$(xy)^z\ln(xy)+\frac{z^{xy}\,xy}{z}$	1

	f_{xx}	in a	f_{xy}	in a	f_{yy}	in a
(a)	$-\frac{2(x+y+z)}{(1+(x+y+z)^2)^2}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$
(b)	2	2	-2	-2	6	6
(c)	$-\frac{y^2}{4(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{16}$	$\frac{xy+2z^2}{4(xy+z^2)^{\frac{3}{2}}}$	$\frac{3\sqrt{2}}{16}$	$-\frac{x^2}{4(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{16}$
(d)	$ye^{xy+z}\,(xy+2)$	0	$xe^{xy+z}\,(xy+2)$	2	x^3e^{xy+z}	1
(e)	$e^{x+y^2+z^3}$	1	$2ye^{x+y^2+z^3}$	0	$2\left(2y^2+1\right)e^{x+y^2+z^3}$	2
(f)	0	0	$\cos(y)$	-1	$-x\sin(y)$	0
(g)	(*)	0	(*)	1	(*)	0

(*): These expressions are too long to fit in the table

	f_{xz}	in a	f_{yz}	in a	f_{zz}	in a
(a)	$=f_{xx}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$
(b)	6	6	7	7	12	12
(c)	$-\frac{yz}{2(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{8}$	$-\frac{xz}{2(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{8}$	$\frac{xy}{(xy+z^2)^{\frac{3}{2}}}$	$\frac{\sqrt{2}}{4}$
(d)	$e^{xy+z}(xy+1)$	1	x^2e^{xy+z}	1	xe^{xy+z}	1
(e)	$3z^2e^{x+y^2+z^3}$	0	$6yz^2e^{x+y^2+z^3}$	0	$3z\left(3z^3+2\right)e^{x+y^2+z^3}$	0
(f)	0	0	$\frac{1}{z}$	1	$-\frac{y}{z^2}$	$-\pi$
(g)	(*)	2	(*)	2	(*)	0

(*): These expressions are too long to fit in the table

Assignment 16.6 —

$$\frac{\partial z}{\partial x}=\frac{x}{x^2+y^2},\quad \frac{\partial z}{\partial y}=\frac{y}{x^2+y^2}$$

Assignment 16.7 —

$$\frac{\partial z}{\partial x}=\frac{e^{\frac{x}{y}}}{y}\left(\sin\left(\frac{x}{y}\right)+\cos\left(\frac{x}{y}\right)\right)-\frac{ye^{\frac{y}{x}}}{x^2}\left(\cos\left(\frac{y}{x}\right)-\sin\left(\frac{y}{x}\right)\right)$$

$$\frac{\partial z}{\partial y}=-\frac{xe^{\frac{x}{y}}}{y^2}\left(\sin\left(\frac{x}{y}\right)+\cos\left(\frac{x}{y}\right)\right)+\frac{e^{\frac{y}{x}}}{x}\left(\cos\left(\frac{y}{x}\right)-\sin\left(\frac{y}{x}\right)\right)$$

Assignment 16.8 —

$$\frac{\partial z}{\partial x}=f'(x)g(y),\quad \frac{\partial z}{\partial y}=f(x)g'(y),\quad \frac{\partial^2 z}{\partial x\partial y}=f'(x)g'(y)$$

Assignment 16.9 — $\frac{\partial^2 z}{\partial x^2}=-A\lambda^2\sin(\lambda x)\sin(a\lambda y+\varphi),\quad \frac{\partial^2 z}{\partial y^2}=-a^2A\lambda^2\sin(\lambda x)\sin(a\lambda y+\varphi)$

Assignment 16.10 —

$$\frac{\partial u}{\partial t}=\frac{e^{-(x^2+y^2)/(4t)}}{t^2}\left(-1+\frac{x^2+y^2}{4t}\right)$$

$$\frac{\partial^2 u}{\partial x^2}=\frac{e^{-(x^2+y^2)/(4t)}}{2t^2}\left(-1+\frac{x^2}{2t}\right)$$

$$\frac{\partial^2 u}{\partial y^2}=\frac{e^{-(x^2+y^2)/(4t)}}{2t^2}\left(-1+\frac{y^2}{2t}\right)$$

Assignment 16.11 —

$$\frac{\partial u}{\partial x} = 2xy + z^2, \quad \frac{\partial u}{\partial y} = x^2 + 2yz, \quad \frac{\partial u}{\partial z} = y^2 + 2zx$$

Assignment 16.12 —

(a) $$\left(\frac{\partial f}{\partial T}\right)_{p,n} = \left(-\frac{2n^2a}{V^3}\right)\left(\frac{\partial V}{\partial T}\right)_{p,n}(V-nb) + \left(p + \frac{n^2a}{V^2}\right)\left(\frac{\partial V}{\partial T}\right)_{p,n} - nR = 0$$

$$\Leftrightarrow \quad \left(\frac{\partial V}{\partial T}\right)_{p,n} = \frac{nR}{p - \frac{n^2a}{V^2} + \frac{2n^3ab}{V^3}}$$

(b) $$\left(\frac{\partial f}{\partial p}\right)_{T,n} = \left(1 - \frac{2n^2a}{V^3}\left(\frac{\partial V}{\partial p}\right)_{T,n}\right)(V-nb) + \left(p + \frac{n^2a}{V^2}\right)\left(\frac{\partial V}{\partial p}\right)_{T,n} = 0$$

$$\Leftrightarrow \quad \left(\frac{\partial V}{\partial p}\right)_{T,n} = \frac{nb - V}{p - \frac{n^2a}{V^2} + \frac{2n^3ab}{V^3}}$$

(c) $$\left(\frac{\partial f}{\partial T}\right)_{V,n} = \left(\frac{\partial p}{\partial T}\right)_{V,n}(V-nb) - nR = 0 \quad \Leftrightarrow \quad \left(\frac{\partial p}{\partial T}\right)_{V,n} = \frac{nR}{V-nb}$$

(d) $$\left(\frac{\partial f}{\partial V}\right)_{T,n} = \left(\left(\frac{\partial p}{\partial V}\right)_{T,n} - \frac{2n^2a}{V^3}\right)(V-nb) + \left(p + \frac{n^2a}{V^2}\right) = 0 \quad \Leftrightarrow \quad \left(\frac{\partial p}{\partial V}\right)_{T,n} = \frac{2n^2a}{V^3} - \frac{p + \frac{n^2a}{V^2}}{V-nb}$$

Assignment 16.13 —

(a) Find $f(tx, ty)$ and separate the appropriate power of t.

(b) The equation follows directly by differentiating $f(tx, ty) = t^n f(x, y)$ with respect to t, using the chain rule, and assuming $t = 1$.

Assignment 16.14 —

	f_x	in a	f_y	in a
(a)	$2x+2y-2$	4	$2x+2y+3$	9
(b)	$2xy^5+3x^2y+y^2$	34	$5x^2y^4+x^3+2xy$	78
(c)	$-\dfrac{y\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}$	0	$\dfrac{x\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}$	0
(d)	$\dfrac{x^y y}{x}$	1	$x^y \ln(x)$	0
(e)	$\dfrac{2}{2x-3y}$	2	$-\dfrac{3}{2x-3y}$	-3
(f)	$-\dfrac{e^y}{x^2}$	$-e$	$\dfrac{e^y}{x}$	e
(g)	$-\dfrac{ye^{\frac{y}{x}}}{x^2}$	$-e$	$\dfrac{e^{\frac{y}{x}}}{x}$	e
(h)	$-3\sin(3x+2y)$	0	$-2\sin(3x+2y)$	0
(i)	$\dfrac{1}{1+x^2+2xy+y^2}$	$\dfrac{1}{2}$	$\dfrac{1}{1+x^2+2xy+y^2}$	$\dfrac{1}{2}$
(j)	$\sinh(y)+y\sinh(x)$	0	$x\cosh(y)+\cosh(x)$	1
(k)	$f(x,y)\left(\ln(2x+y)+\dfrac{2(x+3y)}{2x+y}\right)$	6	$f(x,y)\left(3\ln(2x+y)+\dfrac{x+3y}{2x+y}\right)$	3

	f_{xx}	in a	f_{xy}	in a	f_{yy}	in a
(a)	2	2	2	2	2	2
(b)	$2y^5+6xy$	20	$10xy^4+3x^2+2y$	59	$20x^2y^3+2x$	186
(c)	$\frac{2xy\left(x^2-3y^2\right)}{\left(x^2+y^2\right)^3}$	$-\frac{1}{2}$	$\frac{6x^2y^2-x^4-y^4}{\left(x^2+y^2\right)^3}$	$\frac{1}{2}$	$-\frac{2xy\left(3x^2-y^2\right)}{\left(x^2+y^2\right)^3}$	$-\frac{1}{2}$
(d)	$\frac{x^y y(y-1)}{x^2}$	0	$\frac{x^y(y\ln(x)+1)}{x}$	1	$x^y\ln^2(x)$	0
(e)	$-\frac{4}{(2x-3y)^2}$	-4	$\frac{6}{(2x-3y)^2}$	6	$-\frac{9}{(2x-3y)^2}$	-9
(f)	$\frac{2e^y}{x^3}$	$2e$	$-\frac{e^y}{x^2}$	$-e$	$\frac{e^y}{x}$	e
(g)	$\frac{ye^{\frac{y}{x}}(2x+y)}{x^4}$	$3e$	$-\frac{e^{\frac{y}{x}}(x+y)}{x^3}$	$-2e$	$\frac{e^{\frac{y}{x}}}{x^2}$	e
(h)	$-9\cos(3x+2y)$	-9	$-6\cos(3x+2y)$	-6	$-4\cos(3x+2y)$	-4
(i)	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$
(j)	$y\cosh(x)$	0	$\cosh(y)+\sinh(x)$	1	$x\sinh(y)$	0
(k)	(*)	28	(*)	19	(*)	12

(*): These expressions are too long to fit in the table

Total differentials ine the given points :

(a) $df=(2x+2y-2)\,dx+(2x+2y+3)\,dy \qquad \Rightarrow df(1,2)=4dx+9dy$

$d^2f=2\,dx^2+4\,dxdy+2\,dy^2 \qquad \Rightarrow d^2f(1,2)=2\,dx^2+4\,dxdy+2\,dy^2$

(b) $df=(2xy^5+3x^2y+y^2)\,dx+(5x^2y^4+x^3+2xy)\,dy \qquad \Rightarrow df(3,1)=34dx+78dy$

$d^2f=(2y^5+6xy)\,dx^2+2(10xy^4+3x^2+2y)\,dxdy+(20x^2y^3+2x)\,dy^2$

$\Rightarrow d^2f(3,1)=20\,dx^2+118\,dxdy+186\,dy^2$

(c) $df=-\frac{y(x^2-y^2)}{(x^2+y^2)^2}\,dx+\frac{x(x^2-y^2)}{(x^2+y^2)^2}\,dy \qquad \Rightarrow df(1,1)=0$

$$d^2f=\frac{2xy(x^2-3y^2)}{(x^2+y^2)^3}\,dx^2+2\,\frac{6x^2y^2-x^4-y^4}{(x^2+y^2)^3}\,dxdy-\frac{2xy(3x^2-y^2)}{(x^2+y^2)^3}\,dy^2$$

$$\Rightarrow d^2f(1,1)=\frac{1}{2}\left(-dx^2+2\,dxdy-dy^2\right)$$

(d) $df=\frac{x^y y}{x}\,dx+x^y\ln(x)\,dy \qquad \Rightarrow df(1,1)=dx$

$$d^2f=\frac{x^y y(y-1)}{x^2}\,dx^2+\frac{2x^y(y\ln(x)+1)}{x}\,dxdy+x^y\ln^2(x)\,dy^2 \qquad \Rightarrow d^2f(1,1)=2\,dxdy$$

(e) $df = \frac{2}{2x-3y}\,dx - \frac{3}{2x-3y}\,dy \qquad \Rightarrow df(2,1) = 2dx - 3dy$

$d^2f = -\frac{4}{(2x-3y)^2}\,dx^2 + \frac{12}{(2x-3y)^2}\,dxdy - \frac{9}{(2x-3y)^2}\,dy^2$

$\Rightarrow d^2f(2,1) = -4\,dx^2 + 12\,dxdy - 9\,dy^2$

(f) $df = -\frac{e^y}{x^2}\,dx + \frac{e^y}{x}\,dy \qquad \Rightarrow df(1,1) = -e\,dx + e\,dy$

$d^2f = \frac{2e^y}{x^3}\,dx^2 - \frac{2e^y}{x^2}\,dxdy + \frac{e^y}{x}\,dy^2 \qquad \Rightarrow d^2f(2,1) = 2e\,dx^2 - 2e\,dxdy + e\,dy^2$

(g) $df = -\frac{ye^{\frac{y}{x}}}{x^2}\,dx + \frac{e^{\frac{y}{x}}}{x}\,dy \qquad \Rightarrow df(1,1) = -e\,dx + e\,dy$

$d^2f = \frac{ye^{\frac{y}{x}}(2x+y)}{x^4}\,dx^2 - \frac{2e^{\frac{y}{x}}(x+y)}{x^3}\,dxdy + \frac{e^{\frac{y}{x}}}{x^2}\,dy^2 \qquad \Rightarrow d^2f(1,1) = 3e\,dx^2 - 4e\,dxdy + e\,dy^2$

(h) $df = -3\sin(3x+2y)\;dx - 2\sin(3x+2y)\;dy \qquad \Rightarrow df(0,\pi) = 0$

$d^2f = -9\cos(3x+2y)\,dx^2 - 12\cos(3x+2y)\,dxdy - 4\cos(3x+2y)\,dy^2$

$\Rightarrow d^2f(0,\pi) = -9\,dx^2 - 12\,dxdy - 4\,dy^2$

(i) $df = \frac{1}{1+(x+y)^2}(dx+dy) \qquad \Rightarrow df(1,0) = \frac{1}{2}(dx+dy)$

$d^2f = -\frac{2(x+y)}{\left(1+(x+y)^2\right)^2}\left(dx^2 + 2\,dxdy + dy^2\right) \qquad \Rightarrow d^2f(1,0) = -\frac{1}{2}\left(dx^2 + 2\,dxdy + dy^2\right)$

(j) $df = \left(\sinh(y) + y\sinh(x)\right)\,dx + \left(x\cosh(y) + \cosh(x)\right)\,dy \qquad \Rightarrow df(0,0) = dy$

$d^2f = y\cosh(x)\,dx^2 + 2\left(\cosh(y) + \sinh(x)\right)\,dxdy + x\sinh(y)\,dy^2 \qquad \Rightarrow d^2f(0,0) = 2\,dxdy$

(k) $df = f(x,y)\left(\ln(2x+y) + \frac{2(x+3y)}{2x+y}\right)dx + f(x,y)\left(3\ln(2x+y) + \frac{x+3y}{2x+y}\right)$

$\Rightarrow df(0,1) = 6dx + 3dy$

d^2f: expression too long.

Assignment 16.15 —

(a) 3% (b) 2% (c) 1%

Assignment 16.16 —

$$\frac{dz}{dt} = 1$$

Assignment 16.17 —

$$u_t = \frac{\partial u}{\partial t} = \frac{xse^{st} - ys^2\sin(t)}{\sqrt{x^2+y^2}}$$

Assignment 16.18 —

$$\frac{dz}{dx} = 2x + 2e^{ax} + 2xae^{ax} + 8ae^{2ax}$$

Assignment 16.19 —

(a) $$f_x = \frac{20x - 4y - 5)}{(x+2y+1)^2 + (3x-y-1)^2}$$

$$f_y = \frac{-2(x-5y-3)}{(x+2y+1)^2 + (3x-y-1)^2}$$

(b) $$f_x = 2\,(-5x+6y)\sinh\Big((2x-3y)^2 - (3x-4y)^2\Big)$$

$$f_y = 2\,(6x-7y)\sinh\Big((2x-3y)^2 - (3x-4y)^2\Big)$$

(c) $$f_x = \frac{e^{xy}}{\sin^2(x^2-y^2)}\Big(y\sin\big(x^2-y^2\big) - 2x\cos\big(x^2-y^2\big)\Big)$$

$$f_y = \frac{e^{xy}}{\sin^2(x^2-y^2)}\Big(x\sin\big(x^2-y^2\big) + 2y\cos\big(x^2-y^2\big)\Big)$$

(d) $$f_x = y^2 e^{x^3y^3}\left(3x^3y^3+1\right)$$

$$f_y = xye^{x^3y^3}\left(3x^3y^3+2\right)$$

(e) $$f_x = \left(x^2+y^2\right)^{xy}\left(\frac{2x^2y}{x^2+y^2} + y\ln\left(x^2+y^2\right)\right) f_y = \left(x^2+y^2\right)^{xy}\left(\frac{2xy^2}{x^2+y^2} + x\ln\left(x^2+y^2\right)\right)$$

Assignment 16.20 —

$$\frac{\partial^2 z}{\partial s^2} = 4f_{xx}(x,y) + 12f_{xy}(x,y) + 9f_{yy}(x,y)$$
$$\frac{\partial^2 z}{\partial s\partial t} = 6f_{xx}(x,y) + 5f_{xy}(x,y) - 6f_{yy}(x,y)$$
$$\frac{\partial^2 z}{\partial t^2} = 9f_{xx}(x,y) - 12f_{xy}(x,y) + 4f_{yy}(x,y)$$

Assignment 16.21 —

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial^2 z}{\partial x^2}e^{2s}\cos^2(t) + 2\frac{\partial^2 z}{\partial x\partial y}e^{2s}\sin(t)\cos(t) + \frac{\partial^2 z}{\partial y^2}e^{2s}\sin^2(t) + \frac{\partial z}{\partial x}e^{s}\cos(t) + \frac{\partial z}{\partial y}e^{s}\sin(t)$$
$$\frac{\partial^2 z}{\partial t^2} = \frac{\partial^2 z}{\partial x^2}e^{2s}\sin^2(t) - 2\frac{\partial^2 z}{\partial x\partial y}e^{2s}\sin(t)\cos(t) + \frac{\partial^2 z}{\partial y^2}e^{2s}\cos^2(t) - \frac{\partial z}{\partial x}e^{s}\cos(t) - \frac{\partial z}{\partial y}e^{s}\sin(t)$$

Assignment 16.22 —

$$r^2 = x^2+y^2 \;\Rightarrow\; \frac{\partial}{\partial x}\left(r^2\right) = \frac{\partial}{\partial x}\left(x^2+y^2\right) \;\Leftrightarrow\; \frac{\partial}{\partial r}\left(r^2\right)\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}\left(x^2+y^2\right) \;\Leftrightarrow\; 2r\frac{\partial r}{\partial x} = 2x \;\Leftrightarrow\; \frac{\partial r}{\partial x} = \frac{x}{r}$$
$$r^2 = x^2+y^2 \;\Rightarrow\; \frac{\partial}{\partial y}\left(r^2\right) = \frac{\partial}{\partial y}\left(x^2+y^2\right) \;\Leftrightarrow\; \frac{\partial}{\partial r}\left(r^2\right)\frac{\partial r}{\partial y} = \frac{\partial}{\partial y}\left(x^2+y^2\right) \;\Leftrightarrow\; 2r\frac{\partial r}{\partial y} = 2y \;\Leftrightarrow\; \frac{\partial r}{\partial y} = \frac{y}{r}$$
$$\frac{\partial u}{\partial x} = x(2\ln(r)+1), \quad \frac{\partial u}{\partial y} = y(2\ln(r)+1)$$

$$\frac{\partial^2 u}{\partial x^2} = 1 + 2\ln(r) + \frac{2x^2}{r^2}, \quad \frac{\partial^2 u}{\partial y^2} = 1 + 2\ln(r) + \frac{2y^2}{r^2} \quad \Rightarrow \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 4 + 4\ln(r)$$

$$\frac{\partial}{\partial x}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{4x}{r^2}, \quad \frac{\partial}{\partial y}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{4y}{r^2}$$

$$\frac{\partial^2}{\partial x^2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{4}{r^2} - \frac{8x^2}{r^4}, \quad \frac{\partial^2}{\partial y^2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{4}{r^2} - \frac{8y^2}{r^4}$$

Assignment 16.23 —

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

Assignment 16.24 —

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{dy}{dx}\frac{\partial x}{\partial u}$$

Assignment 16.25 —

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\left[\frac{\partial x}{\partial r}\left(\frac{\partial r}{\partial s}\frac{ds}{dt} + \frac{\partial r}{\partial t}\right) + \frac{\partial x}{\partial s}\frac{ds}{dt}\right] + \frac{\partial w}{\partial y}\left[\frac{\partial y}{\partial r}\left(\frac{\partial r}{\partial s}\frac{ds}{dt} + \frac{\partial r}{\partial t}\right) + \frac{\partial y}{\partial t}\right]$$

Assignment 16.26 —

(a) $\dfrac{dx}{dy} = -\dfrac{x^4 + 3xy^2}{y^3 + 4x^3 y}$.

The given equation has a solution in $x = x(y)$ near each point where $y^3 + 4x^3 y \neq 0$ or: $y \neq 0$ and $y^2 + 4x^3 \neq 0$.

(b) $\dfrac{\partial z}{\partial y} = \dfrac{xz + 3xy^4}{xy - 2y^2 z}$.

The given equation has a solution in $z = z(x,y)$ near each point where $y \neq 0$ and $x \neq 2yz$.

(c) $\dfrac{\partial y}{\partial z} = \dfrac{x^2 y \ln(y) - y^2 e^{yz}}{yze^{yz} - x^2 z}$.

The given equation has a solution in $y = y(x,z)$ near each point where $y > 0$, $z \neq 0$ and $ye^{yz} \neq x^2$.

Assignment 16.27 — Prove yourself

Assignment 16.28 —

(a) The given point corresponds to the given formula: $F(0,1,0) = 0$.

(b) $\dfrac{\partial F}{\partial y}(0,1,0) = 1 \neq 0 \rightarrow$ the implicit function theorem guarantees that there exists a unique function $y = f(x,z)$ such that $F(x, f(x,z), z) = 0$ close to $(0,1,0)$.

(c) $\dfrac{\partial F}{\partial z}(0,1,0) = 0 \rightarrow$ condition not met. We cannot conclude that there exists a unique function $z = f(x,y)$.

Assignment 16.29 —

(a) Assume $G(x,y) = F(F(x,y),y)$ and $z = F(x,y)$.

(b) $\dfrac{\partial G}{\partial y} = \dfrac{\partial F}{\partial z}\dfrac{\partial F}{\partial y} + \dfrac{\partial F}{\partial y} = \dfrac{\partial F}{\partial y}\left(\dfrac{\partial F}{\partial x} + 1\right) \neq 0$ with $\dfrac{\partial F}{\partial z} = \left.\dfrac{\partial F}{\partial x}\right|_{x=z}$

(c) It should apply that $\dfrac{\partial F}{\partial y} \neq 0$ en $\dfrac{\partial F}{\partial x} \neq -1$.

Assignment 16.30 — $D_{\hat{\boldsymbol{u}}} f(-1,-1) = \dfrac{4}{\sqrt{5}}$ with $\hat{\boldsymbol{u}}$ the unit vector in the direction of $\vec{\boldsymbol{v}}$

Assignment 16.31 —

$\nabla f(1,1) = (-2,4) = -2\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}}$

Assignment 16.32 —

$D_{\hat{\boldsymbol{u}}} f(P) = \dfrac{11}{3}$

Assignment 16.33 —

$D_{\hat{\boldsymbol{u}}} G(\vec{\boldsymbol{a}}) = \dfrac{-1}{12} + \dfrac{2}{3}\ln(2)$

Assignment 16.34 — The horse either has to escape at $(1,0)$ in the direction of $\hat{\boldsymbol{\imath}}$ or at $(-1,0)$ in the direction $-\hat{\boldsymbol{\imath}}$.

Assignment 16.35 —

(a) Tangent plane: $3x - 2y - z = 4$. Normal: $\dfrac{x-1}{3} = \dfrac{y+1}{-2} = \dfrac{z-1}{-1}$

(b) Tangent plane: $3x + 2y - 6z = -5$. Normal: $\dfrac{x-3}{3} = \dfrac{y+1}{2} = \dfrac{z-2}{-6}$

(c) Tangent plane: $z = x$. Normal: $\begin{cases} y = 2 \\ x = -z. \end{cases}$

(d) Tangent plane: $x - y - z + 1 = 0$. Normal: $4x - \pi = \pi - 4y = 4 - 4z$

Assignment 16.36 — The distance form the point $(1,1,0)$ to the paraboloid is $\dfrac{\sqrt{3}}{2}$.

Assignment 16.37 — An equation for the tangent plane is $2x + 4y + 6z = 11\pi$.

Assignment 16.38 —

(a) $xe^{xy+y} = e^2\left(1 + 2(x+y-2) + \dfrac{3}{2}(x-1)^2 + 5(x-1)(y-1) + 2(y-1)^2 + \cdots\right)$

(b) $x\ln(y) = x(y-1) + \cdots$

(c) $xy + \ln(xy) = 1 + 2(x+y-2) - \frac{1}{2}(x-y)^2 + \cdots$

(d) $x\sin(y) = xy + \cdots$

(e) $xy\cos(x+y) = \frac{\pi^2}{4}x - \frac{\pi}{2}x^2 - \frac{\pi}{2}xy + \ldots$

(f) $\arctan\left(\frac{x}{y}\right) = 2x - xy + \cdots$

(g) $\sin(xe^y) = x + xy + \cdots$

(h) $\frac{\sin(x)}{y} = 1 - (y-1) + (y-1)^2 - \frac{1}{2}\left(x - \frac{\pi}{2}\right)^2 + \cdots$

Assignment 16.39 — We give an approximation of the second degree.

(a) $\arctan\left(\frac{1.02}{0.95}\right) = \arctan\left(\frac{x}{y}\right) = f(x,y)$ with $x = 1+0.02$ and $y = 1-0.05$

$$f(1+h,1+k) = \frac{\pi}{4} + \frac{1}{2}h - \frac{1}{2}k + \frac{1}{2!}\left(-\frac{1}{2}h^2 + \frac{1}{2}k^2\right)$$

$$\Rightarrow f(1+0.02, 1-0.05) \approx \frac{\pi}{4} + \frac{0.02}{2} - \frac{(-0.05)}{2} + \frac{1}{2!}\left(-\frac{0.02^2}{2} + \frac{(-0.05)^2}{2}\right)$$

(b) $\sqrt{3.99 \times 4.02} = \sqrt{xy} = f(x,y)$ with $x = 4-0.01$ and $y = 4+0.02$

$$f(4+h,4+k) = 4 + \frac{1}{2}h + \frac{1}{2}k + \frac{1}{2!}\left(-\frac{1}{16}h^2 + \frac{1}{8}hk - \frac{1}{16}k^2\right)$$

$$\Rightarrow f(4-0.01, 4+0.02) = 4 + \frac{(-0.01)}{2} + \frac{0.02}{2} + \frac{1}{2!}\left(-\frac{(-0.01)^2}{16} + \frac{1}{8}(-0.01)(0.02) - \frac{0.02^2}{16}\right)$$

Assignment 16.40 —

(a) no extrema, saddle point at $(0,0)$

(b) no extrema, saddle point at $(0,0)$

(c) minimum at $\left(\frac{1}{4},-\frac{1}{8}\right)$, saddle point at $(1,0)$

(d) minimum at $(0,1)$, saddle points at $(\pm 1,2)$

(e) minima at $\left(\sqrt{2},-\sqrt{2}\right)$ and $\left(-\sqrt{2},\sqrt{2}\right)$, no conclusion for $(0,0)$

(f) no extrema, no conclusion for $(0,0)$

(g) minimum at $(1,1)$

(h) saddle points at $\left(k\pi,(-1)^{k+1}\right)$ with $k \in \mathbb{Z}$

(i) saddle points at $\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$ and $\left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)$

(j) no conclusion for $(x,0)$ with $x \in \mathbb{R}$ and $(0,y)$ with $y \in \mathbb{R}$

(k) no conclusion for $(x,k\pi-x)$ with $x \in \mathbb{R}$ and $k \in \mathbb{Z}$

(l) Local minimum at $(2,-1)$.

(m) Saddle point at $(-1,1)$.

(n) Saddle point at $(0,0)$ and Local minimum at $(1,1)$.

(o) Saddle point at $(0,0)$ and Local minima at $(-1,-1)$ and $(1,1)$.

(p) Local maximum at $(-4,2)$.

(q) Saddle point at $(0,n\pi)$, with $n \in \mathbb{N}$.

(r) Saddle points at $(m\pi,n\pi)$ if $m+n$ is odd, Local minima at $(m\pi,n\pi)$ if m and n are odd, Local maxima at $(m\pi,n\pi)$ if m and n are even.

(s) Local minima at $(0,y)$ if $y>0$ and at $\left(\pm 1,-1/\sqrt{2}\right)$, Local maxima at $(0,y)$ if $y<0$ and at $\left(\pm 1,1/\sqrt{2}\right)$, Saddle point at $(0,0)$.

(t) Saddle point at $(0,0)$, Local maxima at $(1,1)$ and $(-1,-1)$, Local minima at $(1,-1)$ and $(-1,1)$.

(u) Saddle point at $(3^{-1/3},0)$.

(v) Local minima at $(0,y)$ if $y \neq 0$ and Local maxima at $(x,0)$ if $x \neq 0$.

(w) Local minima at $(x,-x)$ if $x \neq 0$ and Local maxima at (x,x) if $x \neq 0$.

(x) Saddle point at $(1,1,1/2)$.

Assignment 16.41 — Maxima at $(1,\pm 1,1)$ and minima at $(-1,\pm 1,-1)$.

Assignment 16.42 — $x=1$ and $y=\sqrt{2}$

Assignment 16.43 — The box has a minimum area if the base of the box has dimensions 4 dm and 4 dm and the box has a height 2 dm.

Assignment 16.44 — The numbers for which ab^2c^3 is maximal, are $a=5$, $b=10$ and $c=15$

Assignment 16.45 — The dimensions of the base and cover should be 2 m and 2 m, while the height of the box should be 3 m.

Assignment 16.46 —

$$\left(\frac{4}{3},-\frac{4}{3},-\frac{4}{3}\right)$$

Assignment 16.47 — The box has a maximum volume if the base of the box has dimensions 16 cm and 16 cm and the box has a height of 8 cm.

Assignment 16.48 —

(a) Global maximum: $f\left(\dfrac{1}{2},1\right)=\dfrac{5}{4}$. Global minimum: $f(2,0)=-2$.

(b) Global maximum: $f(-1,0)=2$. Global minimum: $f(1,0)=-2$.

(c) Global maximum: $f\left(\dfrac{\sqrt{3}}{3},1\right)=\dfrac{2\sqrt{3}}{9}$. Global minima: $f(x,0)=f(0,y)=f(1,1)=0$, with $0\leq x,y\leq 1$.

(d) Global maximum: $f\left(\dfrac{1}{3},\dfrac{1}{3}\right)=\dfrac{1}{27}$. Global minima: $f(x,0)=f(0,y)=f(x,1-x)=0$, with $0\leq x,y\leq 1$.

(e) Global maximum: $f\left(\dfrac{\pi}{2},0\right)=1$. Global minima: $f\left(\dfrac{3\pi}{2},0\right)=f\left(\dfrac{\pi}{2},\pi\right)=-1$.

(f) Global maximum: $f(2,1)=\dfrac{4}{e^3}$. Global minima: $f(x,0)=f(0,y)=0$, with $0\leq x,y\leq 4$.

Assignment 16.49 —

(a) Maximum: $f(0,3)=21$.

(b) Maximum: $f\left(\dfrac{7}{4},5\right)=\dfrac{37}{2}$.

Assignment 16.50 — Minimum: $f(1,1,1)=9$.

Assignment 16.51 — The manufacturer should produce 20000/3 kg ≈ 6667 kg of each textile for a maximum profit of \$100 000/3 ≈ \$33 333.

Assignment 16.52 — The profit is maximal (€66) with 6 circuits of type *A* and three circuits of type *B*.

Assignment 16.53 — There should be 4 hours of pop music, 4 hours of oldies and 12 hours of information programs to maximize the ratin, which then will be 160.

Assignment 16.54 — There should be 40 detached houses, no half open buildings and 40 apartments. The profit will then be 2 €240 000.

Assignment 16.55 — The profit is maximum for 10 suits, 30 jackets and 40 pants.

Assignment 16.56 — The profit is maximum for 75 kg of tea and 15 kg of sugar.

Assignment 16.57 — The cost is minimum for 300 Silver balls, 300 Yellow balls and 1200 Gold balls. The cost is then €2460.

Assignment 16.58 —

(a) $\nabla f(1,-2) = \left(\dfrac{2}{5}, -\dfrac{4}{5}\right) = \dfrac{2}{5}(\hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}})$

(b) $2x - 4y - 5z + 5\ln 5 - 10 = 0$

(c) $x - 2y - 5 = 0$

Assignment 16.59 —

(a) The isotherms that are found by $T(x,y) = c > 0$ are hyperbolas with their real vertices on the x-axis. The isotherms found by $T(x,y) = c < 0$ are hyperbolas with their real vertices on the y-axis. The ones belonging to $T(x,y) = c = 0$ are the lines $x = \pm\sqrt{2}y$.

(b) $-\nabla T(2,-1) = (-4,-4) = -4\hat{\boldsymbol{\imath}} - 4\hat{\boldsymbol{\jmath}}$ $\quad\rightarrow$ direction $(-1,-1)$

(c) $yx^2 = -4$

Chapter 17

Assignment 17.1 —

(a) $\displaystyle\int_1^2 \int_y^{3y} (x+y)\, dx\, dy = 14$

(b) $\displaystyle\int_{-1}^2 \int_{2x^2-2}^{x^2+x} x\, dy\, dx = \frac{9}{4}$

(c) $\displaystyle\int_0^{\pi} \int_0^{\cos(\theta)} r\sin(\theta)\, dr\, d\theta = \frac{1}{3}$

(d) $\displaystyle\int_0^{\frac{\pi}{2}} \int_2^{4\cos(\theta)} r^3\, dr\, d\theta = 10\pi$

(e) $\displaystyle\int_0^1 \int_0^1 \frac{x^2}{1+y^2}\, dy\, dx = \frac{\pi}{12}$

(f) $\displaystyle\int_{-1}^2 \int_0^1 (xy^2 + x^2y)\, dx\, dy = 2$

(g) $\displaystyle\int_0^1 \int_{x^2}^{2-x^2} \sqrt{x}y\, dy\, dx = \frac{16}{21}$

(h) $\displaystyle\int_0^{\pi/4} \int_{\sin(x)}^{\cos(x)} xy\, dy\, dx = \frac{\pi-2}{16}$

Assignment 17.2 —

(a) $\displaystyle\iint_R dA = \int_0^1 \int_{\sqrt{x^3}}^{x} dy\, dx = \int_0^1 \int_y^{\sqrt[3]{y^2}} dx\, dy = \frac{1}{10}$

(b) $\displaystyle\iint_R x^2\, dA = \int_0^4 \int_0^x x^2\, dy\, dx + \int_4^8 \int_0^{\frac{16}{x}} x^2\, dy\, dx = \int_0^2 \int_y^8 x^2\, dx\, dy + \int_2^4 \int_y^{\frac{16}{y}} x^2\, dx\, dy = 448$

(c) $\displaystyle\iint_R y\, dA = \int_0^1 \int_{x^3}^{x^2} y\, dy\, dx = \int_0^1 \int_{\sqrt{y}}^{\sqrt[3]{y}} y\, dx\, dy = \frac{1}{35}$

(d) $$\iint_R \frac{1}{\sqrt{2y-y^2}}\,dA = \int_0^2 \int_0^{\frac{4-x^2}{2}} \frac{1}{\sqrt{2y-y^2}}\,dy\,dx = \int_0^2 \int_0^{\sqrt{4-2y}} \frac{1}{\sqrt{2y-y^2}}\,dx\,dy = 4$$

(e) $$\iint_R e^{\frac{x}{y}}\,dA = \int_0^1 \int_{\sqrt{x}}^1 e^{\frac{x}{y}}\,dy\,dx = \int_0^1 \int_0^{y^2} e^{\frac{x}{y}}\,dx\,dy = \frac{1}{2}$$

(f) $$\iint_R e^{-(x+y)}\,dA = \int_{1/2}^1 \int_0^x e^{-(x+y)}\,dy\,dx = \int_0^{1/2} \int_{1/2}^1 e^{-(x+y)}\,dx\,dy + \int_{1/2}^1 \int_y^1 e^{-(x+y)}\,dx\,dy$$
$$= \frac{1}{2e^2} - \frac{3}{2e} + \frac{1}{\sqrt{e}}$$

(g) $$\iint_R dA = \int_0^a \int_{a-y}^{\sqrt{a^2-y^2}} dx\,dy = \int_0^a \int_{a-x}^{\sqrt{a^2-x^2}} dy\,dx = \frac{a^2}{2}\left(\frac{\pi}{2}-1\right)$$

(h) $$\iint_R \frac{dA}{x+y} = \int_0^1 \int_y^1 \frac{1}{x+y}\,dx\,dy = \int_0^1 \int_0^x \frac{1}{x+y}\,dy\,dx = \ln(2)$$

Assignment 17.3 —

(a) $$\int_0^3 \int_1^{\sqrt{4-y}} f(x,y)\,dx\,dy = \int_1^2 \int_0^{4-x^2} f(x,y)\,dy\,dx$$

(b) $$\int_0^1 \int_{\arccos(y)}^{\frac{\pi}{2}} f(x,y)\,dx\,dy = \int_0^{\frac{\pi}{2}} \int_{\cos(x)}^1 f(x,y)\,dy\,dx$$

(c) $$\int_{-6}^2 \int_{\frac{x^2}{4}}^{3-x} f(x,y)\,dy\,dx = \int_0^1 \int_{-2\sqrt{y}}^{2\sqrt{y}} f(x,y)\,dx\,dy + \int_1^9 \int_{-2\sqrt{y}}^{3-y} f(x,y)\,dx\,dy$$

(d) $$\int_{\frac{a}{2}}^a \int_0^{\sqrt{2ax-x^2}} f(x,y)\,dy\,dx = \int_0^{\frac{\sqrt{3}a}{2}} \int_{\frac{a}{2}}^a f(x,y)\,dx\,dy + \int_{\frac{\sqrt{3}a}{2}}^a \int_{a-\sqrt{a^2-y^2}}^a f(x,y)\,dx\,dy$$

Assignment 17.4 —

(a) $$A = 2\int_0^{\sqrt{15}} \int_{\frac{25-y^2}{-10}}^{\frac{y^2-9}{-6}} dx\,dy = \frac{16\sqrt{15}}{3}$$

(b) $$A = \int_1^2 \int_{\sqrt{1-(x-1)^2}}^x dy\,dx + \int_2^4 \int_0^{\sqrt{4-(x-2)^2}} dy\,dx = \int_0^1 \int_{1+\sqrt{1-y^2}}^{2+\sqrt{4-y^2}} dx\,dy + \int_1^2 \int_y^{2+\sqrt{4-y^2}} dx\,dy$$

$$= \int_0^{\frac{\pi}{4}} \int_{2\cos(\theta)}^{4\cos(\theta)} r \, dr \, d\theta = \frac{3\pi}{4} + \frac{3}{2}$$

(c) $A = 2\int_0^{\frac{\pi}{3}} \int_{\frac{1}{\cos(\theta)}}^{2} r \, dr \, d\theta = \frac{4\pi}{3} - \sqrt{3}$

(d) $A = 2\int_0^{\pi} \int_0^{1+\cos(\theta)} r \, dr \, d\theta - 2\int_0^{\frac{\pi}{2}} \int_0^{\cos(\theta)} r \, dr \, d\theta = \frac{5\pi}{4}$

Assignment 17.5 —

(a) $V = \int_0^{2} \int_0^{\frac{\sqrt{4-x^2}}{2}} \frac{4-y^2}{2} \, dy \, dx = \int_0^{1} \int_0^{2\sqrt{1-y^2}} \frac{4-y^2}{2} \, dx \, dy = \frac{15\pi}{16}$

(b) $V = \int_0^{1} \int_0^{1-x} (1-x^2-y^2) \, dy \, dx = \int_0^{1} \int_0^{1-y} (1-x^2-y^2) \, dx \, dy = \frac{1}{3}$

(c) $V = \int_0^{1} \int_{x^2}^{\sqrt{x}} (12+y-x^2) \, dy \, dx = \int_0^{1} \int_{y^2}^{\sqrt{y}} (12+y-x^2) \, dx \, dy = \frac{569}{140}$

(d) $V = \int_0^{2} \int_0^{1} \frac{4+x-y}{2} \, dy \, dx = \int_0^{1} \int_0^{2} \frac{4+x-y}{2} \, dx \, dy = \frac{9}{2}$

(e) $V = 8\int_0^{1} \int_0^{\sqrt{1-x^2}} \sqrt{1-x^2} \, dy \, dx = 8\int_0^{1} \int_0^{\sqrt{1-y^2}} \sqrt{1-x^2} \, dx \, dy = \frac{16}{3}$

Assignment 17.6 —

(a) $V = 8\int_0^{\frac{\pi}{2}} \int_0^{1} \sqrt{4-r^2} \, r \, dr \, d\theta = \frac{4\pi}{3}\left(8-3\sqrt{3}\right)$

(b) $V = \int_0^{a\sqrt{2}} \int_0^{2\pi} \left(\sqrt{3a^2-r^2} - \frac{r^2}{2a}\right) r \, d\theta \, dr = \frac{a^3\pi}{3}\left(6\sqrt{3}-5\right)$

(c) $V = 4\int_0^{\frac{\pi}{2}} \int_0^{a\cos(\theta)} \sqrt{a^2-r^2} \, r \, dr \, d\theta = \frac{4}{3}a^3\left(-\frac{2}{3}+\frac{\pi}{2}\right)$

(d) $V = 4\int_0^{\frac{\pi}{2}} \int_0^{2\sin(\theta)} \sqrt{r\sin(\theta)} \, r \, dr \, d\theta = \frac{64\sqrt{2}}{15}$

Assignment 17.7 —

(a) $\dfrac{d}{dt}\left(\int\limits_0^1 \sin(x-t)\,dx\right) = -\sin(1-t)-\sin(t)$

(b) $\dfrac{d}{dt}\left(\int\limits_1^{t^2} \dfrac{e^{tx}}{x}\,dx\right) = 3\dfrac{e^{t^3}}{t}-\dfrac{e^t}{t}$

(c) $\dfrac{d}{dx}\left(\int\limits_{\frac{1}{x}}^{\frac{2}{x}} \dfrac{\sin(tx)}{t}\,dt\right) = 0$

Assignment 17.8 — First we derive both limits to α:

$$F'(\alpha) = \int\limits_0^{+\infty} \frac{\partial}{\partial\alpha}\left(e^{-x}\frac{\sin(\alpha x)}{x}\right)\,dx = \frac{1}{\alpha^2+1}.$$

Integrating this result to α results in $\arctan(\alpha)+C$. As $F(0)=0$, is $C=0$. Therefor we conclude $F(\alpha)=\arctan(\alpha)$.
$I = \arctan\left(\dfrac{1}{a}\right)$
$I = \ln|a+1|$

Assignment 17.9 —

(a)
- $M = \int\limits_0^{\pi}\int\limits_0^{\sin(x)} ky\,dy\,dx = \dfrac{k\pi}{4}$
- $M_y = k\int\limits_0^{\pi}\int\limits_0^{\sin(x)} xy\,dy\,dx = \dfrac{k\pi^2}{8} \quad \Rightarrow \overline{x} = \dfrac{\pi}{2}$
- $M_x = k\int\limits_0^{\pi}\int\limits_0^{\sin(x)} y^2\,dy\,dx = \dfrac{4k}{9} \quad \Rightarrow \overline{y} = \dfrac{16}{9\pi}$

Center of mass: $\left(\dfrac{\pi}{2}, \dfrac{16}{9\pi}\right)$

(b)
- $M = \int\limits_{-2}^{2}\int\limits_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} 1\,dx\,dy = 8$
- $M_y = \int\limits_{-2}^{2}\int\limits_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} x\,dx\,dy = \dfrac{16}{5} \quad \Rightarrow \overline{x} = \dfrac{2}{5}$

- $M_x = \int_{-2}^{2} \int_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} y \, dx \, dy = 0 \quad \Rightarrow \overline{y} = 0$

Center of mass: $\left(\frac{2}{5}, 0\right)$

(c)
- $M = \int_{0}^{4} \int_{0}^{\frac{12-3y}{2}} 1 \, dx \, dy = 12$
- $M_y = \int_{0}^{4} \int_{0}^{\frac{12-3y}{2}} x \, dx \, dy = 24 \quad \Rightarrow \overline{x} = 2$
- $M_x = \int_{0}^{4} \int_{0}^{\frac{12-3y}{2}} y \, dx \, dy = 16 \quad \Rightarrow \overline{y} = \frac{4}{3}$

Center of mass: $\left(2, \frac{4}{3}\right)$

Assignment 17.10 —

(a) $SA = 2 \int_{0}^{\frac{\pi}{2}} \int_{0}^{4\sin(\theta)} \frac{2}{\sqrt{3}} r \, dr \, d\theta = \frac{8\pi}{\sqrt{3}}$

(b) $SA = 8 \int_{0}^{4} \int_{0}^{\sqrt{16-x^2}} \sqrt{\frac{16}{16-x^2}} \, dy \, dx = 128$

(c) $SA = 4 \int_{0}^{1} \int_{0}^{2\sqrt{x}} \sqrt{\frac{x+1}{x}} \, dy \, dx = \frac{16}{3}\left(2\sqrt{2}-1\right)$

(d) $SA = \int_{0}^{2\pi} \int_{0}^{1} \sqrt{1+4r^2} \, r \, dr \, d\theta = \frac{\pi}{6}\left(5\sqrt{5}-1\right)$

(e) $SA = 8 \int_{0}^{a} \int_{0}^{\frac{\sqrt{a^2-x^2}}{2}} \sqrt{\frac{a^2}{a^2-x^2-y^2}} \, dy \, dx = \frac{4a^2\pi}{3}$

Assignment 17.11 —

(a) $\int_{0}^{1} \int_{0}^{1-z} \int_{0}^{1-y-z} xyz \, dx \, dy \, dz = \frac{1}{720}$

(b) $\int_{0}^{1} \int_{0}^{\sqrt{1-z^2}} \int_{0}^{\sqrt{1-y^2-z^2}} (x+z) \, dx \, dy \, dz = \frac{\pi}{8}$

(c) $\displaystyle\int_0^1\int_0^1\int_0^{\sqrt{1-y^2}} xy\;dx\;dy\;dz = \frac{1}{8}$

(d) $\displaystyle 8\int_0^1\int_0^{\sqrt{1-z^2}}\int_0^{\sqrt{1-y^2-z^2}} \left(x^2+y^2+z^2\right)^{3/2}\;dx\;dy\;dz = \frac{2\pi}{3}$

Assignment 17.12 —

(a) $\displaystyle V = \int_0^2\int_0^{2\pi}\int_{2\left(r^2-1\right)}^{10-r^2} r\;dz\;d\theta\;dr = 24\pi$

(b) $\displaystyle V = \int_{-a}^{+a}\int_{-b\sqrt{1-\frac{x^2}{a^2}}}^{b\sqrt{1-\frac{x^2}{a^2}}}\int_{-c\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}}}^{c\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}}} dz\;dy\;dx = \frac{4}{3}\pi abc$

(c) $\displaystyle V = \int_0^2\int_0^{2\pi}\int_0^{\frac{16-r^2}{2}} r\;dz\;d\theta\;dr = 28\pi$

(d) $\displaystyle V = 2\int_0^{2\pi}\int_0^{a}\int_0^{\frac{\pi}{6}} \rho^2\sin(\varphi)\;d\varphi\;d\rho\;d\theta = \frac{4\pi a^3}{3}\left(1-\frac{\sqrt{3}}{2}\right)$

(e) $\displaystyle V = \int_0^{2\pi}\int_0^{\sqrt{3}}\int_{\frac{r^2}{3}}^{\sqrt{4-r^2}} r\;dz\;dr\;d\theta = \frac{19\pi}{6}$

Assignment 17.13 — $\displaystyle V = \int_0^1\int_0^{\frac{\pi}{2}} \theta r\;d\theta\;dr = \frac{\pi^2}{16},\quad SA = \int_0^1\int_0^{\frac{\pi}{2}} \sqrt{1+\frac{1}{r^2}}\;r\;d\theta\;dr = \frac{\pi}{4}\left(\sqrt{2}+\ln\left(\sqrt{2}+1\right)\right)$

Assignment 17.14 — $\displaystyle V = \int_0^a\int_{-\sqrt{\frac{a-z}{a}}}^{\sqrt{\frac{a-z}{a}}}\int_{-\sqrt{\frac{a-z}{a}}}^{\sqrt{\frac{a-z}{a}}} dxdydz = 2a$

Assignment 17.15 —

(a) Cylindrical coordinates:

- $\displaystyle M = \int_0^{2\pi}\int_0^{h}\int_r^{h} \left(1-\frac{z}{h}\right) r\;dz\;dr\;d\theta = \frac{h^3\pi}{12}$
- $\displaystyle M_{yz} = \int_0^{2\pi}\int_0^{h}\int_r^{h} r\cos(\theta)\left(1-\frac{z}{h}\right) r\;dz\;dr\;d\theta = 0 \quad \Rightarrow \overline{x} = 0$

- $M_{xz} = \int_0^{2\pi}\int_0^h\int_r^h r\sin(\theta)\left(1-\frac{z}{h}\right)r\,dz\,dr\,d\theta = 0 \quad \Rightarrow \overline{y} = 0$
- $M_{xy} = \int_0^{2\pi}\int_0^h\int_r^h z\left(1-\frac{z}{h}\right)r\,dz\,dr\,d\theta = \frac{h^4\pi}{20} \quad \Rightarrow \overline{z} = \frac{3h}{5}$

Center of mass: $\left(0, 0, \frac{3h}{5}\right)$

Spherical coordinates:

- $M = \int_0^{2\pi}\int_0^{\frac{\pi}{4}}\int_0^{\frac{h}{\cos(\varphi)}}\left(1-\frac{\rho\cos(\varphi)}{h}\right)\rho^2\sin(\varphi)\,d\rho\,d\varphi\,d\theta = \frac{h^3\pi}{12}$
- $M_{yz} = \int_0^{2\pi}\int_0^{\frac{\pi}{4}}\int_0^{\frac{h}{\cos(\varphi)}}\rho\sin(\varphi)\cos(\theta)\left(1-\frac{\rho\cos(\varphi)}{h}\right)\rho^2\sin(\varphi)\,d\rho\,d\varphi\,d\theta = 0$
- $M_{xz} = \int_0^{2\pi}\int_0^{\frac{\pi}{4}}\int_0^{\frac{h}{\cos(\varphi)}}\rho\sin(\varphi)\sin(\theta)\left(1-\frac{\rho\cos(\varphi)}{h}\right)\rho^2\sin(\varphi)\,d\rho\,d\varphi\,d\theta = 0$
- $M_{xy} = \int_0^{2\pi}\int_0^{\frac{\pi}{4}}\int_0^{\frac{h}{\cos(\varphi)}}\rho\cos(\varphi)\left(1-\frac{\rho\cos(\varphi)}{h}\right)\rho^2\sin(\varphi)\,d\rho\,d\varphi\,d\theta = \frac{h^4\pi}{20}$

(b)

- $M = \int_0^{2\pi}\int_0^2\int_{r^2}^4 kr\,r\,dz\,dr\,d\theta = \frac{128k\pi}{15}$
- $M_{yz} = \int_0^{2\pi}\int_0^2\int_{r^2}^4 r\cos(\theta)kr\,r\,dz\,dr\,d\theta = 0 \quad \Rightarrow \overline{x} = 0$
- $M_{xz} = \int_0^{2\pi}\int_0^2\int_{r^2}^4 r\sin(\theta)kr\,r\,dz\,dr\,d\theta = 0 \quad \Rightarrow \overline{y} = 0$
- $M_{xy} = \int_0^{2\pi}\int_0^2\int_{r^2}^4 zkr\,r\,dz\,dr\,d\theta = \frac{512k\pi}{21} \quad \Rightarrow \overline{z} = \frac{20}{7}$

Center of mass: $\left(0, 0, \frac{20}{7}\right)$

(c)

- $M = \int_0^{2\pi}\int_0^2\int_0^{\sqrt{1+z^2}} r\,dr\,dz\,d\theta = \frac{14\pi}{3}$
- $M_{yz} = \int_0^{2\pi}\int_0^2\int_0^{\sqrt{1+z^2}} r\cos(\theta)\,r\,dr\,dz\,d\theta = 0 \quad \Rightarrow \overline{x} = 0$

- $M_{xz} = \int_0^{2\pi}\int_0^{2}\int_0^{\sqrt{1+z^2}} r\sin(\theta)\, r\, dr\, dz\, d\theta = 0 \quad \Rightarrow \overline{y} = 0$
- $M_{xy} = \int_0^{2\pi}\int_0^{2}\int_0^{\sqrt{1+z^2}} z\, r\, dr\, dz\, d\theta = 6\pi \quad \Rightarrow \overline{z} = \dfrac{9}{7}$

Center of mass: $\left(0, 0, \dfrac{9}{7}\right)$

Chapter 18

Assignment 18.1 —

(a) $\int_C x^2\, ds = 3\sqrt{14}$

(b) $\int_C y\, ds = 156$

(c) $\int_C (x+y)\, ds = 2\sqrt{2}$

(d) $\int_C \dfrac{ds}{x^2+y^2+z^2} = \dfrac{\sqrt{65}}{8}\arctan\left(\dfrac{\pi}{4}\right)$

(e) $\int_C \sqrt{2y^2+z^2}\, ds = 8\pi$

(f) $\int_C e^z\, ds = \dfrac{e^{2\pi}\sqrt{1+2e^{4\pi}}-\sqrt{3}}{2} + \dfrac{1}{2\sqrt{2}}\ln\left(\dfrac{\sqrt{2}e^{2\pi}+\sqrt{1+2e^{4\pi}}}{\sqrt{2}+\sqrt{3}}\right)$

(g) $\int_C \sqrt{1+4x^2z^2}\, ds = 3\pi$

(h) $\int_C x\, ds = \dfrac{a^2}{2}\left(\sqrt{2}+\ln\left(1+\sqrt{2}\right)\right)$

(i) $\int_C z\, ds = 1$

Assignment 18.2 —

(a) $\vec{\boldsymbol{F}}_1 = \dfrac{\vec{\boldsymbol{r}}}{\|\vec{\boldsymbol{r}}\|} \quad \rightarrow$ graph *II*

(b) $\vec{\boldsymbol{F}}_2 = \vec{\boldsymbol{r}} \quad \rightarrow$ graph *I*

(c) $\vec{\boldsymbol{F}}_3 = y\hat{\boldsymbol{\imath}} - x\hat{\boldsymbol{\jmath}} \quad \rightarrow$ graph *IV*

(d) $\vec{\boldsymbol{F}}_4 = x\hat{\boldsymbol{\jmath}} \quad \rightarrow$ graph *III*

Assignment 18.3 — $\nabla f = (2xyz^3, x^2z^3, 3x^2yz^2)$, $\nabla \cdot \vec{\boldsymbol{F}} = z - 2y$ and $\nabla \times \vec{\boldsymbol{F}} = (2x^2, -4xy + x, 0)$

Assignment 18.4 — $\nabla f = (y+z, x+z, y+x) \quad \Rightarrow \quad \nabla f(3,-1,2) = (1,5,2)$, $\nabla \cdot \vec{\boldsymbol{F}} = 10$,
$\nabla \times \vec{\boldsymbol{F}} = (y^2, -z^2, -x^2) \quad \Rightarrow \quad (\nabla \times \vec{\boldsymbol{F}})(3,-1,2) = (1,-4,-9)$ and
$(\vec{\nabla} f) \times \vec{\boldsymbol{F}} = (x^2z^2 + xz^3 + y^3z + xy^2z, -xyz^2 - xz^3 + x^2y^2 + x^3y, -y^3z - y^2z^2 - x^3y - x^2yz)$
$\Rightarrow \quad ((\vec{\nabla} f) \times \vec{\boldsymbol{F}})(3,-1,2) = (64,-30,43)$

Assignment 18.5 —

(a) The direction of the vector field changes in all points.

(b) In points on the y- or z-axis, the direction of the vector field does not change.

Assignment 18.6 — $\nabla \times \vec{\boldsymbol{F}} = \left(7z, -\dfrac{3x}{z^2}, 2\right) \quad \Rightarrow$ on the curve C: $\nabla \times \vec{\boldsymbol{F}} = \big(-7\cos(t), -3, 2\big)$

The length of the curl is maximum ifcos$(t) = \pm 1$. This corresponds to the points $(1,0,\pm 1)$ on C. The length of the curl is minimal if $\cos(t) = 0$. This corresponds to the points $(0,\pm 1,0)$ on C.

Assignment 18.7 — Let $\overrightarrow{OP} = (x,y,z)$ with $r = \sqrt{x^2+y^2+z^2}$, so $V = \dfrac{q}{4\pi\epsilon_0\sqrt{x^2+y^2+z^2}}$. We find that

$$\vec{\boldsymbol{E}} = -\nabla V = -\frac{q}{4\pi\epsilon_0\left(\sqrt{x^2+y^2+z^2}\right)^3}(x,y,z) = \frac{q\vec{\boldsymbol{r}}}{4\pi\epsilon_0\, r^3}$$

Assignment 18.8 —

(a) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 0$

(b) $\displaystyle\int_{C_1} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{1}{3}, \quad \int_{C_2} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{1}{12}, \quad \int_{C_3} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{17}{30}$

(c) $\displaystyle\int_{C_i} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 2$ for $i = 1\ldots 4$

(d) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 2\pi$

(e) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{76}{35}$

(f) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 0$

(g) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 5$

(h) $\displaystyle\int_{C_1} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{49}{3}, \quad \int_{C_2+C_3} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{49}{3}, \quad \oint \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = 0$

(i) $\displaystyle\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{32}{3}$

Assignment 18.9 —

(a) $R = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, x > 0,\ y \geq 0\right\}$ is a simply connected region.

(b) $R = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, x = 0,\ y \geq 0\right\}$ is not a region.

(c) $R = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, x \neq 0,\ y > 0\right\}$ is a region, but is not connected. There is no path in R from $(-1,1)$ to $(1,1)$.

(d) $R = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2 > 1\right\}$ a region, but is not connected. There is no path in R from $(-2,0,0)$ to $(2,0,0)$.

(e) $R = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2+y^2 > 1\right\}$ but cannot be shrunk to a point within $x^2+y^2 = 2, z = 0$ lies in R, but cannot be shrunk to a point within this region.

(f) $R = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2+y^2+z^2 > 1\right\}$ is a simply connected region.

Assignment 18.10 —

(a) $\vec{F} = \left(xy, \frac{1}{2}x^2 - y^2\right)$ is conservative. A potential function is $f(x,y) = \frac{x^2y}{2} - \frac{y^3}{3}$.

(b) $\vec{F} = (y,x,-2z)$ is conservative. A potential function is $f(x,y,z) = xy - z^2$.

(c) $\vec{F} = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$ is conservative. A potential function is $f(x,y) = \frac{\ln\left(x^2+y^2\right)}{2}$.

(d) $\vec{F} = \left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2}\right)$ is not conservative.

(e) $\vec{F} = \left(2xy - z^2, 2yz + x^2, -2zx + y^2\right)$ is conservative. A potential function is $f(x,y,z) = x^2y - xz^2 + y^2z$.

(f) $\vec{F} = e^{x^2+y^2+z^2}(xz, yz, xy)$ is not conservative.

(g) $\vec{F} = \left(xy - \sin(z), \frac{1}{2}x^2 - \frac{e^y}{z}, \frac{e^y}{z^2} - x\cos(z)\right)$ is conservative. A potential function is $f(x,y,z) = \frac{x^2y}{2} - \frac{e^y}{z} - x\sin(z)$.

Assignment 18.11 — $\oint_C \left(\left(\sin(x) + 3y^2\right)\, dx + \left(2x - e^{-y^2}\right)\, dy\right) = \int_0^{\pi}\int_0^{a} \left(2 - 6r\sin(\theta)\right)\, r\, dr\, d\theta = \pi a^2 - 4a^3$

Assignment 18.12 —

(a) $\oint_C \left(y^2\, dx - xy\, dy\right) = -\frac{1}{2}$

(b) $\oint_C \left(\left(2xy - x^2\right)\, dx + \left(x + y^2\right)\, dy\right) = \frac{1}{30}$

(c) $\oint_C \left(\ln\left(1 + y^{2/3}\right)\, dx + x^2\, dy\right) = \frac{84}{5} - 2\ln(5) - 2\arctan(2)$

(d) $\oint_C \left(\left(x^2-xy\right)\ dx+\left(xy-y^2\right)\ dy\right)=-\frac{4}{3}$

(e) $\oint_C \left(\left(x\sin\left(y^2\right)-y^2\right)\ dx+\left(x^2y\cos\left(y^2\right)+3x\right)\ dy\right)=9$

Assignment 18.13 — We choose as vector field $\vec{\boldsymbol{F}}=\left(\frac{-y}{2},\frac{x}{2}\right)$.

$$\iint_R\ dA=\oint_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}}=\frac{1}{2}\oint_C(-y\ dx+x\ dy)=\frac{3ab}{2}\int\limits_0^{2\pi}\sin^2(t)\cos^2(t)\ dt=\frac{3\pi ab}{8}$$

Assignment 18.14 —

Assignment 18.15 —

Assignment 18.16 —

Assignment 18.17 —

Assignment 18.18 —

Assignment 18.19 —

Assignment 18.20 —

Assignment 18.21 —

Assignment 18.22 —

Assignment 18.23 —

Assignment 18.24 —

E

Review exercises

E.1 Exam May 2019

Assignment E.1 — For each statement, indicate whether it is true or false. Also briefly motivate your answer.

(a) The function

$$f(x) = \frac{\sinh^2(x) + \cosh(x)}{\coth(x)\,\mathrm{sech}(x)}$$

is not even

(b) Let f be continuous and strictly positive ($f(x) > 0$) on $[a,b]$, then $1/f(x)$ takes all values between $1/f(a)$ and $1/f(b)$.

(c) The function $f(x) = e^x|x-2|$ is not differentiable in $x = 2$.

(d) For the function $y = x^{1/3}$ considered on the interval $[-2,2]$, we can find a $c \in \mathbb{R}$ such that

$$f'(c) = \frac{f(2) - f(-2)}{2-(-2)}\,.$$

(e) If f is a continuous function over $[a,b]$, then no constants m and M exist such that

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)\,.$$

(f) If all contours of $f(x,y)$ are parallel lines, then the graph of f is a plane.

Assignment E.2 — Consider the following theorem

Theorem E.1 (Limits and one-sided limits)
Let f be a function defined on an open interval I containing c, then

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

if and only if

$$\lim_{\substack{x\to c\\<}} f(x) = L \quad \text{and} \quad \lim_{\substack{x\to c\\>}} f(x) = L.$$

Below is an attempt to prove this theorem, but it contains a couple of logical errors. Correct these errors on this sheet.

We first prove the necessary condition ($\Rightarrow$). For that purpose, let $\lim_{x\to c} f(x) = L$, so that, according to the definition of the limit of a function, it holds that

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : 0 \le |x-c| < \delta \Rightarrow f(x) - L \le \epsilon\big)\,.$$

The antecedent $0 \le |x-c| < \delta$ can also be written as

$$-\delta < x - c \le 0 \quad \wedge \quad 0 \le x - c < \delta\,,$$

from which, by addition of c, it immediately follows that

$$c - \delta < x \le c \quad \wedge \quad c \le x < c + \delta\,.$$

Consequently, we see that

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : c - \delta \le x < c \Rightarrow f(x) - L \le \epsilon\big)$$

en

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : c \le x < c + \delta \Rightarrow f(x) - L \le \epsilon\big)\,.$$

Consequently, $\lim_{\substack{x\to c\\<}} f(x) = L$ and $\lim_{\substack{x\to c\\>}} f(x) = L$.

To prove the sufficient condition ($\Leftarrow$), we assume that $\lim_{\substack{x\to c\\<}} f(x) = L$ and $\lim_{\substack{x\to c\\>}} f(x) = L$, so that

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : c - \delta \le x < c \Rightarrow f(x) - L \le \epsilon\big)$$

and

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : c \le x < c + \delta \Rightarrow f(x) - L \le \epsilon\big)\,.$$

From this, we conclude that

$$c - \delta \le x < c \qquad \vee \qquad c \le x < c + \delta\,,$$

from which it follows that $0 < |x-c| < \delta$. We may therefore conclude that

$$\forall \epsilon > 0, \exists\, \delta(\epsilon) > 0 : \big(\exists x \in I \setminus \{c\} : 0 \le |x-c| < \delta \Rightarrow f(x) - L \le \epsilon\big)\,.$$

Assignment E.3 — Consider the midpoint method on the interval $[a,b]$, which we subdivide into n subintervals as

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b\,,$$

where $\Delta x = x_{i+1} - x_i = \frac{b-a}{n}$ for all $i = 1, \dots, n+1$. Using this method, we approximate

$$S = \int_a^b f(x)dx$$

as

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f(m_i)\,,$$

where $m_i = \frac{x_i + x_{i+1}}{2}$ and the total error on the approximation of S is nothing but

$$E = \left| S - \hat{\boldsymbol{S}} \right|\,.$$

In what follows we will prove that the following holds for the upper bound on the total error E of the midpoint approximation:

$$E \leq \frac{B(b-a)^3}{24n^2}\,,$$

where B is a constant.

Complete in this bundle where you find a dotted line.

This total error cannot be greater than the sum of the approximation errors E_i for the subintervals, so it holds that

$$E \leq \sum_{i=1}^{n} E_i\,.$$

The local error E_i is nothing but the net area between the tangent to f in $x = m_i$ and the graph of f on $[x_i, x_{i+1}]$. Let $l(x)$ be the linear approximation of f in $x = m_i$, i.e.

$$l(x) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

then

$$E_i = \left| \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \right|,$$

from which it follows that

$$E_i \leq \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \tag{E.1}$$

Essentially, the linear function $l(x)$ is nothing but the first-order Taylor polynomial in $x = m_i$ of f, for which we know that the remainder is given by

....

for $z \in [x_i, x_{i+1}]$. If we now define the upper bound of $\left|f''(z)\right|$ for $z \in [x_i, x_{i+1}]$ as B, we immediately find that

$\left|f(x) - l(x)\right| \leq$ ....

Consequently, we find as upper bound for the right-hand side of the inequality in Equation (E.1)

.... (E.2)

Or more explicitly, after calculating the integral in Equation (E.2) and taking into account that $x_{i+1} - m_i = \frac{\Delta x}{2}$ en $x_i - m_i = -\frac{\Delta x}{2}$

....

....

Using this in Equation (E.2), we obtain as upper bound for E_i

....

From this, it follows for the total error of the midpoint approximation E that

....

Assignment E.4 — The height of the groundwater table is described by the so-called diffusion equation, which in Cartesian coordinates is given by

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0\,. \tag{E.5}$$

Here $G = u(x, y)$ is nothing but the groundwater table height. To describe G over a circular surface, we reformulate the diffusion equation in polar coordinates using the transformation $r^2 = x^2 + y^2$ en

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

Show that Equation (E.5) in polar coordinates becomes

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0\,.$$

Assignment E.5 — Determine the corresponding function rule for each graph in Figure E.1.

(a) $y = \sin(2\arcsin(x))$

(b) $y = \sin(2\arcsin(|x|))$

(c) $y = \sin(2|\arcsin(x)|)$

(d) $y = \arcsin(2\sin(x))$

(e) $y = \arcsin(2\sin(|x|))$

(f) $y = \arcsin(2|\sin(x)|)$

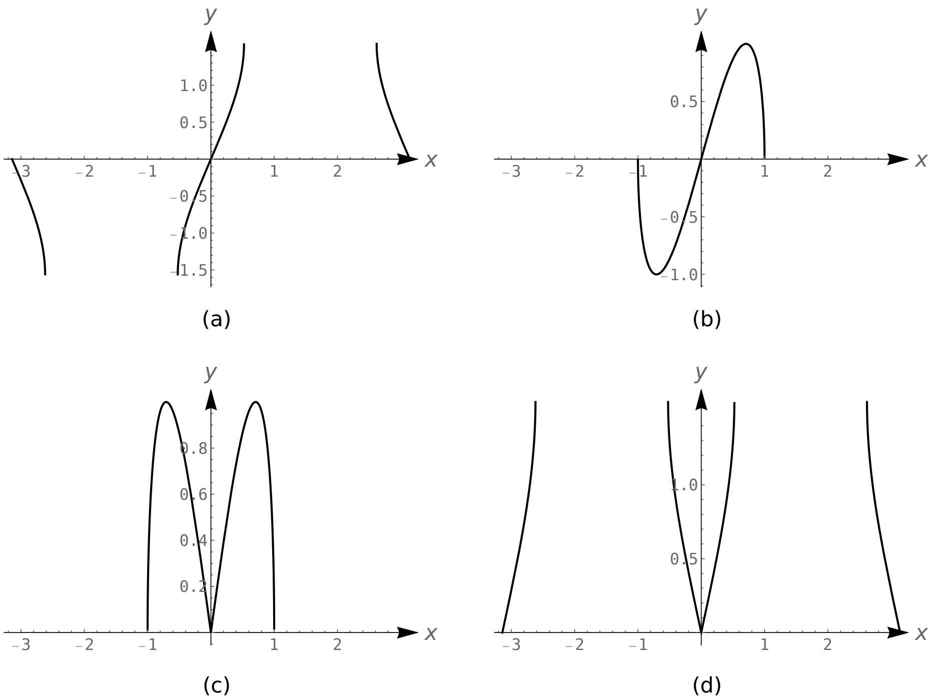


Figure E.1

Assignment E.6 — Determine the value(s) of the parameter p for which the following function is increasing on $]-\infty,+\infty[$

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

Assignment E.7 — For each of the problems below, construct the integral that allows you to calculate what is asked:

(a) the area of the region bounded by $y = x^3$, $y = 0$ and the tangent line in $(-1,-1)$ to the graph of $y = x^3$.

(b) the volume of the body created by rotating the region enclosed between $y = -x^2+1$ and $y = x^2-3$ about the line $y = -4$.

(c) the distance traveled by a particle moving between $t = 0$ and $t = 1$ along the curve C parameterised by $\vec{r}(t) = \left(2t^{3/2}, \cos(2t), \sin(2t)\right)$;

(d) the work done by a particle moving from $(1,1,1)$ to $(2,4,8)$ along the curve C parametrised by $\vec{r}(t) = \left(t, t^2, t^3\right)$ and subject to the field $\vec{F} = \left(\sin(x), \sin(y), \sin(z)\right)$.

Assignment E.8 — Calculate

$$I = \int_0^1 \int_{\arcsin(y)}^{\frac{\pi}{2}} \cos(x)\sqrt{1+\cos^2(x)}\,dx\,dy\,.$$

Assignment E.9 — Change the order of integration in

$$I = \int_{-1}^{1}\int_{x^2}^{1}\int_{0}^{1-y} f(x,y,z)\,dz\,dy\,dx$$

to $dx\,dy\,dz$.

Assignment E.10 — Determine the value(s) of the parameter p for which the following series converges

$$\sum_{n=3}^{+\infty} \frac{1}{n\ln(n)\Big[\ln\big(\ln(n)\big)\Big]^p}\,.$$

Assignment E.11 — Show that for $n \in \mathbb{Z}$ it holds that

$$\int_0^1 \frac{\cos(nx)}{x+1}\,dx \le \ln(2)\,.$$

Assignment E.12 — Determine the MacLaurin series expansion of the second order of

$$f(x) = \int_{-x}^{3\cos(x)} \cos\left(xt^2\right)dt\,.$$

Assignment E.13 — Show that $f(x,y) = x^2 + 4y^2 - 4xy + 2$ has an infinite number of critical points and that these are all minima.

Assignment E.14 — The cat Kamu is sleeping peacefully after a strenuous day. However, a mouse has planned an attack on Kamu and is moving with a helicoidal trajectory given by

$$\vec{r}(t) = 6\cos(\pi t)\hat{\imath} + 6\sin(\pi t)\hat{\jmath} + 2t\hat{k}\,,$$

where $t \ge 0$, in the direction of Kamu, who can be represented as a sphere with cartesian equation

$$x^2 + y^2 + z^2 = 100\,.$$

(a) Determine when the mouse arrives at Kamu's skin surface.

(b) Determine where the mouse bites Kamu.

(c) Determine at which angle this happens, i.e. which angle does the skin surface make with the mouse when it inflicts the bite.

E.2 Exam August 2019

Assignment E.15 — For each statement, indicate whether it is true or false. Also briefly motivate your answer.

(a) Let A and B be non-empty, bounded subsets of $\mathbb{R}$. If $\sup A = \inf B$, then for every element a of A there exists an element b of B such that $a \leq b$.

(b) The graph of $r(\theta) = 4\cos(\theta - \pi/3)$, $0 \leq \theta \leq \pi$, is not a circle.

(c) Each contour surface of a function of three variables can be seen as a surface in three-dimensional space.

(d) Figure E.2 shows the surface $z = f(x, y)$. We can conclude that $f_{xx}(P) > 0$ and $f_{yy}(P) < 0$.

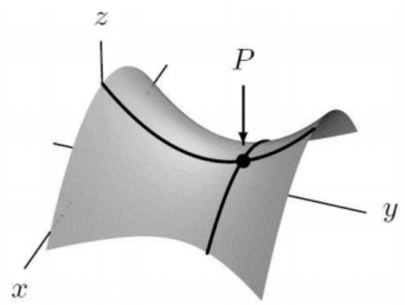


Figure E.2: Surface $z = f(x, y)$ from Question 4.

(e) Consider a point P on a curve C for which there are two different parameterisations, $\vec{\boldsymbol{r}}_1(t)$ and $\vec{\boldsymbol{r}}_2(t)$. Then the tangent vector of $\vec{\boldsymbol{r}}_1(t)$ in P is the same as the one of $\vec{\boldsymbol{r}}_2(t)$.

(f) Consider the vector field $\vec{\boldsymbol{F}}(x, y, z) = (x + \sin(y), y - \sin(z), z)$. The divergence of this vector field is $(1, 1, 1)$.

Assignment E.16 — Determine the corresponding function rule for each graph in Figure E.3.

(a) $x(t) = t^4 - t + 1, \quad y(t) = t^2$

(b) $x(t) = t^2 - 2t, \quad y(t) = \sqrt{t}$

(c) $x(t) = \sin(2t), \quad y(t) = \sin(t + \sin(2t))$

(d) $x(t) = \cos(5t), \quad y(t) = \sin(2t)$

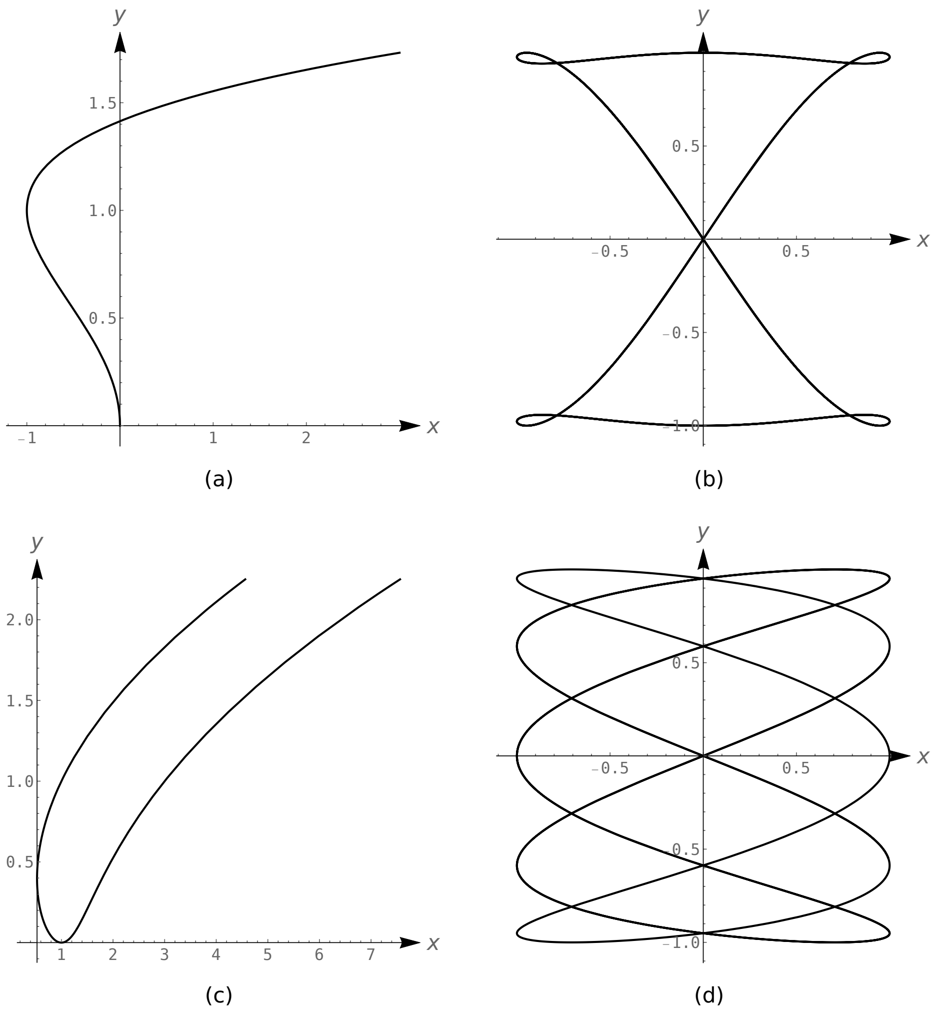


Figure E.3

Assignment E.17 — Find out where the f function below is continuous and differentiable.

$$f(x) = \begin{cases} \tan(x), & \text{if} \quad x < 0, \\ x, & \text{if} \quad 0 \le x < 1, \\ 1 + \ln(x), & \text{if} \quad 1 \le x. \end{cases}$$

Assignment E.18 — Discuss

$$\lim_{x \to 0} x^p \left(1 - \sqrt{1+x}\right)$$

as a function of the parameter $p \in \mathbb{Z}$.

Assignment E.19 — A water reservoir has a parabolic shape and is formed by the rotation of the parabola $y = ax^2$ about the y-axis. The maximum water depth in this reservoir is 8 m and when the reservoir is filled to a height of 5 m, the water surface has a diameter of 20 m. What is the maximum volume of water that this reservoir can hold?

Assignment E.20 — Consider the following theorem.

Theorem E.2 (Generalised mean value theorem)
Let f and g be two continuous functions on $[a,b]$ and differentiable on $]a,b[$. If $g'(x) \neq 0$ for every $x \in]a,b[$, then there exists a $c \in]a,b[$ such that

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

Below is an attempt to prove this theorem, but it contains a couple of logical errors. Correct these errors.

From $g'(x) \neq 0$, for every $x \in]a,b[$, it follows that $g(b) \neq g(a)$. Otherwise, g' would equal 0 for every $c \in]a,b[$.
We apply the mean value theorem to the function

$$h(x) = \big(f(b)-f(a)\big)\big(g(x)-g(a)\big) - \big(g(b)-g(a)\big)\big(f(x)-f(a)\big).$$

Since $h(a) = h(b) \neq 0$, for every $c \in]a,b[$, it will hold that $h'(c) = 0$. We obtain

$$f'(c)\big(g(x)-g(a)\big) - g'(c)\big(f(x)-f(a)\big) = 0,$$

From which it follows that

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

Assignment E.21 — For each of the problems below, construct the integral that allows you to calculate what is asked. Write the integrandum as simply as possible.

(a) the area of the region outside $r = a$ and bounded by $r = 2a\cos(3\theta)$.

(b) the surface area of $2z = x^2 + y^2$, cut off by $x^2 + y^2 = 1$.

(c) the line integral of $\vec{\boldsymbol{V}} = \left(x^2 + y, xy + 2y^2\right)$ along the closed curve K where K connects the points $A(0,0)$, $B(5.0)$ and $C(0\sqrt{20})$. AB and CA are line segments and BC is the part of the ellipse $4x^2 + 5y^2 = 100$ in the first quadrant.

Assignment E.22 — Determine the convergence of

$$\sum_{n=1}^{+\infty} \frac{(-1)^n}{\cosh(n)}.$$

Assignment E.23 — We consider the function $f(x)$ which we can approximate with a Taylor series of the n-th order in the point a. We write f as

$$f(x) = T_n(x) + R_n(x),$$

where $T_n(x)$ is the Taylor polynomial of the n-th order of f in a and $R_n(x)$ is the remainder of this sequence which can be written as follows.

Theorem E.4 (Alternative formulation for the remainder)
Let $f^{(n+1)}$ be continuous on an open interval I containing a and x, then

$$R_n(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt.$$

We prove this theorem below by means of mathematical induction. Complete in this bundle where you find a dotted line.

For $n=1$ it holds that

$$R_1(x) = f(x) - T_1(x) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

and the integral in the theorem becomes

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

We calculate this integral by using integration by parts with

$$u(t) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{and} \quad dv(t) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

from which it follows that

$$du(t) = -dt \quad \text{and} \quad v(t) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.$$

Thus, the integral becomes

$$\int_a^x \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\, dt$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$= R_1(x).$$

This proves the theorem for $n=1$.

We assume that the theorem holds for $n = k$, so

$$R_k(x) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

We then prove the theorem for $n = k + 1$, so

$$R_{k+1}(x) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

We again use integration by parts to calculate this last integral. We set

$$u(t) = \dots\dots\dots\dots\dots\dots\dots\dots \quad \text{and} \quad dv(t) = \dots\dots\dots\dots\dots\dots\dots\dots,$$

from which it follows that

$$du(t) = -(k+1)(x-t)^k\,dt \quad \text{and} \quad v(t) = \dots\dots\dots\dots\dots\dots\dots\dots.$$

The integral is thus

$$\begin{aligned}
&\frac{1}{(k+1)!}\int_a^x \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\, dt \\
&= \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&= \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&= -\frac{f^{(k+1)}(a)}{(k+1)!}(x-a)^{k+1} + R_k(x) \\
&= f(x) - \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&= f(x) - \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&= R_{k+1}(x).
\end{aligned}$$

The theorem is valid for $n = k + 1$ if it is valid for $n = k$, therefore it is valid for all $n \in \mathbb{N}$.

Assignment E.24 — Consider the surface

$$z + \ln(z) - xy = 0\,.$$

(a) Show that the above relation determines z as an implicit function of x and y in an environment of $(1/2,2,1)$.

(b) Determine the equation of the tangent plane to the surface at the point $(1/2,2,1)$.

(c) Determine

$$\frac{\partial^2 z}{\partial x \partial y}.$$

Assignment E.25 — Consider the double integral

$$I = \int_0^2 \int_0^{\sqrt{4-y^2}} (x^2+y^2)\,dx\,dy\,.$$

(a) Rewrite I in polar coordinates.

(b) Rewrite I as a triple integral in cylinder coordinates.

Assignment E.26 — Consider the double integral

$$I = \iint_D f(x,y)\,dy\,dx$$

with D the area within $(x-1)^2+(y-1)^2=2$ ánd on the convex side of $(y-1)^2=1-2x$. Give x and y concrete boundaries.

Assignment E.27 — Calculate

$$I = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \sqrt{x^2+y^2+z^2}\,e^{-(x^2+y^2+z^2)}\,dx\,dy\,dz.$$

E.3 Exam January 2020

Assignment E.28 — For each statement, indicate whether it is true or false. Also briefly motivate your answer.

(a) If

$$f(x) = \frac{x^2-1}{x+1} \qquad \text{and} \qquad g(x) = x-1\,,$$

then it holds that $f(x)=g(x)$.

(b) If $x=1$ is a vertical asymptote of the graph of $y=f(x)$, then the function $f(x)$ is not defined at $x=1$.

(c) The curves $4 = r\cos(\theta - 2\pi/5)$ and $\theta = 2\pi/5$ determine (half-)lines that are perpendicular to each other.

(d) If $f(x)$ has a Taylor series expansion around $x_0 = 0$ with radius of convergence R, than

$$g(x) = f\left(\frac{x-1}{2}\right)$$

has a Taylor series expansion around $x_0 = 1$ with radius of convergence $2R$.

Assignment E.29 — Use the (ϵ, δ) definition to show that

$$\lim_{\substack{x \to 4 \\ >}} \sqrt{x-4} = 0\,.$$

Assignment E.30 — Consider the real-valued function f:

$$f(x) = \frac{\ln\left(\ln\left(x^2\right)\right)}{2x^3 - 4x}\,.$$

(a) Determine the domain of f.

(b) Determine whether f is an even or odd function.

(c) Determine all roots of f.

(d) Determine the asymptotes of f.

(e) Discuss the behaviour, extrema and concavity of f on the interval $[0.3]$, given the graphs of $f'(x)$ and $f''(x)$ in Figure E.4(a) and E.4(b) and Figure E.4(c) and E.4(d), respectively.

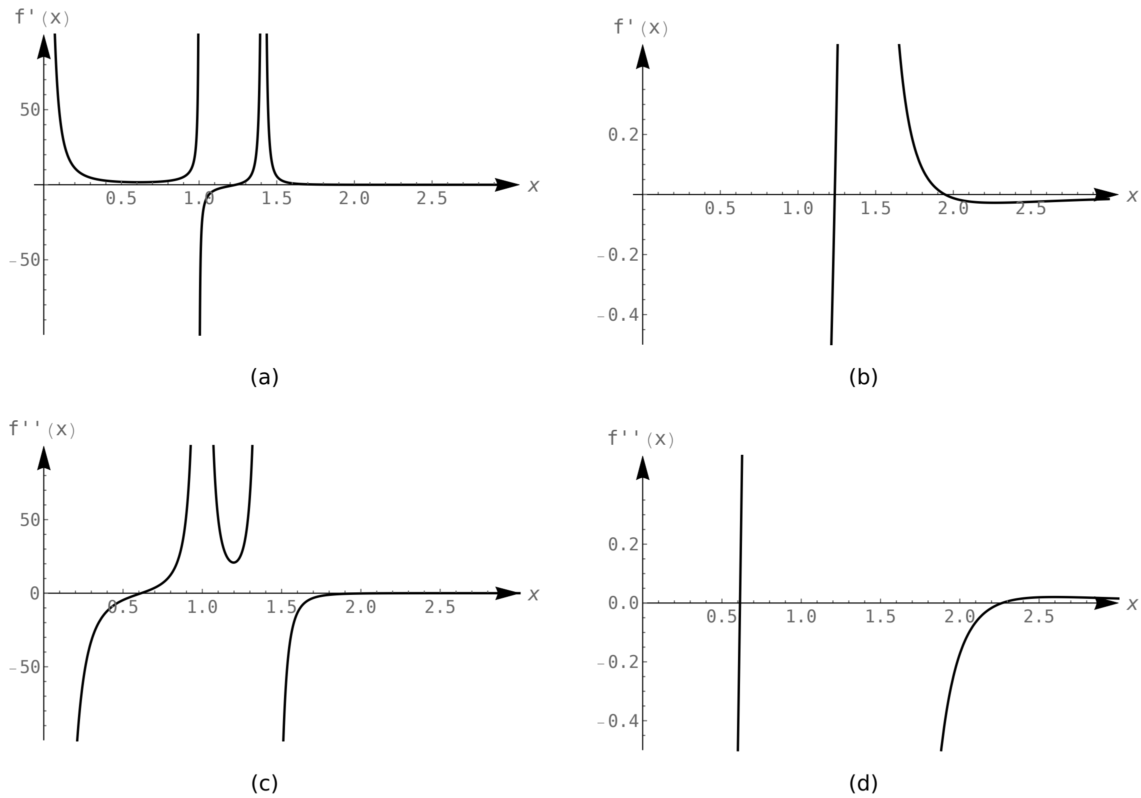


Figure E.4

Assignment E.31 — Construct up the integral to calculate the volume of solid of revolution obtained by rotating the region within $x^2+y^2=2$ AND within $(x-1)^2+y^2=1$ about the line with equation $x=2$.

Assignment E.32 — Consider the curve in $\mathbb{R}^3$ parameterised as

$$\vec{r}(t) = \big(2t, 3\sin(2t), 3\cos(2t)\big).$$

At what point in $\mathbb{R}^3$ do we arrive after traveling a distance of $\sqrt{10}\pi/3$ along this curve?

Assignment E.33 — Determine the parameter p so that

$$f(\mathbf{x}) = \left(\sum_{i=1}^{n} x_i^2\right)^p$$

is a solution of

$$\sum_{i=1}^{n} f_{x_i x_i} = 0.$$

Assignment E.34 — Consider a slab of negligible thickness with mass density $\delta(x,y,z)$ whose shape is described by $z=f(x,y)=\sqrt{x^2+y^2}$.

(a) Given the area of an infinitesimal particle of this plate in cartesian coordinates, i.e.

$$dA = \sqrt{1+f_x^2+f_y^2}\, dx\, dy,$$

Determine a formula for the mass of this plate above a region R in the xy plane.

(b) Determine the mass of the plate for which $z=f(x,y)=\sqrt{x^2+y^2}$, $1 \le z \le 4$ en $\delta(x,y,z) = x^2 z$.

Assignment E.35 — Cats have a very sensitive sense of smell that allows them to detect prey. During the hunt they typically follow the path along which the concentration of odour molecules increases the most. The cats Kamu and Kamiel are in their garden where the concentration of odour molecules is given by

$$z = f(x,y) = e^{\cos(x)-y^2}.$$

(a) Is the function f bounded or unbounded?

(b) Suppose that Kamu and Kamiel are at the point $(x,y)=(3\pi/2,1)$ at the start of the hunt. In which direction will they then move?

(c) What is the rate at which the concentration of odour molecules increases in the direction determined under (b)?

(d) Determine the rule of the implicit function $F(x,y)=0$ that describes the traveled hunting trajectory.

(e) Does $F(x,y)=0$ define an explicit function $y=g(x)$ in the neighbourhood of $(x,y) = (3\pi/2,1)$?

(f) If possible, determine $y=g(x)$.

Assignment E.36 — Consider the graph of the function $f(x,y)$ on the region $[-2,2] \times [-2,2]$ in Figure E.5.

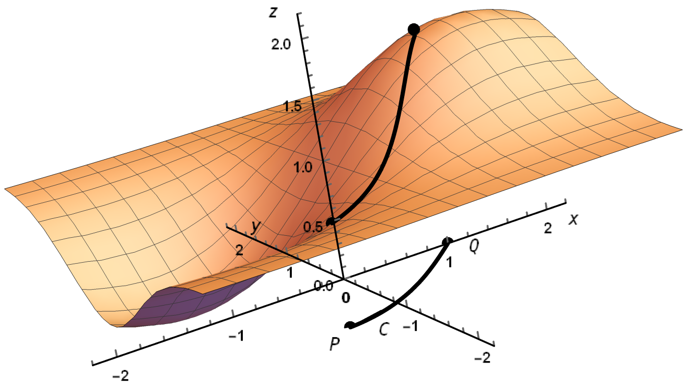


Figure E.5: Graph of the function $f(x,y)$ from Question E.36.

Circle the correct statement.

(a) At the point $P(-0.5,-1)$ it holds that

(i) $\dfrac{\partial f}{\partial y} > 0$, (ii) $\dfrac{\partial f}{\partial y} = 0$, (iii) $\dfrac{\partial f}{\partial y} < 0$.

(b) At the point $Q(1,0)$ it holds that

(i) $\dfrac{\partial^2 f}{\partial x^2} > 0$, (ii) $\dfrac{\partial^2 f}{\partial x^2} = 0$, (iii) $\dfrac{\partial^2 f}{\partial x^2} < 0$.

(c) The line integral $\displaystyle\int_C f(s)ds$ is

(i) strictly positive, (ii) zero, (iii) strictly negative.

(d) The vector field $\vec{\nabla} f$ is shown in

(i) Figure E.6(a), (ii) Figure E.6(b), (iii) Figure E.6(c), (iv) Figure E.6(d).

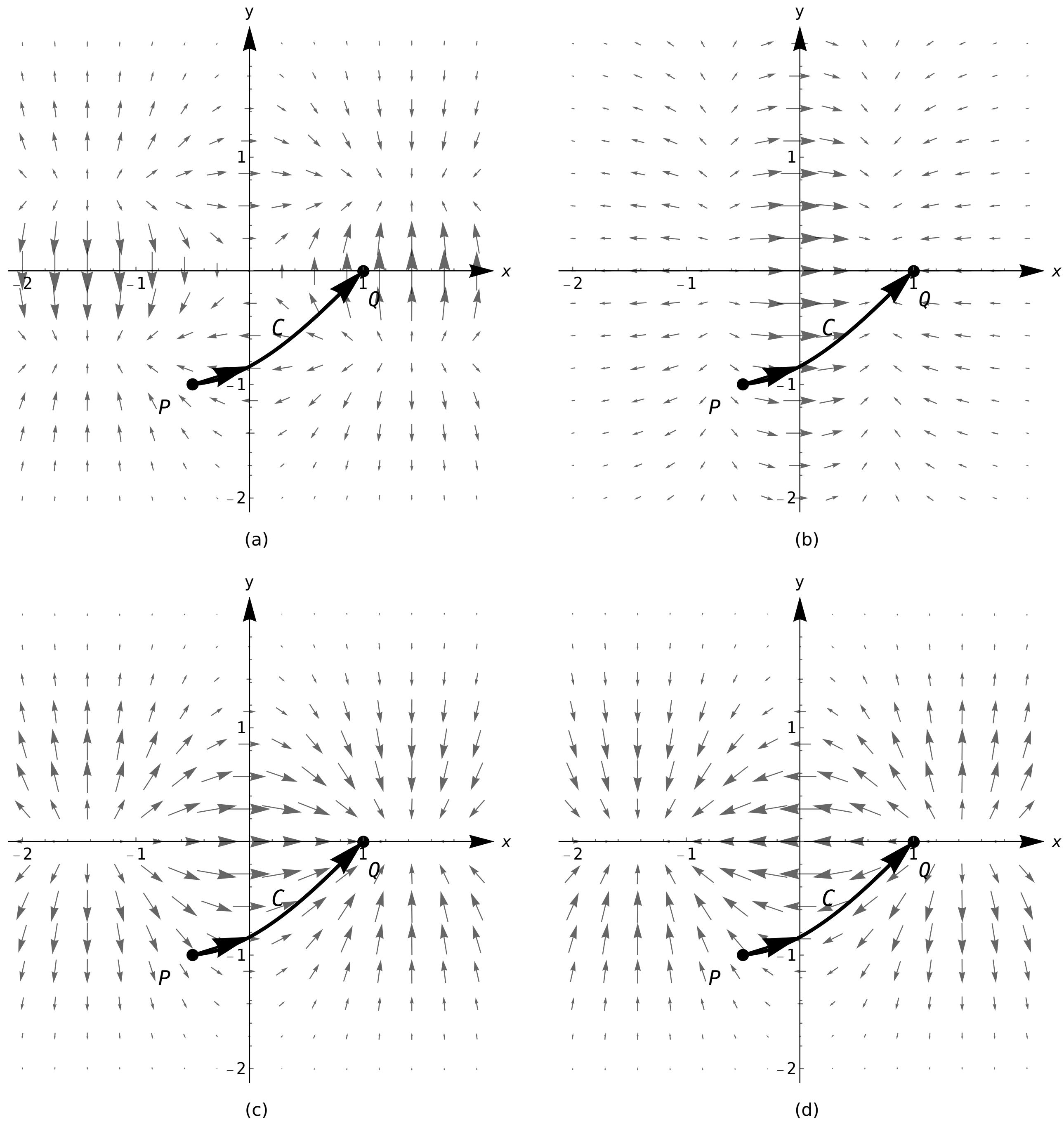


Figure E.6

E.4 Exam August 2020

Assignment E.37 — Suppose

$$|f(x) - f(y)| \leq |x - y|^n$$

where $x, y \in \mathbb{R}$ and $n > 1$. Prove that f is a constant function.

Assignment E.38 — Consider the real-valued function f:

$$f(x) = \csc(x) - 2\sin(x).$$

(a) Determine the domain and image of f.

(b) Determine the periodicity of f.

(c) Determine the symmetry of f.

(d) Determine all roots of f.

(e) Determine the asymptotes of f.

(f) Determine the extrema and inflection points of f.

(g) Examine the behavior of f and sketch its graph.

Assignment E.39 — Calculate the volume of the solid of revolution resulting from the rotation of the region enclosed by $y = 0$, $y = x$, $xy = 9$ and $x + y = 10$ for $x \geq y$ about the y axis.

Assignment E.40 — Determine the radius of convergence and the convergence interval of:

$$\sum_{n=1}^{+\infty} \frac{\ln(n+1)}{\sqrt{n}} (2x-3)^n .$$

Assignment E.41 — Consider the squeeze theorem for functions of n variables.

Theorem E.6 (Generalised squeeze theorem)
Let f, g and h be functions on an open ball B in $\mathbb{R}^n$ containing the point $P = \mathbf{x}_0$ and for which it holds that

$$f(\mathbf{x}) \leq g(\mathbf{x}) \leq h(\mathbf{x})$$

for all $\mathbf{x}$ in B.

It then holds that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} g(\mathbf{x}) = L$$

if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = L = \lim_{\mathbf{x} \to \mathbf{x}_0} h(\mathbf{x}) .$$

(a) Prove this theorem.

(b) Illustrate this statement and your proof with a sketch.

(c) Use this theorem to calculate the following limit:

$$\lim_{(x,y) \to (0,0)} x^4 \sin\left(\frac{1}{x^2 + |y|}\right) .$$

Assignment E.42 — Consider the curve with polar equation $r(\theta) = \sin(\theta)\cos(\theta)$ whose graph is shown in Figure E.7.

(a) Construct the double integral to find the area A_1 of the region outside $r(\theta) = \sin(\theta)\cos(\theta)$ and inside $r = 0.5$.

(b) Construct the double integral to determine the area A_2 of the region inside $r(\theta) = \sin(\theta)\cos(\theta)$ and outside $r = \sqrt{2}/4$.

(c) Calculate A_1 *or* A_2.

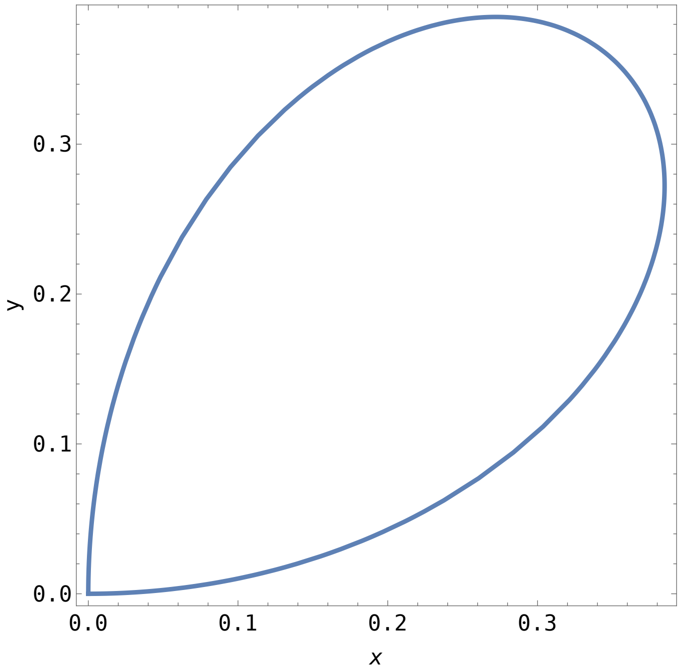


Figure E.7

Assignment E.43 — Calculate the line integral on the *conservative* vector field $\vec{\boldsymbol{F}}\left(\frac{1}{x^2+1}, 2yz, y^2\right)$ along the curve C with parameterisation $(t, 2t, \arcsin(t))$ for $t \in [0,1]$.

E.5 Exam January 2021

Assignment E.44 — The proportion of the population p that adheres to the corona measures is described by:

$$p(t) = \frac{e^{\beta_0+\beta_1(t-t^*)}}{1+e^{\beta_0+\beta_1(t-t^*)}},$$

where t represents the time in days, $t^* > 0$ the day on which measures take effect and $\beta_0 > 0$ and $\beta_1 > 0$ parameters. These parameters influence the time instant when we transition from an increase in the acceptance of the measures with increasing speed to an increase with decreasing speed, i.e. the inflection point of the graph of $p(t)$. The faster this time instant is reached, the more efficiently we can slow down the spread of the virus.

Show that awareness campaigns to expedite this time instant should aim to increase β_0 or decrease β_1.

Assignment E.45 — If we use the ϵ-δ definition of the limit in our definition for the continuity of a function of one variable f in a point c, then we arrive at the following definition.

Theorem E.7 (Continuity in a point)
Let I be an open interval and $c \in I$, then a function $f : I \to \mathbb{R}$ is continuous in the point c if for all $\epsilon > 0$ a $\delta > 0$ exists such that $|f(x) - f(c)| < \epsilon$ if $|x - c| < \delta$ for all $x \in I$.

However, we can also define a stronger form of continuity.

Theorem E.8 (Uniform continuity)
A function $f : I \to \mathbb{R}$ is uniformly continuous over I if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ if $|x - y| < \delta$ for all $x, y \in I$.

(a) Explain what uniform continuity means.

(b) How does uniform continuity (Definition E.8) differ from continuity (Definition E.7)?

(c) Explain why uniform continuity implies continuity.

(d) We know that $f(x) = x^2$ is continuous on its domain. Show that this function is not uniformly continuous on its domain.

Assignment E.46 — Consider the real-valued function f:

$$f(x) = \arccos\left(\ln\left(x^2 + 1\right)\right).$$

(a) Determine the domain and image of f.

(b) Examine the symmetry of f.

(c) Calculate the derivative f'.

(d) How does f' behave at the boundary points of the domain?

(e) Show that f reaches a maximum on its domain.

Assignment E.47 — Calculate the following integral:

$$I = \int \ln\left(x^2 + 2x + 2\right) dx\,.$$

Assignment E.48 — Consider the circle with equation $r = 3\sin(\theta)$ and the cardioid with equation $r = 1 + \sin(\theta)$ shown in Figure E.8.

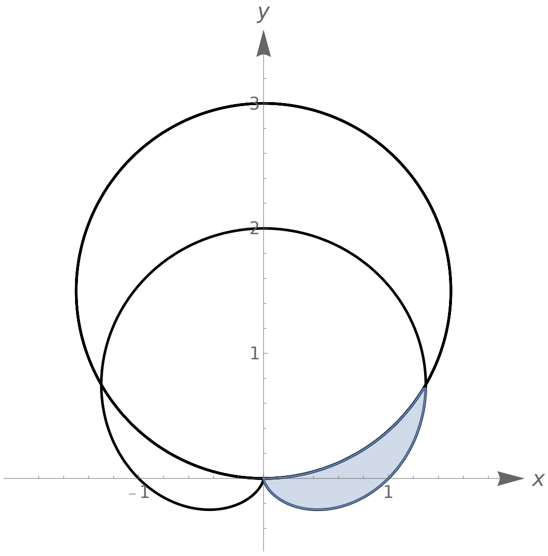


Figure E.8: Graph of the circle $r = 3\sin(\theta)$ and the cardioid $r = 1 + \sin(\theta)$.

(a) Construct the integral that allows you to calculate the area of the shaded region.

(b) Derive a formula for the volume of the solid of revolution obtained by rotating the area contained between the graphs of $r = f(\theta)$ and $\theta = 0$ about the x axis.

(c) Construct the integral that allows you to determine the volume of the solid of revolution resulting from rotation about the x-axis of the region enclosed by $r = 3\sin(\theta)$, $y = 0$ and $x = 3/2$.

Assignment E.49 — The vector-valued functions below describe the paths of two flying insects.

$$\vec{\boldsymbol{r}}_1(t) = \left(t^2, 2t+3, t^2\right) \quad \text{and} \quad \vec{\boldsymbol{r}}_2(t) = \left(5t-6, t^2, 9\right)$$

(a) At what time instant will the insects collide?

(b) Where will the collision occur?

(c) Determine the angle between the paths at the point where the collision occurs.

Assignment E.50 — Given is the function of three variables:

$$E(N, V, S) = \frac{3h^2 N}{4\pi m}\left(\frac{N}{V}\right)^{2/3} e^{\left(\frac{2S}{3Nk} - \frac{5}{3}\right)},$$

where h, m en k are known constants. Now introduce the temperature via $T = \frac{\partial E}{\partial S}$ and the pressure via $p = -\frac{\partial E}{\partial V}$

(a) Show that

$$p(N, T, V) = \frac{kNT}{V}.$$

(b) Calculate $\vec{\nabla} p$.

(c) Calculate the directional derivative of p in the direction determined by a directional vector of the line that is the intersection of $V = 0$ and $N = T$.

(d) Consider the level surface $p = k$. Calculate the tangent plane to this level surface at the point $(1,1,1)$.

E.6 Exam Augustus 2021

Assignment E.51 — Consider the real-valued function

$$f(x) = \begin{cases} \dfrac{e^x - 1}{|x|}, & x \neq 0, \\ 1, & x = 0. \end{cases}$$

(a) Examine the continuity of f. Also specify where f is left or right continuous.

(b) Classify any discontinuities.

(c) Calculate $f'(x)$ for all $x \neq 0$. Is $f'(x)$ (strictly) positive or negative for $x > 0$?

(d) Now, let $\operatorname{dom} f = \mathbb{R}_0^+$.

(i) Why does the inverse function f^{-1} exist on $]0, +\infty[$?

(ii) Calculate $\left(f^{-1}\right)'(e-1)$.

Assignment E.52 — If we use the (ϵ,δ) definition of the limit in our definition for the continuity of a function of one variable f at a point c, we arrive at the following definition.

Theorem E.9 (Continuity in a point)
Let I be an open interval and $c \in I$, then the function $f : I \to \mathbb{R}$ is continuous in the point c if for every $\epsilon > 0$ there exist a $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ if $|x - c| < \delta$ for every $x \in I$.

However, we can also define an alternative form of continuity.

Theorem E.10 (Lipschitz continuity)
A function $f : I \to \mathbb{R}$ is Lipschitz continuous on I if there exists a number L such that for all $x, y \in \mathbb{R}$ it holds true that

$$|f(x) - f(y)| \leq L|x - y|.$$

The smallest number L for which this inequality holds true, is called the Lipschitz constant of f.

(a) Explain what Lipschitz continuity means.

(b) Wherein does Lipschitz continuity differ (Definition E.10) from continuity (Definition E.9)?

(c) Show that the sine and consine functions are Lipschitz continuous and determine their Lipschitz constants.

(d) Give an example of a function that is over its domain

1. differentiable, but not Lipschitz continuous.
2. not differentiable, but Lipschitz continuous.

Assignment E.53 — Determine the volume of the body of revolution created by rotating the area within $x^2+y^2=2$ and within $(x-1)^2+y^2=1$ around the line $x=2$.

Assignment E.54 — Determine the convergence radius and convergence interval of the power series below:

$$\sum_{n=1}^{+\infty} \frac{\ln(n+1)}{\sqrt{n}}(2x-3)^n$$

Thoroughly motivate your decisions.
The convergence interval is $[1,2[$. The convergence radius is $1/2$.

Assignment E.55 — A function $f:\mathbb{R}^n \to \mathbb{R}$ is called f homogeneous of degree $p \in \mathbb{R}$ if

$$\forall \vec{\mathbf{x}} \in \mathbb{R}^n,\ \forall t>0: \quad f(t\vec{\mathbf{x}}) = t^p f(\vec{\mathbf{x}}). \tag{E.6}$$

Euler's homogeneous function theorem applies to such functions.

Theorem E.11 (Euler's homogeneous function theorem)
If $f:\mathbb{R}^n \to \mathbb{R}$ is differentiable over $\mathbb{R}^n$, then it holds true that f is homogeneous of degree p if and only if

$$\vec{\mathbf{x}} \cdot \nabla f(\vec{\mathbf{x}}) = pf(\vec{\mathbf{x}}), \quad \forall \vec{\mathbf{x}} \in \mathbb{R}^n. \tag{E.7}$$

(a) Prove the sufficient condition ($\Longrightarrow$).

(b) Complete the proof below for the necessary condition ($\Longleftarrow$).

Prove Assume that eq. (E.7) holds true and choose $\vec{\mathbf{x}} \in \mathbb{R}^n$ at random, but fixed. We define the following function:

$$g(t) = f(t\vec{\mathbf{x}}) - t^p f(\vec{\mathbf{x}}),$$

for all $t>0$. The theorem is proven if we can show that $g(t) = 0$ for all $t>0$.

Apparently $g(1)=0$. Derivation to t gives us:

$g'(t) =$ ....

As $t\vec{\mathbf{x}} \in \mathbb{R}^n$, for all $t>0$, applies to the understated

$$(t\vec{\mathbf{x}}) \cdot \nabla f(t\vec{\mathbf{x}}) = pf(t\vec{\mathbf{x}}),$$

such that we can rewrite $g'(t)$ as a function of $g(t)$ if

$g'(t) =$ .... (E.8)

Next, we define the function h as follows.:

$$g(t) = t^p h(t),$$

for all $t>0$. Deriving the above relationship gives

$g'(t) =$ .... (E.9) from which by equating Equations (E.8) and (E.9) follows

....

such that ....

....

....

....

....

Assignment E.56 — Calculate the volume of the body located outside $x^2+y^2=1$, inside $x^2+y^2=4$, under $z=\sqrt{x^2+y^2}$ and above $z=-\sqrt{x^2+y^2}$.

Assignment E.57 — In each case, perform the requested transformation without calculating the integral.

(a) Rewrite the triple integral below using spherical coordinates.

$$I_1 = \int_0^3 \int_0^{\sqrt{9-y^2}} \int_{\sqrt{x^2+y^2}}^{\sqrt{18-x^2-y^2}} \left(x^2+y^2+z^2\right) dz\,dx\,dy$$

(b) Rewrite

$$I_2 = \int_C xy\,dx + x^2y^3\,dy$$

as a double integral if C is given by the triangle with vertices in $(0,0)$, $(1,0)$ and $(1,2)$ and C is run through in a clockwise direction.

E.7 Oplossingen

Assignment E.1 —

(a) True. Determine $f(-x)$ and check whether or not it is equal to $f(x)$.

(b) True. This is a consequence of the intermediate value theorem that is useful here since $1/f(x)$ is continuous on the considered interval.

(c) True. Determine the left and right derivative in $x=2$ and observe that they are not equal. The function is thus not differentiable in $x=2$.

(d) False. The function is not differentiable in $x=0$, so the mean value theorem cannot be used.

(e) False. f is continuous on a closed interval, so f is bounded on $[a,b]$ and thus has a minimum and maximum here. If we choose m to be equal to this minimum and M to be equal to the maximum, then the inequality is satisfied.

(f) False. We can also encounter this with other surfaces such as a sinusoidal cylinder, for example $z=\sin(x)$.

Assignment E.2 — We first prove the necessary condition (⇒). For that purpose, let $\lim_{x\to c} f(x)=L$, so that, according to the definition of the limit of a function, it holds that

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : 0<|x-c|<\delta \Rightarrow |f(x)-L|<\epsilon\right).$$

The premise $0<|x-c|<\delta$ can also be written as

$$-\delta<x-c<0 \quad\vee\quad 0<x-c<\delta,$$

from which, by addition of c, it immediately follows that

$$c-\delta<x<c \quad\vee\quad c<x<c+\delta.$$

Consequently, we see that

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : c-\delta<x<c \Rightarrow |f(x)-L|<\epsilon\right)$$

en

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : c<x<c+\delta \Rightarrow |f(x)-L|<\epsilon\right).$$

Consequently, $\lim_{\substack{x\to c\\ <}} f(x)=L$ and $\lim_{\substack{x\to c\\ >}} f(x)=L$.

To prove the sufficient condition (⇐), we assume that $\lim_{\substack{x\to c\\ <}} f(x)=L$ en $\lim_{\substack{x\to c\\ >}} f(x)=L$, so that

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : c-\delta<x<c \Rightarrow |f(x)-L|<\epsilon\right)$$

and

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : c<x<c+\delta \Rightarrow |f(x)-L|<\epsilon\right).$$

From this, we conclude that

$$c-\delta<x<c \qquad\wedge\qquad c<x<c+\delta,$$

from which it follows that $0<|x-c|<\delta$. We may therefore conclude that

$$\forall\epsilon>0, \exists\,\delta(\epsilon)>0 : \left(\forall x\in I\setminus\{c\} : 0<|x-c|<\delta \Rightarrow |f(x)-L|<\epsilon\right).$$

Assignment E.3 — Consider the midpoint method on the interval $[a, b]$, which we subdivide into n subintervals as

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b\,,$$

where $\Delta x = x_{i+1} - x_i$ for all $i = 1, \ldots, n+1$. Using this method, we approximate

$$S = \int_a^b f(x)dx$$

as

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f(m_i)\,,$$

where $m_i = \frac{x_i + x_{i+1}}{2}$ and the total error on the approximation of S is nothing but

$$E = \left| S - \hat{\boldsymbol{S}} \right|\,.$$

Moreover, this total error cannot be greater than the sum of the approximation errors for the subintervals E_i, so it holds that

$$E \leq \sum_{i=1}^{n} E_i\,.$$

The local error E_i is nothing but the net area between the tangent to f in $x = m_i$ and the graph of f on $[x_i, x_{i+1}]$. Let $l(x)$ be the linear approximation of f in $x = m_i$, i.e.

$$l(x) = f(m_i) + f'(m_i)(x - x_i)\,,$$

then

$$E_i = \left| \int_{x_i}^{x_{i+1}} \big(f(x) - l(x)\big)\,dx \right|\,,$$

from which it follows that

$$E_i \leq \int_{x_i}^{x_{i+1}} \left| f(x) - l(x) \right| dx\,. \tag{E.3}$$

Essentially, the linear function $l(x)$ is nothing but the first-order Taylor polynomial in $x = m_i$ of f, for which we know that the remainder is given by

$$\frac{\max \left| f''(z) \right|}{2} (x - m_i)^2$$

for $z \in [x_i, x_{i+1}]$. If we now define the upper bound of $\left| f''(z) \right|$ for $z \in [x_i, x_{i+1}]$ as B, we immediately find that

$$\left| f(x) - l(x) \right| \leq \frac{B}{2} (x - m_i)^2\,.$$

Consequently, as an upper bound for the right-hand side of the inequality in Equation (E.1) we find that

$$\frac{B}{2} \int_{x_i}^{x_{i+1}} (x - m_i)^2 dx\,. \tag{E.4}$$

Or more explicitly, after calculating the integral in Equation E.1 and taking into account that $x_{i+1} - m_i =$

$\frac{\Delta x}{2}$ en $x_i - m_i = -\frac{\Delta x}{2}$

$$\int_{x_i}^{x_{i+1}} (x-m_i)^2 dx = \frac{1}{3}(x_{i+1}-m_i)^3 - \frac{1}{3}(x_i-m_i)^3 = \frac{\Delta x^3}{12}.$$

Using this in Equation (E.2), we obtain as upper bound for E_i

$$\int_{x_i}^{x_{i+1}} \left|f(x)-l(x)\right| dx \le \frac{B\Delta x^3}{24}.$$

Since $\frac{B\Delta x^3}{24} = \frac{B(b-a)^3}{24n^3}$, we finally find for the upper bound of the total error on the midpoint approximation E that

$$E \le n\left(\frac{B(b-a)^3}{24n^3}\right) = \frac{B(b-a)^3}{24n^2}.$$

Assignment E.4 — You have to calculate the occurring partial derivatives. When partially differentiating U with respect to x, y respectively, you should take into account that U actually depends on r and θ. To partially differentiate U with respect to x, you must first partially differentiate U with respect to r and then r partially with respect to x and analogously for θ. Make use of the symmetry in x and y.

Assignment E.5 —

(a) $y = \sin(2\arcsin(x))$ - graph b

(b) $y = \sin(2\arcsin(|x|))$ - graph c

(c) $y = \sin(2|\arcsin(x)|)$ - graph c

(d) $y = \arcsin(2\sin(x))$ - graph a

(e) $y = \arcsin(2\sin(|x|))$ - graph d

(f) $y = \arcsin(2|\sin(x)|)$ - graph d

Assignment E.6 — $p \ge \dfrac{1}{8}$

Assignment E.7 —

(a) There are two possible areas. Giving one of them is enough.

$$A = \int_{-1}^{0} \int_{\frac{y-2}{3}}^{\sqrt[3]{y}} dx dy \quad \text{or} \quad A = \int_{-2/3}^{0} \int_{0}^{3x+2} dy dx + \int_{0}^{2} \int_{x^3}^{3x+2} dy dx$$

(b) $V = 2x \displaystyle\int_0^{\sqrt{2}} \pi\left(-x^{*2}+5\right)^2 dx^* - 2x \int_0^{\sqrt{2}} \pi\left(x^{*2}+1\right)^2 dx^*$ where $y^* = y+4$ and $x^* = x$

(c) $L = \displaystyle\int_0^1 \sqrt{9t+4\sin^2(2t)+4\cos^2(2t)}\, dt = \int_0^1 \sqrt{9t+4}\, dt$

(d) $W = \displaystyle\int_1^2 \left(\sin(t), \sin\left(t^2\right), \sin\left(t^3\right)\right) \cdot \left(1, 2t, 3t^2\right) dt$

Assignment E.8 — $I = -\frac{1}{3} + \frac{2^{3/2}}{3}$

Assignment E.9 — $I = 2\int_0^1 \int_0^{1-z} \int_0^{\sqrt{y}} f(x,y,z)\,dx\,dy\,dz$

Assignment E.10 — The series converges for $p > 1$.

Assignment E.11 — $\int_0^1 \frac{\cos(nx)}{x+1}\,dx \leq \int_0^1 \frac{1}{x+1}\,dx \leq \ln(2)$. We can make this estimate, independent of the argument of the cosine, because the image of the cosine is limited to $[-1,1]$.

Assignment E.12 — $f(x) = 3 + x - \frac{1}{2}\frac{258}{5}x^2 + \ldots$

Assignment E.13 — The critical points are solutions of $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$. We find infinitely many critical points along the line $y = x/2$. If we write $f(x,y)$ as $f(x,y) = (x-2y)^2 + 2$, then it immediately follows that the function values in points not located on the line are higher than those that are reached for points on the line.

Assignment E.14 —

(a) $t = 4$

(b) $\vec{r}(4) = (6,0,8)$

(c) $\frac{\pi}{2} - \arccos\left(\frac{-8}{5\sqrt{36\pi^2+4}}\right)$

Assignment E.15 —

(a) True. Let a be any element of A. Since $\sup A$ is an upper bound for A, $a \leq \sup A = \inf B$. Since $\inf B$ is a lower bound for B, a will be smaller than or equal to any element of B. So there exists a $b \in B$ such that $a \leq b$.

(b) False. The graph is in fact a circle with center along $\theta = \pi/3$ and radius 2.

(c) False. A contour surface can be a surface, but also, for example, a point.

(d) False. $f_{xx}(P) < 0$ and $f_{yy}(P) > 0$.

(e) False. As an example, take the circle $x^2 + y^2 = 4$ and parameterise it in two different ways. Also calculate the tangent vector.

(f) False. The divergence is a number and not a vector. Verify that it is equal to 3.

Assignment E.16 —

(a) $x(t) = t^4 - t + 1, \quad y(t) = t^2$ - graph c

(b) $x(t) = t^2 - 2t, \quad y(t) = \sqrt{t}$ - graph a

(c) $x(t) = \sin(2t), \quad y(t) = \sin(t + \sin(2t))$ - graph b

(d) $x(t) = \cos(5t), \quad y(t) = \sin(2t)$ - graph d

Assignment E.17 —

(a) $x < 0$: the function is continuous and differentiable on the domain of the tangent.

(b) $0 \leq x < 1$: the function is continuous and differentiable.

(c) $x \geq 1$: the function is continuous and differentiable.

Assignment E.18 —

(a) $p \geq 0 : \lim\limits_{x\to 0} x^p \left(1 - \sqrt{1+x}\right) = 0$

(b) $p < 0$:

- $p < -1 : \lim\limits_{x\to 0} x^p \left(1 - \sqrt{1+x}\right) = \pm\infty$
- $p = -1 : \lim\limits_{x\to 0} x^p \left(1 - \sqrt{1+x}\right) = -\frac{1}{2}$
- $p > -1 : \lim\limits_{x\to 0} x^p \left(1 - \sqrt{1+x}\right) = 0$

Assignment E.19 — $V = 8\pi\left(\sqrt{160}\right)^2 - 2\pi \int\limits_0^{\sqrt{160}} \frac{x^3}{20}\, dx$

Assignment E.20 — Correct proof of the theorem below.

Theorem E.3 (Generalised mean value theorem)
Let f and g be two continuous functions on $[a,b]$ and differentiable on $]a,b[$. If $g'(x) \neq 0$ for every $x \in]a,b[$, then there exists a $c \in]a,b[$ such that

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

From $g'(x) \neq 0$, for every $x \in]a,b[$, it follows that $g(b) \neq g(a)$. Otherwise, there would exist a number in $]a,b[$ for which g' equals 0.
We apply Rolle's theorem to the function

$$h(x) = \big(f(b)-f(a)\big)\big(g(x)-g(a)\big) - \big(g(b)-g(a)\big)\big(f(x)-f(a)\big).$$

Since $h(a) = h(b) = 0$, there exists a $c \in]a,b[$ such that $h'(c) = 0$. We obtain

$$\big(f(b)-f(a)\big)g'(c) - \big(g(b)-g(a)\big)f'(c) = 0,$$

from which it follows that

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

Assignment E.21 —

(a) $A = 6.\dfrac{1}{2}\displaystyle\int_0^{\pi/9}\left(\left(2a\cos(3\theta)\right)^2 - a^2\right)d\theta$

(b) $SA = \displaystyle\int_0^{2\pi}\int_0^1 \sqrt{1+r^2}\,r\,dr d\theta$

(c) $\displaystyle\oint_K \vec{\boldsymbol{V}}\cdot d\vec{\boldsymbol{r}} = \int_0^5 \int_0^{\sqrt{\frac{100-4x^2}{5}}} (y-1)\,dydx$ (Green's theorem)

Assignment E.22 — The alternating series is absolutely convergent.

Assignment E.23 — We consider the function $f(x)$ which we can approximate with a Taylor series of the n-th order in the point a. We write f as

$$f(x) = T_n(x) + R_n(x),$$

where $T_n(x)$ is the Taylor polynomial of the n-th order of f in a and $R_n(x)$ is the remainder term of this sequence which can be written as follows.

Theorem E.5 (The remainder term of a Taylor series)
Let $f^{(n+1)}$ be continuous on an open interval I containing a and x, then

$$R_n(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt.$$

We prove this theorem below by means of mathematical induction.

For $n = 1$ it holds that

$$R_1(x) = f(x) - T_1(x) = f(x) - f(a) - f'(a)(x-a),$$

and the integral in the theorem becomes

$$\int_a^x (x-t)\,f''(t)\,dt.$$

We calculate this integral by using integration by parts with

$$u(t) = x - t \quad \text{en} \quad dv(t) = f''(t)\,dt,$$

from which it follows that

$$du(t) = -dt \quad \text{en} \quad v(t) = f'(t).$$

Thus, the integral becomes

$$\begin{aligned}\int_a^x (x-t)\,f''(t)\,dt &= \left[(x-t)\,f'(t)\right]_a^x + \int_a^x f'(t)\,dt \\ &= 0 - (x-a)\,f'(a) + f(x) - f(a)\end{aligned}$$

$$= \quad f(x) - f(a) - f'(a)(x-a) = R_1(x).$$

This proves the theorem for $n = 1$.

We assume that the theorem holds for $n = k$, so

$$R_k(x) = \frac{1}{k!}\int_a^x (x-t)^k f^{(k+1)}(t)\,dt.$$

We then prove the theorem for $n = k+1$, so

$$R_{k+1}(x) = \frac{1}{(k+1)!}\int_a^x (x-t)^{k+1} f^{(k+2)}(t)\,dt.$$

We again use integration by parts to calculate this last integral. We set

$$u(t) = (x-t)^{k+1} \quad \text{and} \quad dv(t) = f^{(k+2)}(t)\,dt,$$

from which it follows that

$$du(t) = -(k+1)(x-t)^k\,dt \quad \text{and} \quad v(t) = f^{(k+1)}(t).$$

The integral is thus

$$\begin{aligned}
\frac{1}{(k+1)!}\int_a^x (x-t)^{k+1} f^{(k+2)}(t)\,dt &= \frac{1}{(k+1)!}\Big[(x-t)^{k+1} f^{(k+1)}(t)\Big]_a^x + \frac{k+1}{(k+1)!}\int_a^x (x-t)^k f^{(k+1)}(t)\,dt \\
&= 0 - \frac{1}{(k+1)!}(x-a)^{k+1} f^{(k+1)}(a) + \frac{1}{k!}\int_a^x (x-t)^k f^{(k+1)}(t)\,dt \\
&= -\frac{f^{(k+1)}(a)}{(k+1)!}(x-a)^{k+1} + R_k(x) \\
&= f(x) - T_k(x) - \frac{f^{(k+1)}(a)}{(k+1)!}(x-a)^{k+1} \\
&= f(x) - T_{k+1}(x) = R_{k+1}(x).
\end{aligned}$$

The theorem is valid for $n = k+1$ if it is valid for $n = k$, therefore it is valid for all $n \in \mathbb{N}$.

Assignment E.24 —

(a) Verify that the point is on the surface. Then verify the condition $\partial F/\partial z \neq 0$ in the given point.

(b) $\left(x - \frac{1}{2}\right) + \frac{1}{4}(y-2) - (z-1) = 0$

(c) $\frac{\partial^2 z}{\partial x \partial y} = \frac{z(z+1)^2 + xyz}{(z+1)^3}$

Assignment E.25 —

(a) $I = \int\limits_0^{\pi/2} \int\limits_0^2 r^2 \, r \, dr d\theta$

(b) $I = \int\limits_0^{\pi/2} \int\limits_0^2 \int\limits_0^{r^2} r \, dz dr d\theta$

Assignment E.26 —

(a) You can consider three regions.

(b) Region 1: $x : 0 \to \frac{1}{2}, \quad y : 1 - \sqrt{2-(x-1)^2} \to 1 - \sqrt{1-2x}$

(c) Region 2: $x : 0 \to \frac{1}{2}, \quad y : 1 + \sqrt{1-2x} \to 1 + \sqrt{2-(x-1)^2}$

(d) Region 3: $x : \frac{1}{2} \to 1 + \sqrt{2}, \quad y : 1 - \sqrt{2-(x-1)^2} \to 1 + \sqrt{2-(x-1)^2}$

Assignment E.27 — $I = 2\pi$

Assignment E.28 —

(a) False. The functions are equal for $x \neq -1$. At $x = -1$ a perforation occurs.

(b) True. $f(x)$ is not defined at $x = 1$, because $\lim\limits_{x \to 1} f(x) = \pm\infty$.
OR: False. There may be a removable discontinuity. Find a counterexample yourself.

(c) True. Determine the Cartesian equation and/or direction and conclude that the product of the slopes equals -1. This means that the (half-)lines are perpendicular to each other.

(d) True. Convert the convergence interval for x to one for $(x-1)/2$.

Assignment E.29 — Do this yourself. In the proof, choose $\delta = \epsilon^2$.

Assignment E.30 —

(a) $\operatorname{dom} f =]-\infty, -\sqrt{2}[\cup]-\sqrt{2}, -1[\cup]1, \sqrt{2}[\cup]\sqrt{2}, +\infty[$

(b) f is odd.

(c) $x = \pm\sqrt{e}$

(d) There are four vertical asymptotes: $x = \pm 1$ and $x = \pm\sqrt{2}$ (boundary points of the domain). There are two horizontal asymptotes $y = 0$ for x going to positive and negative infinity.

(e) f descreases on $]1, 1.2[$ en $]1.9, 3]$ and increases on $]1.2, 1.9[$. At $x = 1.2$ a minimum is reached and at $x = 1.9$ a maximum is reached. On $]1, 1.5[$, f is convex and on $]1.5, 3[$ it is concave.

Assignment E.31 —

(a) With slices

$$V = \pi \int_{-1}^{1} \left(1+\sqrt{1-y^2}\right)^2 dy - \pi \int_{-1}^{1} \left(2-\sqrt{2-y^2}\right)^2 dy = \pi \int_{-1}^{1} \left(-4+2\sqrt{1-y^2}+4\sqrt{2-y^2}\right) dy$$

(b) With cylindrical shells

$$V = 4\pi \int_{0}^{1} \sqrt{2x-x^2}(2-x)\,dx - 4\pi \int_{1}^{\sqrt{2}} \sqrt{2-x^2}(2-x)\,dx$$

Assignment E.32 — We are in the point $\vec{r}\left(\dfrac{\pi\sqrt{10}}{3}\right) = \left(\dfrac{\pi}{3}, \dfrac{3\sqrt{3}}{2}, \dfrac{3}{2}\right)$.

Assignment E.33 — $p = 0 \quad \vee \quad p = \dfrac{-2n+4}{4} \ (n > 0)$

Assignment E.34 —

(a) $M = \displaystyle\int\int_R \delta\big(x, y, f(x,y)\big)\sqrt{1+f_x^2+f_y^2}\,dx\,dy$

(b) $M = \dfrac{1023\pi\sqrt{2}}{5}$

Assignment E.35 —

(a) $f(x,y) \in]0, e]$, so f is bounded.

(b) $\vec{\nabla} f\,(3\pi/2, 1) = e^{-1}(1, -2)$

(c) $||\vec{\nabla} f\,(3\pi/2, 1)|| = e^{-1}\sqrt{5}$

(d) $F(x,y) = \ln|y| - \ln\left|\dfrac{\cos(x)-1}{\cos(x)+1}\right|$

(e) Yes, because $\dfrac{\partial F}{\partial y} = \dfrac{1}{|y|} \overset{(3\pi/2,1)}{=} 1 \neq 0$

(f) $y = \dfrac{\cos(x)-1}{\cos(x)+1}$

Assignment E.36 —

(a) $\dfrac{\partial f}{\partial y} < 0$

(b) $\dfrac{\partial^2 f}{\partial x^2} < 0$

(c) strictly positive

(d) Figure E.6(c),

Assignment E.37 — Prove this yourself.

Assignment E.38 —

(a) $\operatorname{dom} f =]0, \pi[+ k\pi, k \in \mathbb{Z}$

(b) The function is periodic with period 2π.

(c) The function is odd.

(d) $f(x) = 0 \quad \Leftrightarrow \quad x = \dfrac{\pi}{4} + k\pi, k \in \mathbb{Z} \quad \vee \quad x = \dfrac{3\pi}{4} + k\pi, k \in \mathbb{Z}$

(e) There are only vertical asymptotes at $x = k\pi, k \in \mathbb{Z}$.

(f) The extrema are found at $x = \dfrac{\pi}{2} + k\pi, k \in \mathbb{Z}$. There are no inflection points.

(g) Onderzoek zelf het gedrag van f en maak een schets.

Assignment E.39 — With cylindrical shells:

$$V = 2\pi \int_0^3 x^2\, dx + 2\pi \int_3^9 9\, dx + 2\pi \int_9^{10} x(10 - x)\, dx = \frac{406\pi}{3}$$

Assignment E.40 — The convergence interval is $[1, 2[$ and the radius of convergence is $1/2$.

Assignment E.41 — Prove this yourself.

Assignment E.42 —

(a) $A_1 = 2 \displaystyle\int_0^{\pi/4} \int_{1/2\sin(2\theta)}^{1/2} r\, dr d\theta$

(b) $A_2 = 2 \displaystyle\int_{\pi/8}^{\pi/4} \int_{\sqrt{2}/4}^{1/2\sin(2\theta)} r\, dr d\theta$

(c) $A_2 = \dfrac{1}{32}$

Assignment E.43 — $\displaystyle\oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = f(B) - f(A) = \frac{9\pi}{4}$

Assignment E.44 — You can check the impact of β_i on the inflection point by using differentials. To do this, you must of course first find out the time instant at which the inflection point manifests itself. So you first calculate the second derivative. Then you determine for which t_{BP} there is an inflection point. Finally, check the impact of β_i on the inflection point by differentiating t_{BP} with respect to β_0 and β_1.

Assignment E.45 — Show this yourself.

Assignment E.46 —

(a) $\operatorname{dom} f = \left[-\sqrt{e-1}, \sqrt{e-1}\right], \quad \operatorname{im} f = \left[0, \dfrac{\pi}{2}\right]$

(b) The function is even.

(c) $f'(x) = \dfrac{-2x}{\left(x^2+1\right)\sqrt{1-\left(\ln(x^2+1)\right)^2}}$

(d) $f'\left(\pm\sqrt{e-1}\right) \to \mp\infty$. There is a vertical tangent to the graph of the function at the boundary points.

(e) The domain of the function is a closed interval. f is also continuous on this interval. It follows from the extreme value theorem that f reaches a maximum on its domain.

Assignment E.47 — $I = x\ln\left(x^2+2x+2\right) - 2(x+1) + 2\arctan(x+1) + \ln\left|(x+1)^2+1\right| + C$

Assignment E.48 —

(a) $A = \dfrac{1}{2}\displaystyle\int_{-\pi/2}^{\pi/6} \left(1+\sin(\theta)\right)^2\, d\theta - \dfrac{1}{2}\int_{0}^{\pi/6} \left(3\sin(\theta)\right)^2\, d\theta$

(b) $V = \pi\displaystyle\int_{\theta_1}^{\theta_2} r^2(\theta)\sin^2(\theta)\left(r'(\theta)\cos(\theta) - r(\theta)\sin(\theta)\right) d\theta$

(c) $V = \pi\displaystyle\int_{0}^{\pi/4} 9(\theta)\sin^4(\theta)\left(3\cos^2(\theta) - 3\sin^2(\theta)\right) d\theta$

Assignment E.49 —

(a) The insects will collide at $t = 3$.

(b) $\vec{\boldsymbol{r}}_1(3) = \vec{\boldsymbol{r}}_2(3) = (9, 9, 9)$

(c) The angle between the paths at the point where the collision occurs is $\arccos\left(\dfrac{21}{\sqrt{19}\sqrt{61}}\right)$.

Assignment E.50 —

(a) Show this yourself.

(b) $\vec{\nabla} p = \left(\dfrac{kT}{V}, \dfrac{kN}{V}, -\dfrac{kNT}{V^2}\right)$

(c) $D_{\vec{\boldsymbol{u}}}\, p = \pm\dfrac{k(T+N)}{\sqrt{2}V}$

(d) $N + T - V = 1$

Assignment E.51 —

(a) The function f is continuous for all $x \neq 0$. f is right continuous in $x = 0$.

(b) There is a jump discontinuity at $x = 0$.

(c) $f'(x) = \dfrac{e^x|x| - (e^x - 1)(\pm 1)}{x^2} = \begin{cases} \dfrac{e^x}{x} - \dfrac{e^x}{x^2} + \dfrac{1}{x^2}, & x > 0, \\ -\dfrac{e^x}{x} + \dfrac{e^x}{x^2} - \dfrac{1}{x^2}, & x < 0. \end{cases}$

f' is strictly positive for $x > 0$.

(d) (i) Since $f' > 0$ if $x > 0$, f is strictly increasing if $x > 0$ and thus is injective for these x values. Therefore the inverse function f^{-1} exists on $]0, +\infty[$.

(ii) $\left(f^{-1}\right)'(e-1) = 1$.

Assignment E.52 — Prove for yourself.

Assignment E.53 — Via shells: $V = 2\displaystyle\int_0^1 2\pi(2-x)\sqrt{1-(x-1)^2}\,dx + 2\int_1^{\sqrt{2}} 2\pi(2-x)\sqrt{2-x^2}\,dx = 3\pi^2 - 4\pi$

Assignment E.55 — Assume that eq. (E.7) holds true and choose $\vec{\mathbf{x}} \in \mathbb{R}^n$ at random, but fixed. We define the following function:

$$g(t) = f(t\vec{\mathbf{x}}) - t^p f(\vec{\mathbf{x}}),$$

for all $t > 0$. The theorem is proven if we can show that $g(t) = 0$ for all $t > 0$.

Apparently $g(1) = 0$. Derivation to t yields, again using the chain rule:

$$g'(t) = \nabla f(t\vec{\mathbf{x}}) \cdot \vec{\mathbf{x}} - pt^{p-1} f(\vec{\mathbf{x}}), \quad \forall t > 0.$$

As $t\vec{\mathbf{x}} \in \mathbb{R}^n$, $\forall t > 0$, applies to understated

$$(t\vec{\mathbf{x}}) \cdot \nabla f(t\vec{\mathbf{x}}) = pf(t\vec{\mathbf{x}}),$$

such that $g'(t)$ can be rewritten as

$$g'(t) = \frac{1}{t}\left[\nabla f(t\vec{\mathbf{x}}) \cdot (t\vec{\mathbf{x}}) - pt^p f(\vec{\mathbf{x}})\right] = \frac{p}{t}\left[f(t\vec{\mathbf{x}}) - t^p f(\vec{\mathbf{x}})\right] = \frac{p}{t} g(t), \quad \forall t > 0.$$

Next, we define the function h as follows:

$$g(t) = t^p h(t), \quad \forall t > 0.$$

Deriving the above relationship gives

$$g'(t) = t^p h'(t) + pt^{p-1} h(t),$$

from which follows by equation

$$\frac{p}{t} g(t) = t^p h'(t) + \frac{p}{t} g(t), \quad \forall t > 0,$$

Such that we can conclude that $h'(t) = 0$, $\forall t > 0$. Consequently, the function h is constant and therefor there exists a number $c \in \mathbb{R}$ such that $g(t) = ct^p$, for all positive t. However, since we already established that $g(1) = 0$, it must be that $c = 0$, so that $g(t) = 0$, $\forall t > 0$.

Assignment E.56 — $V = 2\int_1^2 \int_0^{2\pi} r^2 \, d\theta dr = \dfrac{28\pi}{3}$

Assignment E.57 —

(a) $I_1 = \int_0^{\pi/2} \int_0^{\pi/4} \int_0^{\sqrt{18}} \rho^4 \sin(\phi) \, d\rho d\phi d\theta$

(b) $I_2 = -\int_0^1 \int_0^{2x} \left(2xy^3 - x\right) dydx = -\int_0^2 \int_{y/2}^1 \left(2xy^3 - x\right) dxdy$

F

List with definitions

G

List with theorems

Index