LECTURE NOTES

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Calculus

Abbas Edalat

Version: September 30, 2025

Contents

Preface

These lecture notes are intended for first year undergraduate students of computer science at Imperial College London and cover basic mathematical concepts in calculus that are required in other courses. The lecture notes were originally based on Jeremy Bradley's material from 2014/15, but the course has been restructured since then.

Many people have helped improving these lecture notes by providing comments, suggestions and corrections. In particular, thanks to Romain Barnoud, Mahdi Cheraghchi, Ruhi Choudhury, Marc Deisenroth, Tony Field and Michael Huth for valuable feedback and additions. I am also very grateful to Paul Bilokon who produced the notes in the present format while I was away on sabbatical.

Note: The first four sections of these notes present some historical background and an introduction to several mathematical concepts which may already be known to most of the students. These sections are for preliminary reading and are covered in other courses such as Discrete Mathematics.

The Calculus lecture notes will start properly in **section 5** on sequences. The material presented in these notes is an expanded version of what will be covered in the lectures. Any subsection with a title starting with **Additional Material** would not be examinable.

Abbas Edalat
30 September 2025

1 Course Overview

This introductory course covers background that is essential to many other courses in the Department of Computing. We will cover topics, including

- sequences
- continuous functions
- integration
- series
- differentiation
- power series
- numerical methods
- metric spaces
- vector derivatives
- applications to neural networks

There are many courses in the department that require an understanding of mathematical concepts. These courses range from optimization, operations research and complexity to machine learning, robotics, graphics and computer vision.
The department offers additional courses that cover other and more advanced mathematical concepts: Linear Algebra (1st year), Probability and Statistics (2nd year), Computational Techniques (2nd year), Mathematics for Inference and Machine Learning (4th year).

2 Why Study Calculus in the 21st Century?

2.1 Introduction

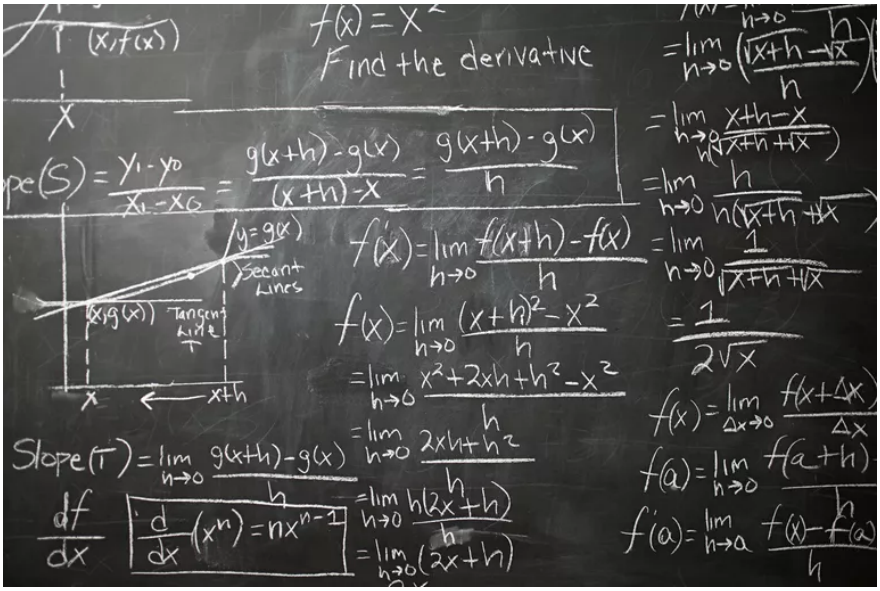


Figure 2.1: Calculus on blackboard.

How do you feel about calculus in the 21st century? Perhaps, you succumb to the mathematical anxiety, "a feeling of tension, apprehension, or fear that interferes with math performance" [Ash02]. Some of you may have felt that you have already studied too much calculus at school, whereas others will feel that you have studied too little. Perhaps the feeling is that of hesitation or doubt: why would we need calculus in the Digital Age? Why should a computer scientist, engineer, or programmer study calculus?
After all, you may have heard Gilbert Strang, the author of [Str17], say in his essay *Too Much Calculus* [Str01]

> All the rest of mathematics is overwhelmed by calculus. The next course might be differential equations (more derivatives), and the previous course is probably pre-calculus. I really think it is our job to adjust this balance, we cannot expect others to do it. We know the central role of linear algebra. It is much more than a random math course, its applications touch many more students than calculus. **We are in a digital world now.**

Let us talk about **calculus** and why it is more important than ever in the 21st century.

Ancient accountants laid pebbles in columns on a sand tray to help them do their sums. It is thought that the impression left in the sand when a pebble is moved to another location is the origin of our symbol for zero.

Figure 2.2: Pebbles.

The word "calculus" has the same source, since it means a "pebble" in Latin. Nowadays it means any systematic way of working out something mathematical. We still speak of a **calculator** when referring to the modern electronic equivalent of an ancient sandtray and pebbles. However, since the English mathematician Isaac Newton (1642–1726/27) invented the differential and integral calculus, the word is seldom applied to anything else. This particular calculus, the one we are about to study, is all about differentiating and integrating.
Have you seen the old Apple logo? Before it became the millennial minimalist logo of today, it was far more elaborate. Designed by Ronald Wayne in 1976, the first Apple logo pictured Isaac Newton, sitting under an apple tree, and an apple dangling precipitously above his head. The phrase on the outside border read,

> Newton... A Mind Forever Voyaging Through Strange Seas of Thought... Alone.

It was William Stukeley, an archaeologist and one of Newton's first biographers, that relayed [Stu52] the story thus told by Newton:

> After dinner, the weather being warm, we went into the garden & drank thea under the shade of some appletrees, only he, & myself. Amidst other discourse, he told me, he was just in the same situation, as when formerly, the notion... came into his mind. "Why should that apple always descend perpendicularly to the ground," thought he to him self: occasion'd by the fall of an apple, as he sat in a contemplative mood: "Why should it not go sideways, or upwards? but constantly to the earths centre?" ... ...

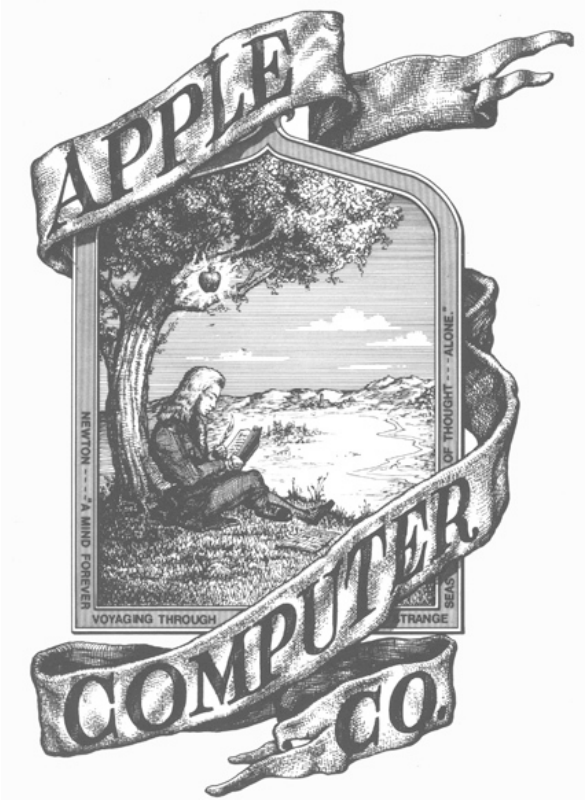


Figure 2.3: The first Apple logo.

On 8th July, 1958, The New York Times published the following article [new58]:

> **NEW NAVY DEVICE LEARNS BY DOING**
>
> **Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser**
>
> WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.
>
> The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.
>
> The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.
>
> Dr. Frank Rosenblatt, designer of the Perceptron conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.
>
> Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers... ... ...

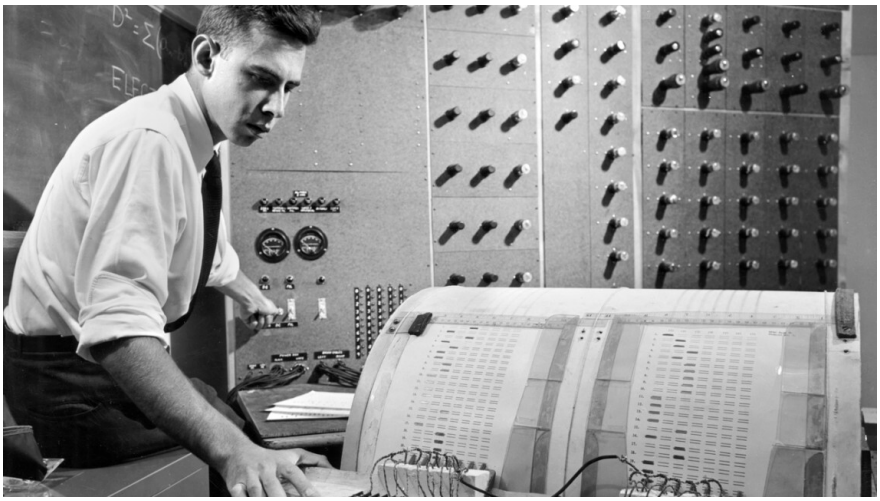


Figure 2.4: Frank Rosenblatt and Mark I Perceptron.

There was a commotion at the Four Seasons Hotel in Seoul on 9th March, 2016. A motley crowd of computer scientists, mathematicians, Go grandmasters, correspondents, and photographers had descended on the hotel. Seoul, out of all places, was no stranger to such diversity, where modern glass-and-steel structures coexisted with grand, historic palaces within a ring of verdant mountains. And yet the event had an air of uniqueness and timelessness about it. Something special was about to happen.

Figure 2.5: The Four Seasons Hotel in Seoul.

Everyone was waiting for one man, Lee Sedol, although some would call him "Ssendol", "The Strong Stone." He had just turned thirty-three years old. Just over two decades ago he was the fifth youngest (12 years 4 months) to become a professional Go player in South Korean history. Today he had an added responsibility that very few would be brave enough to bear.

Go was invented 2,500 years earlier in China. It is played by two players on a 19×19 grid of black lines. Game pieces—**stones**—are placed on the lines' intersections—**points**. One player uses the white stones and the other, black. The players take turns placing the stones on the vacant points. Once placed on the board, stones may not be moved, although they are removed from the board if one or a group of stones are surrounded by opposing stones, in which case the stones are said to

be **captured**. When a game concludes, the winner is determined by counting each player's surrounded territory along with captured stones and **komi**, points added to the score of the player with the white stones as compensation for playing second.
The event's uniqueness was not without two exceptions. Something similar had happened in Philadelphia in 1996 and in New York City in 1997 when Deep Blue (an IBM chess computer) played Garry Kasparov (then the world chess champion). The 1996 game was won by Kasparov; the 1997, by Deep Blue, being the first defeat of a reigning world chess champion by a computer under tournament conditions.
Today it is Sedol's turn to defend humanity's intellectual dignity.

Figure 2.6: AlphaGo versus Lee Sedol.

Compared to chess, Go has both a larger board with more scope for play and longer games and, on average, many more alternatives to consider per move. The number of legal board positions in Go has been calculated to be approximately 2.1×10^{170}, which is vastly greater than the numbers of atoms in the known, observable universe, estimated to be about 1×10^{80}. Despite its relatively simple rules, Go is extremely complex.
Mathematician I. J. Good wrote in 1965 [Goo65]:

> Go on a computer?—In order to programme a computer to play a reasonable game of Go, rather than merely a legal game—it is necessary to formalise the principles of good strategy, or to design a learning programme. The principles are more qualitative and mysterious than in chess, and depend more on judgment. So I think it will be even more difficult to programme a computer to play a reasonable game of Go than of chess.

Sedol arrives. He has already won 18 world championships. He will defeat AlphaGo (a computer Go program developed by Google DeepMind) in a landslide, he

predicts. Some weeks before the match he won the Korean Myungin title, a major championship. There is every reason to be confident.
The match starts at 13:00 KST. Sedol is in control. This is likely to be an easy game for him.
Then comes the game's 102nd stone. Alpha Go makes its move. Sedol examines the board. He mulls over his options. A minute passes by, two minutes, ten, until he finally responds. His confidence begins to evaporate... ... ...

2.2 The Dawn of the Machines

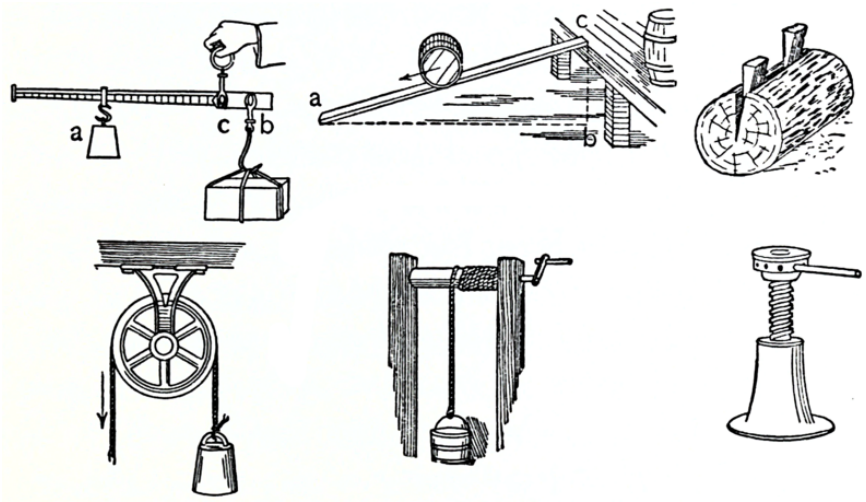


Figure 2.7: Chart of Simple Machines. John Mills. *The Realities of Modern Science*, p. 15, Fig. 3. 31st May, 1919.

The six classical **simple machines** were enumerated by Renaissance scientists:

- The **wedge**, which can be used to separate two objects. Wedges were found among the oldest known Oldowan tools in Gona, Ethiopia, and are dated to about 2.6 million years ago.

- The **inclined plane**, also known as a **ramp**, used as an aid for raising or lowering a load. The heavy stones in Stonehenge (3000 to 2000 BCE) are believed to have been moved and set in place using inclined planes made of earth. The ancient Greeks constructed a paved ramp 6 km long, the Diolkos, to drag ships overland across the Isthmus of Corinth.

- The **lever**, consisting of a beam or rigid rod pivoted at a fixed hinge, or **fulcrum**. This device amplifies an input force to provide a greater output force, which is said to provide **leverage**. The earliest evidence of the lever mechanism dates back to the ancient Near East circa 5000 BCE, when it was first used in a simple balance scale.

- The **wheel and axle**—a wheel attached to a smaller axle so that these two parts rotate together, transferring a force from one to the other. One of the first applications was the potter's wheel, used by prehistoric cultures to fabricate clay pots. The earliest type, known as "tournettes" or "slow wheels," were known in the Middle East by the 5th millenium BCE.

- The **pulley** used for the transfer of power between the shaft and cable or belt. The earliest evidence of pulleys dates back to Ancient Egypt in the Twelfth Dynasty (1991–1802 BCE) and Mesopotamia in the early 2nd millennium BCE.

- The **screw**—a mechanism that converts a torque (rotational force) to a linear force. It was the last of the simple machines to be invented. It first appeared in Mesopotamia during the Neo-Assyrian period (911–609 BCE).

Archimedes' famous remark with regard to the lever: "Give me a place to stand on, and I will move the Earth," expresses his realization that there was no limit to the amount of force amplification that could be achieved by using mechanical advantage.

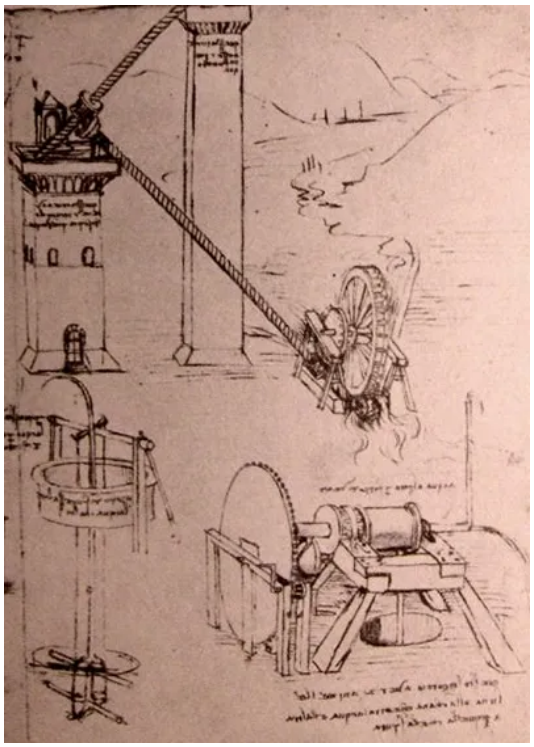
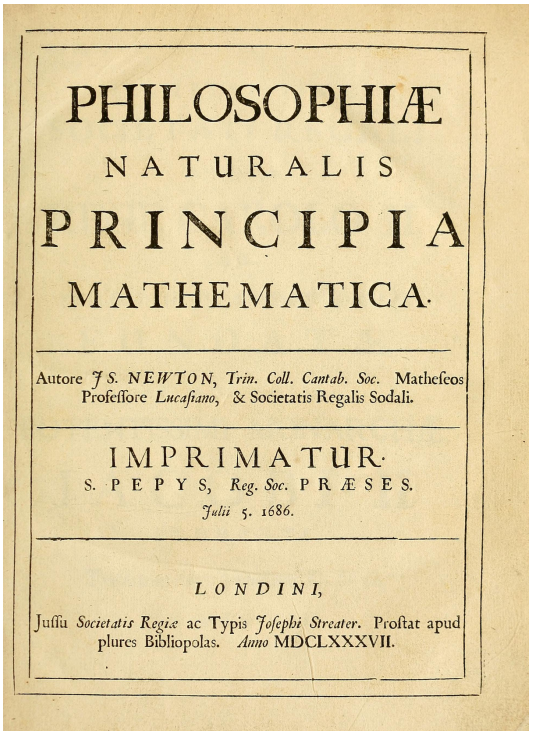


Figure 2.8: Various hydraulic machines by Leonardo da Vinci.

The Italian Leonardo Da Vinci (1452–1519), better known as the painter of Mona Lisa, was also an engineer fascinated with machinery. He invented the strut bridge, the automated bobbin winder, the rolling mill, the machine for testing the tensile strength of wire, and the lens-grinding machine. He preferred to keep some of his inventions a secret:

> How by means of a certain machine many people may stay some time under water. How and why I do not describe my method of remaining under water, or how long I can stay without eating; and I do not publish nor divulge these by reason of the evil nature of men who would use them as means of destruction at the bottom of the sea, by sending ships to the bottom, and sinking them together with the men in them. And although I will impart others, there is no danger in them; because the mouth of the tube, by which you breathe, is above the water supported on bags of corks.

2.3 *Principia*


PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.

Autore *J S. NEWTON*, *Trin. Coll. Cantab. Soc.* Matheſeos Profeſſore *Lucaſiano*, & Societatis Regalis Sodali.

IMPRIMATUR.
S. PEPYS, *Reg. Soc.* PRÆSES.
Julii 5. 1686.

LONDINI,
Juſſu *Societatis Regiæ* ac Typis *Joſephi Streater*. Proſtat apud plures Bibliopolas. *Anno* MDCLXXXVII.

Figure 2.9: Isaac Newton's *Philosophiæ Naturalis Principia Mathematica*.

The early work on the machines was mostly **empirical**: it was based on observation or experience rather than theory or pure logic. But this was about to change. On 5th July, 1687, the humanity's understanding of machinery was radically transformed. The English mathematician Isaac Newton (1642–1726/27) published his revolutionary work, *Philosophiæ Naturalis Principia Mathematica*. *Principia* established the laws of motion, which were henceforth to be known **Newton's laws of motion**. In the words of Newton [New94],

> A Vulgar Mechanick can practice what he has been taught or seen done, but if he is in an error he knows not how to find it out and correct it, and if you put him out of his road, he is at a stand; Whereas he that is able to

> reason nimbly and judiciously about figure, force, and motion, is never at rest till he gets over every rub.

Newton's work was an upgrade on Aristotle's analytical and Galileo's experimental methods. It introduced to the study of motion (and of machines) the apparatus of **infinitesimal calculus**—the mathematical study of continuous change. Calculus aims to study functions, processes that associate each element of a set A to a single element of a set B. For example,

$$f(t) = t^2$$

is a function. It associates with each number t another number, t^2. Thus 0 is associated with 0, 1 with 1, 2 with 4, and 3 with 9. We can plot this function—the x-coordinate can be used to represent t, the y-coordinate, t^2. The result will look as in Figure 5.1.

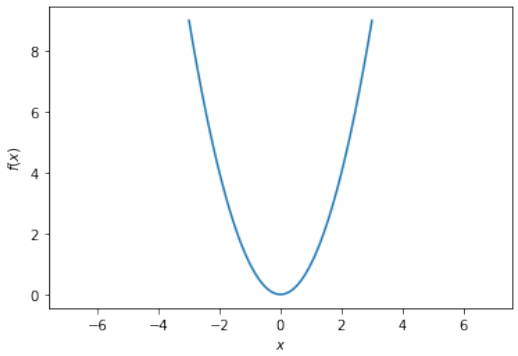


Figure 2.10: A parabola.

This shape is known as a **parabola**. A ball flying through the air will retrace an upturned parabola—it will fall to the ground under the influence of the earth's gravitational field. The trajectory of the ball can be represented by a function: $f(t)$ is the height of the ball above the ground, t is the time since its launch.
Calculus studies the rate of change of f with respect to t. In line with Aristotle's (384–322 BCE) analytical tradition, it breaks a complex topic or substance into small parts in order to gain a better understanding of it. We consider a small change in t, Δt, and reason about what happens to f:

$$\frac{\Delta f}{\Delta t} = \frac{f(t+\Delta t) - f(t)}{\Delta t} = \frac{(t+\Delta t)^2 - t^2}{\Delta t} = \frac{t^2 + 2t\Delta t + \Delta t^2 - t^2}{\Delta t} = 2t + \Delta t.$$

As Δt becomes smaller and smaller (and we approach a "limit") it vanishes and the rate of change becomes the **derivative**

$$\frac{df}{dt}(t) = 2t.$$

The rate of change of f is proportional to t, the constant of proportionality being equal to 2. This rate of change emerges as the **gradient**—the slope—of our parabolical graph.
In the words of George Smith, an expert in the philosophy of science and logic, [Smi07]

> No one could deny that a science had emerged that, at least in certain respects, so far exceeded anything that had ever gone before that it stood alone as the ultimate exemplar of science generally. The challenge to philosophers then became one of spelling out first the precise nature and limits of the knowledge attained in this science and then how, methodologically, this extraordinary advance had been achieved, with a view to enabling other areas of inquiry to follow suit.

Calculus was developed by Gottfried Wilhelm von Leibniz (1646–1716) independently of Newton. The question of which of them had invented calculus first led to the so-called **calculus controversy** or "priority dispute". The notation that we have used here for $\frac{df}{dt}$ is Leibniz's, not Newton's.

2.4 The Industrial Revolution

Figure 2.11: Wentworth Works, File and Steel Manufacturers and Exporters of Iron in Sheffield, England, ca. 1860.

Within the next century the improved understanding of the laws governing motion and, by implication, machines, led to a qualitative change—the Industrial Revolution. It began in Great Britain, which became the world's leading commercial nation by the mid-18th century. The first practical fuel-burning engine was invented in 1712 by Thomas Newcomen (1664–1729). Newcomen's steam engine was improved upon by James Watt (1736–1819) in 1776. The Watt steam engine became the workhorse of the Industrial Revolution. John Wilkinson's (1728–1808) precision boring machine was used to produce cast iron piston cylinders for Watt's invention. The British landscape was being transformed by the machines and the emerging industry. Large factories were springing up all over metropolitan areas as manufacturing capacity exploded, aided by steam engines and copious supplies of coal. In

an 1838 speech in the House of Commons, Benjamin Disraeli (1804–1881) referred to Britain as "the workshop of the world."

Figure 2.12: Exhibition Road.

As you know, the South Kensington campus of Imperial College London is sandwiched between two streets: Queen's Gate in the west and Exhibition Road in the east. Do you know whence its name is derived?

Figure 2.13: The Great Exhibition.

On 1st May, 1851 over half a million people massed in Hyde Park in London to witness the opening of the Great Exhibition of the Works of Industry of All Nations. The

Great Exhibition was organized by the civil servant Henry Cole (1808–1882) and by Prince Albert (1819–1861), husband of the reigning monarch of the United Kingdom, Queen Victoria (1819–1901). Charles Darwin, Karl Marx, Michael Faraday, Samuel Colt, the Orléanist Royal Family, the writers Charlotte Brontë, Charles Dickens, Lewis Carroll, George Eliot, Alfred Tennyson, and William Makepeace Thackeray were all among the guests.
The exhibition was to become the biggest display of objects of industry from all over the world with over half of it given over to all that Britain manufactured. It was a showcase for a hundred thousand objects, of inventions, machines, and creative works; a combination of visual wonder, competition (prizes were awarded among manufacturers), and shopping. The main exhibition hall was a giant glass structure, with over a million square feet of glass. The man who designed it, Joseph Paxton, named it the Crystal Palace. In itself it was a wondrous thing to behold and covered nearly 20 acres, easily accommodating the giant elm trees that grew in the park.
The machines were in the limelight among the exhibits. Frederick Bakewell demonstrated a precursor to today's fax machine. The American Matthew Brady was awarded a medal for his daguerreotypes. William Chamberlin, Jr. of Sussex exhibited what may have been the world's first voting machine. The first modern pay toilets were installed, with 827,280 visitors paying the one-penny fee to use them. Samuel Colt demonstrated his prototype for the 1851 Colt Navy. "The Trophy Telescope," so called because it was considered the "trophy" of the exhibition, was shown. Its main lens of 280 mm aperture and 4.9 m focal length was manufactured by Ross of London. The instrument maker J. S. Marratt exhibited a five-foot achromatic telescope and a transit theodolite used in surveying, tunnelling, and for astronomical purposes.
Mieke Molthof [Mol11] explains:

> The Newtonian cosmology of a mechanical universe promoted a drive to mechanize human activities at large scale, a drive which underlaid the development of new machinery and the capacity to industrialize.

Cliff T. Bekar and Richard G. Lipsey [BL04] put it as follows:

> Indeed, it does not seem an overstatement to say that Newtonian mechanics provided the intellectual basis for the First Industrial Revolution, which in its two stages, was almost wholly mechanical.

Margaret C. Jacob [Jac97] describes how

> Brought together by a shared technical vocabulary of Newtonian origin, engineers, and entrepreneurs—like Boulton and Watt—negotiated, in some instances battled their way through the mechanization of workshops or the improvement of canals, mines, and harbours... By 1750 British engineers could talk the same mechanical talk. They could objectify the physical world, see its operations mechanically and factor their common interests and values into their partnerships. What they said and did changed the Western world for ever.

2.5 Electrification

Figure 2.14: *American Progress* (1872) by John Gast is an allegorical representation of the modernization of the new west. Columbia, a personification of the United States, is shown leading civilization westward with the American settlers. She is laying a telegraph wire with one hand and carries a school book in the other.

At around the same time another revolution was beginning to take place—**electrification**. In 1831–1832, Michael Faraday (1791–1867, one of the judges at the Great Exhibition) discovered the operating principle of electromagnetic generators. The inventions of Hippolyte Pixii (1808–1835), André-Marie Ampère (1775–1836), William Fothergill Cooke (1806–1879), Charles Wheatstone (1802–1875), Zénobe Gramme (1826–1901), R. E. B. Crompton (1845–1940), Humphry Davy (1778–1829), William Petrie (1821–1908), William Edwards Staite (1809–1854), and Pavel Yablochkov (1847–1894) were heralding the new era of machines powered by electricity, the era of Thomas Edison (1847–1931) and Nikola Tesla (1856–1943). In the United States, the new era was becoming an embodiment of the manifest destiny.

2.6 The Digital Revolution

The latter half of the 20th century saw rapid adoption and proliferation of digital computers and digital record-keeping, a trend that continues to the present day. This trend is known as the **Digital Revolution**. It constitutes a shift from mechanical and analogue electronic technology to digital electronics. The figure below shows some important dates in Digital Revolution from 1968 to 2017 as rings of time on a tree.

The language of the Digital Revolution is that of linear algebra. This has prompted Gilbert Strang to write the essay *Too Much Calculus*:

> We know the central role of linear algebra. It is much more than a random math course, its applications touch many more students than calculus. We are in a digital world now.

2.7 Deep Learning Revolution

And yet, we could argue that science and technology are undergoing a different revolution in this day and age—the **Deep Learning Revolution**. Neural networks introduced in the 1960s

2.8 Attitude for Success

How did Newton arrive at his amazing discoveries? When asked, he replied "by always thinking unto them." Elsewhere he explains, "I keep the subject constantly before me, and wait till the first dawnings open slowly, little by little, into a full and clear light." In a letter to Dr. Bentley he says, "If I have done the public any service in this way, it is due to nothing but industry and patient thought." Towards the close of his life, he uttered this memorable sentiment, "I know not what the world will think of my labours, but to myself, it seems that I have been but as a child playing on the sea-shore, now finding some pebble rather more polished, and now some shell rather more agreeably variegated than another, while the immense ocean of truth extended itself unexplored before me."

Part I

Preliminaries

3 Sets

3.1 Introduction

We begin our journey by introducing the language of **set theory**. This theory, originally developed towards the end of the 19th century, has by now become an extensive subject in its own right. More important, however, is the great influence which set theory has exerted and continues to exert on mathematical thought as a whole.

3.2 Basic definitions

Definition 1
A **set** *is simply a collection of objects, which are called its* **members** *or* **elements**.

There are several ways of describing a set. Sometimes the most convenient way is to make a list of all the objects in the set and put curly brackets around the list.

Example 1
Here are some examples of sets:

- $\{1,3,5\}$ *is the set consisting of objects 1, 3, and 5.*
- $\{John, cat, 3.57\}$ *is the set consisting of objects John, cat, and 3.57.*
- $\{1,\{2\}\}$ *is the set consisting of two objects, one being the number 1 and the other the set* $\{2\}$.

Often this isn't a convenient way to describe a set. For example, the set consisting of all people who live in England is for most purposes best described by precisely this phrase (i.e., "the set of people who live in England"); it is unlikely to be useful to describe this set in list form $\{\text{Michael}, \text{Anne}, \text{Abbas}, \ldots\}$. As another example, the set of all real numbers whose square is less than 2 is neatly described by the notation

$$\{x \,|\, x \text{ a real number}, x^2 < 2\}.$$

(This is read: "the set of all x such that x is a real number and $x^2 < 2$." The symbol "|" is the "such that" part of the phrase.)
Likewise,

$$\{x \,|\, x \text{ a real number}, x^2 - 2x + 1 = 0\}$$

denotes the set consisting of all real numbers x such that $x^2 - 2x + 1 = 0$.
As a convention, we define the **empty set** to be the set consisting of no objects at all, and denote the empty set by the symbol $\emptyset$.

Definition 2
If S is a set, and s is an element of S (i.e., an object that belongs to S), we write

$$s \in S$$

and say s **belongs to** *S. If some other object t does not belong to S, we write $t \notin S$.*

Example 2
For example,

- $1 \in \{1,3,5\}$ *but* $2 \notin \{1,3,5\}$.
- *If* $S = \{x \mid x \text{ a real number}, 0 \leq x \leq 1\}$, *then* $1 \in S$ *but Fred* $\notin S$.
- $\{2\} \in \{1,\{2\}\}$ *but* $2 \notin \{1,\{2\}\}$.
- $1 \notin \emptyset$.

Definition 3
Two sets are defined to be **equal** *when they consist of exactly the same elements.*

Example 3
Here are some examples:

- $\{1,3,5\} = \{3,5,1\} = \{1,5,1,3\}$.
- $\{x \mid x \text{ a real number}, x^2 - 2x + 1 = 0\} = \{1\}$.
- $\{x \mid x \text{ a real number}, x^2 + 1 = 0\} =$ *the set of green swans* $= \emptyset$.

Definition 4
We say a set T is a **subset** *of a set S if every element of T also belongs to S (i.e., T consists of some of the elements of S). We write $T \subseteq S$ if T is a subset of S and $T \nsubseteq S$ if not.*

Example 4
For example,

- *If* $S = \{1,\{2\},cat\}$, *then*

$$\{cat\} \subseteq S, \{\{2\}\} \subseteq S, \{2\} \nsubseteq S.$$

- *The subsets of* $\{1,2\}$ *are*

$$\{1,2\},\{1\},\{2\},\emptyset.$$

(By convention, $\emptyset$ is a subset of every set.)
Note that $S = T$ iff[1] S and T iff $S \subseteq T$ and $T \subseteq S$, i.e. iff every element of S is an element of T and every element of T is an element of S. If $S \subseteq T$ but $S \neq B$, we call S a **proper subset** of T.

[1]The abbreviation "iff" stands for "if and only if". It first appeared in print in [Kel55], whereas the invention of this abbreviation is attributed to the Hungarian-born American mathematician and statistician Paul Halmos (1916–2006).

3.3 Russell's paradox

The set theory that we have discovered so far is deceptively simple. Amidst this seeming simplicity lurks an important paradox[2]:

Paradox 1 (Russell's paradox)
Most sets commonly encountered are not members of themselves. For example, consider the set of all squares in the plane. This set is not itself a square in the plane, thus it is not a member of itself. Let us call a set "normal" if it is not a member of itself, and "abnormal" if it is a member of itself. Clearly every set must be either normal or abnormal. The set of squares in the plane is normal. In contrast, the complementary set that contains everything which is **not** *a square in the plane is itself not a square in the plane, and so it is one of its own members and is therefore abnormal.*
Now consider the set of all normal sets R, and try to determine whether R is normal or abnormal. If R were normal, it would be contained in the set of all normal sets (itself), and therefore be abnormal; on the other hand if R were abnormal, it would be contained in the set of all normal sets (itself), and therefore be normal. This leads to the conclusion that R is neither normal nor abnormal: **Russell's paradox**.

Because the set theory that we have considered so far admits paradoxes, such as the one above, it is sometimes called the **naïve set theory**. However, it usually suffices for the everyday use in contemporary mathematics.

3.4 Operations on sets

Definition 5
Let S and T be any two sets. By the **sum** *or* **union** *of S and T, denoted by $S \cup T$, is meant the set consisting of all elements which belong to at least one of the sets S and T (Figure 3.1). More generally, by the sum or union of an* **arbitrary** *number (finite or infinite) of sets S_α (indexed by some parameter α), we mean the set, denoted by $\bigcup_\alpha S_\alpha$, of all elements belonging to at least one of the sets S_α.*

Definition 6
By the **intersection** *$S \cap T$ of two given sets S and T, we mean the set consisting of all elements which belong to both S and T (Figure 3.2). For example, the intersection of the set of all even numbers and the set of all integers divisible by 3 is the set of all integers divisible by 6. By the intersection of an* **arbitrary** *number (finite or infinite) of sets S_α, we mean the set, denoted by $\bigcap_\alpha S_\alpha$, of all elements belonging to every one of the sets S_α.*

[2] A **paradox** is a logically self-contradictory statement or statement that runs contrary to one's expectation; it is a statement that, despite apparently valid reasoning from true premises, leads to a seemingly self-contradictory or a logically unacceptable conclusion. A paradox usually involves contradictory-yet-interrelated elements that exist simultaneously and persist over time.

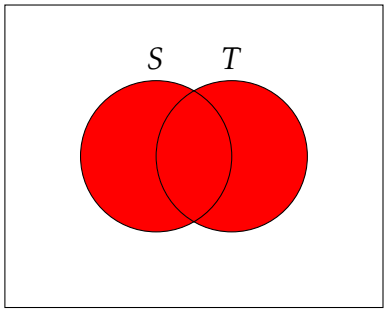


Figure 3.1: $S \cup T$.

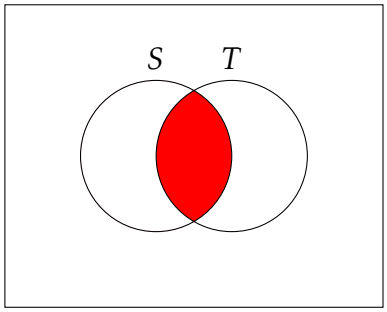


Figure 3.2: $S \cap T$.

Definition 7
Two sets S and T are said to be **disjoint** *if $S \cap T = \emptyset$, i.e. if they have no elements in common. More generally, let $\mathcal{F}$ be a family of sets such that $S \cap T = \emptyset$ for every pair of sets S, T in $\mathcal{F}$. Then the sets in $\mathcal{F}$ are said to be* **pairwise disjoint**.

It is an immediate consequence of the above definitions that the operations $\cup$ and $\cap$ are **commutative**

$$S \cup T = T \cup S,$$
$$S \cap T = T \cap S$$

and **associative**

$$(S \cup T) \cup U = S \cup (T \cup U),$$
$$(S \cap T) \cap U = S \cap (T \cap U).$$

Moreover, the operations $\cup$ and $\cap$ obey the following **distributive laws**:

$$(S \cup T) \cap U = (S \cap U) \cup (T \cap U),$$
$$(S \cap T) \cup U = (S \cup U) \cap (T \cup U).$$

Let us actually **prove** that

$$(S \cup T) \cap U = (S \cap U) \cup (T \cap U). \tag{3.1}$$

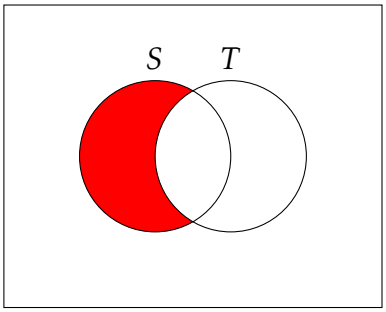


Figure 3.3: $S \setminus T$.

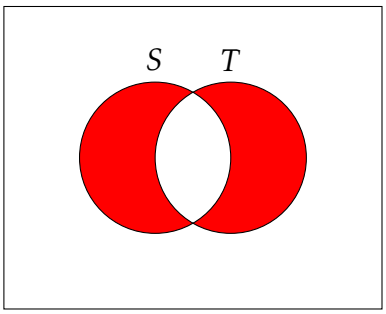


Figure 3.4: $S \triangle T$.

Suppose $x \in (S \cup T) \cap U$, so that x belongs to the left-hand side of (3.1). Then x belongs to both U and $S \cup T$, i.e. x belongs to both U and at least one of the sets S and T. But then x belongs to at least one of the sets $S \cap U$ and $T \cap U$, i.e., $x \in (S \cap U) \cup (T \cap U)$, so that x belongs to the right-hand side of (3.1).
Conversely, suppose $x \in (S \cap U) \cup (T \cap U)$. Then x belongs to at least one of the sets $S \cap U$ and $T \cap U$. It follows that x belongs to both U and at least one of the sets S and T, i.e. $x \in U$ and $x \in S \cup T$, or equivalently $x \in (S \cup T) \cap U$.
This completes the proof. Incidentally, this was the first proof of the course. We'll talk about proofs and their importance in calculus shortly.
By the difference $S \setminus T$ between two sets S and T (in that order), we mean the set of all elements of S which do not belong to T (Figure 3.3). Note that it is not assumed that $S \supseteq T$.
It is sometimes convenient (e.g., in measure theory) to consider the **symmetric difference** of two sets S and T (Figure 3.4), denoted by $S \triangle T$ and defined as the union of the two differences $S \setminus T$ and $T \setminus S$:

$$S \triangle T = (S \setminus T) \cup (T \setminus S).$$

We will often be concerned later with various sets which are all subsets of some underlying set R, for example, various sets of points on the real line. In this case, given a set S, the difference $R \setminus S$ is called the **complement** of A, denoted by $\overline{A}$.
The diagrams in Figures 3.1, 3.2, 3.3, and 3.4 are known as the Venn diagrams, so named after the English mathematician, logician, and philosopher John Venn

Figure 3.5: Stained glass window at Gonville and Caius College, Cambridge, commemorating Venn and the Venn diagram.

(1834–1923). There is even a stained glass window (Figure 3.5) at Gonville and Caius College, Cambridge, commemorating Venn and the Venn diagram.
An important role is played in set theory and its applications by **De Morgan's laws**, so named after the 19th-century British mathematician Augustus De Morgan:

$$\overline{\bigcup_\alpha S_\alpha} = \bigcap \overline{S_\alpha},$$
$$\overline{\bigcap_\alpha S_\alpha} = \bigcup \overline{S_\alpha};$$

as a special case,

$$\overline{S \cup T} = \overline{S} \cap \overline{T},$$
$$\overline{S \cap T} = \overline{S} \cup \overline{T}.$$

In words, the complement of a union equals the intersection of the complement, and the complement of an intersection equals the union of the complements.
Let us prove that

$$\overline{\bigcup_\alpha S_\alpha} = \bigcap \overline{S_\alpha}.$$

Suppose that $x \in R \setminus \bigcup_\alpha S_\alpha$. Then x does not belong to the union $\bigcup_\alpha S_\alpha$, i.e., x does not belong to any of the sets S_α. It follows that x belongs to each of the complements $R - S_\alpha$, and hence

$$x \in \bigcap_\alpha (R - S_\alpha) = \bigcap \overline{S_\alpha}. \tag{3.2}$$

Conversely, suppose (3.2) holds, so that x belongs to every set $R - S_\alpha$. Then x does not belong to any of the sets S_α, i.e. x does not belong to the union $\bigcup_\alpha S_\alpha$.

3.5 Functions

Let S and T be two arbitrary sets. Then a rule associating a unique element $b = f(a) \in T$ for each element $a \in S$ is said to define a **function** f on S (or a function f with **domain** S). f is sometimes also referred to as a **mapping of** S **into** T and is said to **map** S **into** T (and a into b).
If a is an element of S, the corresponding element $b = f(a)$ is called the **image** of a (**under** the mapping f). Every element of S with a given element $b \in T$ as its image is called a **preimage** of b. Note that in general b may have several preimages. Moreover, T may contain elements with no preimages at all. If b has a unique preimage, we denote this preimage by $f^{-1}(b)$.
If A is a subset of S, the set of all elements $f(a) \in T$ such that $a \in A$ is called the **image** of A, denoted by $f(A)$. The set of all elements of S whose images belong to a given set $B \subseteq T$ is called the **preimage** of B, denoted by $f^{-1}(B)$. If no element of B has a preimage, then $f^{-1}(B) = \emptyset$. A function f is said to map S **into** T if $f(S) \subseteq T$, as is always the case, and **onto** T if $f(S) = T$. Thus every "onto mapping" is an "into mapping," but not conversely.
Suppose f maps S **onto** T. Then f is said to be **one-to-one** if each element $b \in T$ has a unique preimage $f^{-1}(b)$. In this case, f is said to establish a **one-to-one correspondence** (or a **bijection**) between S and T and the mapping f^{-1} associating $f^{-1}(b)$ with each $b \in T$ is called the **inverse** of f.

Theorem 2
The preimage of the union of two sets is the union of the preimages of the sets:

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B).$$

Proof *If $x \in f^{-1}(A \cup B)$, then $f(x) \in A \cup B$, so that $f(x)$ belongs to at least one of the sets A and B. But then x belongs to at least one of the sets $f^{-1}(A)$ and $f^{-1}(B)$, i.e., $x \in f^{-1}(A) \cup f^{-1}(B)$.*
Conversely, if $x \in f^{-1}(A) \cup f^{-1}(B)$, then x belongs to at least one of the sets $f^{-1}(A)$ and $f^{-1}(B)$. Therefore, $f(x)$ belongs to at least one of the sets A and B, i.e., $f(x) \in A \cup B$. But then $x \in f^{-1}(A \cup B)$. □

The symbol □ stands for the Latin phrase *quod erat demonstrandum (Q.E.D.)*, meaning "which was to be demonstrated", "what was to be shown". While some authors still use the classical abbreviation, Q.E.D., it is relatively uncommon in modern mathematical texts. Paul Halmos pioneered the use of a solid black square ■ at the end of a proof as a Q.E.D. symbol, a practice which has become standard, although not universal. Halmos adopted this use of a symbol from magazine typography customs in which simple geometric shapes had been used to indicate the end of an article. This symbol was later called the **tombstone**, the **Halmos symbol**, or even a **halmos** by mathematicians. In these notes we use the □ variant of the halmos.

Theorem 3
The preimage of the intersection of two sets is the intersection of the preimages of the sets:

$$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B).$$

Proof *If $x \in f^{-1}(A \cap B)$, then $f(x) \in A \cap B$, so that $f(x) \in A$ and $f(x) \in B$. But then $x \in f^{-1}(A)$ and $x \in f^{-1}(B)$, i.e., $x \in f^{-1}(A) \cap f^{-1}(B)$.*
Conversely, if $x \in f^{-1}(A) \cap f^{-1}(B)$, then $x \in f^{-1}(A)$ and $x \in f^{-1}(B)$. Therefore $f(x) \in A$ and $f(x) \in B$, i.e., $f(x) \in A \cap B$. But then $x \in f^{-1}(A \cap B)$. □

Theorem 4
The image of the union of two sets equals the union of the images of the sets:

$$f(A \cup B) = f(A) \cup f(B).$$

Proof *If $y \in f(A \cup B)$, then $y = f(x)$ where x belongs to at least one of the sets A and B. Therefore $y = f(x)$ belongs to at least one of the sets $f(A)$ and $f(B)$, i.e., $y \in f(A) \cup f(B)$.*
Conversely, if $y \in f(A) \cup f(B)$, then $y = f(x)$ where x belongs to at least one of the sets A and B, i.e., $x \in A \cup B$ and hence $y = f(x) \in f(A \cup B)$. □

Remark 1
Surprisingly enough, the image of the intersection of two sets does not necessarily equal the intersection of the images of the sets. For example, suppose the mapping f projects the xy-plane onto the x-axis, carrying the point (x, y) into the $(x, 0)$. Then the segments $0 \le x \le 1, y = 0$ and $0 \le x \le 1, y = 1$ do not intersect, although their images coincide.

Remark 2
Theorems 2, 3, and 4 hold for unions and intersections of an **arbitrary** *number (finite or infinite) of sets A_α:*

$$f^{-1}\left(\bigcup_\alpha A_\alpha\right) = \bigcup_\alpha f^{-1}(A_\alpha),$$
$$f^{-1}\left(\bigcap_\alpha A_\alpha\right) = \bigcap_\alpha f^{-1}(A_\alpha),$$
$$f\left(\bigcup_\alpha A_\alpha\right) = \bigcup_\alpha f(A_\alpha).$$

3.6 Finite and infinite sets

The set of all vertices of a given polyhedron, the set of all prime numbers less than a given number, and the set of all residents of London (at a given time) have a certain property in common, namely, each set has a definite number of elements which can be found in principle, if not in practice. Accordingly these sets are all said to be **finite**. Clearly, we can be sure that a set is finite without knowing the number of

elements in it. On the other hand, the set of all positive integers, the set of all points on the line, the set of all circles in the plane, and the set of all polynomials with rational coefficients have a different property in common, namely, if we remove one element from each set, then remove two elements, three elements, and so on, there will still be elements left in the set at each stage. Accordingly, sets of this kind are said to be **infinite**.

Given two finite sets, we can always decide whether or not they have the same number of elements and, if not, we can always determine which set has more elements than the other. It is natural to ask whether the same is true of infinite sets. In other words, does it make sense to ask, for example, whether there are more circles in the plane than rational points on the line, or more functions defined in the interval $[0,1]$ than lines in space? As will soon be apparent, questions of this kind can indeed be answered.

To compare two finite sets S and T, we can count the number of elements in each set and then compare the two numbers, but alternatively, we can try to establish a **one-to-one correspondence** between (the elements of) A and B, i.e., a correspondence such that each element in A corresponds to one and only one element in B and vice versa. It is clear that a one-to-one correspondence between two finite sets can be set up iff the two sets have the same number of elements. For example, to ascertain whether or not the number of students in an assembly is the same as the number of seats in the auditorium, there is no need to count the number of students and the number of seats. We need merely observe whether or not there are empty seats or students with no place to sit down. If the students can all be seated with no empty seats left, i.e., if there is a one-to-one correspondence between the set of students and the set of seats, then these two sets obviously have the same number of elements. The important point here is that the first method (counting elements) works only for finite sets, while the second (setting up a one-to-one correspondence) works for infinite sets as well as for finite sets.

3.7 Countable sets

The simplest infinite set is the set $\mathbb{N} = \{1,2,3,\ldots\}$ of all natural numbers. An infinite set is called **countable** if its elements can be put in one-to-one correspondence with those of $\mathbb{N}$. In other words, a countable set is a set whose elements can be numbered $a_1, a_2, \ldots, a_n, \ldots$. By an **uncountable** set we mean, of course, an infinite set which is not countable.

Example 5

The set $\mathbb{Z}$ of all integers, positive, negative and zero, is countable. In fact, we can set up the following one-to-one correspondence between $\mathbb{Z}$ and the set $\mathbb{N}$ of all positive integers:

0	*-1*	*1*	*-2*	*2*	...
1	*2*	*3*	*4*	*5*	...

More explicitly, we associate the nonnegative integer $n \geq 0$ with the odd number $2n+1$,

and the negative integer $n < 0$ with the even number $2|n|$, i.e.

$$n \leftrightarrow \begin{cases} 2n+1, & \text{if } n \geq 0, \\ 2|n|, & \text{if } n < 0 \end{cases}$$

(the symbol $\leftrightarrow$ denotes a one-to-one correspondence).

Example 6
The set of all positive even numbers is countable, as shown by the obvious correspondence $n \leftrightarrow 2n$.

Example 7
The set $2, 4, 8, \ldots, 2^n, \ldots$ of powers of 2 is countable, as shown by the obvious correspondence $n \leftrightarrow 2^n$.

Example 8
The set $\mathbb{Q}$ of all rational numbers is countable. to see this, we first note that every rational number α can be written as a fraction p/q, $q > 0$ in lowest terms with a positive denominator. Call the sum $|p| + q$ the "height" of the rational number α. For example

$$\frac{0}{1} = 0$$

is the only rational number of height 1,

$$\frac{-1}{1}, \frac{1}{1}$$

are the only rational numbers of height 2,

$$\frac{-2}{1}, \frac{-1}{2}, \frac{1}{2}, \frac{2}{1}$$

are the only rational numbers of height 3, and so on. We can now arrange all rational numbers in order of increasing height (with the numerators increasing in each set of rational numbers of the same height). In other words, we first count the rational numbers of height 1, then those of height 2 (suitably arranged), those of height 3, and so on. In this way, we assign every rational number a unique positive integer, i.e., we set up a one-to-one correspondence between the set $\mathbb{Q}$ of all rational numbers and the set $\mathbb{N}$ of natural numbers.

Next we prove some elementary theorems involving countable sets:

Theorem 5
Every subset of a countable set is countable.

Proof *Let S be countable, with elements $s_1, s_2, \ldots$, and let A be a subset of S. Among the elements $s_1, s_2, \ldots$, let $s_{n_1}, s_{n_2}, ldots$ be those in the set A. If the set of numbers $n_1, n_2, \ldots$ has a largest number, then A is finite. Otherwise A is countable (consider the correspondence $i \leftrightarrow a_{n_i}$).* □

Theorem 6
The union of a finite or countable number of countable sets $A_1, A_2, \ldots$ is itself countable.

Proof *We can assume that no two of the sets $A_1, A_2, \ldots$ have elements in common, since otherwise we could consider the sets*

$$A_1, A_2 \setminus A_1, A_3 \setminus (A_1 \cup A_2), \ldots$$

instead, which are countable by Theorem 5 and have the same union as the original sets. Suppose we write the elements of $A_1, A_2, \ldots$ in the form of an infinite table

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{14} & \ldots \\ a_{21} & a_{22} & a_{23} & a_{24} & \ldots \\ a_{31} & a_{32} & a_{33} & a_{34} & \ldots \\ a_{41} & a_{42} & a_{43} & a_{44} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

where the elements of the set A_1 appear in the first row, the elements of the set A_2 appear in the second row, and so on. We now count all the elements in this table "diagonally," i.e., first we choose a_{11}, then a_{12}, then a_{21}, and so on, moving in the way shown in the following table:

$$\begin{array}{ccccccc} a_{11} & \rightarrow & a_{12} & & a_{13} & \rightarrow & a_{14} & \ldots \\ & \swarrow & & \nearrow & & \swarrow & & \\ a_{21} & & a_{22} & & a_{23} & & a_{24} & \ldots \\ \downarrow & \nearrow & & \swarrow & & & & \\ a_{31} & & a_{32} & & a_{33} & & a_{34} & \ldots \\ & \swarrow & & & & & & \\ a_{41} & & a_{42} & & a_{43} & & a_{44} & \ldots \\ \vdots & & \vdots & & \vdots & & \vdots & \ddots \end{array}$$

It is clear that this procedure associates a unique number to each element in each of the sets $A_1, A_2, \ldots$, thereby establishing a one-to-one correspondence between the union of the sets $A_1, A_2, \ldots$ and the set $\mathbb{N}$ of all positive integers. □

3.8 Historical remarks

The German mathematician Georg Cantor (1845–1918) is regarded as the founder of modern set theory.
Not everyone was happy with Cantor's work. Leopold Kronecker remarked:

> I don't know what predominates in Cantor's theory—philosophy or theology, but I am sure that there is no mathematics there.

However, his ideas became so influential that eventually set theory became the underlying language of mathematics. Writing decades after Cantor's death, Ludwig

Wittgestein lamented that mathematics is "ridden through and through with the pernicious idioms of set theory," which he dismissed as "utter nonsense" that is "laughable" and "wrong".
The harsh criticism had been matched by later accolades. In 1904, the Royal Society awarded Cantor its Sylvester Medal, the highest honour it can confer for work in mathematics. David Hilbert (1862–1943) defended Cantor's work from its critics by declaring,

> No one shall expel us from the paradise that Cantor has created.

The successes of set theory led to **logicism**—the belief that some or all of mathematics is reducible to logic. Bertrand Russell (1872–1970) and Alfred North Whitehead (1861–1947) championed this programme, initiated by Gottlob Frege (1848–1925) and subsequently developed by Richard Dedekind (1831–1916) and Giuseppe Peano (1858–1932). Hilbert launched what became known as Hilbert's programme:

1. a quest for a finite, complete set of axioms to build all of mathematics on, and
2. a proof that these axioms were consistent.

Russell discovered Russell's paradox in 1901. In 1908, Ernst Zermelo proposed an axiomatization of set theory that avoided the paradoxes of naïve set theory. Modifications to this axiomatix theory proposed in the 1920s by Abraham Fraenkel, Thoralf Skolem, and by Zermelo himself resulted in the axiomatic set theory called the **Zermelo–Fraenkel set theory**. Today, Zermelo–Fraenkel set theory, with the historically controversial axiom of choice (AC) included, is the standard form of axiomatic set theory and as such is the most common foundation of mathematics. Zermelo–Fraenkel set theory with the axiom of choice included is abbreviated **ZFC**, where C stands for "choice", and **ZF** refers to the axioms of Zermelo–Fraenkel set theory with the axiom of choice excluded.
In 1931 Kurt Gödel (1806–1978) dealt a blow to logicism by publishing his famous **incompleteness theorems** that are concerned with the limits of provability in formal axiomatic theories. These theorems are widely, but not universally, interpreted as showing that Hilbert's programme to find a complete and consistent set of axioms for all mathematics is impossible. In his first theorem, Gödel showed that any consistent system with a computable set of axioms which is capable of expressing arithmetic can never be complete: it is possible to construct a statement that can be shown to be true, but that cannot be derived from the formal rules of the system. In his second theorem, he showed that such a system could not prove its own consistency, so it certainly cannot be used to prove the consistency of anything stronger with certainty.

3.9 Further reading

Paul Halmos's *Naïve Set Theory* [Hal60] is a classical introduction to the foundations of naïve set theory. An account of the Zermelo–Fraenkel set theory can be found in [HK68].

Substantial chapters on set theory are contained in [Kel55, KF57, KF70, Kap72, Bin80, Lie15].

4 Proofs

4.1 Introduction

Consider the following mathematical statements:

- The square of an odd integer is odd. (By an **integer** we mean a whole number, i.e., one of the numbers $\ldots, -2, -1, 0, 1, 2, \ldots$.)
- No real number has square equal to -1.
- Every positive integer is equal to the sum of two integer squares. (The integer squares are $0, 1, 4, 9, 16, 25$, and so on.)

Each of these statements is either true or false. Probably you have quickly formed an opinion on the truth or falsity of each, and regard this as "obvious" in some sense. Nevertheless, to be totally convincing, you must provide clear, logical proofs to justify your opinions.
According to Bell (as cited in [Alm96]), "A proof is a directed tree of statements, connected by implications, whose end point is the conclusion and whose starting points are either in the data or are generally agreed facts or principles."
To clarify what constitutes a proof, we need to introduce a little notation. If P and Q are statements, we write

$$P \Rightarrow Q$$

to mean that statement P implies statement Q. For example,

$$x = 2 \Rightarrow x^2 < 6,$$

$$\text{it is raining} \Rightarrow \text{the sky is cloudy.}$$

Other ways of saying $P \Rightarrow Q$ are:

- if P then Q (e.g., if $x = 2$ then $x^2 < 6$);
- Q if P (e.g., the sky is cloudy if it is raining);
- P only if Q (e.g. $x = 2$ only if $x^2 < 6$; it rains only if the sky is cloudy).

Notice that $P \Rightarrow Q$ does **not** mean that also $Q \Rightarrow P$; for example $x^2 < 6 \not\Rightarrow x = 2$ (where $\not\Rightarrow$ means "does not imply"). However, for some statements, P, Q, it **is** the case that both $P \Rightarrow Q$ and $Q \Rightarrow P$; in such cases we write $P \Leftrightarrow Q$, and say "P if and only if Q." For example,

$$x = 2 \Leftrightarrow x^3 = 8.$$

The negation of a statement P is the opposite statement, "not P," written as the symbol $\overline{P}$. Notice that if $P \Rightarrow Q$ then also $\overline{Q} \Rightarrow \overline{P}$ (since if $\overline{Q}$ is true then P cannot be true, as $P \Rightarrow Q$).
For example, if P is the statement $x = 2$ and Q the statement $x^2 < 6$, then $P \Rightarrow Q$ says "$x = 2 \Rightarrow x^2 < 6$," while $\overline{Q} \Rightarrow \overline{P}$ says "$x^2 \geq 6 \Rightarrow x \neq 2$." Likewise, for the other example above we have "the sky is not cloudy $\Rightarrow$ it is not raining."
Perhaps labouring the obvious, let us now make a list of the deductions that can be made from the implication "it is raining $\Rightarrow$ the sky is cloudy," given various assumptions:

Assumption	Deduction
it is raining	sky is cloudy
it is not raining	no deduction possible
sky is cloudy	no deduction possible
sky is not cloudy	it is not raining

Now let us put together some examples of proofs. In general, a proof will consist of a series of implications, proceeding from given assumptions, until the desired conclusion is reached. As we shall see, the logic behind a proof can take several different forms.

4.2 Examples

Proposition 7
The square of an odd integer is odd.

Proof *Let n be an odd integer. Then n is 1 more than an even integer, so $n = 1 + 2m$ for some integer m. Therefore $n^2 = (1 + 2m)^2 = 1 + 4m + 4m^2 = 1 + 4(m + m^2)$. This is 1 more than $4(m + m^2)$, an even number, hence n^2 is odd.* □

Formally, we could have written this proof as the following series of implications:

$$n \text{ odd} \Rightarrow n = 1 + 2m \Rightarrow n^2 = 1 + 4(m + m^2) \Rightarrow n^2 \text{ odd}.$$

However, this is evidently somewhat terse, and such an approach with more complicated proofs quickly leads to unreadable mathematics; so, as in the above proof, we insert words of English to make the proof readable, including words like "hence", "therefore", "then" and so on, to take the place of implication symbols.
Note that the above proof shows rather more than just the oddness of n^2: it shows that the square of an odd number is always 1 more than a multiple of 4, i.e., is of the form $1 + 4k$ for some integer k.
This proof could be described as a **direct** proof in that it proceeds from the given assumptions directly to the conclusion via a series of implications. We now discuss two other types of proof, both very commonly used.
The first is **proof by contradiction**. Suppose we wish to prove the truth of a statement P. A proof by contradiction would proceed by first assuming that P is false—in

other words, assuming $\overline{P}$. We would try to deduce from this a statement Q that is palpably false (for example Q could be the statement "0 = 1"). Having done this, we have shown

$$\overline{P} \Rightarrow Q.$$

Hence also $\overline{Q} \Rightarrow P$. Since we know Q is false, $\overline{Q}$ is true, and hence so is P, so we have proved P as desired.
The next three propositions illustrate the method of proof by contradiction.

Proposition 8
Let n be an integer such that n^2 is a multiple of 3. Then n is also a multiple of 3.

Proof *Suppose n is not a multiple of 3. Then when we divide n by 3, we get a remainder of either 1 or 2; in other words, n is either 1 or 2 more than a multiple of 3. If the remainder is 1, then $n = 1 + 3k$ for some integer k, so*

$$n^2 = (1 + 3k)^2 = 1 + 6k + 9k^2 = 1 + 3(2k + 3k^2).$$

But this means that n^2 is 1 more than a multiple of 3, which is false, as we are given that n^2 is a multiple of 3. And if the remainder is 2, then $n = 2 + 3k$ for some integer k, so

$$n^2 = (2 + 3k)^2 = 4 + 12k + 9k^2 = 1 + 3(1 + 4k + 3k^2),$$

which is again false as n^2 is a multiple of 3.
Thus we have shown that assuming n is not a multiple of 3 leads to a false statement. Hence, as explained above, we have proved that n is a multiple of 3. □

Usually in a proof by contradiction, when we arrive at our false statement Q, we simply write something like "this is a contradiction" and stop. We do this in the next proof.

Proposition 9
No real number has square equal to -1.

Proof *Suppose the statement is false. This means that there* **is** *a real number, say x, such that $x^2 = -1$. However, it is a general fact about real numbers that the square of any real number is greater than or equal to 0. Hence $x^2 \geq 0$, which implies that $-1 \geq 0$. Contradiction.* □

Proposition 10
We'll prove that

$$\sqrt{2} + \sqrt{6} < \sqrt{15}.$$

Proof *Let us start by giving a non-proof:*

$$\sqrt{2} + \sqrt{6} < \sqrt{15} \Rightarrow \left(\sqrt{2} + \sqrt{6}\right)^2 < 15$$
$$\Rightarrow 8 + 2\sqrt{12} < 15 \Rightarrow 2\sqrt{12} < 7 \Rightarrow 48 < 49.$$

The last statement ($48 < 49$) is true, so why is this not a proof? Because the implication is going the wrong way—we have shown that if P is the statement we want to prove,

Figure 4.1: A black swan.

and Q is the statement that $48 < 49$, then $P \Rightarrow Q$; but this tells us nothing about the truth or otherwise of P.
A cunning change to the above false proof gives a correct proof, by contradiction. So assume the result is false; i.e., assume that $\sqrt{2}+\sqrt{6} \geq \sqrt{15}$. Then

$$\begin{aligned}\sqrt{2}+\sqrt{6} \geq \sqrt{15} &\Rightarrow \left(\sqrt{2}+\sqrt{6}\right)^2 \geq 15 \\ &\Rightarrow 8+2\sqrt{12} \geq 15 \Rightarrow 2\sqrt{12} \geq 7 \Rightarrow 48 \geq 49,\end{aligned}$$

which is a contradiction. Hence we have proved that $\sqrt{2}+\sqrt{6} < \sqrt{15}$. □

The other method of proof we shall discuss is actually a way of proving statements are false—in other words, **disproving** them. We call the method **disproof by counterexample**. It is best explained by examples.

Example 9
Consider the following two statements:

(a) All swans are white.

(b) Every positive integer is equal to the sum of two integer squares.

As the reader will have cleverly spotted, both these statements are false. To disprove (a), we need to prove the negation, which is "not all swans are white," or equivalently, "there exists a swan that is not white"; this is readily done by simply displaying one swan that is not white—this swan will then be a **counterexample** *to statement (a). The point is that to disprove (a), we do not need to consider* **all** *swans, we just need to produce a single counterexample.*
Likewise, to disprove (b) we just need to provide a single counterexample—that is, a positive integer that is **not** *equal to the sum of two squares. The number 3 fits the bill nicely.*

4.3 Quantifiers

We will conclude the chapter by slightly formalizing some of the discussion we have already had about proofs.
Consider the following statements:

(1) There is an integer n such that $n^3 = -27$.

(2) For some integer x, $x^2 = -1$.

(3) There exists a positive integer that is not equal to the sum of three integer squares.

Each of these statements has the form: "there exists some integer with a certain property." This type of statement is so common in mathematics that we represent the phrase "there exists" by a special symbol, namely $\exists$. So, writing $\mathbb{Z}$ for the set of all integers, the above statements can be written as follows:

(1) $\exists n \in \mathbb{Z}$ such that $n^3 = -27$.

(2) $\exists x \in \mathbb{Z}$ such that $x^2 = -1$.

(3) $\exists x \in \mathbb{Z}$ such that x is positive and is not equal to the sum of three integer squares.

The symbol $\exists$ is called the **existential quantifier**. To prove that an existence statement is true, it is enough to find just one object satisfying the required property. So (1) is true, since $n = -3$ has the required property; and (3) is true since $x = 7$ is not the sum of three squares (of course there are many other values of x having this property, but only one value is required to demonstrate the truth of (3)).
Now consider the following statements:

(4) For all integers n, $n^2 \geq 0$.

(5) The cube of any integer is positive.

(6) Every integer is equal to the difference of two positive integers.

All these statements are of the form: "for all integers, a certain property is true." Again, this type of statement is very common in mathematics, and we represent the phrase "for all" by a special symbol, namely $\forall$. So the above statements can be rewritten as follows:

(4) $\forall n \in \mathbb{Z}$, $n^2 \geq 0$.

(5) $\forall n \in \mathbb{Z}$, $n^3 > 0$.

(6) $\forall x \in \mathbb{Z}$, x is equal to the difference of two positive integers.

The symbol $\forall$ is called the **universal quantifier**. To show that a "for all" statement is true, a general argument is required; to show it is false, a single counterexample is all that is needed (this is just proof by counterexample, discussed in the previous section). We will leave you to show that (4) and (6) are true, while (5) is false.
Many mathematical statements involve more than one quantifier. For example, statement (6) above can be rewritten as

(6) $\forall x \in \mathbb{Z}$, $\exists m, n \in \mathbb{Z}$ such that $m > 0, n > 0$ and $x = m - n$.

Here's another example: the statement "for any integer a, there is an integer b such that $a+b=0$" can be rewritten as "$\forall a \in \mathbb{Z}, \exists b \in \mathbb{Z}$ such that $a+b=0$." Notice that the order of quantifiers is important: the statement "$\exists b \in \mathbb{Z}$ such that $\forall a \in \mathbb{Z}, a+b=0$" means something quite different.
Let's finish by seeing how to find the negation of a statement involving quantifiers. Consider statement (1) above: $\exists n \in \mathbb{Z}$ such that $n^3 = -27$. The negation of this is the statement "there does not exist an integer n such that $n^3 = -27$"—in other words, "every integer has cube not equal to -27," or more succinctly "$\forall n \in \mathbb{Z}$, $n^3 \neq -27$." So to form the negation of the original statement, we have changed $\exists$ to $\forall$ and negated the conclusion (i.e. changed $n^3 = -27$ to $n^3 \neq -27$).
Now consider statement (5): $\forall n \in \mathbb{Z}$, $n^3 > 0$. The negation of this is "not all integers have a positive cube"—in other words, "there is an integer having a non-positive cube," or more succinctly, "$\exists n \in \mathbb{Z}$ such that $n^3 \leq 0$." This time, to form the negation we have replaced $\forall$ by $\exists$ and negated the conclusion.
To summarize: when forming the negation of a statement involving quantifiers, we change $\exists$ to $\forall$, change $\forall$ to $\exists$ and negate the conclusion.
Let's do another example, and negate the following statement:

(7) For any integers x and y, there is an integer z such that $x^2 + y^2 = z^2$.

We can rewrite this as: $\forall x \in \mathbb{Z}$, $\forall y \in \mathbb{Z}$, $\exists z \in \mathbb{Z}$ such that $x^2 + y^2 = z^2$. Hence the negation is

$$\exists x \in \mathbb{Z}, \exists y \in \mathbb{Z}, \text{ such that } \forall z \in \mathbb{Z}, x^2 + y^2 \neq z^3.$$

In other words, there exist integers x, y such that for all integers z, $x^2 + y^2 \neq z^2$. Can you decide whether (7) or its negation is true?
Finally, let us make an observation for you to be wary of or amused by (or both). Here are a couple of strange statements involving the empty set:

(8) $\forall a \in \{x \mid x \text{ a real number}, x^2 + 1 = 0\}$, we have $a^{17} - 72a^12 + 39 = 0$.

(9) $\exists b \in \{x \mid x \text{ a real number}, x^2 + 1 = 0\}$ such that $b^2 \geq 0$.

You will have noticed that the set $\{x \mid x \text{ a real number}, x^2 + 1 = 0\}$ is equal to the empty set. Hence the statement in (8) says that all elements of the empty set have a certain property; this is true, since there are no elements in the empty set! Likewise, any similar "for all" statement involving the empty set is true. On the other hand, the statement (9) says that there exists an element of the empty set with a certain property; this must be false, since there are no elements in the empty set.

4.4 Induction

Consider the following three statements, each involving a general positive integer n:

(1) The sum of the first n odd numbers is equal to n^2.

(2) If $p > -1$ then $(1+p)^n \geq 1 + np$.

(3) The sum of the internal angles in an n-sided polygon is $(n-2)\pi$.

[A **polygon** is a closed figure with straight edges, such as a triangle (3 sides), a quadrilateral (4 sides), a pentagon (5 sides), etc.]
We can check that these statements are true for various specific values of n. For instance, (1) is true for $n = 2$ as $1 + 3 = 4 = 2^2$, and for $n = 3$ as $1 + 3 + 5 = 9 = 3^2$; statement (2) is true for $n = 1$ as $1 + p \geq 1 + p$, and for $n = 2$ as $(1 + p)^2 = 1 + 2p + p^2 \geq 1 + 2p$; and (3) is true for $n = 3$ as the sum of the angles in a triangle is π, and for $n = 4$ as the sum of the angles in a quadrilateral is 2π.
But how do we go about trying to prove the truth of these statements for **all** values of n?
The answer is that we use the following basic principle. In it we denote by $P(n)$ a statement involving a positive integer n; for example, $P(n)$ could be any of statements (1), (2) or (3) above.

Definition 8 (Principle of Mathematical Induction)
Suppose that for each positive integer n we have a statement $P(n)$. If we prove the following two things:

(a) $P(1)$ is true;

(b) for all n, if $P(n)$ is true then $P(n+1)$ is also true;

then $P(n)$ is true for all positive integers n.

The logic behind this principle is clear: by (a), the first statement $P(1)$ is true. By (b) with $n = 1$, we know that $P(1) \Rightarrow P(2)$, hence $P(2)$ is true. By (b) with $n = 2$, $P(2) \Rightarrow P(3)$ is true; and so on.
The principle may look a little strange at first sight, but a few examples should clarify matters.

Example 10
Let us try to prove statement (1) above using the Principle of Mathematical Induction. Here $P(n)$ is the statement that the sum of the first n odd numbers is n^2. In other words:

$$P(n) : 1 + 3 + 5 + \ldots + 2n - 1 = n^2.$$

We need to carry out parts (a) and (b) of the principle.

(a) $P(1)$ is true, since $1 = 1^2$.

(b) Suppose $P(n)$ is true. Then

$$1 + 3 + 5 + \ldots + 2n - 1 = n^2.$$

Adding $2n + 1$ to both sides gives

$$1 + 3 + 5 + \ldots + 2n - 1 + 2n + 1 = n^2 + 2n + 1 = (n + 1)^2,$$

which is statement $P(n + 1)$. Thus, we have shown that $P(n) \Rightarrow P(n + 1)$.

We have now established parts (a) and (b). Hence by the Principle of Mathematical Induction, $P(n)$ is true for all positive integers n.

The phrase "Principle of Mathematical Induction" is quite a mouthful, and we usually use just the single word "induction" instead.

Example 11
Now let us prove statement (2) above by induction. Here, for n a positive integer $P(n)$ is the statement

$$P(n) : \textit{if } p > -1 \textit{ then } (1+p)^n \geq 1 + np.$$

For (a), observe $P(1)$ is true, as $1 + p \geq 1 + p$.
For (b), suppose $P(n)$ is true, so $(1+p)^n \geq 1 + np$. Since $p > -1$ we know that $1 + p > 0$, so we can multiply both sides of the inequality by $1 + p$ to obrain

$$(1+p)^{n+1} \geq (1+np)(1+p) = 1 + (n+1)p + np^2.$$

Since $np^2 \geq 0$, this implies that $(1+p)^{n+1} \geq 1 + (n+1)p$, which is statement $P(n+1)$. Thus we have shown $P(n) \Rightarrow P(n+1)$.
Therefore, by induction, $P(n)$ is true for all positive integers n.

Next we attempt to prove the statement (3) concerning n-sided polygons. There is a slight problem here. If we naturally enough let $P(n)$ be statement (3), then $P(n)$ makes sense only if $n \geq 3$; $P(1)$ and $P(2)$ make no sense, as there is no such thing as a 1-sided or 2-sided polygon. To take care of such a situation, we need a slightly modified Principle of Mathematical Induction:

Definition 9 (Principle of Mathematical Induction II)
Let k be an integer. Suppose that for each integer $n \geq k$ we have a statement $P(n)$. If we prove the following two things:

(a) $P(k)$ is true;

(b) for all $n \geq k$, if $P(n)$ is true then $P(n+1)$ is also true;

then $P(n)$ is true for all integers $n \geq k$.

The logic behind this is the same as explained before.

Example 12
Now we prove statement (3). Here we have $k = 3$ in the above principle, and for $n \geq 3$, $P(n)$ is the statement

$$P(n) : \textit{ the sum of the internal angles in an } n\textit{-sided polygon is } (n-2)\pi.$$

For (a), observe that $P(3)$ is true, since the sum of the angles in a triangle is $\pi = (3-2)\pi$.
Now for (b). Suppose $P(n)$ is true. Consider an $(n+1)$-sided polygon with corners $A_1, A_2, \ldots, A_{n+1}$:
TODO Figure
Draw the line A_1A_3. Then $A_1A_3A_4 \ldots A_{n+1}$ is an n-sided polygon. Since we are assuming $P(n)$ is true, the internal angles in this n-sided polygon add up to $(n-2)\pi$. From the picture we see that the sum of the angles in the $(n+1)$-sided polygon $A_1A_2 \ldots A_{n+1}$ is equal to the sum of those in $A_1A_3A_4 \ldots A_{n+1}$ plus the sum of those in the triangle $A_1A_2A_3$, hence is

$$(n-2)\pi + \pi = ((n+1)-2)\pi.$$

We have now shown that $P(n) \Rightarrow P(n+1)$. Hence, by induction, $P(n)$ is true for all $n \geq 3$.

4.5 Historical remarks

The concept of mathematical proof had its beginnings with the ancient Greeks, who transformed mathematics from an experimental science to an intellectual science [Gra74]. Thales of Miletus (634–548 B.C.E.) is the first person to be given credit for discoveries in mathematics. He used deductive reasoning for finding new mathematical truths. The first proof in the history of mathematics is considered to be when Thales proved that the diameter of a circle divides a circle into two equal parts.
To learn more about the history of mathematical proof, see [Che12, BD13].
To learn more about the black swans, take a look at [Tal07].

4.6 Further reading

The book *How to Prove It: A Structured Approach* [Vel19] by Daniel J. Velleman is dedicated exclusively to proofs.
Chapters on proofs can be found in [Bin80, Lie15].

Part II

Foundations of analysis

5 Sequences

5.1 Binary relations

A **binary relation** over sets A and B is a subset of the **Cartesian product** $A \times B$; that is, it is a set of ordered pairs (a, b) consisting of elements a in A and b in B. It encodes the common concept of "relation": an element a is **related** to an element b, if and only if the pair (a, b) belongs to the set of ordered pairs that defines the binary relation.
A binary relation is the most studied special case $n = 2$ of an **n-ary relation** over sets $A_1, \ldots, A_n$, which is a subset of the Cartesian product $A_1 \times \ldots \times A_n$.
An example of a binary relation is the "divides" relation over the set of prime numbers $\mathbb{P}$ and the set of integers $\mathbb{Z}$, in which each price p is related to each integer z that is a multiple of p, but not to an integer that is not a multiple of p.
In this relation, for instance, the prime number 2 is related to numbers such as $-4, 0, 6, 10$, but not to 1 or 9, just as the prime number 3 is related to $0, 6$, and 9, but not to 4 or 13. We can put this in the language of sets: $(2, -4), (2, 0), (2, 6)$, and $(2, 10)$ are elements of this binary relation, whereas $(2, 1)$ and $(2, 9)$ are not; likewise, $(3, 0), (3, 6), (3, 9)$ are elements of this binary relation, but $(3, 4)$ and $(3, 13)$ are not.
Binary relations are used in many branches of mathematics to model a wide variety of relationships. These include, among others:

- the "is greater than," "is equal to," and "divides" relations in arithmetic;
- the "is congruent to" relation in geometry;
- the "is adjacent to" relation in graph theory;
- the "is orthogonal to" relation in linear algebra.

5.2 Mathematical functions

Let us recall that a function f from a set A to a set B is a mathematical object that associates to every element a in A a unique element $f(a)$ in B. We denote this by $f : A \to B$. For example, $f : \mathbb{N} \to \mathbb{N}$ with $f(x) = x + 1$ is a function since it maps every natural number $x \geq 1$ to a unique natural number $x+1$. In contrast, the mathematical object that associates to every element x in $\mathbb{N}$ both x and $-x$ is not a function $\mathbb{N} \to \mathbb{N}$, since it associates to 3, for example, two elements 3 and -3.[1] Such an object is a binary relation and so functions are special binary relations.

[1] If we take the same map $x \mapsto \{-x, x\}$ and give it the type $\mathbb{N} \to 2^{\mathbb{N}}$ where $2^{\mathbb{N}}$ is the power set of $\mathbb{N}$, the set of all subsets of $\mathbb{N}$, then this **is** a function.

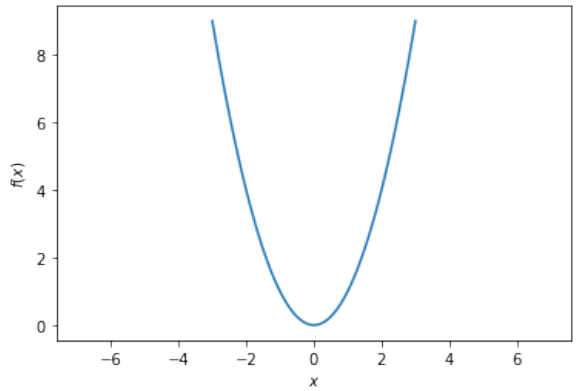


Figure 5.1: A parabola.

Not all binary relations are functions: for example "divides" associates to 2 both 4 and 6 and is therefore not a function; "equals" associates each element of a set (only) to itself and therefore **is** a function (called an **identity function**).
Here is another example of a function: $f : \mathbb{R} \to \mathbb{R}$, $f : x \mapsto x^2$. We can plot the graph of this function (Figure 5.1); this graph is called a **parabola**.

5.3 Functions in programming

Functions are a core concept in mathematics and computer science.
When we program, we can also **define** functions, for example

```
def square(x):
    return x * x
```

We can then **call** (**evaluate**) this function

```
square(3)
```

and obtain the answer 9. In programming, functions can take multiple values as **arguments** (**parameters**), for example

```
def add(x, y):
    return x + y


add(3, 5)
```

(that gives the answer 8). This function corresponds to the mathematical function of type add : $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$.
In programming, functions can have **side effects**; for example, they can output text to the console (write a record to a database, send data over the network, etc.):

```
def greet(name):
    print('Hello,', name)

greet('Paul')
```

Hello , Paul

In mathematics we don't consider functions with side effects. We are solely concerned with the input and output of a function; a mathematical function can have no side effects.

5.4 Sequences

We begin this course with the study of functions of the type $f : \mathbb{N} \to \mathbb{R}$, that associate real numbers to natural numbers; we will call these functions **sequences**, and the outputs $f(n)$ will be called **terms** of such sequences.
Thus far we have spoken about sequences informally; let us give a formal definition:

Definition 10 (Sequence)

1. *A* **sequence** *is a function $f : \mathbb{N} \to \mathbb{R}$ that maps natural numbers to real numbers. Often it is convenient to define a sequence as a function $f : \mathbb{N}^+ \to \mathbb{R}$ that maps* **positive** *natural numbers to real numbers, which is the terminology we adopt in these notes.*

2. *We will mostly write $(a_n)_{n\geq 1}$ for such a sequence, where $a_n = f(n)$.*

3. *By abuse of notation, we may write a_n or $a_1, a_2, \ldots$ for such a sequence if the context makes it clear that we refer to a sequence and not to one of its elements.*

Example 13
Here are some examples of sequences:

- $f : \mathbb{N}^+ \to \mathbb{R}$, $f : n \mapsto 1$, *also written* $(1)_{n\geq 1}$, *is a sequence:*

$$1, 1, 1, 1, 1, \ldots.$$

- $f : \mathbb{N}^+ \to \mathbb{R}$, $f : n \mapsto \frac{1}{n}$, *also written* $(\frac{1}{n})_{n\geq 1}$, *is a sequence:*

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots.$$

- $f : \mathbb{N}^+ \to \mathbb{R}$, $f : n \mapsto n^2 - n$, *also written* $(n^2 - n)_{n\geq 1}$, *is also a sequence:*

$$0, 2, 6, 12, 20, \ldots.$$

- $f : \mathbb{N}^+ \to \mathbb{R}$, $f : n \mapsto (-1)^n$, *also written* $((-1)^n)_{n\geq 1}$ *is yet another sequence:*

$$-1, 1, -1, 1, -1, \ldots.$$

- $f : \mathbb{N}^+ \to \mathbb{R}$ *defined by*

$$f(n) = \begin{cases} 10^{-6}, & \text{if } n \text{ is prime,} \\ \frac{1}{n}, & \text{if } n \text{ is not prime} \end{cases}$$

  is yet another sequence:

$$1, 10^{-6}, 10^{-6}, \frac{1}{4}, 10^{-6}, \frac{1}{6}, 10^{-6}, \frac{1}{8}, \frac{1}{9}, \frac{1}{10}, 10^{-6}, \ldots.$$

Remark 3 (Index Parameter)
Note that $(a_n)_{n\geq 1}$ and $(a_i)_{i\geq 1}$ refer to the same sequence as index variables are bound parameters that do not change meaning.

Example 14
An **arithmetic sequence** *(also known as an* **arithmetic progression***) is the sequence $f : \mathbb{N}^+ \to \mathbb{R}$ defined by*

$$f : n \mapsto \begin{cases} a_1, & n = 1, \\ a_1 + (n-1)d, & \text{otherwise.} \end{cases}$$

where a_1 is the first term and d the **common difference***. In general,*

$$f(n) = f(m) + (n-m)d.$$

Here is an example of an arithmetic sequence with $a_1 = 5$ and $d = 2$:

$$5, 7, 9, 11, 13, 15, \ldots.$$

The sum of the first n terms of an arithmetic sequence can be written in several different ways:

$$S_n = a_1 + (a_1 + d) + (a_1 + 2d) + \ldots + (a_1 + (n-2)d) + (a_1 + (n-1)d),$$
$$S_n = (a_n - (n-1)d) + (a_n - (n-2)d) + \ldots + (a_n - 2d) + (a_n - d) + a_n.$$

Adding both sides of the two equations, all terms involving d cancel:

$$2S_n = n(a_1 + a_n).$$

Dividing both sides by 2 produces a common form of the equation:

$$S_n = \frac{n}{2}(a_1 + a_n).$$

Example 15
A **geometric sequence** *(also known as a* **geometric progression***) is the sequence $f : \mathbb{N}^+ \to \mathbb{R}$ defined by*

$$f : n \mapsto ar^{n-1},$$

where a is a **scale factor***, equal to the sequence's first term, and r the* **common ratio***. Thus the terms of this sequence are*

$$a, ar, ar^2, ar^3, ar^4, \ldots.$$

A geometric sequence follows the recursive relation

$$a_n = ra_{n-1}$$

for every integer $n \geq 2$.
Let us derive the sum of the first n terms of a geometric sequence,

$$S_n = ar^0 + ar^1 + ar^2 + \ldots + ar^{n-1}.$$

Multiply both sides by $(1-r)$*:*

$$\begin{aligned}(1-r)S_n &= (1-r)\left(ar^0 + ar^1 + ar^2 + \ldots + ar^{n-1}\right)\\ &= \left(ar^0 + ar^1 + ar^2 + \ldots + ar^{n-1}\right) - \left(ar^1 + ar^2 + ar^3 + \ldots + ar^n\right)\\ &= a - ar^n,\end{aligned}$$

since all the other terms cancel. If $r \neq 1$*, we can rearrange the above to get the convenient formula*

$$S_n = \frac{a(1-r^n)}{(1-r)}.$$

It is possible to specify a sequence **recursively**, where the latter terms are defined in terms of the previously defined terms. For example:

Example 16
The **Fibonacci sequence** $f : \mathbb{N}^+ \to \mathbb{R}$ *is defined by*

$$f : n \mapsto \begin{cases} 0, & n = 1,\\ 1, & n = 2,\\ f(n-1) + f(n-2), & n \geq 3.\end{cases}$$

The first few terms of this sequence are

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots.$$

We can easily find that this is sequence A000045 in the On-line Encyclopedia of Integer Sequences (OEIS): `https://oeis.org/A000045`

Definition 11
A sequence $(a_n)_{n\geq 1}$ *is* **increasing** *if* $a_{n+1} \geq a_n$ *for* $n \geq 1$*; it is* **decreasing** *if* $a_{n+1} \leq a_n$ *for* $n \geq 1$*. The sequence is called* **monotonic** *if it is either increasing or decreasing.*

The notions of a strictly increasing ($a_{n+1} > a_n$ for $n \geq 1$) and strictly decreasing sequence ($a_{n+1} < a_n$ for $n \geq 1$) are also similarly defined.

Example 17
The sequence $(1/n)_{n\geq 1}$ *is strictly decreasing while* $(n^2 - n)_{n\geq 1}$ *is strictly increasing. They are both monotonic.*

Example 18
The Fibonacci sequence of Example 16 is increasing (but not strictly increasing) and monotonic.

Before we proceed, it's appropriate to consider why sequences are (fundamentally) important.
We learn how to solve some equations analytically at school. For example, we know how to solve a quadratic equation, such as $2x^2 + 4x - 4 = 0$, using the appropriate formula. However, how do we solve something like $\cos(x) = x^3$? Typically we use

a **numerical method**, such as Newton's method, to solve such equations. We will study Newton's method in detail later in this course. For now, let's just say that solving $\cos(x) = x^3$ is equivalent to finding the zero of $f(x) = \cos(x) - x^3$. We have $f'(x) = -\sin(x) - 3x^2$. (If this is your first encounter with a derivative, don't worry—in this course you will see many of them.)
Since $-1 \leq \cos(x) \leq 1$ for all x, whereas $x^3 < 1$ for $x < 1$ and $x^3 > 1$ for $x > 1$, we know that our solution must lie between -1 and 1.
For example, with an initial guess $x_0 = 0.5$, the **sequence** (!) given by Newton's method is

$$\begin{array}{rcccccl}
x_1 & = & x_0 - \dfrac{f(x_0)}{f'(x_0)} & = & 0.5 - \dfrac{\cos 0.5 - 0.5^3}{-\sin 0.5 - 3 \times 0.5^2} & = & 1.112\,141\,637\,097\ldots \\
x_2 & = & x_1 - \dfrac{f(x_1)}{f'(x_1)} & = & \vdots & = & \underline{0.}909\,672\,693\,736\ldots \\
x_3 & = & \vdots & = & \vdots & = & \underline{0.867}\,263\,818\,209\ldots \\
x_4 & = & \vdots & = & \vdots & = & \underline{0.865\,477}\,135\,298\ldots \\
x_5 & = & \vdots & = & \vdots & = & \underline{0.865\,474\,033\,1}11\ldots \\
x_6 & = & \vdots & = & \vdots & = & \underline{0.865\,474\,033\,102}\ldots
\end{array}$$

Thus a numerical method for finding roots (and minima, and maxima) gives us not the true solution, but a **sequence** of approximate solutions. This sequence (hopefully!) has a special property that the approximate solutions **get closer and closer** to the true solution. We nearly said "that the approximate solutions **converge** to the true solution", but we haven't defined that term yet.
When we train a neural network (another numerical method), we find a sequence of approximate weights and biases, which hopefully get closer and closer to the globally optimal weights and biases (although often we end up getting stuck in a local optimum, which is unfortunate).
Furthermore, every real number has a decimal expression. The **decimal expression**

$$b_0.b_1b_2b_3\ldots,$$

represents the real number that is the "sum to infinity" of the series

$$b_0 + \frac{b_1}{10} + \frac{b_2}{10^2} + \frac{b_3}{10^3} + \ldots.$$

In other words, the real number $b_0.b_1b_2b_3\ldots$ is the sum of the sequence

$$b_0, \frac{b_1}{10}, \frac{b_2}{10^2}, \frac{b_3}{10^3}, \ldots.$$

In order to make this statement precise, we need to introduce one of the most fundamental concepts in mathematics—that of the **convergence to a limit** of a sequence of real numbers.

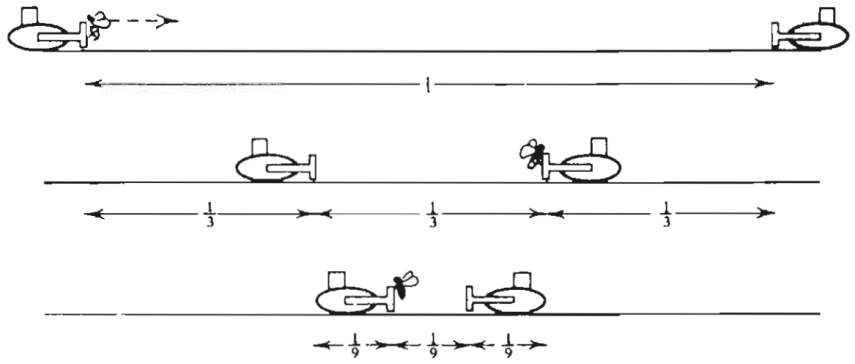


Figure 5.2: Two bulldozers and the bee.

5.5 The bulldozers and the bee

Two bulldozers are moving towards each other at a speed of one mile per hour on a collision course. When they started they were one mile apart and a bee was perched on the front of one of them. The bee began to fly back and forward between the bulldozers at a constant speed of two miles per hour, vainly seeking to avoid his fate. How far from his starting point will the unfortunate insect be crushed (Figure 5.2)? This riddle is usually posed in the hope that the victim will embark on some complicated calculations involving the flight of the bee. But it is quite obvious that the bee will be crushed when the bulldozers collide. Since the bulldozers travel at the same speed, this will happen halfway between their starting points. The answer is therefore one half mile.

Let us, however, take up the role of the riddler's victim and examine the flight of the bee. Let x_n denote the distance the bee is from his starting point when he makes his nth landing. Then

$$x_1 = \frac{2}{3},$$
$$x_2 = \frac{2}{3} - \frac{2}{9},$$

and, in general,

$$\begin{aligned} x_n &= \frac{2}{3} - \frac{2}{9} + \frac{2}{27} - \ldots + (-1)^{n-1} 2\left(\frac{1}{3}\right)^n \\ &= \frac{2}{3}\left\{1 + \left(-\frac{1}{3}\right) + \left(-\frac{1}{3}\right)^2 + \ldots + \left(-\frac{1}{3}\right)^{n-1}\right\} \\ &= \frac{2}{3}\left\{\frac{1 - \left(-\frac{1}{3}\right)^n}{1 - \left(-\frac{1}{3}\right)}\right\} = \frac{1}{2}\left(1 - \left(-\frac{1}{3}\right)^n\right). \end{aligned}$$

How is the answer $\frac{1}{2}$ to be extracted from this formula?

It seems clear that, as n gets larger and larger, $\left(-\frac{1}{3}\right)^n$ gets closer and closer to zero and so x_n gets closer and closer to $\frac{1}{2}$. We say that "x_n tends to $\frac{1}{2}$ as n tends to infinity"—or, in symbols, "$x_n \to \frac{1}{2}$ as $n \to \infty$".
This idea is of the greatest importance, but, before we can make proper use of it, it is necessary to give a **precise** formulation of what it means to say that $x_n \to l$ as $n \to \infty$.

5.6 The definition of convergence

Given a sequence of real numbers $(a_n)_{n\geq 1}$, we would like to have a precise definition of what it means for this sequence to converge to a limit l, a given real number or $\pm\infty$. You may already have a good intuitive understanding of what it means for a sequence to have a limit. Below is the mathematical definition that formally represents the concept.

Definition 12 (Convergence to a limit, divergence)
Let $(a_n)_{n\geq 1}$ be a sequence.

1. *The sequence $(a_n)_{n\geq 1}$* **converges to a limit** *l in $\mathbb{R}$ iff* [2] *the following statement can be shown to be true:*

   $$\text{For all } \epsilon > 0, \text{ we can find a } N \text{ in } \mathbb{N} \text{ such that, for all } n > N\colon\ |a_n - l| < \epsilon \quad (5.1)$$

   In that case, we denote this fact as $\lim_{n\to\infty} a_n = l$ or $(a_n)_{n\geq 1} \to l$.

2. *The sequence $(a_n)_{n\geq 1}$* **converges to ∞** *(in the extended real line $\mathbb{R}\cup\{\infty\}$) iff for all r in $\mathbb{R}$, there is an N in $\mathbb{N}$ such that for all $n \geq N$ we have $a_n > r$. We write $\lim_{n\to\infty} a_n = \infty$ or $(a_n)_{n\geq 1} \to \infty$ in that case.*

3. *The sequence $(a_n)_{n\geq 1}$* **converges to $-\infty$** *(in the extended real line $\mathbb{R}\cup\{-\infty\}$) iff for all r in $\mathbb{R}$, there is an N in $\mathbb{N}$ such that for all $n \geq N$ we have $a_n < r$. We write $\lim_{n\to\infty} a_n = -\infty$ or $(a_n)_{n\geq 1} \to -\infty$ in that case.*

4. *The sequence $(a_n)_{n\geq 1}$* **converges** *if it converges either to a real number or to ∞ or to $-\infty$.*

5. *The sequence $(a_n)_{n\geq 1}$* **diverges** *if it does not converge.*

The statement about convergence is best understood by taking the last part first and working backwards.

The limit inequality $|a_n - l| < \epsilon$ is called the *limit inequality* and says that the distance from the nth term in the sequence to the limit should be less than ϵ.

For all $n > N\ldots$ There should be some point N after which all the terms a_n in the sequence are within ϵ of the limit.

[2] We write "iff" as a shorthand for "if and only if".

For all $\epsilon > 0 \ldots$ No matter what value of ϵ we pick, however tiny, we will still be able to find some value of N such that all the terms to the right of a_N in that sequence are within ϵ of the limit.

Note that divergence is simply expressing that the sequence is not converging to any limit l in $\mathbb{R}$ or to $\pm\infty$.

Example 19 (Convergence and Divergence)

- *The sequence* $(n)_{n\geq 1}$ *converges to* ∞.
- *The sequence* $(-\ln n)_{n\geq 1}$ *converges to* $-\infty$.
- *Consider the sequence* $((-1)^n \cdot n)_{n\geq 1}$. *It grows without bounds but it always jumps between negative and positive values in that growth. Therefore, this sequence diverges.*

How do we prove convergence The above definition of convergence also tells us how to prove that a sequence converges to a real number l:

- Someone gives us a positive ϵ, we have no control over that value.
- For this value of ϵ, we have to then find a natural number N and demonstrate that $|a_n - l| < \epsilon$ for all $n > N$.

It may be helpful to think of the person who gives you that ϵ acting in the role of a Refuter who claims that this sequence does not have limit l. So you, as the Verifier of the opposite claim (that the sequence has l as limit), then need to find a value N that will convince the Refuter. In particular, you cannot first determine a value N and then let the refuter choose ϵ. Rather, the value N is a function of the choice of ϵ. We have an example below for $a_n = 1/n$.
We can similarly formulate how we can prove the convergence of a sequence to ∞ or to $-\infty$.

5.7 An illustration of convergence

A demonstration of the convergence of the sequence with $a_n = 1/n$ for $n \geq 1$ is shown in Figure 5.3. The *game* is to find a value of N for whatever value of ϵ is chosen. N represents the point after which all a_n will be within ϵ of the limit. Clearly the smaller the value of ϵ chosen, the further to the right in Figure 5.3 we will have to go to achieve this, thus the larger the value of N we will have to choose.

Applying the limit inequality In the example in Figure 5.3, we have taken a particular value of ϵ to demonstrate the concepts involved. Taking $\epsilon = 0.28$ (picked arbitrarily), we apply the $|a_n - l| < \epsilon$ inequality, to get:

$$|1/n| < 0.28$$
$$1/n < 0.28$$

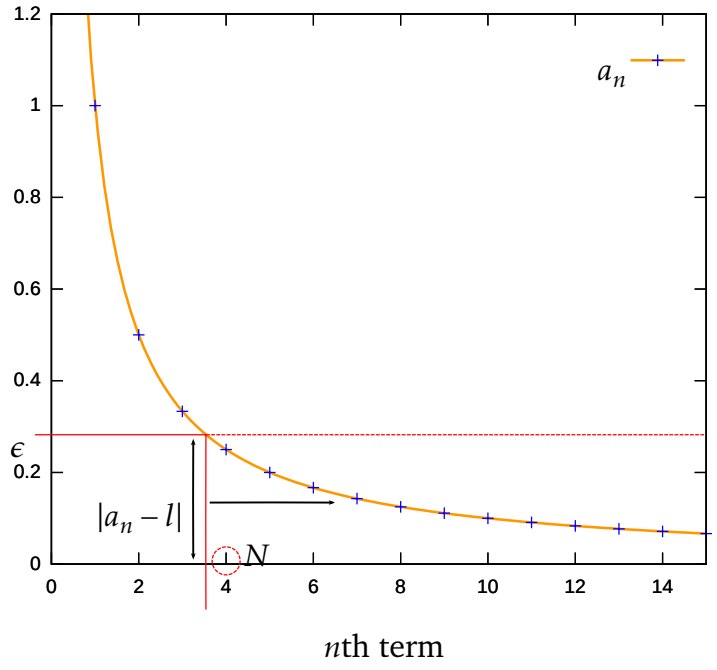


Figure 5.3: The convergence of the sequence $a_n = 1/n$ for $n \geq 1$ tending to 0

$$n > 1/0.28$$
$$n > 3.57$$

This states that for $n > 3.57$ the function $1/n$ will be less than 0.28 from the limit. However, we are dealing with a sequence which is only defined at integer points on this function. Hence we need to find the Nth term, a_N, which is the first point in the sequence to also be less than this value of ϵ. In this case, we can see that the next term above 3.57, that is $N = 4$, satisfies this requirement.

For all $n > N$ We are further required to ensure that all other points to the right of a_N in the sequence $(a_n)_{n>N}$ are also within 0.28 of the limit. We know this to be true because the condition on n we originally obtained was $n > 3.57$, so we can be sure that a_N, a_{N+1}, a_{N+2},... are all closer to the limit than 0.28.

For all $\epsilon > 0$ The final point to note is that it is not sufficient to find N for just a single value of ϵ. We need to find N for every value of $\epsilon > 0$. Since N will vary with ϵ as mentioned above, we are clearly required to find a function of type $\mathbb{R}^+ \to \mathbb{N}$ that maps any such ϵ to a suitable $N(\epsilon)$.
In the case of $a_n = 1/n$ for all $n \geq 1$, this is straightforward. We apply the limit inequality in general and derive a condition for n.

$$\frac{1}{n} < \epsilon \quad \Leftrightarrow \quad n > \frac{1}{\epsilon}$$

We are looking for a greater-than inequality. So long as we get one, we can select

the next largest integer value; we use the ceiling function to do this, giving:

$$N(\epsilon) = \left\lceil \frac{1}{\epsilon} \right\rceil$$

Now whatever value of ϵ we are given, we can find a value of N using this function and we are done. The notation $N(\epsilon)$ reminds us that the value of N is indeed a function of the chosen value of ϵ.

Exercise 1 (Divergence)
Show that $((-1)^n)_{n\geq 1}$ *diverges.*

5.8 Common convergent sequences

It is very useful to have an intuition about certain common sequences and whether they converge. Below is a list of useful results, which can be proved directly. We would recommend you get practice at direct convergence proofs on some of the sequences below.

1. $(a_n)_{n\geq 1} = \left(\frac{1}{n}\right)_{n\geq 1} \to 0$, also $(a_n)_{n\geq 1} = \left(\frac{1}{n^2}\right)_{n\geq 1} \to 0$, $(a_n)_{n\geq 1} = \left(\frac{1}{\sqrt{n}}\right)_{n\geq 1} \to 0$
2. In general, $(a_n)_{n\geq 1} = \left(\frac{1}{n^c}\right)_{n\geq 1} \to 0$ whenever c is a positive real constant $c > 0$
3. $(a_n)_{n\geq 1} = \left(\frac{1}{2^n}\right)_{n\geq 1} \to 0$, also $(a_n)_{n\geq 1} = \left(\frac{1}{3^n}\right)_{n\geq 1} \to 0$, $(a_n)_{n\geq 1} = \left(\frac{1}{e^n}\right)_{n\geq 1} \to 0$
4. In general, $(a_n)_{n\geq 1} = \left(\frac{1}{c^n}\right)_{n\geq 1} \to 0$ whenever c is a real constant with $|c| > 1$;
   or equivalently $(a_n)_{n\geq 1} = \left(c^n\right)_{n\geq 1} \to 0$ whenever c is a real constant with $|c| < 1$
5. $(a_n)_{n\geq 1} = \left(\frac{1}{n!}\right)_{n\geq 1} \to 0$
6. $(a_n)_{n\geq 1} = \left(\frac{1}{\log n}\right)_{n\geq 1} \to 0$

5.9 Combinations of sequences

It is fortunate that the convergence of sequences to limits enjoys good compositionality properties. By this we mean that if sequence elements get combined by some algebraic operator such as addition, and the argument sequences converge to a limit, then the combined sequence will converge to the combined limits. Let us see how this works concretely:
Given sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ which converge to limits $a \in \mathbb{R}$ and $b \in \mathbb{R}$ respectively, and a real constant λ, we have

1. $\lim_{n\to\infty} \lambda a_n = \lambda a$

2. $\lim_{n\to\infty} a_n + b_n = a + b$
3. $\lim_{n\to\infty} a_n - b_n = a - b$
4. $\lim_{n\to\infty} a_n b_n = ab$
5. $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}$ provided that $b \neq 0$

Note that the first item refers to a sequence $(\lambda a_n)_{n\geq 1}$ and that its limit equals λ times the limit of sequence $(a_n)_{n\geq 1}$. Note also that we use the convention that in rule 5, it is implicitly assumed that $b_n \neq 0$ for $n \geq 1$.

Exercise 2
Which of the five rules above would hold if a_n or b_n converge to $\pm\infty$?

We will next demonstrate how to prove the first of these rules. But first, we need a lemma:

Lemma 1 (triangle inequality)
For any real numbers a and b

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

Proof

$$\begin{aligned} |a+b|^2 = (a+b)^2 &= a^2 + 2ab + b^2 \\ &= |a|^2 + 2ab + |b|^2 \\ &\leq |a|^2 + 2|ab| + |b|^2 \\ &= |a|^2 + 2|a||b| + |b|^2 \\ &= (|a| + |b|)^2. \end{aligned}$$

We can show (do it!) that if x and y are positive, then $x \leq y$ iff $x^2 \leq y^2$.
Therefore

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

□

Proposition 11
Given sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ which converge to limits a and b respectively, and a real constant λ, we have

$$\lim_{n\to\infty} a_n + b_n = a + b.$$

Proof *Let $\epsilon > 0$ be given. Then $\frac{1}{2}\epsilon > 0$. Since $a_n \to a$ as $n \to \infty$, it follows that we can find an N_1 such that, for any $n > N_1$,*

$$|a_n - a| < \frac{1}{2}\epsilon. \tag{5.2}$$

Similarly, we can find an N_2 such that, for any $n > N_2$,

$$|b_n - b| < \frac{1}{2}\epsilon. \tag{5.3}$$

Let N be whichever of N_1 and N_2 is the larger, i.e. $N = \max\{N_1, N_2\}$. Then, if $n > N$, both inequalities (5.2) and (5.3) are true simultaneously. Thus, for any $n > N$,

$$\begin{aligned} |(a_n + b_n) - (a + b)| &= |(a_n - a) + (b_n - b)| \\ &= |a_n - a| + |b_n - b| \quad \textit{(triangle inequality)} \\ &= \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon. \end{aligned}$$

Given any $\epsilon > 0$ we have found a value of N (namely $N = \max\{N_1, N_2\}$) such that, for any $n > N$, $|(a_n + b_n) - (a + b)| < \epsilon$. Hence $a_n + b_n \to a + b$ as $n \to \infty$. □

5.10 Bounded sequences

A subset $A \subset \mathbb{R}$ is *bounded* if there exists $k > 0$ such that $|a| \leq k$ for all $a \in A$. Applying this to sequences, we have that a sequence $(a_n)_{n\geq 1}$ is bounded if there exists a real number $k > 0$ such that $|a_n| \leq k$ for all $n \geq 1$.

Proposition 12
If $(a_n)_{n\geq 1}$ converges to $a \in \mathbb{R}$, then it is bounded.

Proof Let $\epsilon = 1$ in the definition of convergence of $(a_n)_{n\geq 1}$ to a. Then there exists $N > 0$ such that $|a_n - a| < 1$ for $n > N$. Thus, $|a_n| = |a_n - a + a| \leq |a| + |a_n - a| < |a| + 1$ for all $n > N$. Put $k = \max\{|a_i| : 1 \leq i \leq N\} \cup \{|a| + 1\}$.

Exercise 3
Prove the rules 3 and 4 of combination of sequences. (Hint: For 4, write $|a_n b_n - ab| \leq |a_n b_n - a_n b + a_n b - ab|$ and use the triangle inequality and the boundedness of $(a_n)_{n\geq 1}$.)

Can you think how to prove rule 5? Try to prove that if $b_n \to b \neq 0$ as $n \to \infty$, then the sequence $(1/b_n)_{n\geq 1}$ is bounded.

Example 20 (Sequence Combination)
These sequence combination results are useful for being able to understand convergence properties of combined sequences.
As an example, if we take the sequence defined by

$$a_n = \frac{3n^2 + 2n}{6n^2 + 7n + 1}$$

for all $n \geq 1$, then we can transform this sequence into a combination of sequences that we know how to deal with, by dividing the numerator and denominator by n^2:

$$a_n = \frac{3 + \frac{2}{n}}{6 + \frac{7}{n} + \frac{1}{n^2}} = \frac{b_n}{c_n},$$

where

$$b_n = 3 + \frac{2}{n} \quad \textit{and} \quad c_n = 6 + \frac{7}{n} + \frac{1}{n^2}.$$

We know from rule (5) above that if $(b_n)_{n\geq 1} \to b$ *and* $(c_n)_{n\geq 1} \to c \neq 0$*, then*

$$\left(\frac{b_n}{c_n}\right)_{n\geq 1} \to \frac{b}{c}$$

This means that we can investigate the convergence of $(b_n)_{n\geq 1}$ *and* $(c_n)_{n\geq 1}$ *separately. Similarly we can rewrite* b_n *as* $3 + d_n$*, for* $d_n = \frac{2}{n}$*. We know that* $(d_n)_{n\geq 1} = (\frac{2}{n})_{n\geq 1} \to 0$ *as it is one of the common sequence convergence results, thus* $(b_n)_{n\geq 1} \to 3$ *by rule (2) above. By a similar argument, we can see that* $(c_n)_{n\geq 1} \to 6$ *using composition rule (2) and common results.*
Finally, we get the result that

$$(a_n)_{n\geq 1} \to \frac{3}{6} = \frac{1}{2}.$$

5.11 Cauchy sequences

The definition of a sequence $(a_n)_{n\geq 1}$ to have a limit l is certainly useful, as we have already appreciated above. However, there are two issues with this that we now want to resolve:

- **Uniqueness of limit:** What if a sequence could converge to more than one limit? In other words, is it possible that a sequence converges to more than one limit?

- **Convergence as guessing:** Our definition of convergence assumes that we guessed the value of a limit, which may be a hard undertaking. Can we come up with a definition of convergence that appeals only to the sequence itself and does not require such guesswork?

Exercise 4 (Uniqueness of limits)
Show that a sequence $(a_n)_{n\geq 1}$ *can only have one limit.*

If we want to get rid of having to guess the limit for a definition of convergence, we will require a somewhat more complex definition, since we no longer can work with a claimed limit value $l \in \mathbb{R}$. The notion of a **Cauchy Sequence** turns out to be the appropriate concept. Its intuition is that we keep the roles of ϵ and N as in the definition with a limit l. But that then we need to argue that all points in the sequence that come after a_N are within ϵ of each other:

Definition 13 (Cauchy sequence)
A sequence $(a_n)_{n\geq 1}$ *is a* **Cauchy sequence** *iff for all* $\epsilon > 0$*, there is some* N *in* $\mathbb{N}$ *such that for all* $n, m > N$ *we have* $|a_n - a_m| < \epsilon$*.*

Remark 4
It is not sufficient for each term to become arbitrarily close to the **preceding** *term. For instance, in the sequence of square roots of natural numbers*

$$a_n = \sqrt{n}$$

the consecutive terms become arbitrarily close to each other:

$$a_{n+1} - a_n = \sqrt{n+1} - \sqrt{n} = \frac{1}{\sqrt{n+1} + \sqrt{n}} < \frac{1}{2\sqrt{n}}.$$

However, with growing values of the index n, the terms a_n become arbitrarily large. So for any index n and distance d there exists an index m big enough such that $a_m - a_n > d$. (Actually, any $m > (\sqrt{n} + d)^2$ suffices.) As a result, despite how far one goes, the remaining terms of the sequence never get close to **each other**, *hence the sequence is not Cauchy.*

Exercise 5
Show that every sequence that converges to a real number is a Cauchy sequence. Is this also true of a sequence that converges to $\pm\infty$?

Here is an example of a sequence that converges to a real number and is thus Cauchy.

Example 21
[Hard] Let $a_1 = 1$ and set $a_{n+1} = \frac{1}{2} \cdot (a_n + 2/a_n)$ for all $n \geq 1$. This defines a sequence of rational numbers that converges to $\sqrt{2}$.

In particular, it follows from the uniqueness of limits that the Cauchy sequence in the above example does not have a limit over the rational numbers $\mathbb{Q}$.

Definition 14
A subset $A \subset \mathbb{R}$ is said to be **complete** *if any Cauchy sequence in A converges to a limit in A.*

This means that the rational numbers are not complete since, by Example 21, not all Cauchy sequences of rational numbers have limits in $\mathbb{Q}$. In contrast, the set $\mathbb{R}$ of real numbers **is** complete as we will prove later.

Exercise 6 (Cauchy Sequence)
Show that $(1/n)_{n\geq 1}$ is a Cauchy sequence.

Cauchy sequences can be defined similarly when elements a_n are vectors and the expression $|a_n - a_m|$ refers to a metric that generalizes the absolute value. The real vector spaces that we cover in Linear Algebra are then also complete in this sense.

5.12 The sandwich theorem

The sandwich theorem is an alternative, and often a simpler, way of proving that a sequence converges to a limit. In order to use the sandwich theorem, you need to have two sequences (one of which can be constant) which bound the sequence you wish to reason about and which converge to the same limit. The two bounding sequences should be given as known results, or proved to converge prior to the use of the sandwich theorem. The bounding sequences should be easier to reason about than the sequence that we want to reason about.

Theorem 13 (Sandwich theorem)
Let $(l_n)_{n\geq 1}$ *and* $(u_n)_{n\geq 1}$ *be sequences, and* l *a real number where both* $\lim_{n\to\infty} l_n = l$ *and* $\lim_{n\to\infty} u_n = l$*. Suppose that for a third sequence* $(a_n)_{n\geq 1}$ *we have:*

$$\text{there is some } N \in \mathbb{N} \text{ such that } l_n \leq a_n \leq u_n \text{ for all } n \geq N \tag{5.4}$$

Then $\lim_{n\to\infty} a_n = l$ *as well.*

Example 22 (Applying the sandwich theorem)
We provide an illustration of the sandwich theorem in Figure **??**. *Here, we have a sequence*

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

which oscillates around its limit of 0*. We construct two sandwiching sequences where* $l_n = -1/n$ *and* $u_n = 1/n$ *for all* $n \geq 1$*. Since we know (or can easily prove) that* $(l_n)_{n\geq 1}$ *and* $(u_n)_{n\geq 1}$ *both tend to* 0*, we only need to show that there is some* N *in* $\mathbb{N}$ *such that* $l_n \leq a_n$ *and* $a_n \leq u_n$ *from all* n *with* $n > N$*. In our particular case, we can show it for all* $n > 0$*: we can prove this in one line by referring to a property of the* sin *function:*

$$-\frac{1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n} \quad \Leftrightarrow \quad -1 \leq \sin n \leq 1,$$

the inequalities on the right-hand side being true for all n *by the range of* $\sin n$*.*

Proof of the sandwich theorem We know that $(l_n)_{n\geq 1}$ and $(u_n)_{n\geq 1}$ tend to the same limit l. We use the definition of convergence for both of these sequences. Given some $\epsilon > 0$, we can find an N_1 and N_2 such that for all $n > N_1$, $|l_n - l| < \epsilon$ and for all $n > N_2$, $|u_n - l| < \epsilon$ since these sequences have l as limit.
If we want both sequences to be simultaneously less than ϵ from the limit l for a single value of N', then we have to pick $N' = \max(N_1, N_2)$. So let n be such that $n > N'$. Then we know that

$$|l_n - l| < \epsilon \text{ and } |u_n - l| < \epsilon.$$

Therefore, this together with the fact that $l_n < u_n$ gives us:

$$l - \epsilon < l_n < u_n < l + \epsilon.$$

By the assumption of the sandwich theorem, there is some N in $\mathbb{N}$ such that $l_n \leq a_n \leq u_n$ for all $n > N$. Hence, for all $n > \max(N, N')$ we conclude that:

$$\begin{aligned} l - \epsilon < l_n \leq a_n \leq u_n < l + \epsilon \Rightarrow \\ l - \epsilon < a_n < l + \epsilon \Rightarrow \\ -\epsilon < a_n - l < \epsilon \Rightarrow \\ |a_n - l| < \epsilon. \end{aligned}$$

Since n was an arbitrary natural number larger than $\max(N, N')$, this proves that $(a_n)_{n\geq 1}$ converges to l as a result. We note that the $N(\epsilon)$ we determined here is $\max(N, N')$, which combines two such "N" values from the convergence of the two bounding sequences. □

5.13 Ratio tests for sequences

A third technique (and a very simple one) for proving sequence convergence is called the **ratio test**. Its name stems from the fact that it compares the ratio of consecutive terms in the sequence. The test can be used to verify whether a sequence converges to 0 or diverges. This is not really restricting its usage: If a sequence $(b_n)_{n\geq 1}$ has a nonzero limit, $l \neq 0$, then the ratio test can still be applied to a modified sequence $(a_n)_{n\geq 1}$ where a_n equals $b_n - l$ for all $n \geq 1$. If the modified sequence passes the test, we know it converges to 0 and so sequence $(b_n)_{n\geq 1}$ converges to l.
The convergence test checks whether the ratio of consecutive terms in the sequence is eventually below a constant that is strictly less than 1:

The ratio test for sequences Let c in $\mathbb{R}$ be such that $0 \leq c < 1$. Suppose that there is some N in $\mathbb{N}$ such that for all $n \geq N$ we have $\left|\frac{a_{n+1}}{a_n}\right| \leq c$. Then $\lim_{n\to\infty} a_n = 0$.

Example 23 (The ratio test for sequences)
Consider the sequence $(a_n)_{n\geq 1}$ where a_n equals 3^{-n} for all $n \geq 1$. We have

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{3^{-(n+1)}}{3^{-n}}\right| = \frac{1}{3} < 1.$$

So we can choose c to be $1/3$ and N to be 1 and the ratio test for sequences passes for these values. Therefore, $(3^{-n})_{n\geq 1}$ has limit 0.

Limit ratio test The **limit ratio test** is a method that is derived from the ratio test that we just discussed. It works by considering the limit of the ratio of consecutive terms in the sequence. This assumes that this limit exists. This is often the case when the above ratio test is inconclusive, e.g., when the ratios have n as a variable after any simplifications and so guessing a suitable constant c may be hard.

Formally, suppose that the limit

$$r = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists. If $r < 1$, then the sequence $(a_n)_{n\geq 1}$ converges to 0.

Example 24 (Ratio test and limit ratio test)
Consider the sequence $(a_n)_{n\geq 1}$ where a_n equals $\frac{1}{n!}$ for all $n \geq 1$. We compute

$$\left| \frac{a_{n+1}}{a_n} \right| = \left| \frac{1/(n+1)!}{1/n!} \right| = \frac{1}{n+1}$$

Since $c = 1/2$ satisfies $0 \leq c < 1$, and since for all $n > 1$, we have that $1/(n+1) \leq c$, we may apply the ratio test for $c = 1/2$ and $N = 1$ to conclude that the sequence $(1/n!)_{n\geq 1}$ converges to 0.
We can also apply the limit ratio test here since we get that

$$\lim_{n\to\infty} \Big(\frac{1}{n+1} \Big)_{n\geq 1} = 0.$$

So $r < 1$ since $r = 0$. Therefore, the limit ratio test passes and we also conclude with this test that this sequence has limit 0.

Applying these tests and the ones we develop below is something that you will be expected to be able to do in tutorials and exams. Producing more complex proofs such as the ones we discuss next, is not something you are expected to reproduce in an exam or assessed work.

5.14 Proof of correctness of ratio tests

For the sake of completeness, and as a good illustration of applying the concepts we developed, we will now prove the correctness of these ratio tests.

Proof of the correctness of the ratio test We are given c in $\mathbb{R}$ with $c < 1$ and N in $\mathbb{N}$ such that $\left| \frac{a_{n+1}}{a_n} \right| \leq c < 1$ for all $n \geq N$. For the left-hand inequality, we multiply both sides by $|a_n|$ to get $|a_{n+1}| \leq c|a_n|$ for $n \geq N$. Similarly, we get $|a_n| \leq c|a_{n-1}|$ and so forth. In fact, repeating this creates a chain of inequalities down to the point when n equals N, at which point that inequality also holds. Therefore, we obtain:

$$|a_n| \leq c|a_{n-1}| \leq c^2|a_{n-2}| \leq \ldots \leq c^{n-N}|a_N|$$

for all $n \geq N$. In particular, this means that for all $n \geq N$ we have

$$\begin{aligned} |a_n| &\leq c^{n-N}|a_N| \\ &\leq kc^n \text{ where } k = \frac{|a_N|}{c^N} \text{ is a constant.} \end{aligned}$$

Now, since $c < 1$, we know that the sequence $(b_n)_{n\geq 1}$ with $b_n = kc^n$ for all $n \geq 1$ has limit 0, by the standard result (4) from Section 5.8. We also bounded the sequence $(a_n)_{n\geq 1}$ between $(-b_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$. Since both of these sequences have limit 0, we can apply the sandwich theorem to conclude that the sequence $(a_n)_{n\geq 1}$ converges to 0. □

Proof of the correctness of the limit ratio test We are given that the new sequence $(b_n)_{n\geq 1}$ with $b_n = \left|\frac{a_{n+1}}{a_n}\right|$ has limit r. Recall that this means that for all $\epsilon > 0$, we can find an N in $\mathbb{N}$ such that for all $n > N$ we have $|b_n - r| < \epsilon$.
When we unpack that last inequality, we get:

$$r - \epsilon < b_n < r + \epsilon. \tag{5.5}$$

We will now pick specific values of ϵ and then reason about the inequalities in (5.5) in order to reduce the limit ratio test to the corresponding test without the limit.

Proof Assume $r < 1$. Recall that here we have to prove that the sequence $(a_n)_{n\geq 1}$ converges. Since $r < 1$, we know that ϵ defined as $\frac{1-r}{2}$ is positive. Therefore, there is some $N(\epsilon)$ such that for all $n \geq N$ the inequalities in (5.5) are true, and they now read as:

$$r - \frac{1-r}{2} < b_n < r + \frac{1-r}{2}$$

which we can simplify to

$$\frac{3r-1}{2} < b_n < \frac{r+1}{2}.$$

Taking the right-hand side, we see that $b_n < \frac{r+1}{2}$, and, since $r < 1$, we can show that $\frac{r+1}{2} < 1$ also. If we take $c = \frac{r+1}{2} < 1$ and N to be the $N(\epsilon)$ above, then the original ratio test passes for these choices. Since we have proved the correctness of the ratio test already, we can conclude that the original sequence $(a_n)_{n\geq 1}$ converges, as claimed. □

5.15 Subsequences of a sequence

If $f : \mathbb{N} \to \mathbb{R}$ is a sequence and $M \subset \mathbb{N}$ an infinite subset, then the restriction $f : M \to \mathbb{R}$ is called a *subsequence* of f. Using the usual notation $(a_n)_{n\geq 1}$ for a sequence, any subsequence of this sequence would be of the form $(a_{n_i})_{i\geq 1}$ where n_i are positive integers with $n_1 < n_2 < n_3 < ....$, i.e., $M = \{n_i : i \geq 1\}$.

Example 25
Suppose $a_n = 1/n$, then $a_{n_i} = 1/(2i)$ is the subsequence $1/2, 1/4, 1/6, \ldots$ of even terms, whereas $a_{n_i} = 1/(3i-2)$ is the subsequence $a_1, a_4, a_7, a_{10}, \ldots$.

Theorem 14
Any subsequence of a convergent sequence converges to the limit of the sequence.

Proof Suppose $\lim_{n\to\infty} a_n = l$, where l is a real number, and let $(a_{n_i})_{i\geq 1}$ be any subsequence. Let $\epsilon > 0$ be given. Then since $\lim_{n\to\infty} a_n = l$, there exists $N > 0$ such that for $n > N$ we have $|a_n - l| < \epsilon$. Note that by definition we always have $n_i \geq i$. Therefore for $i > N$, we have $|a_{n_i} - l| < \epsilon$. The proof for sequences that converge to $\pm\infty$ is similar. □

Proposition 15
Any sequence of real numbers has a monotonic subsequence.

Proof Consider a sequence $(a_n)_{n\geq 1}$. For any $m \geq 1$, we say that a_m is a **peak** of $(a_n)_{n\geq 1}$ if $a_m \geq a_n$ for all $n \geq m$. (A strictly increasing sequence will have no peaks while in a decreasing sequence every term is a peak.) There are now two cases. Suppose $(a_n)_{n\geq 1}$ has infinitely many peaks denoted by $a_{n_1}, a_{n_2}, \ldots$. Then the subsequence $(a_{n_i})_{i\geq 1}$ is monotonically decreasing. If, on the other hand, $(a_n)_{n\geq 1}$ has only a finite number of peaks $a_{n_1}, a_{n_2}, \ldots, a_{n_k}$, then choose $t_1 = n_k + 1$. Since t_1 is not a peak, there exists $n > t_1$ such that $t_2 := n$ satisfies $a_{t_1} < a_{t_2}$. In this way construct t_i for all $i \geq 1$ such that $(a_{t_i})_{i\geq 1}$ is a monotonically increasing sequence. □

5.16 Manipulating absolute values: useful techniques

Many of the reasoning techniques behind limits require the ability to manipulate the absolute values of expressions. There are some properties about them that are useful to know, and which we already used above. These will come in handy for other topics, including our study of series and reasoning about their radius of convergence. A commonly used decomposition of the absolute value operator is:

$$|x| < a \quad \Leftrightarrow \quad -a < x < a$$

In fact, we may think of this as a definition of the absolute value. Where we are combining absolute values of expressions, the following are useful rules, where x and y are real numbers:

Abs1 $|x \cdot y| = |x| \cdot |y|$.

Abs2 $\left|\dfrac{x}{y}\right| = \dfrac{|x|}{|y|}$.

Abs3 $|x + y| \leq |x| + |y|$, the triangle inequality.

Abs4 $|x - y| \geq ||x| - |y||$, the reverse triangle inequality.

We have already seen a proof of the triangle inequality. Let us prove the reverse triangle inequality.
From the triangle inequality it follows that

$$|y| = |x + y - x| \leq |x| + |y - x|$$

and

$$|x| = |y + x - y| \leq |y| + |x - y|.$$

Hence

$$|y - x| \geq |y| - |x| \text{ and } |x - y| \geq |x| - |y|.$$

From Abs1, $|y - x| = |(-1)(x - y)| = |-1||x - y| = |x - y|$. So we have

$$|x - y| \geq -(|x| - |y|) \text{ and } |x - y| \geq |x| - |y|.$$

Hence

$$|x - y| \geq ||x| - |y||,$$

as required. □

5.17 Properties of real numbers

Now that we have looked at limits of sequences of real numbers, we will see how this relates to some fundamental properties of sets of real numbers. Such considerations will lead us to describe the so called **fundamental axiom of analysis**.
Both the rational numbers $\mathbb{Q}$ and the real numbers $\mathbb{R}$ have a linear ordering $<$, meaning that for numbers x and y we either have $x < y$, $y < x$, or $x = y$. This allows us to reveal an intimate connection between this order structure and the convergence behavior of certain sequences. For that, we need to review concepts from order theory. We do this for real numbers, the definitions for rational numbers are essentially the same ones:

Definition 15 (Concepts from order theory)
Let X be a set of real numbers.

1. *Let l and u be real numbers. Then:*
   (a) *u is an* **upper bound** *of X if $x \leq u$ for all $x \in X$;*
   (b) *l is a* **lower bound** *of X if $l \leq x$ for all $x \in X$;*
   (c) *a* **least upper bound** *(***supremum***, $\sup(X)$) of X is an upper bound s of X such that $s \leq u$ for all upper bounds u of X;*
   (d) *a* **greatest lower bound** *(***infimum***, $\inf(X)$) of X is a lower bound i of X such that $l \leq i$ for all lower bounds l of X.*
2. *We say that such a set X is*
   (a) **bounded above** *if X has an upper bound;*
   (b) **bounded below** *if X has a lower bound;*

Check that if X has an upper bound and a lower bound, then it is a bounded set as defined in Section 5.10.

Proposition 16
The least upper bound of a set X is unique if it exists.

Proof *Let $L, L' \in \mathbb{R}$ such that L, L' are both least upper bounds of X.*
Because L is an upper bound of X and L' is a least upper bound of X,

$$L' \leq L.$$

Because L' is an upper bound of X and L is a least upper bound of X,

$$L \leq L'.$$

Thus $L = L'$. □

The uniqueness of the greatest lower bound (if it exists) is proved similarly.
For example, the set of positive real numbers $\{x \in \mathbb{R} \mid x > 0\}$ does not have any upper bounds and so also no least upper bound. But it has 0 and all negative numbers as lower bounds and 0 is its greatest lower bound.

5.17.1 Axiomatisation of real numbers

The set of real numbers $\mathbb{R}$ is **Dedekind-complete** by the following axiom of the real numbers:

Definition 16 (Axiom of Dedekind-completeness for real numbers)
Every nonempty subset X of the real numbers $\mathbb{R}$ that is bounded above has a least upper bound.

We are now in a position to state that a monotonically increasing sequence that is bounded above has a limit:

Theorem 17 (Fundamental theorem of analysis)
Let $(a_n)_{n\geq 1}$ be a sequence of real numbers that is

1. *increasing, i.e., for all $n \geq m \geq 1$ we have $a_n \geq a_m$, and*
2. *bounded above, i.e., there is some u in $\mathbb{R}$ such that $a_n \leq u$ for all $n \geq 1$.*

Then $s = \sup\{a_n \mid n \geq 1\}$ exists and is the limit of the sequence $(a_n)_{n\geq 1}$.

Proof By the axiom of Dedekind-completeness from Definition 16, we know that the supremum s stated in the theorem exists since $\{a_n \mid n \geq 1\}$ is nonempty and bounded above. We need to show that s is the limit of that sequence. So let $\epsilon > 0$ be given. We prove this in stages:

CLAIM 1: There exists some N in $\mathbb{N}$ such that $|a_N - s| < \epsilon$.
Proof by contradiction: If this claim is false, then we have $|a_n - s| \geq \epsilon$ for all $n \geq 1$. This implies $s - a_n \geq \epsilon$, i.e. $s - \epsilon \geq a_n$ for all $n \geq 1$ since $s \geq a_n$. But then $s - \epsilon$ is an upper bound of the set $\{a_n \mid n \geq 1\}$. Since s is the least such upper bound, we get that $s \leq s - \epsilon$. But this is a contradiction as ϵ is positive. This proves CLAIM 1.

CLAIM 2: For the N whose existence we proved in CLAIM 1, we have that $|a_n - s| < \epsilon$ for all $n > N$.
Proof: Since the sequence is monotonically increasing, we have that $n > N$ implies $a_n \geq a_N$. And so $0 \leq s - a_n \leq s - a_N < \epsilon$ proves the claim, where the first inequality follows since s is an upper bound of $\{a_n \mid n \geq 1\}$ and the second inequality follows from CLAIM 1.

By the definition of convergence, this shows that $(a_n)_{n\geq 1}$ has limit s. □

Similarly, a bounded below decreasing sequence is convergent to a real number.
Recall the definition of completeness.

Theorem 18
The set of real numbers is complete.

Proof Let $(a_n)_{n\geq 1}$ be a Cauchy sequence. We show that it converges to a limit in $\mathbb{R}$ from which the result follows.
CLAIM 1: $(a_n)_{n\geq 1}$ is bounded. Since the sequence is Cauchy, putting $\epsilon = 1$, there exists $N \in \mathbb{N}$ such that for $n, m > N$ we have $|a_n - a_m| < 1$. Put $K =$

$1 + \max\{|a_1|, |a_2|, \ldots, |a_N|, |a_{N+1}|\}$. Then for $m > N$ we have $|a_m| \leq |a_{N+1}| + |a_m - a_{N+1}| < |a_{N+1}| + 1$. Hence, $|a_n| < K$ for all $n \geq 1$.
CLAIM 2: The sequence $(a_n)_{n\geq 1}$ has a convergent subsequence. In fact, by Theorem 15, $(a_n)_{n\geq 1}$ has a monotonic subsequence, $(a_{n_i})_{i\geq 1}$ say, which by CLAIM 1 is also bounded. Hence, by the fundamental theorem of analysis, the subsequence is convergent with, say, $\lim_{i\to\infty} a_{n_i} = l$.
CLAIM 3: $\lim_{n\to\infty} a_n = l$. Let $\epsilon > 0$ be given. Since $(a_n)_{n\geq 1}$ is Cauchy, there exists $N_1 \in \mathbb{N}$ such that for $n, m > N_1$ we have $|a_n - a_m| < \epsilon/2$. Since $\lim_{i\to\infty} a_{n_i} = l$, there exists $N_2 > 0$ such that for $i > N_2$, we have $|a_{n_i} - l| < \epsilon/2$. Let $N = \max(N_1, N_2)$. Then

$$|a_n - l| \leq |a_n - a_{n_i}| + |a_{n_i} - l| < \epsilon/2 + \epsilon/2 = \epsilon$$

for any $n, i > N$. Thus, $|a_n - l| < \epsilon$ for $n > N$. Hence $(a_n)_{n\geq 1}$ is convergent to a real number. □

We already know that the set $\mathbb{Q}$ of rationals is not complete.

Exercise 7
Can you find some other subsets of $\mathbb{R}$ *that are complete?*

5.18 Historical remarks

The Greek philosopher Zeno of Elea (c. 495–c. 430 BCE) is famous for formulating paradoxes that involve limiting processes. In the **paradox of Achilles and the tortoise**, Achilles is in a footrace with the tortoise. Achilles allows the tortoise a head start of 100 metres. Each racer starts running at some constant speed, Achilles faster than the tortoise. After some finite time, Achilles will have run 100 metres, bringing him to the tortoise's starting point. During this time, the tortoise has run a much shorter distance, say 2 metres. It will then take Achilles some further time to run that distance, by which time the tortoise will have advanced farther; and then more time still to reach this third point, while the tortoise move ahead. Thus, whenever Achilles arrives somewhere the tortoise has been, he still has some distance to go before he can even reach the tortoise.
The paradox is resolved by observing that the sum of the terms of an infinite sequence (i.e., the sum of an infinite series—we'll consider series soon) need not be infinite. Unfortunately, Zeno and his contemporaries had no access to such mathematical machinery.
Leucippus (460–4th century BCE), Democritus (c. 460–c. 370 BCE), Antiphon (480–411 BCE), Eudoxus (c. 408–c. 355 BCE), and Archimedes (c. 287–c. 212 BCE) developed the method of exhaustion, which uses an infinite sequence of approximations to determine an area or a volume. Archimedes succeeded in summing what is now called a geometric series.
Isaac Newton (1642–1726/27) dealt with series in his work on *De analysi per aequationes numero terminorum infinitas* (written in 1669, circulated in manuscript, published in 1711), *De methodis serierum et fluxionum* (written in 1671, published in English translation in 1736, Latin original being published much later) and *Tractatus de Quadratura Curvarum* (written in 1693, published in 1704 as an Appendix to

his *Optiks*). In the latter work, Newton considers the binomial expansion of $(x+o)^n$, which he then linearizes by taking the limit as o tends to 0.
In the 18th century, mathematicians such as Leonhard Euler (1701–1783) succeeded in summing some divergent series by stopping at the right moment; they did not much care whether a limit existed, as long as it could be calculated. At the end of the century, Joseph-Louis Lagrange (1736–1813) in his *Théorie des fonctions analytiques* (1797) opined that the lack of rigour precluded further development in calculus. Carl Friedrich Gauss (1777–1855) in his etude of hypergeometric series (1813) for the first time rigorously investigated the conditions under which a series converged to a limit.
The modern definition of a limit (for any ϵ there exists an index N so that...) was given by Bernhard Bolzano (1781–1848) (*Der binomische Lehrsatz*, Prague, 1816, which was little noticed at the time), and by Karl Weierstrass (1815–1897) in the 1870s.
One of the problems with deciding if a sequence is convergent is that you need to have a limit before you can test the definition. Bolzano was the first to spot a way around this problem by using an idea first introduced by the French mathematician Augustin Louis Cauchy (1789–1857).
The sandwich theorem was first used geometrically by the mathematicians Archimedes and Eudoxus in an effort to compute π, and was formulated in modern terms by Gauss.

5.19 Further reading

There are excellent chapters on sequences in [Bin82, WS02, Lie15].
David Applebaum's book *Limits, Limits Everywhere: The Tools of Mathematical Analysis* [App12] is dedicated to the concept of a limit in various guises. It covers limits of sequences in Chapter 4.
Lara Alcock's *How to Think about Analysis* [Alc14] "aims to ensure that no student need be unprepared" for the study of analysis. It is not a textbook containing standard content. Rather, it is designed to be read before arriving at university and/or before starting an Analysis course, or as a companion text once a course is begun. It provides a friendly and readable introduction to the subject by building on the student's existing understanding of six key topics: sequences, series, continuity, differentiability, integrability, and the real numbers. It explains how mathematicians develop and use sophisticated formal versions of these ideas, and provides a detailed introduction to the central definitions, theorems, and proofs, pointing out typical areas of difficulty and confusion and explaining how to overcome these.

6 Continuous Functions

6.1 Limits of Functions

Consider a function $f : [a,b] \to \mathbb{R}$ for real values a and b with $a < b$. We can equally assume the domain of the function to be (a,b), $[a,b)$, $(a,b]$, or $\mathbb{R}$. Comparing this with a sequence $a : \mathbb{N} \to \mathbb{R}$, we see the basic difference that while $\mathbb{N}$ is a countably infinite discrete set, the closed interval $[a,b] \subset \mathbb{R}$ is a continuum, consisting of uncountably many points. We can still consider a generalisation of limit of a sequence to define the limit of a function.

Example 26
Let's take a function $f : [0,1] \to \mathbb{R}$ with $f(x) = \sin 1/x$ if $x \neq 0$ and $f(0) = 0$. Intuitively we know that f has a limit at all points in $(0,1]$ but not at 0.

Definition 17
The function $f : [a,b] \to \mathbb{R}$ has a **limit** *$l \in \mathbb{R}$ at $x_0 \in [a,b]$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that whenever $x \in [a,b]$ and $0 < |x - x_0| < \delta$, then $|f(x) - l| < \epsilon$. We write this as $\lim_{x \to x_0} f(x) = l$.*

The function f in Example 26 has a limit for each $x_0 \in (0,1]$ with $\lim_{x \to x_0} f(x) = \sin 1/x_0$ but f has no limit at $x_0 = 0$.

Exercise 8
For a function $f : [a,\infty) \to \mathbb{R}$, define the notion of the limit $\lim_{x \to \infty} f(x)$. Similarly, define $\lim_{x \to -\infty} f(x)$ for a function $f : (-\infty, b] \to \mathbb{R}$.

Exercise 9
Show that the limit of a function at a point, if it exists, is unique.

Exercise 10
Show that $\lim_{x \to x_0} f(x) = l \in \mathbb{R} \cup \{\infty, -\infty\}$ iff for all sequences $(y_n)_{n \geq 1}$ with $\lim_{n \to \infty} y_n = x_0$ we have $\lim_{n \to \infty} f(y_n) = l$.

The properties of limits of combination of functions are similar to the convergence of combination of sequences.

Proposition 19
Suppose the two functions $f, g : [a,b] \to \mathbb{R}$ have limits $k \in \mathbb{R}$ and $l \in \mathbb{R}$ respectively at $x_0 \in [a,b]$.

- *$f \pm g$ has limit $k \pm l$ at x_0.*
- *The product $f \cdot g$ has limit kl at x_0.*

- *If $l \neq 0$, then f/g has limit k/l at x_0.*

Proof Left as an exercise.

Exercise 11

(i) How can the properties in Proposition 19 be extended to the cases when either l or k or both are ∞ or $-\infty$?

(ii) Extend the statements and proofs of Proposition 19, to functions of type $f, g : [a, \infty) \to \mathbb{R}$ keeping the limits $k, l \in \mathbb{R}$.

(iii) How can the properties in Proposition 19 be extended to functions of type $f, g : [a, \infty) \to \mathbb{R}$ when either l or k or both are ∞ or $-\infty$?

6.2 Continuity

Limits are the key tool for defining that functions are continuous. For a function $f : [a, b] \to \mathbb{R}$ to be continuous at a point $x_0 \in [a, b]$, we require that the limit of f at x_0 be $f(x_0)$.

Definition 18 (Continuity of Functions)
Let $f : [a, b] \to \mathbb{R}$ be a function and x in $\mathbb{R}$.

1. *We say f is* **continuous at** *$x_0 \in [a, b]$ if $\lim_{x \to x_0} f(x) = f(x_0)$.*
2. *We say that f is continuous in $[a, b]$ iff f is continuous at all $x_0 \in [a, b]$.*

Exercise 12
Let $f : [0, 1] \to \mathbb{R}$ be given by $f(x) = x \sin 1/x$ if $x \neq 0$ and $f(0) = 0$. Show that $\lim_{x \to 0} f(x)$ exists and prove that f is continuous at 0.

Let $f(x) = |x|$ be a function defined over the reals. Then this function is continuous at all points, including 0. Now consider the sign function $sgn : \mathbb{R} \to \mathbb{R}$ defined by

$$sgn(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } 0 < x \end{cases} \tag{6.1}$$

Exercise 13 (Continuity of Sign Function)
Show that sgn is continuous at all points except at $x = 0$.

Proposition 20
If the two functions $f, g : [a, b] \to \mathbb{R}$ are continuous at $x_0 \in [a, b]$, then we have:

- *$f \pm g$ is continuous at x_0.*
- *The product $f \cdot g$ is continuous at x_0.*
- *If $g(x_0) \neq 0$, then f/g is continuous at x_0.*

Example 27 (Continuous Functions)
Constant functions and, more generally, polynomials are continuous real functions. The sine and cosine functions $\sin, \cos\colon \mathbb{R} \to \mathbb{R}$ *are continuous but* $\tan$ *is not continuous at* $(2n+1)\pi/2$ *for any integer* $n \in \mathbb{Z}$.

Exercise 14 (Show Continuity of a Function)
Consider the real function $f\colon \mathbb{R} \to \mathbb{R}$ *defined by* $f(x) = 2 + 3x$. *Show that this function is continuous at all points* x *in* $\mathbb{R}$.

Proposition 21

(i) Suppose $f, g : (a,b) \to \mathbb{R}$ *are continuous functions. Then, we have:*

- *For any real number* λ, *the function* $\lambda f : (a,b) \to \mathbb{R}$ *with* $(\lambda f)(x) = \lambda f(x)$ *is continuous.*
- $f + g : (a,b) \to \mathbb{R}$ *with* $(f+g)(x) = f(x) + g(x)$ *and* $fg : (a,b) \to \mathbb{R}$ *with* $(fg)(x) = f(x)g(x)$ *are continuous.*
- *If* $f(x_0) \neq 0$ *for some* $x_0 \in (a,b)$ *then* g/f *with* $(g/f)(x) = g(x)/f(x)$ *is continuous at* x_0.

(ii) Suppose $f : (a,b) \to \mathbb{R}$ *is a function with* $x_0 \in (a,b)$ *and* $g : (c,d) \to \mathbb{R}$ *is a function with* $Im(f) \subset (c,d)$. *If* f *is continuous at* x_0 *and* g *is continuous at* $f(x_0)$, *then the composition* $g \circ f : (a,b) \to \mathbb{R}$ *with* $(g \circ f)(x_0) = g(f(x_0))$ *is continuous at* x_0.

6.2.1 Maxima and Minima

A fundamental property of continuous functions is that they attain their supremum and infimum on any bounded closed interval.

Theorem 22
If $f : [a,b] \to \mathbb{R}$ *is a continuous function with* $a.b \in \mathbb{R}$ *then there exist* $r, s \in [a,b]$ *such that* $f(r) = \sup_{x\in[a,b]} f(x)$ *and* $f(s) = \inf_{x\in[a,b]} f(x)$, *i.e.,* f *has a maximum and a minimum in* $[a,b]$.

Proof First we need to prove that f is bounded on $[a,b]$, which is left as an exercise. Let $M = \sup_{x\in[a,b]} f(x)$ and $m = \inf_{x\in[a,b]} f(x)$: If not then for every $n > 0$ the exists $x_n \in [a,b]$ such that $f(x_n) > n$; find a convergent subsequence of x_n and obtain a contradiction using the continuity of f at the limit of the subsequence. We will find $v_M \in [a,b]$ such that $f(v_M) = M$. For each $n > 0$, by the definition of supremum, there exists $x_n \in [a,b]$ with $M \geq f(x_n) > M - 1/n$. Thus, we construct the sequence $(x_n)_{n\geq 1}$ with the property that $\lim_{n\to\infty} f(x_n) = M$. This sequence is bounded and thus has a convergent subsequence $(x_{n_i})_{i\geq 1}$. Put $v_M := \lim_{i\to\infty} x_{n_i}$. Since $(f(x_{n_i})_{i\geq 0}$ is a subsequence of $(f(x_n))_{n\geq 1}$, it follows that $\lim_{i\to\infty} f(x_{n_i}) = M$. By continuity of f at v_M we have $f(v_M) = \lim_{i\to\infty} f(n_i) = M$. Similarly there exists $v_m \in [a,b]$ such that $f(v_m) = m$.

In particular a continuous function on a closed interval $[a,b]$ is bounded, i.e., there exists $K > 0$ such that $|f(x)| < k$ for all $x \in [a,b]$. If however f is only continuous

in $(a,b]$ then it may not attain its supremum or infimum. For example the function $f:(0,1]\to\mathbb{R}$ with $f(x)=1/x$ does not have a maximum and, in fact, no supremum over real numbers. On the other hand the identity function $g:[0,1)\to\mathbb{R}$ with $g(x)=x$ has a supremum 1 in $[0,1)$ but the supremum is not attained in $[0,1)$. In addition, the theorem is in general false if the interval is unbounded such as $[0,\infty)$. For example the identity map has clearly no maximum or supremum in $[0,\infty)$.

6.2.2 Additional Material: Intermediate Value theorem

A continuous function takes all values between any pair of its values. This property is known as the intermediate value theorem.

Theorem 23
If $f:[a,b]\to\mathbb{R}$ is continuous and $s\in\mathbb{R}$ is such that $f(a)<s<f(b)$, then there exists $c\in(a,b)$ such that $f(c)=s$.

Proof Left as an exercise. **Hint**: Let $A=\{x\in[a,b]:f(x)\le s\}$. Then A is bounded above and non-empty, and thus, has a supremum, c, say. Use the continuity of f at c to show, for example through proof by contradiction, that $f(c)=s$.

Exercise 15
Show that the function $f:[0,\pi]\to\mathbb{R}$ with $f(x)=\tan x$ for $x\neq\pi/2$ fails the intermediate value property no matter how it is defined at $\pi/2$.

6.2.3 Uniform Continuity

Another key property of a continuous function on a closed bounded interval $[a,b]$ is the it is **uniformly** continuous. We say that $f:A\to\mathbb{R}$ is **uniformly continuous** on A if for each $\epsilon>0$ there exists $\delta>0$ such that for all $x,x_0\in A$ we have: $|f(x)-f(x_0)|<\epsilon$ if $|x-x_0|<\delta$. In other words $\delta>0$ is independent of $x_0\in A$.

Example 28
The function $f:[a,b]\to\mathbb{R}$ with $f(x)=x^2$ is uniformly continuous. In fact, let $\epsilon>0$ be given. Then $|f(x)-f(x_0)|=|x^2-x_0^2|=|x-x_0||x+x_0|\le 2M|x-x_0|$ where $M=\max(|a|,|b|)$. Thus, for $\delta=\epsilon/(2M)$ we get $|x-x_0|<\delta$ implies $|x^2-x_0^2|<\epsilon$.

Exercise 16
Show that the function $f:(0,1]\to\mathbb{R}$ with $f(x)=1/x$ is not uniformly continuous in $(0,1]$.

Theorem 24
If $f:[a,b]\to\mathbb{R}$, for $a,b\in\mathbb{R}$, is continuous, then it is uniformly continuous on $[a,b]$, i.e., for each $\epsilon>0$ there exists $\delta>0$ such that for all $x,x_0\in[a,b]$ we have $|f(x)-f(x_0)|<\epsilon$ if $|x-x_0|<\delta$.

Proof Suppose, for a contradiction, that f is not uniformly continuous on $[a,b]$. Then, there exists $\epsilon>0$ such that for all $n\ge1$ there exists $x_n,y_n\in[a,b]$ with $|x_n-y_n|<1/n$ but $f(x_n)-f(y_n)|\ge\epsilon$. Since $(x_n)_{n\ge1}$ is a bounded sequence, it has a convergent

subsequence $(x_{n_i})_{i\geq 1}$ with $l := \lim_{i\to\infty} x_{n_i} \in [a,b]$. But then $|l-y_{n_i}| \leq |l-x_{n_i}|+|x_{n_i}-y_{n_i}| < |l-x_{n_i}| + 1/n_i$. Thus, y_{n_i} also converges to l. By the continuity of f at l we have $\lim_{i\to\infty} f(x_{n_i}) = f(l) = \lim_{i\to\infty} f(y_{n_i})$. But this contradicts $|f(x_{n_i}) - f(y_{n_i})| \geq \epsilon$ for all $i \geq 1$. □

Intuitively, you can see, by considering the function $f : (0,1] \to \infty$ with $f(x) = 1/x$, why uniform continuity can fail if the function is continuous on an interval such as $(0,1]$ that is not closed: As x_0 gets close to 0, the difference $|f(x)-f(x_0)|$ can become arbitrary large no matter how close x is to x_0, i.e., no matter how small a value for $\delta > 0$ you choose in $|x-x_0| < \delta$.

7 Integration

7.1 Introduction

We first review the notion of definite integrals that are taught in school. In Figure 7.1, we see the graph of the function $f(x) = x^2$ and the shaded area below the graph of the function and between the lines $y = 2$ and $x = 0$. The size of this area is the intuitive meaning of the definite integral

$$\int_0^2 x^2\,dx \tag{7.1}$$

This intuition is helpful but the definite integral may also contain portions of a graph below the x-axis and then such areas are negative. For example, we have that

$$\int_{-1}^{1} x^3\,dx = \int_{-1}^{0} x^3\,dx + \int_{0}^{1} x^3\,dx = (-1/4) + (1/4) = 0 \tag{7.2}$$

7.2 Lower and Upper Sums, Riemann Sums

To formalise the definition of the integral of a function $f : [a,b] \to \mathbb{R}$, we start by defining the notion of a partition of $[a,b]$ and the corresponding lower and upper sums and Riemann sums with respect to a partition.

Definition 19

- *A* **partition** P *of* $[a,b]$ *is given by a finite set*

  $$P = \{r_i : 0 \le i \le n-1, a = r_0, b = r_n, r_i < r_{i+1}\}$$

  of points in $[a,b]$ *that includes the end points* a *and* b*. We can represent it simply as*

  $$P: \qquad a = r_0 < r_1 < \cdots, < r_i < \cdots, r_{n-1} < r_n = b$$

- *Each closed interval* $[r_i, r_{i+1}]$*,for* $0 \le i \le n-1$*, is called a* **subinterval** *of* P*.*

- *The* **norm** *of* P *is defined as*

  $$\|P\| = \max\{r_{i+1} - r_i : 0 \le i \le n-1\},$$

  i.e., the largest length of the subintervals in P*.*

- *If* P_1 *and* P_2 *are partitions of* $[a,b]$*, we say* P_2 **refines** P_1 *if* $P_1 \subset P_2$*.*

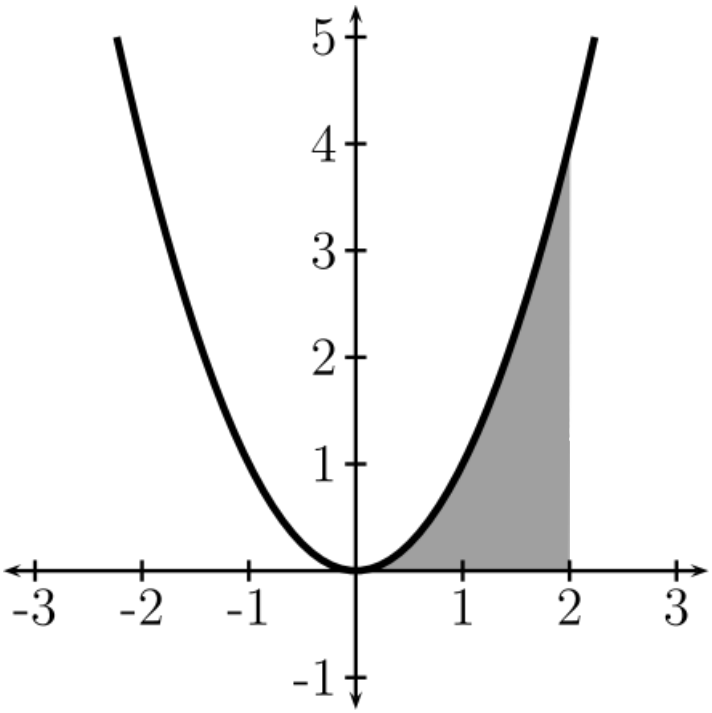


Figure 7.1: An image from Wikipedia that illustrates in the shaded area the value of the definite integral $\int_0^2 x^2\,dx$

Definition 20

- *Given a function* $f : [a,b \to \mathbb{R}$ *and a partition of* $[a,b]$ *given by*

$$P: \qquad a = r_0 < r_1, < \cdots, < r_i < \cdots, r_{n-1} < r_n = b$$

the **Lower sum** $L(f,P)$ *and the* **upper sum** $U(f,P)$ *of* f *wrt* P *are defined as:*

$$L(f,P) = \sum_{i=0}^{n-1}(r_{i+1}-r_i)\inf_{x\in[r_i,r_{i+1}]} f(x), \qquad U(f,P) = \sum_{i=0}^{n-1}(r_{i+1}-r_i)\sup_{x\in[r_i,r_{i+1}]} f(x)$$

- *For any choice of* $s_i \in [r_i, r_{i+1}]$ *for* $0 \le i \le n-1$, *the sum*

$$S(f,P,(s_i)_{0\le i\le n-1}) = \sum_{i=0}^{n-1}(r_{i+1}-r_i)f(s_i)$$

is called a **Riemann sum** *for* P.

Note that if f is actually continuous on $[a,b]$ we can replace sup and inf in the above definition for the upper sum and lower sum by max and min respectively. Clearly, for any partition we have $L(f,P) \le S(f,P,(s_i)_{0\le i\le n-1}) \le U(f,P)$. In addition, as we refine a partition, the lower sum increases while the upper sum decreases:

Exercise 17
If $P_1 \subset P_2$ *then* $L(f,P_1) \le L(f,P_2) \le U(f,P_2) \le U(f,P_1)$.

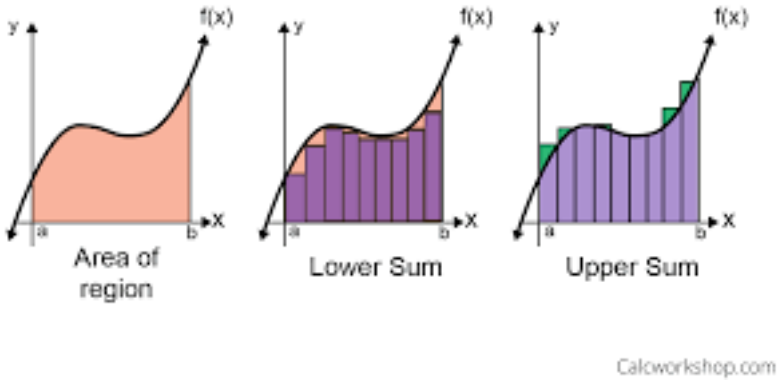


Figure 7.2: Lower and upper sums for the function on the left hand side

This gave the German mathematician Bernard Riemann in the 19th century the key idea to define the notion of what we now call the **Riemann integral** of f.

Definition 21

- *The* **lower** *and* **upper** *integrals of* $f : [a,b] \to \mathbb{R}$ *are defined as*

$$\underline{\int_a^b} f(x)\,dx = \sup_P L(f,P), \qquad \overline{\int_a^b} f(x)\,dx = \inf_P U(f,P)$$

- *We say* f *is Riemann integrable if* $\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx$ *and the common value is called the Riemann integral of* f *written as* $\int_a^b f(x)\,dx$.

Using the definitions of infimum and supremum one can show the following:

Theorem 25

- *A bounded function* $f : [a,b] \to \mathbb{R}$ *is Riemann integrable with Riemann integral* $c \in \mathbb{R}$ *iff for each* $\epsilon > 0$ *there exists a partition* P *of* $[a,b]$ *with* $c - L(f,P) < \epsilon$ *and* $U(f,P) - c < \epsilon$.

- *A bounded function* $f : [a,b] \to \mathbb{R}$ *is Riemann integrable with Riemann integral* $c \in \mathbb{R}$ *iff for each* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that for all partitions* P *of* $[a,b]$ *with* $\|P\| < \delta$ *we have* $|S(f,P,(s_i)_{0\le i\le n-1}) - c| < \epsilon$.

Using uniform continuity of continuous functions and the fact that a continuous function attains its supremum and infimum on any closed bounded subinterval, we can show that any continuous function on$[a,b]$ is Riemann integrable.

Theorem 26 (Riemann Integral of Continuous Functions)
Let $f : [a,b] \to \mathbb{R}$ *be a function that is continuous on the interval* $[a,b]$. *Then the Riemann integral*

$$\int_a^b f(x)\,dx$$

exists.

Proof Let $\epsilon > 0$ be given. There exists, by uniform continuity of f on $[a,b]$ (and using $\epsilon/(2(b-a))$ in the definition of uniform continuity), some $\delta > 0$ such that $|f(x)-f(y)| < \epsilon/(b-a)$ for $x, y \in [a,b]$ with $|x-y| < \delta$. Thus, if $\|P\| < \delta$ then in each given subinterval (i.e., each $[r_i, r_{i+1}]$ with $0 \le i \le n-1$) of P we have $f(x_M) - f(x_m) < \epsilon/(2(b-a))$ where x_M, x_m are the points of the subinterval where f attains its maximum and minimum respectively in the subinterval in question. Hence

$$U(f,P) - L(f,P) = \sum_{i=0}^{n-1} (r_{i+1} - r_i)\Big(\sup_{x\in[r_i,r_{i+1}]} f(x) - \inf_{x\in[r_i,r_{i+1}]} f(x)\Big)$$

$$= \sum_{i=0}^{n-1} (r_{i+1}-r_i)\Big(\max_{x\in[r_i,r_{i+1}]} f(x) - \min_{x\in[r_i,r_{i+1}]} f(x)\Big) < \sum_{i=0}^{n-1} (r_{i+1}-r_i)\epsilon/(2(b-a)) = (b-a)\epsilon/(2(b-a)) = \epsilon/2.$$

Thus, $\overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx \le \epsilon/2 < \epsilon$. Since $\epsilon > 0$ can be arbitrary small, it follows that $\overline{\int_a^b} f(x)\,dx = \underline{\int_a^b} f(x)\,dx$ as required. □

Exercise 18
Show that if a bounded function $f : [a,b] \to \mathbb{R}$ has only a finite set of discontinuities then it is Riemann integrable.

In fact, it can be shown that a bounded function with only a countable set of discontinuities on $[a,b]$ is Riemann integrable.

Example 29 (Analytical Derivation of Riemann Integral)
Consider the continuous function $f : [0,2] \to \mathbb{R}$ with $f(x) = x^2$. For the n subintervals of $[0,2]$ we have $\Delta_n = (2-0)/n = 2/n$. Using the right endpoints of these subintervals for sampling, we get the Riemann sum[1]

$$\begin{aligned} \sum_{i=1}^{n} f(\hat{x}_i)\cdot\Delta_n &= \sum_{i=1}^{n} \Big(\frac{2i}{n}\Big)^2 \cdot \frac{2}{n} \\ &= \frac{8}{n^3}\cdot(1^2 + 2^2 + \cdots + n^2) \\ &= \frac{8}{n^3}\cdot\frac{n\cdot(n+1)\cdot(2n+1)}{6} \\ &= \frac{8}{3} + \frac{4}{n} + \frac{4}{3n^2} \end{aligned} \tag{7.3}$$

If we then take the limit of the latter expression as Δ_n tends to 0 (that is as n tends to ∞), we have

$$\lim_{\Delta_n\to 0} \frac{8}{3} + \frac{4}{n} + \frac{4}{3n^2} = \lim_{\Delta_n\to 0} \frac{8}{3} + \frac{4}{n} + \frac{4}{3n^2} = \frac{8}{3} \tag{7.4}$$

One can see that this limit does not depend on where in the subintervals we sample and so we get that

$$\int_0^2 x^2\,dx = \frac{8}{3} \tag{7.5}$$

[1]Note that the sum began at $i = 1$ to accommodate the rightmost endpoint. Alternatively, you could sum from $i = 0$, but then $2i$ would need to be changed to $2\cdot(i+1)$.

Some additional and useful properties of the Riemann integral are given below.

Theorem 27

- $\int_a^b sf(x)+tg(x)\,dx = s\int_a^b f(x)\,dx + t\int_a^b g(x)\,dx$ *if* f *and* g *are integrable.*
- $\int_a^b c\,dx = c(b-a)$*, where* $c \in \mathbb{R}$ *is a constant.*
- $\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$ *whenever the integrals exist for* $a < b < c$*.*
- *If* $f(x) \geq 0$ *for* $x \in [a,b]$ *then* $\int_a^b f(x)\,dx \geq 0$ *if the integral exists.*

Exercise 19
Show that

$|\int_a^b f(x)\,dx| \leq \int_a^b |f(x)|\,dx$ *when the integrals exist.*

7.2.1 Improper Riemann Integral

Consider a function of type $f : [a,\infty) \to \mathbb{R}$, respectively $f : [a,b) \to \mathbb{R}$. We say that f has improper Riemann integral (or that the integral converges) if the limit $\lim_{x\to\infty}\int_a^x f(y)\,dy$, respectively $\lim_{x\to b}\int_a^x f(y)\,dy$, exists as real numbers. If the limit does not exist or it exists and is $\pm\infty$, then we say that the improper Riemann integral **diverges**.

Example 30
The improper integral $\int_1^\infty 1/x^2\,dx = 1 - \lim_{x\to\infty} 1/x = 1$ *exists but* $\int_1^\infty 1/x\,dx = \lim_{x\to\infty}\ln x = \infty$ *diverges. Similarly* $\int_0^1 1/\sqrt{x}\,dx = 2 - 2\lim_{x\to 0}\sqrt{x} = 2$ *exists but* $\int_0^1 1/x\,dx$ *diverges.*

.

8 Series

Series are formal infinite sums of real numbers:

$$\sum_{i=1}^{\infty} a_i \tag{8.1}$$

Generally, it is not clear whether or under which circumstances such sums have a meaning in that they actually represent another real number. And whether they converge or not has many important applications, including the assurance that iterative numerical algorithms terminate.
We can say that such a series is the formal summation of all elements from a sequence $(a_n)_{n\geq 1}$ that determines its summands as well as the order in which they are summed up.
It turns out that the key to understanding such series is to derive another sequence of real numbers from them, and this is explained best through the concept of **partial sums**.

Partial sums Consider a series of the form given in (8.1). We can associate a sequence $(S_n)_{n\geq 1}$ of real numbers to that series where for each $n \geq 1$ the real number S_n is defined as the partial sum

$$S_n = \sum_{i=1}^{n} a_i \tag{8.2}$$

of the first N summands in that series. This gives us two transformations, one that maps a sequence $(a_n)_{n\geq 1}$ to its series, and one that maps series to a sequence of its partial sums:

$$(a_n)_{n\geq 1} \quad \mapsto \quad \sum_{i=1}^{\infty} a_i \quad \mapsto \quad (S_n)_{n\geq 1} \qquad \text{where } S_n = \sum_{i=1}^{n} a_i \text{ for all } n \geq 1 \tag{8.3}$$

The transformations in (8.3) are useful in several ways. For one, we will use the second transformation to formally define next when a series converges. For another, the limit behavior of the summands a_n itself can help us with understanding whether its series (the result of the first transformation) diverges. We can now define what it means for such a series to converge.

Definition 22 (Convergence and Divergence of Series)
Let $\sum_{i=1}^{\infty} a_i$ be the series determined by the sequence $(a_n)_{n\geq 1}$. Then:

1. *Series $\sum_{i=1}^{\infty} a_i$ has* **limit** *l in $\mathbb{R}$ or* **converges** *to l in $\mathbb{R}$ iff its associated sequence of partial sums $(S_n)_{n\geq 1}$ as defined in (8.2) has limit l.*

2. *Series* $\sum_{i=1}^{\infty} a_i$ *diverges if it does not converge to some* l *in* $\mathbb{R}$.

Observe the resulting convention that if the sequence of partial sums converges to ∞ or $-\infty$ in the extended real line $\mathbb{R} \cup \{\infty, -\infty\}$ then, by definition, the series diverges.

Convergence and being bounded above When $a_i \geq 0$ for all $i \geq 1$, then the partial sum $(S_n)_{n\geq 1}$ as defined in (8.2) is **increasing**, that is:

$$S_1 \leq S_2 \leq S_3 \leq \cdots$$

From Theorem 17, we then know that $(S_n)_{n\geq 1}$ (and therefore the associated series $\sum_{i=1}^{\infty} a_i$) has a limit whenever that sequence is bounded above.

Summands of converging series tend to 0 Whenever the series converges, we show below that the sequence $(a_n)_{n\geq 1}$ of summands does converges to 0. Therefore, for a series to have any chance of converging, the sequence $(a_n)_{n\geq 1}$ has to have limit 0.
The converse is not true, the series can diverge even when the sequence $(a_n)_{n\geq 1}$ of summands does converge to 0, we will cover such an example further below. This also means that we will need to develop a deeper understanding of the convergence of series since we cannot simply reduce it to the understanding of convergence of the sequences of summands.
We state and prove the discussed necessary condition for a series to converge:

Lemma 2
Suppose that the series of real numbers $\sum_{i=1}^{\infty} a_i$ *converges. Then the sequence of summands has limit* 0*:* $\lim_{n\mapsto\infty} a_n = 0$.

Proof Let $\epsilon > 0$ be given. We have to find $N \in \mathbb{N}$ such that for $n > N$ we have $|a_n| < \epsilon$. Since the sequence $(S_n)_{n\geq 1}$ converges to a limit in $\mathbb{R}$, it is a Cauchy sequence. Thus, there exists $N_0 \in \mathbb{N}$ such that for $n, m > N_0$ we have $|S_m - S_n| < \epsilon$. In particular, put $m = n + 1$ to have $|S_{n+1} - S_n| < \epsilon$ for $n > N_0$, i.e., $|a_{n+1}| < \epsilon$ for $n > N_0$. Now let $N = N_0 + 1$ to get $|a_n| < \epsilon$ for $n > N$. □

We stress that $\lim_{n\to\infty} a_n = 0$ is a necessary but not a sufficient condition for $\sum_{i=1}^{\infty} a_i$ to converge. To see this, we note that $\lim_{n\to\infty} 1/n = 0$ but we will show below that the series $\sum_{i=1}^{\infty} 1/n$ diverges!

The tail of a series Since the convergence of a series is defined in terms of the convergence of its partial sums, we know that this convergence does not depend on any initial, finite part of that sequence of partial sums. In other words, this does not depend on any initial, finite set of summands $a_1, \ldots, a_N$ for any fixed N in $\mathbb{N}$. In particular, a series $S = \sum_{n=1}^{\infty} a_n$ converges iff $\sum_{n=10}^{\infty} a_n$ converges or $\sum_{n=1000}^{\infty} a_n$

converges or, in general, iff $\sum_{n=N}^{\infty} a_n$ converges for any natural number N. This explains why, in comparison tests, we will also only have to take into account later summands in the series from a certain point N onward, rather than from the whole series.

8.1 Geometric Series

This is an example of the use of partial sums to determine the convergence of a series. A common and well-known example is the *geometric series*

$$G(x) = \sum_{n=1}^{\infty} x^n \tag{8.4}$$

where x is a real constant. If the limit $G(x)$ for this series does exist, we might reason that

$$G(x) = x + \sum_{n=2}^{\infty} x^n = x + x \cdot \sum_{n=1}^{\infty} x^n = x + x \cdot G(x)$$

Therefore, we would have

$$G(x) = \frac{x}{1-x} \tag{8.5}$$

But this is an informal reasoning that extended our composition results for limits to infinite sums. So how can we more formally determine under which conditions $G(x)$ does exist? By definition, this is exactly when the sequence of its partial sums converges. Let us write $G_n(x)$ for the partial sum $S_n(x) = \sum_{i=1}^{n} x^i$ for geometric series. In this case, we can find a closed expression of the nth partial sum $G_n(x)$ of the geometric series as follows:

$$\begin{aligned} G_n(x) &= x + \sum_{i=2}^{n} x^i && (8.6)\\ &= x + x \cdot \sum_{i=1}^{n-1} x^i && (8.7)\\ &= x + x \cdot (G_n(x) - x^n) && (8.8) \end{aligned}$$

from which we get by elementary algebra that

$$G_n(x) = \frac{x - x^{n+1}}{1-x} = \frac{x}{1-x} - \frac{1}{1-x} \cdot x^{n+1} \tag{8.9}$$

From equation (8.9) and the composition results for limits, we learn that the sequence $(G_n)_{n\geq 1}$ converges iff the non-constant summand x^{n+1} of $G_n(x)$ converges. This we know to be the case iff $|x| < 1$, and then $\lim_{n\to\infty} x^{n+1} = 0$. Again, by the composition results for limits we then have that

$$\lim_{n\to\infty} G_n = \frac{x}{1-x} \tag{8.10}$$

Similarly, for $|x| > 1$, we have that $(x^{n+1})_{n\geq 1}$ and so also $(G_n(x))_{n\geq 1}$ diverges.
And what about the boundary case $x = 1$? In that case, $G_n(x)$ equals n and so $(G_n(x))_{n\geq 1}$ also diverges. We summarise:

Theorem 28 (Geometric Series Converges)
For the convergence behavior of the geometric series, we have that $\sum_{i=1}^{\infty} x^i$ *converges iff* $|x| < 1$, *in which case its limit equals* $\frac{x}{1-x}$.

8.2 Harmonic Series

The harmonic series is given by

$$S = \sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots \tag{8.11}$$

We will now informally show that this series diverges to ∞. This is an important result. For one, it implies that the converge of the sequence of summands to 0 is not sufficient for a series to converge. For another, the divergence of the harmonic series can be used to prove the divergence of many other series through comparison tests covered below.
Let us now see why the harmonic series diverges. By grouping summands in the series, we can see this intuitively:

$$S = 1 + \frac{1}{2} + \underbrace{\left(\frac{1}{3} + \frac{1}{4}\right)}_{>\frac{1}{4}+\frac{1}{4}} + \underbrace{\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)}_{>\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}} + \cdots > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

So if we consider the partial sums at powers of two, we get that

$$S_{2^n} = \sum_{i=1}^{2^n} \frac{1}{i} > 1 + \frac{n}{2} \tag{8.12}$$

This means that the strictly increasing sequence $(S_n)_{n\geq 1}$ is not bounded above. But then this sequence cannot have a limit. For if l were such a limit, then consider $n = 2 \cdot \lceil l \rceil$. From (8.12), we then get that $S_{2^m} > l$ for all $m \geq n$ as the sequence is increasing. But this contradicts the convergence of that sequence to l, choose $\epsilon = S_{2^n} - l$ to see this. By definition, this divergence means that the Harmonic Series diverges.

8.3 Series of Inverse Squares

The **series of inverse squares** is defined as

$$S = \sum_{i=1}^{\infty} \frac{1}{i^2} \tag{8.13}$$

We will now show that this series converges. Similarly to the divergence of the harmonic series, this result is important since it allows us to show the convergence of other series by comparison.
In order to understand the convergence of the series in (8.13), we will first study another series that we can use to bound the series of inverse squares from below and above. That series is:

$$T = \sum_{i=1}^{\infty} \frac{1}{i(i+1)}$$

Using partial fractions, we can rewrite

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

The nth partial sum T_n of T can therefore be written as:

$$T_n = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i+1} \right) = 1 - \frac{1}{n+1} \tag{8.14}$$

(To see why the last equation holds, try writing out the first and last few summands of the sum and observe which ones cancel out). From this, we see that $(T_n)_{n\geq 1}$ converges and converges to limit $T = 1$.
We will use this result to reason about the convergence or divergence of the series of inverse squares in (8.13). We start by considering summands of the partial sum

$$S_n = \sum_{i=1}^{n} \frac{1}{i^2}$$

and we notice that:

$$\frac{1}{i(i+1)} < \frac{1}{i^2} < \frac{1}{(i-1)i} \qquad \text{for } i \geq 2$$

We sum from $i = 2$ and not from $i = 1$ (to avoid a 0 denominator on the right hand side) to n to get:

$$\frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \cdots + \frac{1}{n(n+1)} < \sum_{i=2}^{n} \frac{1}{i^2} < \frac{1}{2\cdot 1} + \frac{1}{3\cdot 2} + \frac{1}{4\cdot 3} + \cdots + \frac{1}{n(n-1)}$$

We note that the left hand and right hand series above differ only in the first and last summands, so can be rewritten as:

$$\left(\sum_{i=1}^{n} \frac{1}{i(i+1)} \right) - \frac{1}{2} < \sum_{i=2}^{n} \frac{1}{i^2} < \sum_{i=1}^{n-1} \frac{1}{i(i+1)}$$

Now we add 1 across all terms of these two inequalities to allow us to sum $\frac{1}{i^2}$ from 1 and get the required partial sum S_n as our middle expression:

$$\frac{1}{2} + \sum_{i=1}^{n} \frac{1}{i(i+1)} < S_n < 1 + \sum_{i=1}^{n-1} \frac{1}{i(i+1)}$$

This effectively bounds the sequence of partial sums, S_n. We use our result of (8.14) to get:

$$\frac{3}{2} - \frac{1}{n+1} < S_n < 2 - \frac{1}{n} \qquad \text{for } n \geq 2$$

Since the sequence of partial sums is increasing and bounded above, we know by Theorem 17 that this sequence converges to $\sup\{S_n \mid n \geq 2\}$ and that this limit is ≤ 2 as the limit is the least upper bound.

Exercise 20 (Continuity at Bounds)
Show that the limit of the series of inverse squares is greater or equal to $3/2$.

Therefore, the limit l of the series of inverse squares satisfies $3/2 \leq l \leq 2$. In the year 1650, the question was first raised what the exact value of l is. It was not before 1734 that the mathematician Euler proved that this value of l is $\pi^2/6$.

8.4 Common Series and Convergence

As with sequences, it is extremely useful to have a good knowledge of whether some common series converge or not. Such knowledge will be very helpful in comparison tests for both convergence and divergence arguments.

Diverging series We recall Lemma 2, that the sequence $(a_n)_{n\geq 1}$ has limit 0 whenever the series $S = \sum_{n=1}^{\infty} a_n$ converges. So one way of showing divergence of a series is to show that the limit of the sequence of its summands either does not exist or does not equal 0. This can be applied to the third sequence below:

1. Harmonic series: $S = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.
2. Harmonic primes: $S = \displaystyle\sum_{p \,:\, \text{prime}} \frac{1}{p}$ diverges.
3. Geometric series: $S = \displaystyle\sum_{n=1}^{\infty} x^n$ diverges for $|x| \geq 1$.

Note that the divergence of the second sequence above is an even stronger result than the divergence of the harmonic series: we already have divergence when we add up the inverses of prime numbers only.

Converging series We already mentioned the value of the limit for the inverse squares series in passing and now state it for the record. While a proof that this is indeed the limit is of interest, it is beyond the scope of this course module.

1. Geometric series: $S = \displaystyle\sum_{n=1}^{\infty} x^n$ converges to $\dfrac{x}{1-x}$ for all x in $\mathbb{R}$ with $|x| < 1$.

2. Inverse squares series: $S = \sum_{n=1}^{\infty} \frac{1}{n^2}$ converges to $\frac{\pi^2}{6}$.

3. $\frac{1}{n^c}$ series: $S = \sum_{n=1}^{\infty} \frac{1}{n^c}$ converges for all c in $\mathbb{R}$ with $c > 1$.

Exercise 21
Consider the **alternating series** $S = a_1 - a_2 + a_3 - a_4 + \cdots + (-1)^{n+1} a_n \cdots$ *such that* $a_n \geq 0$ *and* $a_{n+1} \leq a_n$ *for all* $n \geq 1$ *with* $\lim_{n\to\infty} a_n = 0$. *Show that the series* S *converges by proving the following statements:*

(i) The sequence $(S_{2n})_{n\geq 1}$ *of even partial sums is an increasing sequence with* $S_{2n} \leq a_1$ *for* $n \geq 1$. *Deduce that* $\lim_{n\to\infty} S_{2n} = \ell$ *exists.*

(ii) Show that $S_{2n+1} \to \ell$ *as* $n \to \infty$.

(iii) Conclude that $S_n \to \ell$ *as* $n \to \infty$.

8.5 Convergence Tests for Series of Positive Terms

In this section, we assume that the summands of a series are positive. In particular, for the convergence tests covered below, the sequence of partial sums is strictly increasing. This means that either the sequence S_n of partial sums is bounded above in which case the series is convergent or the sequence S_n is not bounded above in which case $S_n \to \infty$ as $n \to \infty$ and the series diverges.
In formulating these comparison tests, we use the following notational conventions:

1. $a_i > 0$ denotes a summand in the series $\sum_{i=1}^{\infty} a_i$ whose convergence or divergence we wish to establish,

2. $\sum_{i=1}^{\infty} c_i$ denotes a series with $c_i > 0$ for which we already have established convergence to the sum c, and

3. $\sum_{i=1}^{\infty} d_i$ denotes a series with $d_i > 0$ for which we have already established its divergence.

The first comparison test uses a constant λ in $\mathbb{R}$ for scaling the summands of the sequence with which we do the comparison. A parameter N in $\mathbb{N}$ is used to specify that this comparison is allowed to fail on an initial, finite part of the series:

Lemma 3 (Comparison Test)
Let $\lambda > 0$ *and* $N \in \mathbb{N}$. *Further, let* $\sum_{i=1}^{\infty} c_i$ *be a converging series and* $\sum_{i=1}^{\infty} d_i$ *a diverging series. Then we have:*

1. *If* $a_i \leq \lambda c_i$ *for all* $i > N$, *then* $\sum_{i=1}^{\infty} a_i$ *converges*

2. *If* $a_i \geq \lambda d_i$ *for all* $i > N$, *then* $\sum_{i=1}^{\infty} a_i$ *diverges*

Sometimes, it is easier to not have to rely on a scaling constant λ but to rather consider a quotient between summands of the compared series:

Lemma 4 (Limit Comparison Test)
As in the previous lemma, let us suppose that $\sum_{i=1}^{\infty} c_i$ *is a converging series and that* $\sum_{i=1}^{\infty} d_i$ *is a diverging series. Then we have:*

1. *If* $\lim_{i\to\infty} \frac{a_i}{c_i} \in \mathbb{R}$ *exists, then* $\sum_{i=1}^{\infty} a_i$ *converges*

2. *If* $\lim_{i\to\infty} \frac{d_i}{a_i} \in \mathbb{R}$ *exists, then* $\sum_{i=1}^{\infty} a_i$ *diverges*

The D'Alembert ratio test is a very useful test for quickly determining the convergence or divergence of a series. It exists in many subtly different forms, and we offer two of these here:

Lemma 5 (D'Alembert's Ratio Test)
Let N *be in* $\mathbb{N}$. *Furthermore, we consider a series* $\sum_{i=1}^{\infty} a_i$ *for which we mean to determine whether it converges or diverges. Then we have:*

1. *If* $\frac{a_{i+1}}{a_i} \geq 1$ *for all* $i \geq N$, *then* $\sum_{i=1}^{\infty} a_i$ *diverges.*

2. *If there exists some* k *in* $\mathbb{R}$ *with* $k < 1$ *such that* $\frac{a_{i+1}}{a_i} \leq k$ *for all* $i \geq N$, *then* $\sum_{i=1}^{\infty} a_i$ *converges.*

The above test can be used to prove the correctness of an easier-to-use and more often quoted version of the test, which looks at the limit of the ratio of successive summands:

Lemma 6 (D'Alembert Limit Ratio Test)
As in the previous lemma, we consider a series $\sum_{i=1}^{\infty} a_i$ *for which we mean to determine whether it converges or diverges. Suppose that* $\lim_{i\to\infty} \frac{a_{i+1}}{a_i}$ *exists. Then we have:*

1. *If* $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} > 1$, *then* $\sum_{i=1}^{\infty} a_i$ *diverges.*

2. *If* $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} = 1$, *then* $\sum_{i=1}^{\infty} a_i$ *may converge or diverge, so the test is inconclusive.*

3. *If* $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} < 1$, *then* $\sum_{i=1}^{\infty} a_i$ *converges.*

Note that this test will not always succeed: in the second case above, its outcome is inconclusive. And this is necessary as the series may indeed converge or diverge given that information, which we now show:

- The series whose summands a_n equal 1 for all $n \geq 1$ clearly diverges but its limit of successive summands equals 1.
- On the other hand, the series whose summands a_n equal $\frac{1}{n^2}$ for $n \geq 1$ converges, yet the limit of successive summands also has limit 1.

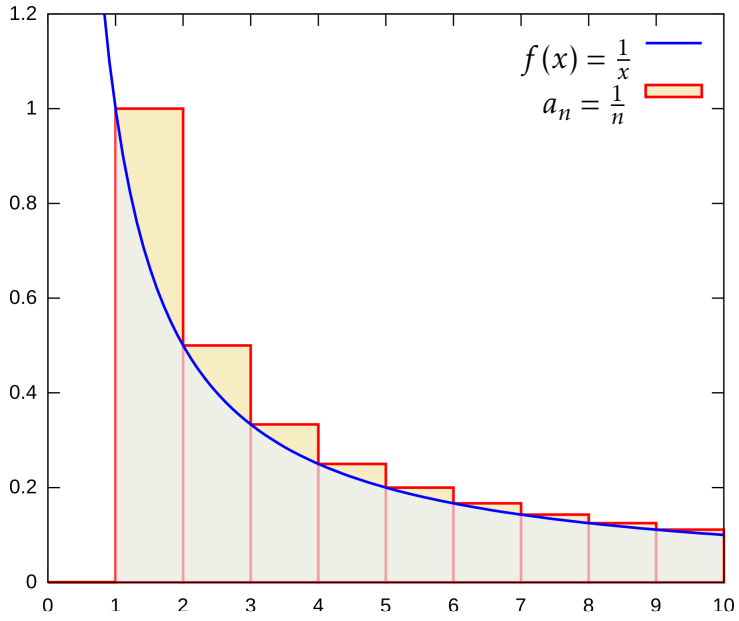


Figure 8.1: A divergence test: the series $\sum_{n=1}^{\infty} \frac{1}{n}$ as compared to $\int_1^{\infty} \frac{1}{x}\,\mathrm{d}x$, showing that the series is bounded below by the integral.

There are also comparison tests that make use of integrals, assuming that the underlying functions are integrable. We will be concerned with improper integrals of the form

$$\int_1^{\infty} f(x)dx \tag{8.15}$$

where $f : \mathbb{R} \to \mathbb{R}$ is continuous on the positive reals and strictly decreasing. Improper integrals cannot be formally defined through Riemann sums, but we will assume that

we can compute then by first computing a definite integral

$$a_y = \int_1^y f(x)dx \tag{8.16}$$

with an upper bound y as formal parameter. The resulting integral value a_y will then be a function of y for which we can compute the limit as y tends to infinity. You may think of y as an index ranging over the natural numbers so that the notion of limit is the one we studied for sequences already. The sequence $(a_y)_{y\geq 1}$ is then increasing and so converges if it is bounded above. In that case, its limit is then defining the value of the improper integral in (8.15), which is then said to converge to that limit. Otherwise, the limit is not defined and the improper integral in (8.15) is said to diverge.
We can now use this somewhat informal definition to reason about the convergence or divergence of a series $\sum_{i=1}^{\infty} a_i$ where a_i equals $f(i)$ for all $i \geq 1$ and f is a continuous, decreasing function on the domain $[1, \infty)$.

Lemma 7 (Integral Test)
Let $f\colon \mathbb{R} \to \mathbb{R}^+$ be a function that is continuous and decreasing and positive on the interval $[1, \infty)$. Let $a_n = f(n)$ for all n in $\mathbb{N}$. Then we have:

1. *If $\int_N^{\infty} f(x)\,\mathrm{d}x$ converges, then the series $\sum_{i=1}^{\infty} a_i$ converges as well.*
2. *If $\int_N^{\infty} f(x)\,\mathrm{d}x$ diverges, then the series $\sum_{i=1}^{\infty} a_i$ diverges as well.*

Example: Integral test divergence We can apply the integral test of Lemma 7 to show that the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. Figure 8.1 shows $\sum_{n=1}^{\infty} \frac{1}{n}$ as a sum of $1 \times a_n$ rectangular areas for $n \geq 1$. When displayed like this, it is intuitive that the area covered by the series is no larger than the area beneath the graph of the function. Formally, if the indefinite integral exists, then that series converges. Let us therefore, evaluate this indefinite integral:

$$\begin{aligned}\int_1^{\infty} f(x)\,\mathrm{d}x &= \lim_{b\to\infty} \int_1^b \frac{1}{x}\,\mathrm{d}x \\ &= \lim_{b\to\infty} [\ln x]_1^b \\ &= \lim_{b\to\infty} (\ln b - \ln 1) \\ &= \lim_{b\to\infty} \ln b\end{aligned}$$

which diverges as the logarithm is a strictly increasing function whose image is not bounded above. Therefore, by Lemma 7 we know that the series $\sum_{n=1}^{\infty} a_n$ must also diverge.

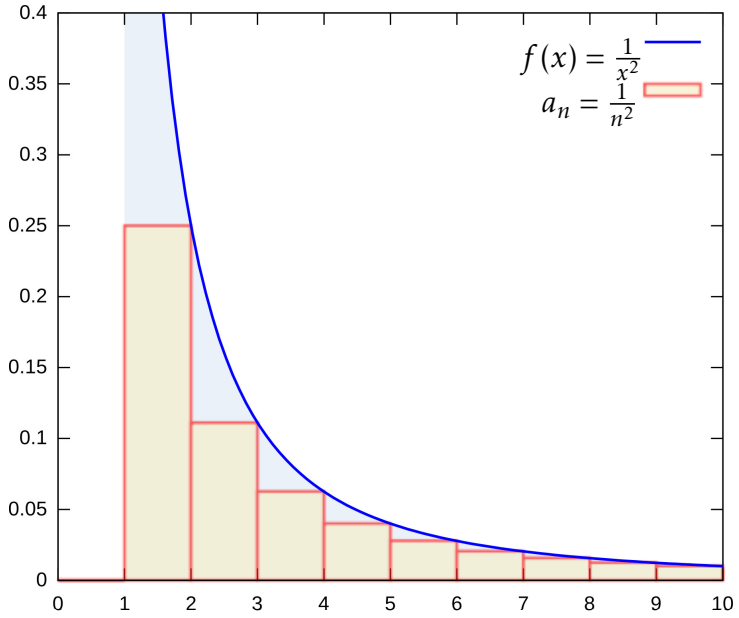


Figure 8.2: A convergence test: the series $\sum_{n=2}^{\infty} \frac{1}{n^2}$ as compared to $\int_1^{\infty} \frac{1}{x^2}\,\mathrm{d}x$, showing that the series is bounded above by the integral.

Example: Integral test convergence Let us now apply the same integral test to show that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges. Figure 8.2 shows similarly the series $\sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$ as a sum of $1 \times a_{n+1}$ rectangular areas for $n \geq 1$. When displayed like this, and again only because function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = \frac{1}{x^2}$ is decreasing, is it intuitive that the area for that series is no larger than the area beneath the function of that graph. Since that series converges iff the original series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges (as the latter only adds one summand to the former), this also gives us an intuition of how to prove that part of Lemma 7. Let us evaluate the indefinite integral:

$$\begin{aligned}
\int_1^{\infty} f(x)\,\mathrm{d}x &= \lim_{b\to\infty} \int_1^b \frac{1}{x^2}\,\mathrm{d}x \\
&= \lim_{b\to\infty} \left[-\frac{1}{x}\right]_1^b \\
&= \lim_{b\to\infty} \left(1 - \frac{1}{b}\right) \\
&= 1
\end{aligned}$$

Since this indefinite integral thus converges, Lemma 7 tells us that the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ must also converge.

8.6 Some Proofs of Ratio Tests

As with the proofs given for ratio tests for sequences, we do not expect you to come up with proofs of this complexity yourself in an exam or assessed work. However, you may find it useful to study and understand such proofs as they also reinforce concepts we have covered.

Proof of D'Alembert Ratio Test (Lemma 5) Let N be in $\mathbb{N}$. We need to prove both items of the lemma:

1. **Case $\frac{a_{i+1}}{a_i} \geq 1$ for all $i \geq N$:** $a_N \leq a_{N+1} \leq a_{N+2} \leq \ldots$. Since $a_N > 0$, it follows that $a_n \geq a_N$ for $n > N$ and thus a_n does not tend to zero, hence the series is not convergent.

2. **Case when there is some k in $\mathbb{R}$ with $\frac{a_{i+1}}{a_i} \leq k < 1$ for all $i \geq N$:** We can see that this series converges by performing a summand-by-summand comparison with a geometric series with ratio $x = k$. From $i \geq N$, we get:

$$\begin{aligned} a_{N+1} &\leq k a_N \\ a_{N+2} &\leq k a_{N+1} \leq k^2 a_N \\ a_{N+3} &\leq k a_{N+2} \leq k^2 a_{N+1} \leq k^3 a_N \\ &\vdots \\ a_{N+m} &\leq k^m a_N \end{aligned}$$

   We can rewrite this last inequality as $a_{N+m} \leq \lambda k^{N+m}$ where $\lambda = \frac{a_N}{k^N}$ is a constant. Letting $n = N + m$, allows us to write that $a_n \leq \lambda k^n$ for all n in $\mathbb{N}$ with $n \geq N$. Since we know that $k < 1$, we infer that the geometric series $(k^n)_{n\geq 1}$ and so also $(\lambda \cdot k^n)_{n\geq 1}$ converge. The latter is precisely what we require under the conditions of the comparison test, to show that $\sum_{n=1}^{\infty} a_n$ converges. □

Proof of D'Alembert Limit Ratio Test (Lemma 6) There is nothing to show for the second item as a series either converges or diverges. We prove the remaining items:

1. **Case $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} = l > 1$:** We need to show that $\sum_{i=1}^{\infty} a_i$ diverges. Using the definition of the limit, we know that for all $\epsilon > 0$ there is some $N(\epsilon)$ in $\mathbb{N}$ such that for all $i > N(\epsilon)$ we have:

$$\left| \frac{a_{i+1}}{a_i} - l \right| < \epsilon$$

   which is equivalent to

$$l - \epsilon < \frac{a_{i+1}}{a_i} < l + \epsilon \tag{8.17}$$

We here want to apply item 1 of Lemma 5 and so we want to use the left hand inequality in (8.17) to show that $\frac{a_{i+1}}{a_i} > 1$ for all $i \geq N(\epsilon)$. Now let $\epsilon = l - 1$. Since $l > 1$, we infer that ϵ is positive. Let N be the $N(\epsilon)$ above. Then for all $i > N$ we have

$$1 < \frac{a_{i+1}}{a_i}$$

By item 1 of Lemma 5, we conclude that the series $\sum_{i=1}^{\infty} a_i$ diverges.

3. **Case** $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} = l < 1$**:** We need to show that $\sum_{i=1}^{\infty} a_i$ converges. Using that l is the limit of sequence $(\frac{a_{i+1}}{a_i})_{n\geq 1}$, we know that for all $\epsilon > 0$ there is some $N(\epsilon)$ in $\mathbb{N}$ such that for all $i > N(\epsilon)$ we have:

$$\left| \frac{a_{i+1}}{a_i} - l \right| < \epsilon$$

and the latter inequality is equivalent to

$$l - \epsilon < \frac{a_{i+1}}{a_i} < l + \epsilon \tag{8.18}$$

Our strategy is to show that there is some k in $\mathbb{R}$ with $k < 1$ such that $\frac{a_{i+1}}{a_i} \leq k < 1$ for all $i \geq N(\epsilon)$. If we achieve that, then we can apply item 2 of Lemma 5. So we are looking to use the right hand inequality in (8.18) for this. Because of this, we need to set $k = l + \epsilon$ but this requires that $k < 1$. Since ϵ also has to be positive, we further need that $l < k$. To solve these constraints, we can place k to be strictly between l and 1, for example by setting

$$k = \frac{l+1}{2}$$

This means that we choose $\epsilon = \frac{1-l}{2}$ which is then positive as $l < 1$. Note that this choice of ϵ then determines the $N(\epsilon)$ we introduced above. So we get that for all $i \geq N(\epsilon)$ we have

$$\frac{a_{i+1}}{a_i} < k < 1$$

So, by item 2 of Lemma 5, the series $\sum_{i=1}^{\infty} a_i$ converges. □

8.7 Absolute Convergence

In the previous subsection, we worked with the assumption that all summands of a series are positive. We also remarked that there is then no loss of generality to assume that all such summands are positive. However, series that occur in practice may have negative summands as well. On the face of it, this does not pose a problem: the partial sums are still well defined (although they may be 0 or negative) and so we can define convergence as before in terms of convergence of the sequence of corresponding partial sums.

But the problem with that is that that sequence of partial sums may depend on the order in which summands of the series are listed. Consider the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \tag{8.19}$$

in which each second summand from the harmonic series is made negative. In the next chapter, we will see that this series converges to $\ln 2$. However, if we rearrange the above order of these summands so that two positive summands are followed by a negative summand, then we arrive at the series

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} - \frac{1}{7} - \frac{1}{4} + \dots \tag{8.20}$$

and it turns out that this sequence also converges, but to the value $\frac{3}{2} \cdot \ln 2$. This dependency on the order in which summands are listed is hardly desirable. Let us define the desirable concept formally:

Definition 23 (Additional material: Unconditional Convergence of Series)

1. *A permutation π over the natural numbers $\mathbb{N}$ is a function $\pi \colon \mathbb{N} \to \mathbb{N}$ that has an inverse, i.e. it is injective and surjective.*

2. *A series $\sum_{i=1}^{\infty} a_i$ is* **unconditionally convergent** *iff it converges and the permuted series*

$$\sum_{i=1}^{\infty} a_{\pi(i)} \tag{8.21}$$

converges to that same limit, for all permutations $\pi \colon \mathbb{N} \to \mathbb{N}$.

While this is an intuitive concept, it seems hard to verify, since there are uncountably many permutations on $\mathbb{N}$! Fortunately, it turns out that for series with summands in $\mathbb{R}$, and indeed for series with summands in finite-dimensional real vector spaces such as those studied in the second part of this course module, there is an equivalent but much simpler definition.

Definition 24 (Absolute Convergence of Series)
A series $\sum_{i=1}^{\infty} a_i$ is **absolutely convergent** *iff the corresponding series of absolute values of summands*

$$\sum_{i=1}^{\infty} |a_i| \tag{8.22}$$

converges.

If we apply the latter, equivalent, concept to the alternating harmonic series, we get the harmonic series which diverges. And so we know that the limit behavior of the alternating harmonic series is sensitive to permutations of summands.

You may wonder how we can formulate the latter concept for vector spaces, i.e. how (8.22) is then interpreted. Then, the absolute value $|\cdot|$ is replaced with a **norm** of a vector, a concept covered in the second part of this course module.
We already saw that convergence of a series does not imply absolute convergence of that series, the alternating harmonic series served as an example therefore. However, a series $\sum_{i=1}^{\infty} a_i$ that converges absolutely, i.e. where $\sum_{i=1}^{\infty} |a_i|$ converges, will also converge itself.

Theorem 29
(Absolute Value Comparison Test) *Suppose b_i is a non-negative sequence such that $\sum_{i=1}^{\infty} b_i$ converges and suppose a_i is a sequence such that $|a_i| \leq b_i$ for all $i \geq 1$. Then $\sum_{i=1}^{\infty} a_i$ converges.*

Proof Denote the partial sums of $\sum_{i=1}^{\infty} a_i$ by S_n and that of $\sum_{i=1}^{\infty} b_i$ by S'_n. We will show that S_n is a Cauchy sequence from which the result follows. Let $\epsilon > 0$. Since S'_n is a Cauchy sequence, there exists N such that for all $m \geq n > N$ we have $S'_m - S'_n < \epsilon$. Thus, for $m \geq n > N$, we have:

$$|S_m - S_n| = \left| \sum_{i=n+1}^{m} a_i \right| \leq \sum_{i=n+1}^{m} |a_i| \leq \sum_{i=n+1}^{m} b_i = S'_m - S'_n < \epsilon.$$

This shows that S_n is a Cauchy sequence.

Corollary 1
If $\sum_{i=1}^{\infty} a_i$ is absolutely convergent then it is convergent.

Proof Put $b_n := |a_n|$ for all $n \geq 1$ in the theorem.

If a series is absolutely convergent, then it is unconditionally convergent, i.e., any rearrangement of the series also converges to the same limit as the original series. We will skip the proof of this result which is somewhat longer than other proofs in these notes.

Example 31
Consider the series

$$S = \sum_{n=1}^{\infty} (-1)^{n+1}/n^2 = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots$$

This series is absolutely convergent since we have already proved that the series $\sum_{n=1}^{\infty} 1/n^2$ is convergent say to a limit $l \in \mathbb{R}$. It follows that any rearrangement of the series S is also convergent to l.

Tests for absolute convergence. Where we have a series with a mixture of positive and negative terms, a_n, we can test for absolute convergence by applying any of the above tests to the absolute value of the terms $|a_n|$. We present a key example.

Theorem 30
(Limit absolute value ratio test) *Consider the series* $\sum_{i=1}^{\infty} a_i$ *with* $a_i \neq 0$ *for* $i \geq 1$.

1. *If* $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$ *then* $\sum_{n=1}^{\infty} a_n$ *converges absolutely (and thus also converges).*

2. *If* $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| > 1$ *then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

3. *If* $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = 1$ *then* $\sum_{n=1}^{\infty} a_n$ *may converge or diverge.*

Proof 1. Suppose $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = l < 1$. Then choosing $\epsilon = (1-l)/2$, there exists N such that for $n > N$ we have $\left|\frac{a_{n+1}}{a_n}\right| < (1+l)/2 < 1$. Putting $c = (1+l)/2$, we have, $|a_{m+N+2}| < c^m |a_{N+1}|$ for $m \geq 1$. Thus, $\sum_{n=1}^{\infty} a_n$ absolutely converges by comparison with the geometric series $|a_{N+1}| \sum_{i=1}^{\infty} c^i$ while ignoring the terms a_i for $i \leq N+2$.
2. Suppose $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = l > 1$. Choosing $\epsilon = (l-1)/2$, there exists N such that for $n > N$ we have $\left|\frac{a_{n+1}}{a_n}\right| > (1+l)/2 > 1$. Put $c = (1+l)/2 > 1$, we get $|a_{m+N+2}| > c^m |a_{N+1}|$ for $m \geq 1$. Thus, $|a_i| \to \infty$ as $i \to \infty$ and the series diverges (since for a convergent series we must have $|a_i| \to 0$ as $i \to \infty$).
3. The two series $\sum_{n=1}^{\infty} 1/n$ and $\sum_{n=1}^{\infty} 1/n^2$ both satisfy $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = 1$, but the first diverges whereas the second converges.

8.7.1 The nth-root test for convergence

The ratio test for absolute convergence is a powerful method, but it does not always work because the sequence a_{n+1}/a_n may not have a limit. The nth root test overcomes this problem but it is somewhat more involved as it relies on the notion of the limit superior (greatest limit) of a sequence, which we now define.

Definition 25 (Limit superior)
Let $(a_n)_{n\geq 1}$ *be a sequence. This determines another sequence* $(b_n)_{n\geq 1}$ *where*

$$b_n = \sup\{a_m \mid m \geq n\} \tag{8.23}$$

Note that $(b_n)_{n\geq 1}$ *is a non-increasing sequence. The* **limit superior** *of* $(a_n)_{n\geq 1}$ *is defined to be the ordinary limit of the sequence* $(b_n)_{n\geq 1}$ *in* $\mathbb{R} \cup \{-\infty, \infty\}$. *We denote this limit as* $\limsup_{n\to\infty} a_n$.

Note that the limit superior is either a real number, or ∞ or $-\infty$. Similarly, we have the definition of limit inferior:

Definition 26 (Limit inferior)
Let $(a_n)_{n\geq 1}$ be a sequence. This determines another sequence $(c_n)_{n\geq 1}$ where

$$c_n = \inf\{a_m \mid m \geq n\} \tag{8.24}$$

Note that $(c_n)_{n\geq 1}$ is a non-decreasing sequence. The **limit inferior** *of $(a_n)_{n\geq 1}$ is defined to be the ordinary limit of the sequence $(c_n)_{n\geq 1}$ in $\mathbb{R}\cup\{-\infty,\infty\}$. We denote this limit as $\liminf_{n\to\infty} a_n$.*

Note that the limit inferior is again either a real number, or ∞ or $-\infty$.

Example 32
For the sequence $a_n = (-1)^n$, we have that $\limsup_{n\to\infty} a_n = 1$ and $\liminf_{n\to\infty} a_n = -1$.

Example 33
Find the $\limsup$ and the $\liminf$ of the sequence $a_n = \sin(n\pi/2)$.

Exercise 22
Given a sequence $(a_n)_{n\geq 1}$, show that there exist subsequences $(a_{m_i})_{i\geq 1}$ and $(a_{p_i})_{i\geq 1}$ such that $\limsup_{n\to\infty} a_n = \lim_{i\to\infty} a_{m_i}$ and $\liminf_{n\to\infty} a_n = \lim_{i\to\infty} a_{p_i}$. Show also that if $(a_{q_i})_{i\geq 1}$ is any convergent subsequence of $(a_n)_{n\geq 1}$, then we have:

$$\liminf_{n\to\infty} a_n \leq \lim_{i\to\infty} a_{q_i} \leq \limsup_{n\to\infty} a_n.$$

It follows from Exercise 22 that the set of limits of all subsequences of a sequence $(a_n)_{n\geq 1}$ is contained in the closed interval of the extended real line:

$$[\liminf_{n\to\infty} a_n, \limsup_{n\to\infty} a_n] \subseteq [-\infty,\infty].$$

In addition, if the sequence does have a limit then $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = \lim_{n\to\infty} a_n$
We can now state the nth root test.

Proposition 31
Consider the series $\sum_{n=1}^{\infty} a_n$. If $\limsup_{n\to\infty} |a_n|^{1/n} < 1$ then the series converges absolutely. If $\limsup_{n\to\infty} |a_n|^{1/n} > 1$ then the series diverges. If $\limsup_{n\to\infty} |a_n|^{1/n} = 1$ then the series can absolutely converge, or just converge or diverge.

Proof Let $l = \limsup_{n\to\infty} |a_n|^{1/n} < 1$. Then $l < (l+1)/2 < 1$ and, by the definition of the limit superior, there exists N such that for all $n > N$ we have $|a_n|^{1/n} < (l+1)/2$, i.e., $|a_n| < c^n$ for $n > N$ where $c = (l+1)/2 < 1$. Hence the series converges absolute by comparison with the geometric series $\sum_{n=1}^{\infty} c_n$.
Suppose now $l = \limsup_{n\to\infty} |a_n|^{1/n} > 1$ and thus $c = (l+1)/2 > 1$. Since l is the limit superior of $(|a_n|)^{1/n}$ there exists a subsequence $(|a_{n_i}|^{1/n_i})_{i\geq 1}$ such that $\lim_{i\to\infty} |a_{n_i}|^{1/n_i} = l$ and thus there exists N such that for $i > N$ we have $|a_{n_i}|^{1/n_i} > c > 1$ or $|a_{n_i}| > c^{n_i} > 1$. Hence, $|a_n|$ does not converge to zero and thus the series diverges. The last statement is left as an exercise below.

Example 34
Consider the series

$$1+1+\frac{1}{2}+\frac{1}{3}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{2^3}+\frac{1}{3^3}+\cdots$$

with $a_{2n} = 1/2^n$ *and* $a_{2n+1} = 1/3^n$. *Then check that* $\limsup_{n\to\infty}(a_n)^{1/n} = 1/\sqrt{2} < 1$. *Hence, the series converges. However, the ratio test fails since* $\liminf_{n\to\infty} a_{n+1}/a_n = 0$ *and* $\limsup_{n\to\infty} a_{n+1}/a_n = \infty$ *and thus* $\lim_{n\to\infty} a_n$ *does not exist.*

Exercise 23
Verify the last statement of Proposition 31, by applying the root test to each of the the following three cases (i) $\sum_{n=1}^{\infty} 1/n^2$, *(ii)* $\sum_{n=1}^{\infty}(-1)^{n+1}/n$, *and (ii)* $\sum_{n=1}^{\infty} 1/n$.

Exercise 24
Check using the nth root test if the series

$$\sum_{n=1}^{\infty}\left(\frac{n}{n+1}\right)^{n^2} 4^n$$

converges or not and verify that applying the ratio test is far more involved.

It can be shown that for any sequence $(a_n)_{n\geq 1}$ of positive terms, we always have the following inequalities

$$\liminf_{n\to\infty}|a_{n+1}/a_n| \leq \liminf_{n\to\infty}|a_n|^{1/n} \leq \limsup_{n\to\infty}|a_n|^{1/n} \leq \limsup_{n\to\infty}|a_{n+1}/a_n|.$$

Thus, if $\lim_{n\to\infty}|a_{n+1}/a_n| = l$ exists then all the four terms above collapse to a single extended real number and the ratio test is a good as the nth root test. But if $\lim_{n\to\infty}|a_{n+1}/a_n|$ fails to exist, then the nth root test will provide the definite information about the convergence of the series.

9 Differentiation

Differentiation is the inverse operation of integration. But the derivative of a function can be defined independently of integration.

9.1 Differentiation of Real Functions

We already discussed the continuity of functions $f : \mathbb{R} \to \mathbb{R}$ at a point x in $\mathbb{R}$. Intuitively, this meant that we can control the size of the change in outputs of f (the ϵ) obtained by moving away from x by limiting how much we can move away from that x (the δ). We now want to see whether we can strengthen this concept so that we can understand how a function changes its output by making infinitesimally small changes to the input. It should be no surprise that we can define this using limits of sequences:

Definition 27 (Differentiation)
Let $f : \mathbb{R} \to \mathbb{R}$ be a function and x in $\mathbb{R}$ and let $h > 0$ be given. Then:

1. *The* **Newton's difference quotient** *at x for f is given by*

$$\frac{\Delta f(x)}{\Delta(x)} = \frac{f(x+h) - f(x)}{(x+h) - x} = \frac{f(x+h) - f(x)}{h} \tag{9.1}$$

2. *Function f is* **differentiable** *at x iff the limit*

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{9.2}$$

exists and is a real number. In that case, this limit is the **derivative** *of f at x and denoted as $f'(x)$ or equivalently $\frac{dy}{dx}$ when $y = f(x)$.*

We note that Newton's difference quotient is implicitly a function of the value h for a given x in $\mathbb{R}$. Also, the limit in (9.2) is strictly speaking not the one defined for a sequence in Chapter 5. Rather, it means that **for all ways in which h converges to 0** (seen as a usual sequence), the limit of Newton's difference quotient exists and it is the same value for all such sequences of h. Let us illustrate the latter point:

Example 35 (Derivative)
Recall the function **absolute value** *$f(x) = |x|$. We argue that this function has a derivative at all positive and all negative points x. For example, consider $x = -2$. If h represents any sequence that converges to 0, then $x + h$ represents a corresponding*

sequence that will consist of negative numbers only, from some point onward. But then Newton's difference quotient will be

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{|x+h|-x}{h}=\lim_{h\to 0}\frac{(-x-h)-(-x)}{h}=-1 \tag{9.3}$$

This reasoning remains valid for any negative number x. A similar analysis shows that this function has derivative 1 at positive x. But what about $x=0$? Here we see that it is important for the existence of a derivative that the limit is always the same value, regardless of which sequence for h is chosen:

- *Let h converge to 0 from below, for example as a sequence $(-1/n)_{n\geq 0}$. Then the limit $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$ computes to -1.*
- *Let h converge to 0 from above, for example as a sequence $(1/n)_{n\geq 0}$. Then the limit $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$ computes to 1.*

Therefore, the limit is not consistently the same value and so the function $|x|$ does **not** *have a derivative at $x=0$.*

In Figure 9.1, we see a schematic taken from Wikipedia that illustrates that the derivative, as a real number, is the **slope** of the tangent of the curve of function f at point x.
We can use our knowledge of convergence of sequences to calculate the derivatives of functions. Let us illustrate this with an example.

Example 36 (Derivative of Quadratic Function)
Consider the function $f:\mathbb{R}\to\mathbb{R}$ with $f(x)=x^2$. We compute

$$\frac{f(x+h)-f(x)}{h}=\frac{(x+h)^2-x^2}{h}=\frac{(x^2+2xh+h^2)-x^2}{h}=\frac{2xh}{h}+\frac{h^2}{h}=2x+h$$

Taking the limit as h tends to 0 we get that

$$(x^2)'=\lim_{h\to 0}2x+h=2x$$

This example generalized to polynomials where we have that

$$(a_0+a_1\cdot x+a_2\cdot x^2+\cdots+a_n\cdot x^n)'=a_1+2a_2\cdot x+\cdots+n\cdot a_n\cdot x^{n-1} \tag{9.4}$$

We conclude this discussion of derivatives by recording some basic but important facts about them:

Theorem 32 (Properties of Derivatives)
Let $f,g:(a,b)\to\mathbb{R}$ be two functions.

1. *Polynomials have derivatives at all points, given by (9.4).*
2. *If f is differentiable at x, then f is also continuous at x.*

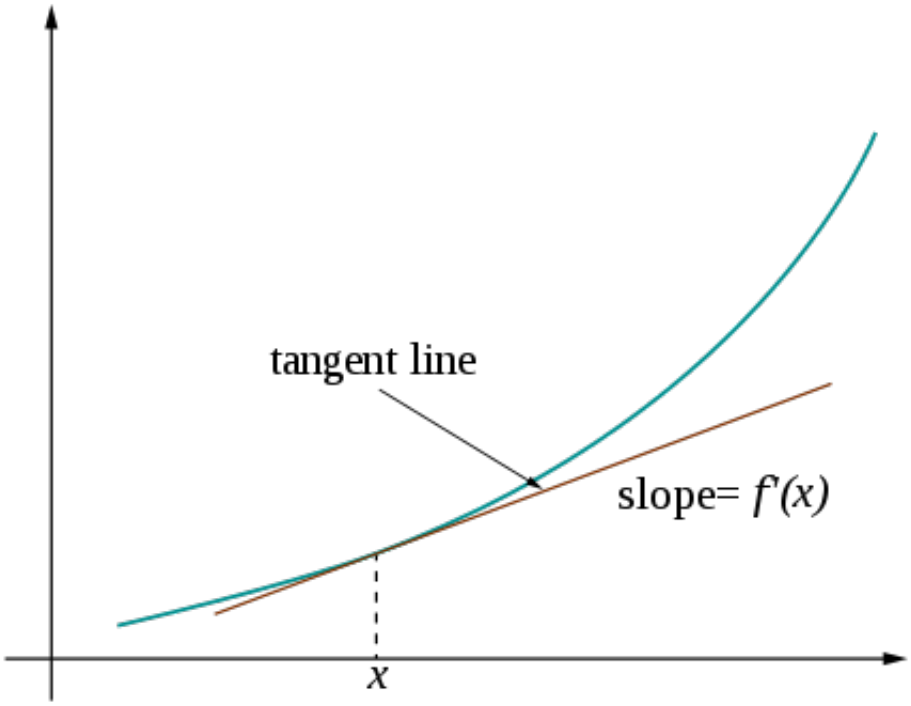


Figure 9.1: Illustration that the derivative $f'(x)$ of function f at point x is the slope of the tangent to the graph of function f at point x (source: Wikipedia)

3. *If f is differentiable in (a,b) then $f'(x_0) = 0$ for any point x_0 at which f is maximum or minimum.*

   Proof *In fact, if $f'(x_0) > 0$ then there exists a $\delta > 0$ such that for $|x - x_0| < \delta$ we have $f(x) - f(x_0)/(x - x_0) > 0$ and thus for $x > x_0$ we have $f(x) - f(x_0) > 0$ and for $x - x_0 < 0$ we have $f(x) - f(x_0) < 0$. Thus, f is increasing at x_0. Similary, if $f'(x_0) < 0$, then f is decreasing at x_0. Therefore at a point of maximum we must have $f'(x_0) = 0$. Similarly for a minimum.*

4. *If f and g are differentiable at x, then the product function $f \cdot g$ defined by $(f \cdot g)(x) = f(x) \cdot g(x)$ is also differentiable at x and*

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x) \tag{9.5}$$

5. *If g and f are, respectively, differentiable at x and $g(x)$, then the composition function $f \circ g$ defined by $(f \circ g)(x) = f(g(x))$ is also differentiable at x and*

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x) \tag{9.6}$$

6. *Differentiation $f \mapsto f'$ is a* **linear function***: for all f and g that are differentiable at x, and for all a and b in $\mathbb{R}$ we have that the function $a \cdot f + b \cdot g$ defined by $(a \cdot f + b \cdot g)(x) = a \cdot f(x) + b \cdot g(x)$ is differentiable at x and*

$$(a \cdot f + b \cdot g)'(x) = a \cdot f'(x) + b \cdot g'(x) \tag{9.7}$$

Equation (9.5) is known as the **product rule** and (9.6) as the **chain rule**. Equation (9.7) states that the function $f \mapsto f'$ that maps differentiable functions to their derivative is a **linear map**. Linear maps over vector spaces are the subject of the second half of this course module.
We refer to the recommended reading for more facts about which functions have derivatives and what those derivatives are. For example, we have that the sine function has a derivative at all points x in $\mathbb{R}$:

$$\sin'(x) = \cos(x) \tag{9.8}$$

Exercise 25
Using the rule of differentiation of the product and Fundamental Theorem of Calculus, derive the formula for integration by parts:

$$\int_a^b u(x)v'(x)\,dx = u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\,dx,$$

assuming $u', v' : [a,b] \to \mathbb{R}$ are continuous.

9.1.1 Mean Value Theorem and Taylor's Theorem

Exercise 26
*(**Rolle's Theorem**) If $f : [a,b] \to \mathbb{R}$ is continuous and $f : (a,b) \to \mathbb{R}$ is differentiable with $f(a) = f(b)$, then there exists $c \in (a,b)$ such that $f'(c) = 0$. (**Hint** We know that f being continuous in $[a,b]$ must have a minimum and a maximum. Consider the two possible cases: (i) Either the minimum or the maximum occurs in (a,b). (ii) Both of them occur at the end points a and b.)*

Theorem 33
If $f : [a,b] \to \mathbb{R}$ is continuous and $f : (a,b) \to \mathbb{R}$ is differentiable, then there exists $c \in (a,b)$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

Proof Consider the function $g : [a,b] \to \mathbb{R}$ with $g(x) = f(x) - sx$ where $s \in \mathbb{R}$ is a constant. Choose s such that $g(a) = g(b)$ and apply Exercise 26.

We can rewrite the statement of the mean value theorem as follows:

$$f(b) = f(a) + (b - a)f'(c) \tag{9.9}$$

If $f : (a,b) \to \mathbb{R}$ is differentiable at $x_0 \in (a,b)$, then the tangent to the graph of f at the point $(x_0, f(x_0))$ is the line $y = f(x_0) + f'(x_0)(x - x_0)$. Thus if x is close to x_0 we expect $f(x) = f(x_0) + f'(x_0)(x - x_0) + E$ where the error E is small.
Taylor's theorem formalises the above idea of approximation of a function in the neighbourhood of a point using the derivative and the higher order derivatives of the function at that point, and gives an expression of the error E. Denote the nth derivative of f at a point x, if it exists, by $f^{(n)}(x_0)$, e.g., $f'(x) = f^{(1)}(x)$ and $f''(x) = f^{(2)}(x)$.

Theorem 34
If f is n times differentiable in (a,b) with $x_0 \in (a,b)$, then for any $x \in (a,b)$ we have:

$$f(x) = f(x_0) + \frac{1}{1!}(x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) + \cdots + \frac{1}{(n-1)!}(x - x_0)^{n-1} f^{(n-1)}(x_0) + E_n,$$

where $E_n = \frac{1}{n!}(x - x_0)^n f^{(n)}(x^*)$ for some x^* between x and x_0.

Proof For y between x and x_0, let

$$F(y) = f(x) - f(y) - (x - y)f'(y) - \cdots - \frac{(x - y)^{n-1}}{(n-1)!} f^{(n-1)}(y),$$

which satisfies $F(x) = 0$. It is easy to check that

$$F'(y) = -\frac{1}{(n-1)!}(x - y)^{n-1} f^{(n)}(y).$$

Consider the function

$$G(y) = F(y) - \left(\frac{x-y}{x-x_0}\right)^n F(x_0)$$

which satisfies $G(x_0) = G(x) = 0$. By Rolle's Theorem (Exercise 26), there exists x^* between x_0 and x such that

$$0 = G'(x^*) = F'(x^*) + \frac{n(x-x^*)^{n-1}}{(x-x_0)^n}F(x_0) = -\frac{(x-x^*)^{n-1}}{(n-1)!}f^{(n)}(x^*) + \frac{n(x-x^*)^{n-1}}{(x-x_0)^n}F(x_0).$$

Therefore, $F(x_0) = (x-x_0)^n f^{(n)}(x^*)/n!$ which gives the required result □.

The term

$$\frac{f^{(n+1)}(x^*)}{(n+1)!}(x-x_0)^{n+1} \tag{9.10}$$

is known as the **Lagrange error term**. Although x^* exists as we have shown, there is unfortunately no easy way to find that value of x^*. So in practice, the bound $x_0 < x^* < x$ or $x < x^* < x_0$ is used to generate a worst-case error for the term in (9.10). For example, if $x_0 < x^* < x$, we may want to compute

$$\max_{y\in(x_0,x)} \frac{f^{(n+1)}(y)}{(n+1)!}(x-x_0)^{n+1} \tag{9.11}$$

using techniques from mathematical optimisation or real analysis.

9.1.2 L'Hopital's rule

Recall that in deriving the rule for the limit $\lim_{x\to c} f(x)/g(x)$ of the ratio of two functions, we assumed that $\lim_{x\to c} g(x) \neq 0$. For functions that are continuously differentiable, i.e., their derivatives are also continuous, we can in certain cases find $\lim_{x\to c} f(x)/g(x)$ when $\lim_{x\to c} g(x) = 0$.

Proposition 35
Suppose $f, g : (a,b) \to \mathbb{R}$ have derivatives $f', g' : (a,b) \to \mathbb{R}$ that are continuous in (a,b). If $f(c) = g(c) = 0$ for some $c \in (a,b)$ and $g'(c) \neq 0$, then

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}.$$

Proof We have:

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f(x)-f(c)}{g(x)-g(c)} = \lim_{x\to c}\frac{\left(\frac{f(x)-f(c)}{x-c}\right)}{\left(\frac{g(x)-g(c)}{x-c}\right)} = \frac{\lim_{x\to c}\left(\frac{f(x)-f(c)}{x-c}\right)}{\lim_{x\to c}\left(\frac{g(x)-g(c)}{x-c}\right)} = \frac{f'(c)}{g'(c)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}$$

Exercise 27
Find the following limits using L'Hopital's rule:

1. $\lim_{x\to 0}\frac{e^x-1}{-x^3+x}$.

2. $\lim_{x\to 0} \frac{2\sin x - \sin 2x}{x - \sin x}$ *(you need to apply the rule three times)*.

The rule can also be extended to the case when $\lim_{x\to c}|f(x)| = \lim_{x\to c}|g(x)| = \infty$ by writing $\frac{f(x)}{g(x)} = \frac{1/g(x)}{1/f(x)}$. It can also be extended to the case when $x \to \infty$ by writing $\lim_{x\to\infty} f(x)/g(x) = \lim_{y\to 0} f(1/y)/g(1/y)$, where we restrict f and g to the non-negative real numbers, $f, g : [0,\infty) \to \infty)$.

9.1.3 Additional material: Fundamental Theorem of Calculus

Theorem 36
If $f : [a,b] \to \mathbb{R}$ is continuous and the function $F : [a,b] \to \mathbb{R}$ is defined by $F(y) = \int_a^y f(x)\,dx$ then F is uniformly continuous on $[a,b]$ and $F'(x) = f(x)$ for $x \in (a,b)$.

Proof Since f is continuous on $[a,b]$, it is bounded by the remark after Theorem 22: Thus, there exists $K > 0$ such that $|f(x)| < K$ for $x \in [a,b]$. Let $\epsilon > 0$ be given for proving the uniform continuity of F. For any h with $0 < h < \epsilon/K$, we have

$$F(y+h) - F(y) = \int_y^{y+h} f(x)\,dx$$

Thus, using the inequalities in Exercise 19 for the Riemann integral, we obtain:

$$|F(y+h) - F(y)| = |\int_y^{y+h} f(x)\,dx| \le \int_y^{y+h} |f(x)|\,dx \le \int_y^{y+h} K\,dx = Kh < \epsilon.$$

To show that $F'(x) = f(x)$ for $x \in (a,b)$, let $\epsilon > 0$ be given. Since f is continuous at x, there exists $\delta > 0$ such that $f(x) - \epsilon < f(y) < f(x) + \epsilon$ for $|y - x| < \delta$. Thus, choosing $0 < h < \delta$ and using the inequality for integrals again, we have

$$hf(x) - \epsilon < \int_x^{x+h} f(y)\,dy < hf(x) + \epsilon h.$$

Or $|F(x+h) - F(x) - hf(x)| < \epsilon h$, i.e., $|\frac{F(x+h)-F(x)}{h} - f(x)| < \epsilon$. Similarly, we obtain $|\frac{F(x)-F(x-h)}{h} - f(x)| < \epsilon$.

Corollary 2
*(**Change of Variable Integration**) Let $g : [a,b] \to [c,d]$ be a differentiable function with $g' : [a,b] \to \mathbb{R}$ continuous, and let $f : [c,d] \to \mathbb{R}$ be a continuous function. Let $y = g(x)$. Then*

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(y)dy.$$

10 Power Series

In this chapter, we will explore how functions $f : \mathbb{R} \to \mathbb{R}$ that are sufficiently "smooth" may be represented, at least locally, as series that contain the function argument as a formal parameter. Such power series are an important tool in real analysis, with many applications in engineering and beyond. In the last chapter, we studied the Taylor series of a function with derivatives of order up to $n+1$ for some positive integer n. The idea of a power series is to see if we can express a function that has derivatives of all orders as a series.

10.1 Basics of Power Series

We already studied the geometric series $\sum_{i=1}^{\infty} x^i$ where x is a formal parameter. And we saw that this series converges when $|x| < 1$. Now, we want to study a slightly more general version of this where the summands are weighted with a scalar and the arguments of the power functions are shifted by a fixed constant. This notion of **power series** we define formally:

Definition 28 (Power Series)
A **power series** *is a series of form*

$$\sum_{i=0}^{\infty} a_i \cdot (x-c)^i \tag{10.1}$$

where x is a a real variable, c is a constant in $\mathbb{R}$, and $(a_n)_{n\geq 0}$ is a sequence of reals.

We note that the sequence of weights a_n now starts counting at index 0. This is merely a notational convention and helps us to align the index with the exponent in the power function of such series.
Power series generalise the geometric series: the argument x to the power function x^i can be shifted by a constant c, and each of the summands $(x-c)^i$ can have a weight a_n. When c equals 0, $a_1 = 0$, and $a_n = 1$ for all $n \geq 1$, this is the geometric series.
Here are some questions that we mean to at least partially answer about power series:

Q1 For which values of x does a given power series of the form in (10.1) converge?

Q2 Suppose that a power series converges for all x in some open interval (a,b). What properties does this function of type $(a,b) \to \mathbb{R}$ enjoy?

Q3 Conversely, given a function $f : (a,b) \to R$ defined on an open interval (a,b), what properties of f are required to be able to represent f as a power series that converges on the entire interval (a,b)?

Let us start by considering a simple example, the quadratic function

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = 1 + 2x + x^2 = (x - (-1))^2 \tag{10.2}$$

Note that we have represented this function in two equivalent ways. And both are in the correct format to be a power series:

- For $f(x) = 1 + 2x + x^2$ we set $c = 0$, $a_0 = 1$, $a_1 = 2$, $a_2 = 1$, and $a_n = 0$ for all $n \geq 3$.
- For $f(x) = (x - (-1))^2$ we set $c = -1$, $a_2 = 1$, and $a_n = 0$ for all n in $\mathbb{N}$ with $n \neq 2$.

It should be intuitively clear that both of these power series converge for all x (and converge to the same values), as they are essentially polynomials. In general, a polynomial (of degree n) is expressible as

$$p(x) = a_0 + a_1 \cdot x + \cdots + a_n \cdot x^n \qquad \text{where } a_n \neq 0. \tag{10.3}$$

Therefore, polynomials are power series where c equals 0 and there is some N such that $a_n = 0$ for all $n \geq N$. These power series converge for all x in $\mathbb{R}$.
Convergence questions for power series are more complex when they do not represent polynomials, when the weights a_n do not become 0 from some point onward. The geometric series is such an example.
We know that a power series of form (10.1) converges for $x = c$, as then all summands are 0. But for which other points, if any, does this power series converge and does this set of points have some structure? Fortunately, there is a simple answer to this, although its proof is beyond the scope of this course module and stems from analysis for functions over complex numbers:

Definition 29 (Radius of Convergence)
Let c be a constant in $\mathbb{R}$ and $(a_n)_{n\geq 0}$ a sequence of reals. The power series $\sum_{i=0}^{\infty} a_i \cdot (x-c)^i$ has as **radius of convergence** *r in $[0, \infty) \cup \{\infty\}$ such that:*

1. *If $r \neq \infty$, then:*
   (a) *the power series converges for all x in $\mathbb{R}$ such that $|x - c| < r$, and*
   (b) *the power series diverges for all x in $\mathbb{R}$ such that $|x - c| > r$.*
2. *If $r = \infty$, then the power series converges for all x in $\mathbb{R}$.*

A moment's thought reveals that there can be at most one radius of convergence for a given power series: it has to be ∞ if the power series converges for all x in $\mathbb{R}$, and otherwise the radius of convergence is the touching point of two half rays on the real numbers where the power series converges on one half ray $(-\infty, r)$ and diverges on the other one (r, ∞).
It is a perhaps remarkable fact that the radius of convergence exists for power series and has an explicit formula. Using the notion of the limit superior, studied in Subsection 8.7.1, we can state the theorem about the radius of convergence:

Theorem 37 (Radius of Convergence)
Every power series $\sum_{n=0}^{\infty} a_n \cdot (x-c)^n$ has a radius of convergence r which is given by

$$r^{-1} = \limsup_{n\to\infty} |a_n|^{1/n} \tag{10.4}$$

Proof We apply the nth root test to $\sum_{n=0}^{\infty} a_i \cdot (x-c)^i$. Suppose $l = \limsup_{n\to\infty} |a_n|^{1/n}$, where l is a real number. We have

$$\limsup_{n\to\infty}(|a_n||x-c|^n|)^{1/n} = \limsup_{n\to\infty}(|a_n|)^{1/n}|x-c| = l|x-c|.$$

Thus, by Proposition 31, the series converges absolutely if $l|x-c| < 1$ i.e., for $|x-c| < 1/l$ and diverges for $l|x-c| > 1$, i.e., for $|x-c| > 1/l$. (We have assumed that $1/l = \infty$ if $l = 0$.) If however $\limsup_{n\to\infty} |a_n|^{1/n} = \infty$ then the series converges only for $|x-c| = 0$ for $x = c$ (why?).

We are using the usual convention that if $\limsup_{n\to\infty} |a_n|^{1/n} = 0$ then $r = \infty$ and if $\limsup_{n\to\infty} |a_n|^{1/n} = \infty$, then $r = 0$.

Example 37
Consider the power series

$$1 + x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{2^2}x^4 + \frac{1}{3^2}x^5 + \cdots + \frac{1}{2^n}x^{2n} + \frac{1}{3^n}x^{2n+1} + \cdots$$

with $a_{2n} = 1/2^n$ and $a_{2n+1} = 1/3^n$. Check that $\liminf_{n\to\infty}(a_n)^{1/n} = \liminf_{n\to\infty}(1/3^n)^{1/(2n+1)} = 1/\sqrt{3}$ and $\limsup_{n\to\infty}(a_n)^{1/n} = 1/\sqrt{2}$. It now follows that the radius of convergence is: $r = 1/(1/\sqrt{2}) = \sqrt{2}$.

Computing the radius of convergence from the formula in (10.4) can be involved. Fortunately, there is another way of computing this that often works in practice:

Lemma 8 (Ratio Test for Radius of Convergence)
Suppose that the sequence

$$\Big(\frac{|a_{n+1}|}{|a_n|}\Big)_{n\geq 1}$$

has a limit l in $\mathbb{R}$. Then l^{-1} is the radius of convergence of any power series $\sum_{i=0}^{\infty} a_i \cdot (x-c)^i$ in (10.1).

We recognise that this lemma is basically applying a limit ratio test for absolute convergence to the sequence of weights used in the formal series in question!

Example 38 (Calculating the Radius of Convergence)
Consider the power series:

$$S = \sum_{n=1}^{\infty} n^2 x^n$$

and so $a_0 = 0$ for this series. We can apply the D'Alembert ratio test to this as stated in Lemma 8. However, in these computations we have to treat x as a formal variable, and we have to be able to deal with possibly negative values of x.
Applying the test of Lemma 8, we get:

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(n+1)^2x^{n+1}}{n^2x^n}\right| \\ &= \lim_{n\to\infty}\left|x\left(1+\frac{1}{n}\right)^2\right| \\ &= |x|\end{aligned}$$

If we look at the ratio test, it says that if the result is < 1 it will converge. So our general convergence condition on this series, is that $|x| < 1$. Thus the radius of convergence is 1.

10.1.1 Addition and product of power series

Suppose we have two power series

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots$$

$$g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n + \cdots$$

with radii of convergence of r_1 and r_2 respectively. Then we can write

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + R_{1n}(x)$$

$$g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n + R_{2n}(x)$$

where $R_{1n}(x) \to 0$ for $|x| < r_1$ and $R_{2n}(x) \to 0$ for $|x| < r_2$ as $n \to \infty$. Writing

$$S_{1n}(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

and

$$S_{2n}(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n,$$

we have for $|x| < \min\{r_1, r_2\}$:

$$f(x) = S_{1n}(x) + R_{1n}(x)$$

$$g(x) = S_{2n}(x) + R_{2n}(x)$$

Hence, for $|x| < \min\{r_1, r_2\}$, we have $S_{1n}(x) + S_{2n}(x) = f(x) + g(x) - R_{1n}(x) - R_{2n}(x)$, i.e.,

$$f(x) + g(x) - R_{1n}(x) - R_{2n}(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n$$

This gives for $|x| < \min\{r_1, r_2\}$ as $n \to \infty$:

$$f(x) + g(x) = f(x) + g(x) - 0 - 0 = a_0 + b_0 + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots(a_n + b_n)x^n + \cdots$$

since $R_{1n}(x) \to 0$ and $R_{2n}(x) \to 0$ for $|x| < \min\{r_1, r_2\}$. Thus, for the radius of convergence of the power series $f(x) + g(x)$ is at least $\min\{r_1, r_2\}$. Note that the radius of

convergence of the power series $f(x)+g(x)$ can be strictly greater than $min\{r_1, r_2\}$. For example, if $f(x) = \sum_{n\geq 0} a_n x^n$ with radius of convergence $r_1 < \infty$, then $g(x) = \sum_{n\geq 0}(-a_n)x^n$ has clearly also radius of convergence $r_2 = r_1$ with $min\{r_1, r_2\} = r_1$. However, the radius of convergence of $f(x)+g(x) = \sum_{n\geq 0}(a_n - a_n)x^n = \sum_{n\geq 0} 0x^n = 0$ is clearly ∞. We have thus shown:

Proposition 38
Two power series with radii of convergence r_1 and r_2, respectively, can be added term by term to get the sum of the two power series, absolutely convergent with the radius of convergence $r \geq \min\{r_1, r_2\}$.

Exercise 28
Prove, with the notations as above, that

$$f(x)g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + \cdots + (a_0b_n + a_1b_{n-1} + \cdots a_{n-1}b_1 + a_nb_0)x^n + \cdots$$

for $|x| < \min\{r_1, r_2\}$.

Exercise 29

(i) Find the power series of $f(x) = 3x/(1+x^2)$ by rewriting the function as $f(x) = (3x)(1/1-(-x^2))$ and hence using first the product rule and then the power series for $1/(1-y) = 1 + y + y^2 + \cdots + y^n + \cdots$ with $|y| < 1$ for $y = -x^2$.

(ii) Find the power series of $f(x) = 1/(x-1)(x-3)$ by first writing

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

where a and b are real numbers to be found, and then using the sum of power series.

10.2 Maclaurin Series

We will now turn to the question of which types of functions can be represented by power series so that their radius of convergence subsumes the domain of definition of the function. In particular, the power series and the function will have the same "input/output" behavior within that radius of convergence.
First, we will look at a special case of functions $f : \mathbb{R} \to \mathbb{R}$ that are **infinitely differentiable** at $x = 0$. Recalling our definition of a derivative at a point, infinitely differentiable at point 0 means that the derivative f' of f exists at 0, and more generally that for all $k \geq 1$, the kth derivative $f^{(k)}$ of f exists at 0. This makes use of an inductive definition of kth derivative:

Definition 30 (Smoothness of Function)
*A function $f : \mathbb{R} \to \mathbb{R}$ is **smooth** at x_0 if for all $k \geq 1$ the kth derivative of f exists at x_0. These kth derivatives are defined inductively by*

$$f^{(1)} = f' \qquad f^{(k+1)} = \left(f^{(k)}\right)' \text{ for all } k \geq 1. \tag{10.5}$$

Given such a function $f : \mathbb{R} \to \mathbb{R}$ that is smooth at 0, we can develop a power series for f with $c = 0$ that has a positive radius of convergence. If this power series has the same outputs as function f within that radius of convergence, then f is called a **real analytical function**. Not every smooth real function is analytical.[1]

Example 39
Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = e^{-1/x^2}$ for $x \neq 0$ and $f(0) = 0$. Then it can be checked easily that f is continuous. By induction, it can be shown that $f^{(n)}(x)$ tends to 0 as $x \to 0$. Thus, by defining $f^{(n)}(0) = 0$, it follows that f is a smooth function. However f has no Maclaurin series as the Taylor series for $f(x)$ reduces to

$$e^{-1/x^2} = 0 + 0 + \cdots + 0 + x^n f^{(n)}(x^*)/n!,$$

for some $0 < x^* < x$ and clearly the remainder $E_n(x) = x^n f^{(n)}(x^*)/n!$ cannot tend to zero, since otherwise we will get the series

$$e^{-1/x^2} = 0 + 0 + 0 + 0 + 0 + 0 + \cdots,$$

which is a contradiction. In fact, we have $E_n(x) = f(x)$ for all n and x, showing that as $n \to \infty$ the remainder $E_n(x)$ does not tend to 0 for any non-zero value of x. Thus, f is not an analytic function although it is smooth.

The power series in Equation 10.1 is called the **Maclaurin Series** when c equals 0.[2] Let us see how to compute the Maclaurin series by writing down its formal power series:

$$f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \ldots \tag{10.6}$$

This means that we now view the weights (or coefficients) a_i as unknown or variables for which we mean to solve. That is, we mean to solve the equation in (10.6) so that its infinitely many variables a_i are determined by f and its infinitely many derivatives $f^{(i)}$ for all $i \geq 1$.
In doing this, we of course assume that we have access to function f and its derivatives so that we know the values $f(0)$, $f'(0)$, and generally $f^{(i)}(0)$ for all $i \geq 1$. Let us see how we can solve for a_0. This is relatively easy since for $x = 0$, equation (10.6) then reads as $f(0) = a_0$ and so we may define a_0 to be $f(0)$.
Next, we differentiate both sides of (10.6) and obtain a new equation

$$f'(x) = a_1 + 2 \cdot a_2 \cdot x + 3 \cdot a_3 \cdot x^2 + \ldots \tag{10.7}$$

by assuming that differentiation is also a linear map over infinite sums (which it is). Again, we evaluate this equation at $x = 0$ and get that $a_1 = f'(0)$. We will repeat this process and this will reveal that the values of a_i also have a weight that stems from the repeated differentiation of powers x^n. Let us see this for a_2. Differentiating both sides in (10.7), we get

$$f^{(2)}(0) = 2 \cdot 1 \cdot a_2 + 3 \cdot 2 \cdot a_3 \cdot x + 4 \cdot 3 \cdot a_4 \cdot x^2 + \ldots \tag{10.8}$$

[1] For complex functions this is true through.
[2] Later, we consider the general case of c and the derived power series are then called **Taylor Series**.

Letting $x = 0$ in this equation gives us $2 \cdot 1 \cdot a_2 = f^{(2)}(0)$. Solving this for a_2 gives us

$$a_2 = \frac{f^{(2)}(0)}{2!} \tag{10.9}$$

We can now repeat this procedure and will see that the solution for a_n with $n \geq 0$ satisfies

$$a_n = \frac{f^{(n)}(0)}{n!} \qquad \text{for } n \geq 0 \tag{10.10}$$

Note that this is also the correct formula for $n = 0$ as $0!$ is defined to be 1.

Maclaurin series: summary Suppose that $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable at $x = 0$ and that f has a formal power series representation, also known as *series expansion*, of the form

$$f(x) = \sum_{i=0}^{\infty} a_i x^i$$

as above.
Differentiating both sides of this equation n times gives another equation

$$f^{(n)}(x) = \sum_{i=n}^{\infty} a_i i(i-1)\dots(i-n+1)x^{i-n}$$

Setting $x = 0$, we then have that $f^{(n)}(0) = n!a_n$ because all terms but the first have x as a factor and so cancel out.
Hence we obtain Maclaurin's series for f:

$$f(x) = \sum_{n=0}^{\infty} f^{(n)}(0)\frac{x^n}{n!} \tag{10.11}$$

For this power series, we know by Theorem 37 that its radius of convergence r exists and is given by

$$r^{-1} = \limsup_{n\to\infty} |f^{(n)}(0)/n!|^{1/n} \tag{10.12}$$

In practice, we will typically apply Lemma 8 instead or use some ad-hoc reasoning to determine that radius of convergence, as done in the next example. That example also demonstrates that the addition of each successive term in the Maclaurin series creates a closer approximation to the original function around the point $x = 0$. For the latter, it is useful to consider the partial sums of a Maclaurin series, which are

$$f_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!}x^i \qquad \text{for all } n \geq 0 \tag{10.13}$$

Example 40 (Maclaurin Series Expansion)
Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = (1 + x)^3$. This is clearly infinitely differentiable. We repeat the process of differentiating repeatedly and setting x to 0 in each derivative expression:

Figure 10.1: Successive approximations from the Maclaurin series expansion of $f(x) = (1+x)^3$ where $f_0(x) = 1$, $f_1(x) = 1 + 3x$ and $f_2(x) = 1 + 3x + 3x^2$.

1. $f(0) = 1$ *so* $a_0 = 1$,
2. $f'(x) = 3 \cdot (1+x)^2$ *so* $f'(0) = 3$ *and* $a_1 = \frac{3}{1!} = 3$,
3. $f''(x) = 3 \cdot 2(1+x)$ *so* $f''(0) = 6$ *and* $a_2 = \frac{6}{2!} = 3$, *and*
4. $f'''(x) = 3 \cdot 2 \cdot 1$ *so* $f'''(0) = 6$ *and* $a_3 = \frac{6}{3!} = 1$.

Higher derivatives are all equal to the constant 0 *function and so* a_n *equals* 0 *for all* $n \geq 4$. *Therefore, the series expansion of* f *is the following:*

$$(1+x)^3 = 1 + 3x + 3x^2 + x^3$$

In this case, since the series expansion is finite, it involves only finitely many summands.[3] *In such cases, we know that the Maclaurin series will be an accurate representation of function* f *and will converge for all* x. *Therefore, the radius of convergence for this series is infinite.*
We next consider the partial sums for this series expansion:

$$\begin{aligned} f_0(x) &= 1 \\ f_1(x) &= 1 + 3x \\ f_2(x) &= 1 + 3x + 3x^2 \\ f_3(x) &= 1 + 3x + 3x^2 + x^3 \end{aligned}$$

We can plot these along with $f(x)$ *in Figure 10.1, and we see that each partial sum is a successively better approximation to* $f(x)$ *around the point* $x = 0$.

Next, we study the series expansion of a function that captures the behavior of the geometric series.

Example 41 (Series Expansion for Geometric Series)
Let $f : (-\infty, 1) \to \mathbb{R}$ *be given by* $f(x) = (1-x)^{-1}$. *Noting that* $1 + f(x) = x/(1-x)$ *and that* 1 *equals* x^0, *we would expect that this function captures the geometric series that starts with a term* x^0 *and not with* x^1. *We then might expect that* a_i *equals* 1 *for all* i *in the series expansion of this function. Let us verify this:*

1. $f(0) = 1$, *so far so good!*
2. $f'(x) = -(1-x)^{-2}(-1) = (1-x)^{-2}$ *so* $f'(0) = 1$

[3] Of course, this power series representation could also have been computed here using the Binomial Theorem. But the power series expansion works for all functions that are infinitely differentiable.

3. $f''(x) = -2!(1-x)^{-3}(-1) = 2!(1-x)^{-3}$ *so* $f''(0) = 2!$

Differentiating repeatedly, we get:

$$f^{(n)}(x) = n!(1-x)^{-(n+1)}.$$

Therefore, we conclude that

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{n!(1)^{-(n+1)}}{n!} = 1$$

Thus $(1-x)^{-1} = \sum_{i=0}^{\infty} x^i$ *confirms that* f *can be represented as an infinite geometric series in* x.
We can check convergence of this series expansion by using the absolute convergence version of D'Alembert's ratio test in Lemma 8, where $b_n = x^n$ *are the series terms:*

$$\lim_{n\to\infty}\left|\frac{b_{n+1}}{b_n}\right| = \lim_{n\to\infty}\left|\frac{x^{n+1}}{x^n}\right| = |x|.$$

Thus convergence is given by $|x| < 1$ *and a radius of convergence of 1 for this power series.*

Example 42 (Series Expansion for $\ln(1+x)$)
Let $f : (-1, \infty) \to R$ *be given by* $f(x) = \ln(1+x)$. *We compute the derivatives at* 0*:*

1. $f(0) = 0$ *because* $\ln 1 = 0$,
2. $f'(x) = (1+x)^{-1}$ *so* $f'(0) = 1$, *and*
3. $f''(x) = (-1)(1+x)^{-2}$ *so* $f''(0) = -1$.

For the n*th derivative of* f *we compute that:*

$$f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n}$$

and from this we infer that

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{(-1)^{n-1}}{n}$$

Therefore, the Maclaurin series for $f(x) = \ln(1+x)$ *is:*

$$\ln(1+x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

Considering the summands of this series $b_n = \dfrac{(-1)^{n-1}x^n}{n}$, *we apply Lemma 8 and see that*

$$\lim_{n\to\infty}|\frac{b_{n+1}}{b_n}| = |x|$$

Therefore, the convergence condition is $|x| < 1$ *and the radius of convergence is therefore* 1.

We can also see that this radius cannot be extended. For example, if we were to set $x = -1$*, then we would get* $\ln 0 = -\infty$ *and a corresponding power series of:*

$$\ln 0 = -\sum_{n=1}^{\infty} \frac{1}{n}$$

which we already know diverges.
On the other hand, when we consider $x = 1$*, the series expansion results in the alternating harmonic series and so we get a very nice convergence result:*

$$\ln 2 = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \tag{10.14}$$

This means that the convergence and divergence behavior is not the same on the entire boundary of the convergence radius, here $\{-1, 1\}$*.*

10.3 Taylor Series

The Taylor series generalise the Maclaurin series so that c in the powers $(x-c)^n$ can be non-zero. This is useful as it allows us to approximate a function at a point c through partial sums of the Taylor series or to represent a function at point c through the entire Taylor series. This is a win-win, since the techniques and approaches for the Maclaurin series transfer very well to the Taylor series:
The general form of the Taylor series for a function $f\colon \mathbb{R} \to \mathbb{R}$ that is infinitely differentiable at point c in $\mathbb{R}$ is derived as

$$\begin{aligned} f(x) &= f(c) + \frac{f^{(1)}(c)}{1!}(x-c) + \frac{f^{(2)}(c)}{2!}(x-c)^2 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n \end{aligned} \tag{10.15}$$

Setting $c = 0$ in (10.15) gives us a series expansion around 0 and this then recovers the Maclaurin series as a special case. Note that the coefficients for the Taylor series are very similar to those of the Maclaurin series, expect that the derivatives are evaluated at c, rather than at 0.

Example 43 (Taylor Series of a Trigonometric Function and ln)
Let's compute the Maclaurin series of $\sin x$*. We have:*

$$\sin' x = \cos x, \quad \sin'' x = -\sin x, \quad \sin^{(3)} x = -\cos x, \quad \sin^{(4)}(x) = \sin x.$$

Thus generally:

$$\sin^{(4n+1)}(0) = 1, \quad \sin^{(4n+2)}(0) = 0, \quad \sin^{(4n+3)}(0) = -1, \quad \sin^{(4n)}(0) = 0.$$

Therefore, the Maclaurin series for $\sin x$ *is given by*

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots.$$

Figure 10.2: Successive approximations from the Taylor series expansion of $f(x) = \ln x$ around the point $x = 2$.

Next consider the function given by $f(x) = \ln x$ and defined on the positive reals. We would not be able to create a Maclaurin series for this, since $\ln 0$ is not defined (the function has a singularity at that point). So this is an example where an expansion around another point is required. We compute here the Taylor expansion around point $x = 2$. First, the derivatives are:

1. $f(x) = \ln x$
2. $f'(x) = \frac{1}{x}$
3. $f''(x) = -\frac{1}{x^2}$
4. $f'''(x) = 2!(-1)^2\frac{1}{x^3}$
5. $f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{x^n}$ *for* $n > 0$

Second, we can evaluate these derivatives at $x = 2$ and get

$$f(x) = \ln 2 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n2^n}(x-2)^n$$

We can now show how this Taylor series approximates the function $f(x)$ around the point $x = 2$ in the same way as the Maclaurin series does around the point $x = 0$. Figure 10.2 shows the first four partial sums

$$f_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(2)}{i!}(x-2)^i$$

from the Taylor series for $f(x) = \ln x$ around the point $x = 2$, where:

$$\begin{aligned}
f_0(x) &= \ln 2 \\
f_1(x) &= \ln 2 + \frac{x-2}{2} \\
f_2(x) &= \ln 2 + \frac{x-2}{2} - \frac{(x-2)^2}{8} \\
f_3(x) &= \ln 2 + \frac{x-2}{2} - \frac{(x-2)^2}{8} + \frac{(x-2)^3}{24}
\end{aligned}$$

As was the case for Maclaurin series, we still need to be aware of the radius of convergence for this series. Fortunately, the techniques for calculating this radius of convergence are identical for Taylor series!

Taking the absolute ratio test for terms

$$b_n = \frac{(-1)^{n-1}}{n2^n}(x-2)^n$$

where we can ignore the $\ln 2$ *term since we are looking for the limit as* $n \to \infty$*:*

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{b_{n+1}}{b_n}\right| &= \lim_{n\to\infty}\left|\frac{(x-2)^{n+1}/(n+1)2^{n+1}}{(x-2)^n/n2^n}\right| \\ &= \lim_{n\to\infty}\frac{|x-2|}{2}\frac{n}{n+1} \\ &= \frac{|x-2|}{2}\end{aligned}$$

This gives us a convergence condition of $|x-2| < 2$ *and a radius of convergence of* 2 *for this power series. In general, the radius of convergence is limited by the nearest singularity of the function (if there is one), such as* $x = 0$ *in this case.*

10.3.1 Differentiation and Integration of Power Series

Within the radius of convergence, a power series is continuous and it can be differentiated and integrated term by term. We state the theorem, skipping its proof.

Theorem 39
Suppose

$$f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$$

with radius of convergence $r > 0$*, i.e., the power series converges absolutely for* $|x| < r$*. Then* $f(x)$ *is continuous for* x *with* $|x-x_0| < r$ *and moreover* f *is differentiable and integrable with*

$$f'(x) = \sum_{n=0}^{\infty} n a_n (x-x_0)^{n-1}$$

$$\int_{x_0}^{x} f(t)\,dt = \sum_{n=0}^{\infty} a_n (x-x_0)^{n+1}/(n+1),$$

for $|x-x_0| < r$*, i.e., the power series can be differentiated and integrated term by term.*

Corollary 3
The theorem implies, by differentiating $f(x)$ *recursively* n *times, that*

$$f^{(n)}(x_0) = n!a_n,$$

i.e., $a_n = f^{(n)}(x_0)/n!$*. It follows that the power series expansion of* f *around* x_0 *is unique and is given by its Taylor series expansion.*

10.4 Power Series Solution of ODEs

Consider the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = ky \tag{10.16}$$

for a constant k in $\mathbb{R}$ where $y = f(x)$ and we mean to solve equation (10.16) for f, given the **boundary condition** that $f(0) = 1$, i.e. that $y = 1$ when $x = 0$. The idea is to assume that f satisfies the equation in (10.6) so that we seek the **series solution**

$$y = \sum_{i=0}^{\infty} a_i x^i \tag{10.17}$$

This means that we want to determine real numbers for each of the infinitely many coefficients a_i with $i \geq 0$. We note that the boundary condition $f(0) = 1$ determines the value of a_0 to be 1 since $f(0) = a_0$. To determine values for the remaining coefficients, we can differentiate the right hand side in (10.17) and set it equal to ky since the equation in (10.16) it meant to hold:

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \Big(\sum_{i=0}^{\infty} a_i \cdot x^i\Big)' && \text{by (10.17)} \\
&= \sum_{i=1}^{\infty} (i \cdot a_i) \cdot x^{i-1} && \text{differentiating each summand} \\
&= \sum_{i=0}^{\infty} ((i+1) \cdot a_{i+1}) \cdot x^i && \text{changing the index range} \\
&= k \cdot \sum_{i=0}^{\infty} a_i \cdot x^i && \text{by (10.16)} \\
&= \sum_{i=0}^{\infty} (k \cdot a_i) \cdot x^i && \text{moving scalar } k \text{ under the infinite sum}
\end{aligned} \tag{10.18}$$

Matching coefficients Two power series are equal, by Corollary 3, iff all of their coefficients are equal. Note that this is equivalent to saying that the constant 0 function has only one representation as a power series, the one in which all coefficients a_i equal 0. Therefore, from (10.18) we get the recurrence relations:

$$(i+1) \cdot a_{i+1} = k \cdot a_i \qquad \text{for all } i \geq 0 \tag{10.19}$$

We can use this equation repeatedly to determine the value of a_i for $i \geq 1$ as a function of the value for a_0:

$$a_i = \frac{k}{i} a_{i-1} = \frac{k}{i} \cdot \frac{k}{i-1} a_{i-2} = \ldots = \frac{k^i}{i!} a_0 \tag{10.20}$$

Using our boundary condition $f(0) = 1$, we already saw that $a_0 = 1$. This, together with (10.20) gives us the formal solution

$$y = \sum_{i=0}^{\infty} \frac{(kx)^i}{i!} = e^{kx} \tag{10.21}$$

The differential equations above are called **ordinary**. This is meant to be in contrast to so called **partial differential equations** (PDEs) that consider equations that involve partial derivatives of functions on more than one variable. A partial derivative of a function $u(x, y)$, for example, with respect to variable x is essentially the derivative of u if we think of y as being a constant. The method above does not really work for PDEs, unfortunately.

PDEs are their own area of mathematics. PDEs may not have explicit solutions, whereas for some PDEs only very few explicit solutions are known – for example in PDEs for the general relativity theory in physics.

ODEs can also have more than one solution. In fact, depending on the boundary conditions, one or more coefficients may be unconstrained in a solution, they are then **degrees of freedom** for this solution. In fact, without the boundary condition that $f(0) = 1$, the above ODE would have one degree of freedom as we would be free to choose any value for a_0, which would then determine the values of all other coefficients a_i.

Such degrees of freedom will be revisited in the second part of this course module when we will solve systems of **linear** equations.

11 Multivariate calculus

11.1 Introduction

In this chapter, we extend some of the basic results in the study of functions of a single real variable to real-valued functions of several real variables.
We will study functions of the form $f : \mathbb{R}^n \to \mathbb{R}$ where the function has n input variables. These functions arise in machine learning as loss functions and can have thousands of input variables. We will see such functions in the next section that in the steepest descent algorithm for optimisation.
As a first example, consider $f_1(x, y) = x^2 + y^2$, representing a paraboloid as in Figure 11.1. Each horizontal section $z = c \geq 0$ gives a circle $x^2 + y^2 = c$, whereas intersection with each vertical plane $y = c$ gives $z = x^2 + c^2$, which is a parabola, hence the name paraboloid.

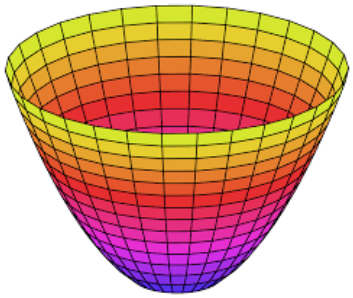


Figure 11.1: A paraboloid: $z = f_1(x, y) = x^2 + y^2$

As a second example, consider $f_2(x, y) = \frac{x^2}{a^2} - \frac{y^2}{b^2}$, representing a hyperboloid as in Figure 11.2. In this case, each horizontal plan $z = c$ intersects the hyperboloid in the hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = c$, hence the name hyperboloid.

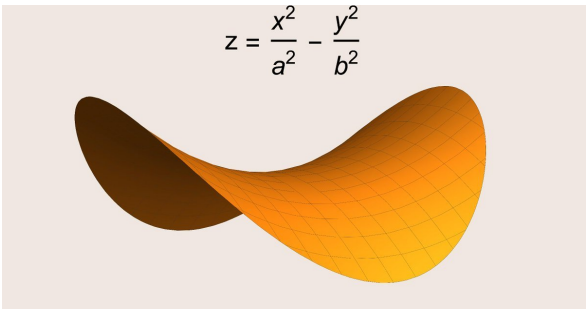


Figure 11.2: A hyperboloid: $z = f_2(x, y) = \frac{x^2}{a^2} - \frac{y^2}{b^2}$

11.1.1 Partial derivatives

If $f : \mathbb{R}^n \to \mathbb{R}$ with $(x_1, x_2, \cdots, x_n) \mapsto f(x_1, x_2, \cdots, x_n)$ then we define its partial derivatives (if they exist) as follows:

$$\frac{\partial f}{\partial x_i} = \frac{df}{dx_i}, \text{ assuming all variables other than } x_i \text{ are fixed.}$$

Example 44
For $f(x,y) = x^2 + y^2$ we have $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = 2y$

Example 45
For $f(x,y) = x^2/a^2 - y^2/b^2$ we have $\frac{\partial f}{\partial x} = 2x/a^2$ and $\frac{\partial f}{\partial y} = -2y/b^2$.

Example 46
For $f(x,y) = x^3y^2$ we have $\frac{\partial f}{\partial x} = 3x^2y^2$ and $\frac{\partial f}{\partial y} = 2x^3y$.

Definition 31
If $f : \mathbb{R}^n \to R$ has partial derivatives, the vector $\nabla f = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n})$ is called the gradient *of f at x.*

The gradient generalises the notion of derivative of a real-valued function of a single variable. It gives the direction of greatest growth of f at x.

Example 47
For the function in Example 44, we have $\nabla f = (2x, 2y)$, whereas for that of Example 46 we have $\nabla f = (3x^2y^2, 2x^3y)$.

The second partial derivatives are defined in a similar manner but now we have mixed second derivatives too. For example, in Example 46, we have:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}(3x^2y^2) = 6xy^2$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}(2x^3y) = 2x^3$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}(2x^3y) = 6x^2y$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}(3x^2y^2) = 6x^2y$$

Notice that in the above example we have

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x},$$

i.e., the order of taking the mixed derivative does not matter in this example. This is not in general true.

Exercise 30
Show that for the function

$$f(x,y) = \begin{cases} xy(x^2-y^2)/(x^2+y^2) & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

we have $\frac{\partial f}{\partial x}(0,y) = -y$ *for all* y *and* $\frac{\partial f}{\partial y}(x,0) = x$ *for all* x*. Deduce that*

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$$

However, under some very mild conditions we do have equality of the mixed derivatives as follows.

Proposition 40
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ *has, for some* $i \neq j$*, first partial derivatives* $\frac{\partial f}{\partial x_i}$ *and* $\frac{\partial f}{\partial x_j}$ *in some open disk* $\{x \in \mathbb{R}^n : |x-y| < r\}$*, where* $y \in \mathbb{R}^n$ *is a given point,* $|x-y| = \sqrt{\sum_{m=1}^n (x_m - y_m)^2}$ *and* $r > 0$*. If the second partial derivatives* $\frac{\partial^2 f}{\partial x_i \partial x_j}$ *and* $\frac{\partial^2 f}{\partial x_j \partial x_i}$ *are continuous at* y*, then* $\frac{\partial^2 f}{\partial x_i \partial x_j}(y) = \frac{\partial^2 f}{\partial x_j \partial x_i}(y)$.

From now on we assume we deal with well-behaved functions for which all mixed derivatives are always the same.

11.2 Additional material: Taylor series for multivariate functions

Recall the Taylor series with remainder for a function f of one variable:

$$f(y) = f(x)+f'(x)(y-x)+f''(x)(y-x)^2/2!+\cdots+(y-x)^{(n-1)}f^{(n-1}(x)/(n-1)!+(y-x)^{(n)}f^{(n}(c)/(n)!$$

where c is between x and y.
Since we are assuming the function has derivatives of all order, in particular the nth derivative $f^{(n)}$ is continuous in a closed interval containing x and y and thus it is bounded above. So the last term is bounded by $(y-x)^n M/n!$ where M is the maximum of $f^{(n)}$ in this closed interval. So, we can write the above Taylor series as

$$f(y) = f(x) + f'(x)(y-x) + f''(x)(y-x)^2/2! + \cdots + \frac{f^{(n-1}(x)(y-x)^{n-1}}{(n-1)!} + \mathcal{O}(|y-x|^n)$$

where $\mathcal{O}(|y-x|^n)$, called the *big O* notation, means that the reminder is bounded by a constant multiple of $|y-x|^n$, i.e., there exists $K > 0$ such that

$$\left| f(y) - \left(f(x) + f'(x)(y-x) + f''(x)(y-x)^2/2! + \cdots + (y-x)^{(n-1)} \frac{f^{(n-1}(x)}{(n-1)!} \right) \right| \leq K(|y-x|^n)$$

Since y is near x, this gives the order of error when y is close to x.
For a multivariate function, we also have a Taylor series. In practice, we will not need to go beyond the second partial derivatives, as we will see in the next section. For a function of two variables, $f(x_1, x_2)$ the Taylor series up to the second order terms can be written as follows. Assume $x, y \in \mathbb{R}^2$, where x is a given point and y is close to x

$$f(y) = f(x) + \left[(y_1 - x_1)\frac{\partial f}{\partial x_1} + (y_2 - x_2)\frac{\partial f}{\partial x_2}\right]$$
$$+\frac{1}{2}\left[(y_1 - x_1)^2\frac{\partial^2 f}{\partial x_1^2} + (y_2 - x_2)^2\frac{\partial^2 f}{\partial x_2^2} + (x_1 - y_1)(x_2 - y_2)\frac{\partial^2 f}{\partial x_1 \partial x_2} + (x_1 - y_1)(x_2 - y_2)\frac{\partial^2 f}{\partial x_2 \partial x_1}\right]$$
$$+O(|y - x|^3).$$

The expression in the first square bracket above can be seen to be simply the scalar product $(y - x) \cdot \nabla f(x)$ of $y - x$ and $\nabla f(x)$. Since the mixed partial derivatives, we assume, are equal, we can thus rewrite the above expression as:

$$f(y) = f(x) + (y - x) \cdot \nabla f(x)$$
$$+\frac{1}{2}\left[(y_1 - x_1)^2\frac{\partial^2 f}{\partial x_1^2} + (y_2 - x_2)^2\frac{\partial^2 f}{\partial x_2^2} + 2(x_1 - y_1)(x_2 - y_2)\frac{\partial^2 f}{\partial x_2 \partial x_1}\right] + O(|y - x|^3). \quad (11.1)$$

11.3 Critical points of a multivariate function

A *critical point* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a point $x \in \mathbb{R}^n$ such that $\nabla f(x) = 0$. This extends the notion of a critical point x of a function of one variable for which $f'(x) = 0$.
Let's recall how we determine the nature of a critical point of a function of a single variable as minimum, maximum or point of inflection. We assume the function is analytic, so it has derivatives of all order.
If x is a critical point of $f : \mathbb{R} \to \mathbb{R}$, i.e., $f'(x) = 0$, we can use the Taylor series to find the value of f at point y near x:

$$f(y) = f(x) + f'(x)(y - x) + f''(x)(y - x)^2/2 + \cdots + (y - x)^n f^{(n)}(x)/(n)! + \cdots$$

Let n be the smallest integer such that $f^{(n)}(x) \neq 0$. (Since f is analytic, such n always exists.) As $f'(x) = 0$ we obtain near x:

$$f(y) \approx (y - x)^n f^{(n}(x)/(n)!$$

So the behaviour of the function at y near the critical point x is determined by the term $f^{(n)}(x)(y - x)^n$.
If n is even, and $f^{(n)}(x) > 0$ it follows that the function is increasing at x_0 since the term $(y - x)^n$ is positive; so x is a minimum. If n is even, and $f^{(n)}(x) < 0$ it follows that the function is decreasing at x, so x is a maximum. If n is odd, then if $f^{(n)}(x) > 0$ then near x the function behaves like the function $(y - x)^n$ equivalently like $(y - x)^3$, whereas if $f^{(n)}(x) < 0$ then near x the function behaves like $-(y-x)^n$ equivalently like $-(y - x)^3$, i.e., it has a point of inflection.

Characterising critical points

In higher dimensions, a generic or non-degenerate critical point of a multivariate function is either a minimum (as in Example 44), a maxima or a saddle as in Example 45. We will define below what we mean by "generic" or "non-degenerate".
We will present the analysis for $n = 2$ as the formulation for $n > 2$ is entirely similar. Notice that in Example 44, we have a critical point at $(0,0)$ where the gradient vanishes, i.e., $\nabla f(0,0) = 0$. As seen in Figure 11.2, this critical point in a minimum. Now observe that we have

$$\frac{\partial^2 f}{\partial x^2}(0,0) = \frac{\partial^2 f}{\partial y^2}(0,0) = 2 > 0$$

Thus, the situation at a minimum critical point is like the minimum x_0 of a function f at one dimensional case where $f'(0) = 0$ and $f''(x_0) > 0$, except that we now have two second order partial derivatives.
For the maximum of the function $f(x,y) = -x^2 - y^2$, the situation is similar with

$$\frac{\partial^2 f}{\partial x^2}(0,0) = \frac{\partial^2 f}{\partial y^2}(0,0) = -2 < 0$$

Turning now to Example 45, we see that for the critical point $x = y = 0$, we have

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 2/a^2, \quad \frac{\partial^2 f}{\partial y^2}(0,0) = -2/b^2,$$

in which along one direction (namely the x-axis or the plane $y = 0$) the function is a minimum while along another direction, namely the y-axis or the plane $x = 0$ the function is a maximum. Clearly, this situation can only occur for $n > 1$.

While in these simple cases the mixed partial derivative is zero, in general this is not the case, and we need to find out conditions in which we have a minimum or maximum when the mixed derivative is not zero.

These conditions can be deduced from the Taylor series, Equation 11.1, which, by ignoring $O(|x-y|^3)$, at y near the critical point $x = (x_1, x_2)$ is reduced to

$$f(y) - f(x) = \frac{1}{2}\left[(y_1 - x_1)^2 \frac{\partial^2 f}{\partial x_1^2} + (y_2 - x_2)^2 \frac{\partial^2 f}{\partial x_2^2} + 2(x_1 - y_1)(x_2 - y_2)\frac{\partial^2 f}{\partial x_2 \partial x_1}\right] \quad (11.2)$$

Therefore, the task is to see if the expression above is positive, negative or changes sign when y is close to the critical point x. Note that the second order partial derivatives in Equation 11.2 are evaluated at the critical point x and thus they are given by three real numbers a, b, c as follows:

$$a := \frac{\partial^2 f}{\partial x_1^2}, \quad b := \frac{\partial^2 f}{\partial x_2 \partial x_1} \quad c := \frac{\partial^2 f}{\partial x_2^2}$$

Thus, Equation 11.2 can be written as

$$f(y) - f(x) = \frac{1}{2}\Big[a(y_1 - x_1)^2 + 2b(x_1 - y_1)(x_2 - y_2) + c(y_2 - x_2)^2\Big] \tag{11.3}$$

Since $y \neq x$, we can assume, say, $y_2 \neq x_2$ (the case in which $y_1 \neq x_1$ is similar). Dividing the RHS of Equation 11.3 by $(y_2 - x_2)^2/2$, we obtain:

$$a[(y_1 - x_1)/(y_2 - x_2)]^2 + 2b[(x_1 - y_1)/(x_2 - y_2)] + c \tag{11.4}$$

Putting $t := (x_1 - y_1)/(x_2 - y_2)$, we obtain the simple and familiar quadratic polynomial in t given by

$$f(y) - f(x) \propto at^2 + 2bt + c$$

where $p \propto q$ means p is positively proportional to q i.e., there exists $k > 0$ such that $p = kq$.

Now the quadratic polynomial $at^2 + 2bt + c$ in the real variable t and thus we have the following conditions to find the nature of the critical point x based on the values of $a,$, b and c.

Proposition 41
The difference $f(y) - f(x)$ will be,

(i) strictly positive, i.e. x is a minimum, if $a > 0$ and $b^2 - ac < 0$. Note that $a > 0$ and $b^2 - ac < 0$ iff $c > 0$ and $b^2 - ac < 0$ iff $a > 0, c > 0$ and $b^2 - ac < 0$ iff $a + c > 0$ and $b^2 - ac < 0$;

(ii) strictly negative, i.e., x is a maximum, if $a < 0$ and $b^2 - ac < 0$. Note that $a < 0$ and $b^2 - ac < 0$ iff $c < 0$ and $b^2 - ac < 0$ iff $a < 0, c < 0$ and $b^2 - ac < 0$ iff $a + c < 0$ and $b^2 - ac < 0$;

(iii) changes sign, i.e., x is a saddle point, if $b^2 - ac > 0$.

Example 48
Consider $f(x_1, x_2) = 3x_1^2 + 7x_1x_2 + 5x_2^2$. Then $\nabla f(x_1, x_2) = 0$ iff $x_1 = x_2 = 0$. We have $\frac{\partial^2 f}{\partial x_1^2} = 6$, $\frac{\partial^2 f}{\partial x_2^2} = 10$ and $\frac{\partial^2 f}{\partial x_1 \partial x_2} = 7$. Thus, $b^2 - ac = 49 - 60 = -11 < 0$ and $a = 6 > 0$. Thus, $(0, 0)$ is a minimum.

Example 49
Consider $f(x_1, x_2) = 3x_1^2 + 8x_1x_2 + 5x_2^2$. Then $\nabla f(x_1, x_2) = 0$ iff $x_1 = x_2 = 0$. We have $\frac{\partial^2 f}{\partial x_1^2} = 6$, $\frac{\partial^2 f}{\partial x_2^2} = 10$ and $\frac{\partial^2 f}{\partial x_1 \partial x_2} = 8$. Thus, $b^2 - ac = 64 - 60 > 0$ Thus, $(0, 0)$ is a saddle.

Returning to Proposition 41, we note that if none of the three conditions hold, then

$$b^2 - ac = 0$$

and the critical point in this case is degenerate. In this case, the second partial derivatives and indeed any higher order derivatives cannot characterise the critical point. One needs to have specific analysis of each such critical point to determine what actually happens at the point. The following examples will illustrate this fact.

Example 50

- *Suppose* $f_1(x_1, x_2) = x_1^2 + x_2^4$. *It is easy to see that the only critical point, namely* $(0,0)$, *in this case is a minimum like the hyperboloid of Example 44. Here, at* $(0,0)$, *we have* $a = 2$, $c = 0$ *and* $b = 0$, *i.e.,* $b^2 - ac = 0$.

- *Consider now* $f_2(x_1, x_2) = x_1^2 - x_2^4$. *Then, it is easy to see that the single critical point at* $(0,0)$ *is a saddle like the hyperboloid in Example 46. At* $(0,0)$, *we have* $a = 2$, $c = 0$ *and* $b = 0$, *i.e.,* $b^2 - ac = 0$.

We see that f_1 *and* f_2 *in the above two examples have completely different critical points, yet have the same values of a,, b and c yielding* $b^2 - ac = 0$. *This shows that when* $b^2 - ac = 0$ *the method of using the Taylor series expansion to determine the nature of a critical point breaks down.*

Extension to $n > 2$: the Hessian matrix

The results stated in Proposition 41 can be extended to higher dimensions, i.e., $n > 2$.

To do this, we first reformulate Proposition 41. We define the *Hessian matrix* H at a critical point (x_1, x_2) as follows:

$$H(x_1, x_2) := \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

In fact, in terms of the Hessian matrix above we can write Equation 11.2 as:

$$f(y) - f(x) = \frac{1}{2} \begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \end{bmatrix}^T \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{bmatrix} \begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \end{bmatrix} + O(|y - x|^3) \quad (11.5)$$

This neat expression is the generalisation of the one dimensional case:

$$f(y) - f(x) = \frac{1}{2} f''(x)(y - x)^2 + O(|y - x|)^3$$

assuming $f'(x) = 0$. The second derivative $f''(x)$ is replaced in the two-dimensional case with the Hessian matrix and $y - x$ with the vector $\begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \end{bmatrix}$ that appears both as a transposed pre-multiplicative factor and a post-multiplicative factor.
Now, we have: $ac - b^2 = \det(H(x_1, x_2)$, i.e., the determinant of $H(x_1, x_2)$ and $a + c = \text{tr}(H(x_1, x_2))$, where *trace* of a square matrix like $H(x_1, x_2)$ is simply the sum of its diagonal elements.
For a 2×2 square matrix such as $H(x_1, x_2)$ the two eigenvalues of the matrix have product $\det(H(x_1, x_2)$ and sum $\text{tr}(H(x_1, x_2))$. In fact, the two eigenvalues of the 2×2 matrix

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

are the roots r_1 and r_2 of the quadratic equation

$$H := \det\begin{bmatrix} a-\lambda & b \\ b & c-\lambda \end{bmatrix} = (\lambda^2 - (a+c)\lambda + (ac-b^2) \equiv (\lambda - r_1)(\lambda - r_2) = \lambda^2 - (r_1+r_2)\lambda + r_1 r_2 = 0,$$

which shows that $r_1 + r_2 = \mathrm{tr}(H)$ and $r_1 r_2 = \det(H)$.

Proposition 41 can now be rephrased as follows:

Proposition 42
The difference $f(y) - f(x)$ will be,

(i) strictly positive, i.e. x is a minimum, if both eigenvalues of $H(x_1, x_2)$ are positive;

(ii) strictly negative, i.e., x is a maximum, if both eigenvalues of $H(x_1, x_2)$ are negative;

(iii) changes sign, i.e., x is a saddle point, if $H(x_1, x_2)$ has a positive and a negative root.

If $f : \mathbb{R}^n \to R$, has a critical point at $(x_i)_{1 \le i \le n}$ then we calculate its $n \times n$ Hessian matrix at $(x_1, \cdots, x_n)$, whose ij entry is given by

$$(H(x_1, \ldots, x_n))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Then we have the following generalisation of Proposition 42:

Proposition 43
The difference $f(y) - f(x)$ will be,

(i) strictly positive, i.e. x is a minimum, if all eigenvalues of $\det(H(x_1, \ldots, x_n))$ are positive;

(ii) strictly negative, i.e., x is a maximum, if all eigenvalues of $H(x_1, \ldots, x_n)$ are negative;

(iii) changes sign, i.e., x is a saddle point, if $\det(H(x_1, \ldots, x_n))$ has both positive and negative eigenvalues with no zero eigenvalue.

Example 51
Consider $f(x_1, x_2) = x_1^3 - 12x_1 + x_2^3 - 3x_2$. Then we have

$$\frac{\partial f}{\partial x_1} = 3x_1^2 - 12 \quad \frac{\partial f}{\partial x_2} = 3x_2^2 - 3$$

Thus, $\nabla f(x_1, x_2) = 0$ for $x_1 = \pm 2$ and $x_2 = \pm 1$.

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{bmatrix} = \begin{bmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{bmatrix}$$

It follows that $(2, 1)$ is a minimum, $(-2, -1)$ is a maximum and $(-2, 1)$ and $(2, -1)$ are saddle points.

12 Numerical Methods

12.1 Introduction

Let us consider a simple **quadratic equation**,

$$5x^2 + 6x + 1 = 0.$$

Such an equation can be solved **analytically**: a solution can be written in "closed form" in terms of known functions, constants, etc. In our example we can apply the quadratic formula

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

with $a = 5$, $b = 6$, and $c = 1$ to find that

$$x = \frac{-6 \pm 4}{10},$$

so there are two solutions, $x_1 = -1$ and $x_2 = -0.2$.
If we are being lazy, we can use SymPy to solve the equation for us:

```
import sympy
x = sympy.symbols('x')
sympy.solve(5. * x**2 + 6. * x + 1., x)
```

The result will be, as expected,

```
[-1.00000000000000, -0.200000000000000]
```

Similar formulae exist for the cubic and quartic equations (although they are quite unwieldy), whereas Galois theory [Ste15] tells us that there is no general formula to solve a quintic equation in terms of radicals (that is, there is no formula using only the arithmetic operations of addition, multiplication, etc., and taking the nth root). In practice it is not always possible to find analytic solutions to mathematical problems. For example, there is no way to analytically derive the optimal weights and biases of a reasonably complex neural network. In these cases we resort to **numerical methods**, on which much of industrial computing relies.

12.2 Additional material: Iteration of functions

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable map with a fixed point x^* satisfying $f(x^*) = x^*$. Under certain conditions, which we discuss below, this fixed point can be obtained by starting with a nearby point x_0 and taking the limit of the sequence of iterates of this point under f:

$$x_0, f(x_0), f(f(x_0)), \cdots, f^n(x_0), \cdots$$

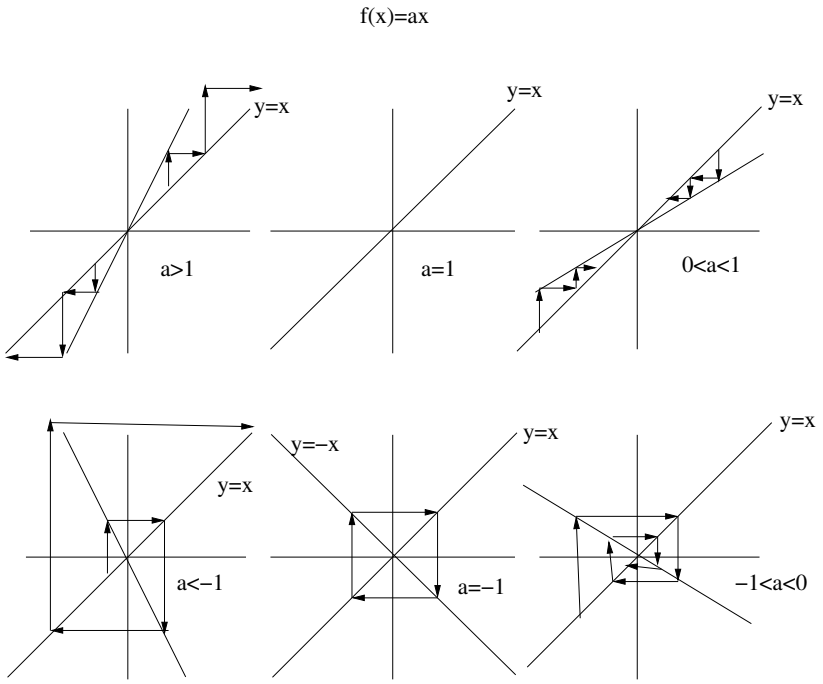


Figure 12.1: Graphical analysis of $x \mapsto ax$ for various ranges of $a \in \mathbb{R}$.

where $f^n(x)$ denotes n application of f on x, $f^n(x) = f(f(\cdots f(x)))$.
There is a simple graphical way to depict the sequence of iterates of f, called *graphical analysis*. Given the graph of a function F we plot the orbit of a point x_0.

- First, superimpose the diagonal line $y = x$ on the graph. (The points of intersection are the fixed points of F.)
- Begin at (x_0, x_0) on the diagonal. Draw a vertical line to the graph of F, meeting it at $(x_0, F(x_0))$.
- From this point draw a horizontal line to the diagonal finishing at $(F(x_0), F(x_0))$. This gives us $F(x_0)$, the next point on the orbit of x_0.
- Draw another vertical line to graph of F, intersecting it at $F^2(x_0))$.
- From this point draw a horizontal line to the diagonal meeting it at $(F^2(x_0), F^2(x_0))$.
- This gives us $F^2(x_0)$, the next point on the orbit of x_0.
- Continue this procedure, known as **graphical analysis**. The resulting "staircase" visualises the orbit of x_0.

The graphical analysis of $f(x) = ax$ for different values of a which lead to different analysis is shown in Figure 12.1. Figure 12.2 depicts the graphical analysis of $f(x) = \cos x$.
Now if x^* is a fixed point of f such that $|f'(x^*)| < 1$ then any sequence of iterates of f starting at a point near x^* will converge to x^*. The fixed point x^* is called a (hyperbolic) *attractor*. This can be seen in Figure 12.1 for $|a| < 1$ and in Figure 12.2

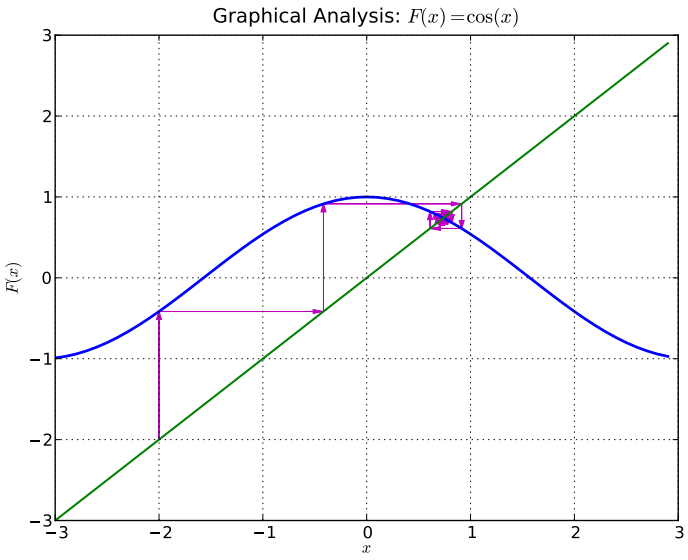


Figure 12.2: Graphical analysis of $C(x) = \cos x$.

for $f(x) = \cos x$ which has a unique attracting fixed point. If on the other hand $|f'(x^*)| > 1$ then starting with a point near x^*, the iterates of f will start to run away from x^* and x^* is called a (hyperbolic) repellor. We see in Figure 12.1 that this happens when $|a| > 1$. If $|f'(x^*)| = 1$, then iterates of nearby point on either side can either get attracted, repelled by x^*, remain stationary or oscillate as in Figure 12.1 when $|a| = 1$.

Exercise 31
Show that $f(x) = x^3$ has three fixed points; determine the nature of each of these fixed points.

Exercise 32
Using the mean value theorem, show that for a fixed point x^* of f with $|f'(x^*)| < 1$, iterates of points starting near x^* converge to x^*.

Exercise 33
Using the mean value theorem, show that for a fixed point x^* of f with $|f'(x^*)| > 1$, iterates of points starting near x^* are repelled by x^*.

12.3 Root finding

Consider a suitably general equation $f(x) = 0$. Let's suppose that we have somehow found an initial approximation to a solution of this equation, x_0. This initial approximation may be poor (it may be a mere guess), so we'll try to improve on it. The equation of a tangent line to $f(x)$ at x_0 is given by

$$y = f(x_0) + f'(x_0)(x - x_0). \tag{12.1}$$

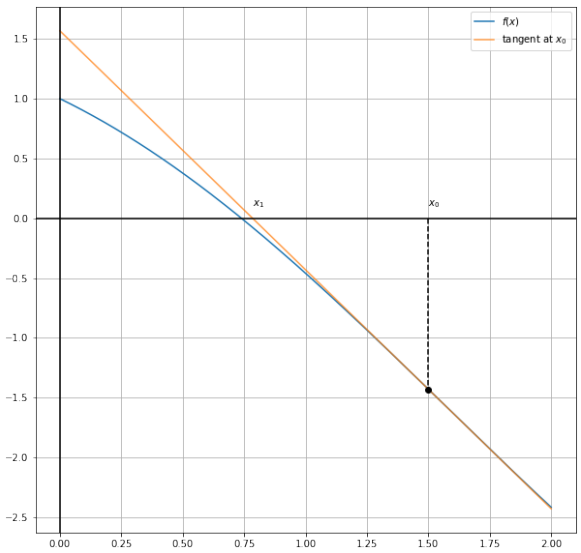


Figure 12.3: Improving on an initial guess.

We can see (Figure 12.3) that this line crosses the x-axis closer to the actual solution to the equation than x_0. Let's call this point where the tangent at x_0 crosses the x-axis x_1. We'll use this point as our new approximation to the solution.
We know the coordinates of this point $(x_1, 0)$, and we know that it lies on the tangent line (12.1). Let us now plug in $x = x_1$ and $y = 0$ into (12.1)

$$0 = f(x_0) + f'(x_0)(x_1 - x_0)$$

and find

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

We would have a problem if $f'(x_0) = 0$. So, we can find the new approximation provided the derivative isn't zero at the original approximation.
We can now repeat the process to find an even better approximation:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Proceeding onwards, we obtain **Newton's method**:

If x_n is an approximation to a solution of $f(x) = 0$ and if $f'(x_n) \neq 0$ the next approximation is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{12.2}$$

12.3.1 Additional material: Root finding by function iteration

Let's now use iteration of functions to capture Newton's method. Define $G_f : (a, b) \to \mathbb{R}$ for an open interval containing the root of f in which $f'(x) \neq 0$ by

$$G_f(x) = x - \frac{f(x)}{f'(x)}.$$

we have $G_f(x) = x$ iff $f(x) = 0$. In addition, the sequence $x_0, x_1 \cdots$ is precisely the sequence $(G_f^n(x_0))_{n\geq 0}$. We have:

$$G_f'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}$$

and thus $G_f'(x) = 0$ if $f(x) = 0$. From our previous work, it follows that the root x with $f(x) = 0$ is an attracting fixed point of G_f.

12.4 Implementation

It is straightforward to implement Newton's method in Python:

```
def newton(f, f_dash, x0, epsilon, max_iter=None):
    xns = [x0]
    iter_count = 0
    while True:
        if iter_count == max_iter: return xns
        f_xn = f(xns[-1])
        if abs(f_xn) < epsilon: return xns
        f_dash_xn = f_dash(xns[-1])
        if f_dash_xn == 0: return None
        xns.append(xns[-1] - f_xn / f_dash_xn)
```

Let's use our implementation to find a solution to

$$f(x) = \cos(x) - x = 0.$$

The derivative of $f(x)$ is given by

$$f'(x) = -\sin(x) - 1.$$

In Python,

```
f = lambda x: np.cos(x) - x
f_dash = lambda x: -np.sin(x) - 1.
```

We have chosen to use $|f(x)| < \epsilon$ as a convergence criterion. Let us see how many iterations we'll need with ϵ = 1e-12:

```
newton(f, f_dash, 1.5, 1e-12)
```

```
[1.5,
 0.7844723977194106,
 0.7395187098320523,
 0.7390851747051963,
 0.739085133215161]
```

Now let's use Newton's method to compute an approximate numeric value of $\sqrt{2}$. (Why approximate? Because we know that $\sqrt{2}$ is irrational, and the decimal expansion of an irrational number never repeats or terminates.)

```
newton(lambda x: x**2 - 2, lambda x: 2. * x, 1.5, 1e-12)
```

```
[1.5,
 1.4166666666666667,
 1.4142156862745099,
 1.4142135623746899,
 1.4142135623730951]
```

12.5 Convergence

According to Taylor's theorem, any function $f(x)$ which has a continuous second derivative can be represented by an expansion about a point that is close to a root of $f(x)$. Suppose this root is α. Then the expansion of $f(\alpha)$ about x_n is

$$f(\alpha) = f(x_n) + f'(x_n)(\alpha - x_n) + R_1 \tag{12.3}$$

where the Lagrange form of the Taylor series expansion remainder is

$$R_1 = \frac{1}{2!} f''(\xi_n)(\alpha - x_n)^2,$$

where ξ_n is between x_n and α.
Since α is the root, (12.3) becomes

$$0 = f(\alpha) = f(x_n) + f'(x_n)(\alpha - x_n) + \frac{1}{2} f''(\xi_n)(\alpha - x_n)^2. \tag{12.4}$$

Dividing (12.4) by $f'(x_n)$ and rearranging gives

$$\frac{f(x_n)}{f'(x_n)} + (\alpha - x_n) = \frac{-f''(\xi_n)}{2f'(x_n)}(\alpha - x_n)^2.$$

Remembering that x_{n+1} is defined by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

we find that

$$\underbrace{\alpha - x_{n+1}}_{\epsilon_{n+1}} = \frac{-f''(\xi_n)}{2f'(x_n)} \underbrace{(\alpha - x_n)}_{\epsilon_n}^2.$$

Taking the absolute value of both sides gives

$$|\epsilon_{n+1}| = \frac{|f''(\xi_n)|}{2|f'(x_n)|} \cdot \epsilon_n^2 \leq M_1 M_2 \epsilon_n^2 = M\epsilon_n^2$$

where $M = M_1 M_2$, M_1 is an upper bound for the continuous function $|f''(x)|/2$ and M_2 is an upper bound for $1/|f'(x)|$ for $x \in [\alpha - r, \alpha + r]$, which exist if the function $f''(x)$ is assumed to be continuous with $f'(x) \neq 0$ on the closed interval $[\alpha - r, \alpha + r]$. This equation shows that the rate of convergence is at least quadratic if

1. $f'(x) \neq 0$ for all $x \in [\alpha - r, \alpha + r] =: I$ for some $r \geq |\alpha - x_0|$;
2. $f''(x)$ is continuous for all $x \in I$;
3. x_0 is "sufficiently" close to the root α.

12.6 Optimization

Let us now consider a different but related problem. Instead of finding the root of $f(x) = 0$, we wish to **minimize** $f(x)$. In symbols, our goal is

$$\min_{x \in \mathbb{R}} f(x).$$

How can we apply Newton's method to this problem?

From our knowledge of calculus, the derivative $f'(x)$ of $f(x)$ is zero at the minimum. So the minimization of $f(x)$ is equivalent to finding the root of $f'(x) = 0$. We can therefore apply Newton's method as follows:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}. \tag{12.5}$$

For example, we know that the minimum of $f(x) = x^2$ is at $x = 0$. Let us confirm this using Newton's method:

$$f'(x) = 2x, \quad f''(x) = 2,$$

and

```
newton(lambda x: 2. * x, lambda x: 2., 1.5, 1e-12)
```

In this case we obtain an exact answer (why?)

```
[1.5, 0.0]
```

12.7 Additional material: Modifications of Newton's method

We have seen that (12.2) involves the first derivative, whereas (12.5) involves the second. In practice these derivatives may be unavailable or too expensive to compute at every iteration (particularly for multivariate problems). We therefore replace the exact $f'(x_n)$ in (12.2) with a **finite-difference approximation**:

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

Approximations of Newton's method, in which the derivatives are replaced by approximate formulae, are known as **quasi-Newton methods**. The particular method that we have just derived, in which

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

is called the **secant method**.
There are many other quasi-Newton methods that work for higher-dimensional problems. We won't consider them in detail here, but simply mention them by name:

- Broyden–Fletcher–Goldfarb–Shanno (BFGS),
- Broyden (and related methods),
- Davidon–Fletcher–Powell (DFP),
- Symmetric Rank 1 (SR1).

In order to better appreciate these methods, we should have an understanding of vector calculus, which we consider in the next chapter (TODO).
Often Newton's method is modified to include a small step size $0 < \gamma \leq 1$ instead of $\gamma = 1$:

$$x_{n+1} = x_n - \gamma \frac{f'(x_n)}{f''(x_n)}.$$

For step sizes other than 1, the method is often referred to as the **relaxed** or **damped** Newton's method.

12.8 Gradient descent

Consider some continuously differentiable real-valued function $f : \mathbb{R} \to \mathbb{R}$. Using a Taylor expansion for $x \in [a, b]$, we obtain

$$f(x + \epsilon) = f(x) + \epsilon f'(x) + \mathcal{O}(\epsilon^2),$$

for all $x \in [a, b]$. Recall that the big O notation $\mathcal{O}$ means that there exists $K > 0$ such that :

$$\left| f(x + \epsilon) - (f(x) + \epsilon f'(x)) \right| \leq K\epsilon^2.$$

That is, in first-order approximation $f(x + \epsilon)$ is given by the function value $f(x)$ and the first derivative $f'(x)$ at x. It is not unreasonable to assume that for small ϵ moving in the direction of negative gradient will decrease f. To keep things simple, we pick a fixed step size $\eta > 0$ and choose $\epsilon = -\eta f'(x)$. Plugging this into the Taylor expansion above we get

$$f(x - \eta f'(x)) = f(x) - \eta f'^2(x) + \mathcal{O}(\eta^2 f'^2(x)).$$

If the derivative $f'(x) \neq 0$ does not vanish, we make progress since $\eta f'^2(x) > 0$. Moreover, we can always choose η small enough for the higher-order terms to become irrelevant. Hence we arrive at

$$f(x - \eta f'(x)) \leq f(x).$$

This means that, if we use

$$x \leftarrow x - \eta f'(x)$$

to iterate x, the value of the function $f(x)$ might decline. Therefore, in gradient descent we first choose an initial value x and a constant $\eta > 0$ and then use them to continuously iterate x until the stop condition is reached, for example, when the magnitude of the gradient $|f'(x)|$ is small enough or the number of iterations has reached a certain value.
For simplicity, we choose the objective function $f(x) = x^2$ to illustrate how to implement gradient descent. Although we know that $x = 0$ is the solution to minimize $f(x)$, we still use this simple function to observe how x changes.

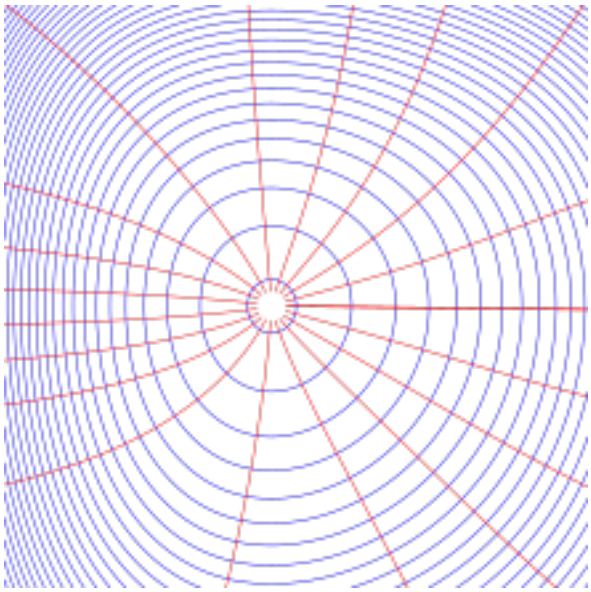


Figure 12.4: The level sets of a function $f : \mathbb{R}^2 \to \mathbb{R}$ in blue colour, sketched in $\mathbb{R}^2$. The red curves are the trajectory of gradient descent along the direction of greatest descent, orthogonal to the level sets. (Source: Wikipedia)

Additional material

We can examine the iterative scheme $x_{n+1} = x_n - \eta f'(x_n)$ as a dynamical system by defining $G_f(x) = x - \eta f'(x)$. The fixed point x^* of G_f is when $f'(x^*) = 0$ and we can determine the nature of this fixed point by looking at $G_f'(x^*) = 1 - \eta f''(x^*)$. Thus, we will have an attractor if $|1 - \eta f''(x^*)| < 1$, equivalently if $-1 < 1 - \eta f''(x^*) < 1$ or $f''(x^*) > 0$ and $\eta f''(x^*) < 2$ as $\eta > 0$.

Higher dimensions

Now consider $f : \mathbb{R}^n \to \mathbb{R}$ and assume it has a continuous derivative with a minimum at x_0. Thus, the gradient descent recursive scheme is given by

$$x_{n+1} = x_n - \eta \nabla f(x_n)$$

The *level sets* of f near a minimum are the surfaces in $\mathbb{R}^n$ given by $f(x) = c$ where $x \in R^n$ and $c \in R$ is a constant. Figure 12.4 shows the level sets in blue of a function $f : \mathbb{R}^2 \to \mathbb{R}$. Each red curve is the envelope of the gradient $\nabla f(x)$, which is orthogonal to all level sets.
Figure 12.5 shows the first iterates of the gradient descent starting at x_0. The sequence converges to the point at which the minimum is attained, where the level set is reduced to a single point at the centre.

Example 52
Consider $f(x,y) = x^2+y^2$, which is a paraboloid with a global minimum at $(x,y) = (0,0)$. The level-sets are concentric circles $x^2 + y^2 = c$, where c

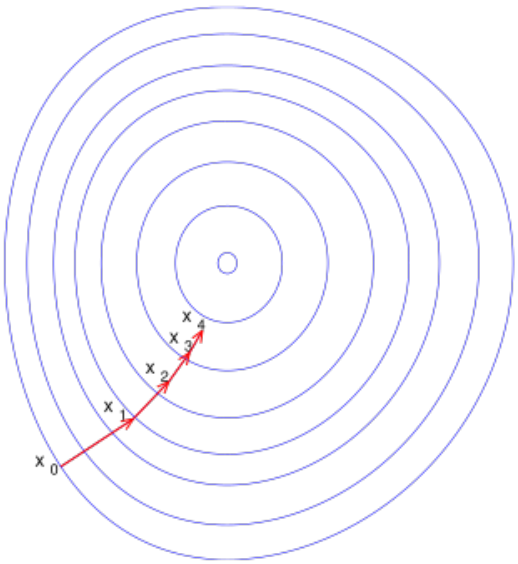


Figure 12.5: The first iterates of the gradient descent algorithm for a function $f : \mathbb{R}^2 \to \mathbb{R}$. (Source: Wikipedia)

$geq 0$. *For any pair* $(x_n, y_n)^T \in \mathbb{R}^2$, *we have*

$$\nabla f(x_n, y_n) = 2(x_n, y_n)^T$$

and thus the recursive expression

$$(x_{n+1}, y_{n+1})^T = (x_n, y_n)^T - \eta \nabla f(x_n, y_n) = (x_n, y_n)^T - 2\eta(x_n, y_n)^T = (1 - 2\eta)(x_n, y_n)^T$$

Exercise 34
Check that for Example 52, the vector $\nabla f(x, y)$ *is perpendicular to the tangent to the level-set of* $f(x, y) = c$ *at the point* $(x, y)^T$.

12.9 History

The name "Newton's method" is derived from Isaac Newton's description of a special case of the method in *De analysi per aequationes numero terminorum infinitas* (written in 1669, published in 1711 by William Jones) and in *De metodis fluxionum et serierum infinitarum* (written in 1671, translated and published as *Method of Fluxions* in 1736 by John Colson).

12.10 Further reading

Numerical Recipes: The Art of Scientific Computing [PTVF07] is a standard reference on numerical methods.

We have also briefly touched on **optimization**—we minimized the loss function of a linear regression and of a neural network. Like linear algebra, optimization is a separate mathematical field, and a prerequisite for a thorough understanding of machine learning. An ambitious reader may wish to consult [GMW82, BV04, SNW12].

12.11 Exercises

Exercise 1

Consider the function

$$f(x) = x^3 - 3x^2 + 3.$$

1. Use Matplotlib to graph f.
2. Use Newton's method to approximate a root of $f(x) = 0$ with $x_0 = 1$ as an initial guess.
3. Use Newton's method to approximate a root of $f(x) = 0$ using $x_0 = \frac{1}{4}$ as an initial guess and observe that the result is different.
4. To understand why the results for $x_0 = 1$ and $x_0 = \frac{1}{4}$ are different, plot the tangent lines to the graph of f at the points $(1, f(1))$ and $(\frac{1}{4}, f(\frac{1}{4}))$. Find the x-intercept of each tangent line and compare the intercepts with the first iteration of Newton's method using the respective initial guesses.

Solution

Our function and its derivative are given by

```
f = lambda x: x**3 - 3. * x**2 + 3.
f_dash = lambda x: 3. * x**2 - 6. * x
```

We will plot the function for $x \in [-10, 10]$:

```
xs = np.linspace(-10., 10., 1000)
```

Here is the code for plotting the function:

```
fs = [f(x) for x in xs]
f_dashs = [f_dash(x) for x in xs]
plt.plot(xs, fs, label='$f(x)$')
plt.plot(xs, f_dashs, label="$f'(x)$")
plt.legend();
```

The result is shown in Figure 12.6.
Applying Newton's method with $x_0 = 1$,

```
newton(f, f_dash, x0=1., epsilon=1e-12)
```

we obtain the root 1.34729:

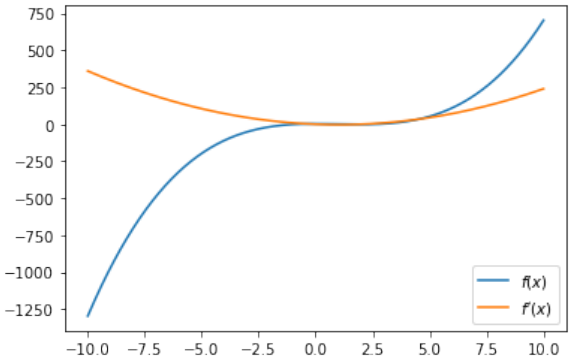


Figure 12.6: A plot of function $f(x)$ and its derivative, $f'(x)$.

```
[1.0,
 1.3333333333333333,
 1.347222222222222,
 1.347296353163868,
 1.3472963553338606]
```

Applying Newton's method with $x_0 = .25$, we obtain a different root, 2.53208:

```
newton(f, f_dash, x0=.25, epsilon=1e-12)
```

```
[0.25,
 2.4047619047619047,
 2.5561934344694976,
 2.5327215804552017,
 2.532089340915494,
 2.5320888862381907]
```

Let us try and understand the reason for this difference. The equations of the tangents at the two initial guesses are given by

```
tangent_at_1 = lambda x: f(1.) + f_dash(1.) * (x - 1.)
tangent_at_0_25 = lambda x: f(.25) + f_dash(.25) * (x - .25)
```

Let's plot them (Figure 12.7):

```
plt.plot(xs, fs, label='$f(x)$')
plt.plot(xs, [tangent_at_1(x) for x in xs], alpha=.5, label="tangent at 1
    ↪ ")
plt.plot(xs, [tangent_at_0_25(x) for x in xs], alpha=.5, label="tangent
    ↪ at 0.25")
plt.axhline(y=0., color='k')
plt.xlim((0., 3.))
plt.ylim((-5., 5.))
plt.legend();
```

The two tangents 'land' on the x-axis next to the two different respective roots, hence the difference of the algorithm's behaviour for the two initial guesses.

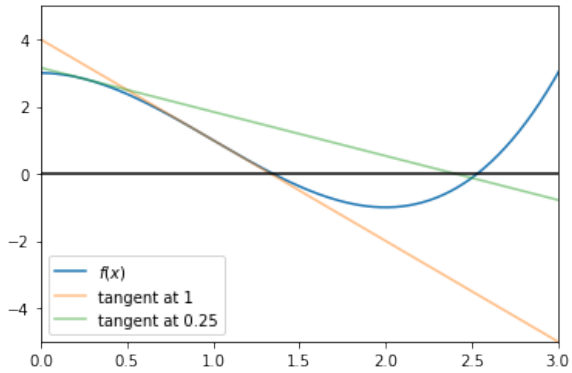


Figure 12.7: The two tangents.

Exercise 2

Use Newton's method to find an approximation to the number π.

12.11.1 Solution

We just need to pick an equation $f(x) = 0$ of which π is a solution. One such choice is $f(x) = \cos\left(\frac{x}{2}\right)$. Its derivative is $f'(x) = -\frac{1}{2}\sin\left(\frac{x}{2}\right)$, and so we can apply `newton` as follows:

```
newton(lambda x: np.cos(.5 * x), lambda x: -5. * np.sin(.5 * x), 1.5, 1e
    ↪ -12)
```

The method converges, but takes 261 iterations.

Exercise 3

The **internal rate of return (IRR)** is the interest rate r that satisfies the equation

$$\frac{F_1}{1+r} + \frac{F_2}{(1+r)^2} + \frac{F_3}{(1+r)^3} + \ldots + \frac{F_N}{(1+r)^N} - C = 0$$

where

$$\begin{aligned} F_t &= \text{cash flow in year } t, \\ N &= \text{number of years}, \\ C &= \text{cost of the investment.} \end{aligned}$$

For most investments, the above equation has a unique solution and therefore the IRR is uniquely defined, but one should keep in mind that this is not always the case. The IRR of a bond is called its **yield**. As an example, consider a 4-year non-callable bond with a 10% coupon rate paid annually and a par value of $1,000. Such a bond has the following cash flows:

Year t	F_t
1	\$100
2	100
3	100
4	1100

Suppose this bond is now selling for \$900. Compute the yield of this bond using Newton's method.

Solution

In order to apply Newton's method, we need the derivative of the objective function:

```
r, F1, F2, F3, F4, C = sympy.symbols('r, F1, F2, F3, F4, C')
sympy.diff(F1 / (1 + r) + F2 / (1 + r)**2 + F3 / (1 + r)**3 + F4 / (1 + r
    ↪ )**4 - C, r)
```

This produces

$$-\frac{F_1}{(r+1)^2} - \frac{2F_2}{(r+1)^3} - \frac{3F_3}{(r+1)^4} - \frac{4F_4}{(r+1)^5}$$

We can now implement both f and f'...

```
def f(x):
    result = 0.
    for i, F in enumerate(Fs):
        result += F / ( (1. + x)**(i + 1) )
    result -= C
    return result

def f_dash(x):
    result = 0.
    for i, F in enumerate(Fs):
        result += ( -(i+1) * F ) / ( (1. + x)**(i + 2) )
    return result
```

...and apply Newton's method:

```
newton(f, f_dash, x0=1., epsilon=1e-12)
```

```
[1.0,
 -2.606060606060606,
 -0.5264454449689513,
 -0.4095240197182348,
 -0.26971984759402673,
 -0.11502093854525602,
 0.026838320916268615,
 0.11177132269329566,
 0.13288395285663232,
 0.13388952282283978,
 0.13389164759298788,
 0.13389164760244182]
```

We see that the yield is equal to 13.38%.

Exercise 4

Implement the secant method in Python. Use it to find the numerical value of $\sqrt{2}$.

Solution

Here is one possible implementation:

```
def secant(f, x0, x1, epsilon, max_iter=None):
    xns = [x0, x1]
    iter_count = 0
    while True:
        if iter_count == max_iter: return xns
        f_xn = f(xns[-1])
        if abs(f_xn) < epsilon: return xns
        xns.append(xns[-1] - f_xn * (xns[-1] - xns[-2]) / (f_xn - f(xns
            ↪ [-2])))
        iter_count += 1
```

We apply it to $f(x) = x^2 - 2$ to find $\sqrt{2}$:

```
secant(lambda x: x**2 - 2, 1.5, 2., 1e-12)
```

```
[1.5,
 2.0,
 1.4285714285714286,
 1.4166666666666667,
 1.4142259414225942,
 1.4142135731001355,
 1.414213562373142]
```

Exercise 5

The well-known Black–Scholes–Merton option pricing formula has the following form for European call option prices:

$$C(K,T) = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}},$$
$$d_2 = d_1 - \sigma\sqrt{T},$$

and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. r in the formula represents the continuously compounded risk-free and constant interest rate and σ is the volatility of the underlying security that is assumed to be constant. Given the market price of a particular option and an estimate for the interest rate r, the unique value of the volatility parameter σ that satisfies the pricing equation above is called the **implied volatility** of the underlying security. Calculate

the implied volatility of a stock currently valued at $20 if a European call option on this stock with a strike price of $18 and a maturity of 3 months (i.e. $T = \frac{1}{4}$ years) is worth $2.20. Assume a zero interest rate and use Newton's method.

Solution

Let us first implement the formula:

```
def d1(S0, K, r, sigma, T):
    return (np.log(S0 / K) + (r + .5 * sigma * sigma) * T) / (sigma * np.
        ↪ sqrt(T))

def d2(S0, K, r, sigma, T):
    return d1(S0, K, r, sigma, T) - sigma * np.sqrt(T)

def C(S0, K, r, sigma, T):
    return S0 * scipy.stats.norm.cdf(d1(S0, K, r, sigma, T)) - K * np.exp
        ↪ (-r * T) * scipy.stats.norm.cdf(d2(S0, K, r, sigma, T))
```

Our objective function is, then

```
def f(x):
    return C(20, 18, 0., x, .25) - 2.20
```

We are lucky that with the secant method we don't need to differentiate the objective function:

```
secant(f, x0=.1, x1=.2, epsilon=1e-12, max_iter=10000)
```

```
[0.1,
 0.2,
 0.24215359973047734,
 0.22404468981138376,
 0.22489211870943285,
 0.22491919581190434,
 0.22491914964087867,
 0.22491914964330545]
```

The implied volatility is, thus, 22.49%.

Exercise 6

Implement the gradient descent method in Python. Use your implementation to find the minimum of $f(x) = x^2$ with $x_0 = 1.5$. Compare the convergence of this method with $\eta = 0.01$ and $\eta = 0.1$. Also compare the convergence of this method with that of Newton's method on this problem.
Now repeat your comparison for $f(x) = x^4 - 3x^3 + 1$ with $x_0 = 2$.

Solution

Here is one possible implementation:

```
def gradient_descent(f_dash, eta, x0, epsilon, max_iter=None):
    xns = [x0]
    iter_count = 0
    while True:
        if iter_count == max_iter: return xns
        f_dash_at_xn = f_dash(xns[-1])
        if abs(f_dash_at_xn) < epsilon: return xns
        xns.append(xns[-1] - eta * f_dash_at_xn)
        iter_count += 1
```

We apply gradient descent with $\eta = 0.01$ and $\eta = 0.1$...

```
gradient_descent_0_01_xns = gradient_descent(lambda x: 2. * x, eta=0.01,
    ↪ x0=1.5, epsilon=1e-12, max_iter=10000)
gradient_descent_0_1_xns = gradient_descent(lambda x: 2. * x, eta=0.1, x0
    ↪ =1.5, epsilon=1e-12, max_iter=10000)
```

...and plot the result (Figure 12.8):

```
plt.plot(gradient_descent_0_01_xns, label='$\eta = 0.01$')
plt.plot(gradient_descent_0_1_xns, label='$\eta = 0.1$')
plt.legend();
```

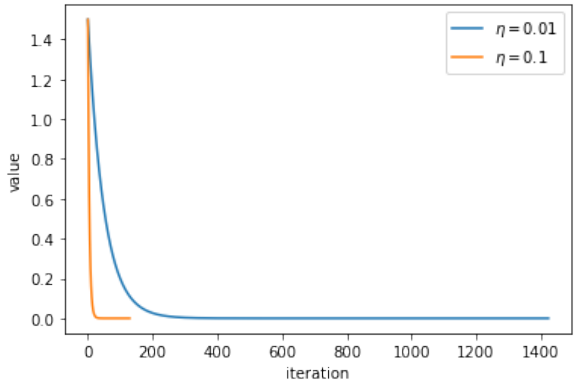


Figure 12.8: The two tangents.

We see that the method with $\eta = 0.1$ converges much faster.
However, Newton's method

```
newton_xns = newton(lambda x: 2. * x, lambda x: 2., x0=1.5, epsilon=1e
    ↪ -12, max_iter=10000)
```

finds the answer in just one iteration:

```
len(newton_xns)
```

```
2
```

We are now going to repeat this procedure for $f(x) = x^4 - 3x^3 + 1$:

```
f = lambda x: x**4 - 3. * x**3 + 1.
f_dash = lambda x: 4. * x**3 - 9. * x**2
```

```
f_dash_dash = lambda x: 12. * x**2 - 18. * x

gradient_descent_0_01_xns = gradient_descent(f_dash, eta=0.01, x0=2.,
    ↪ epsilon=1e-12, max_iter=10000)

len(gradient_descent_0_01_xns)
```

```
132
```

```
gradient_descent_0_1_xns = gradient_descent(f_dash, eta=0.1, x0=2.,
    ↪ epsilon=1e-12, max_iter=10000)

len(gradient_descent_0_1_xns)
```

```
10001
```

```
newton_xns = newton(f_dash, f_dash_dash, x0=2., epsilon=1e-12, max_iter
    ↪ =10000)

len(newton_xns)
```

```
6
```

Let us visualize the results (Figure 12.9):

```
plt.plot(gradient_descent_0_1_xns, label='gradient descent, $\eta=0.1$')
plt.plot(gradient_descent_0_01_xns, label='gradient descent, $\eta=0.01$'
    ↪ )
plt.plot(newton_xns, label="Newton's method")
plt.xlim((0, 125))
plt.legend();
```

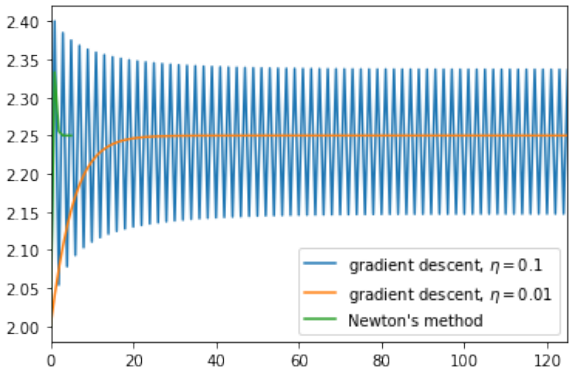


Figure 12.9: The two tangents.

We see that Newton's method converges much faster than gradient descent; and that the step size $\eta = 0.1$ in gradient descent is too large in this instance: the algorithm cannot converge to the solution and instead oscillates around it.

13 Metric Spaces

13.1 Introduction

Recall the definition of convergence of a sequence to a limit:

Definition 32
The sequence $(a_n)_{n\geq 1}$ **converges to a limit** l *in* $\mathbb{R}$ *iff, for all* $\epsilon > 0$*, we can find an* N *in* $\mathbb{N}$ *such that, for all* $n > N$*,* $|a_n - l| < \epsilon$*.*

This definition relies on our ability to express the distance between a_n and l numerically as $|a_n - l|$. Recall also the definition of a limit of a function:

Definition 33
The function $f : [a,b] \to \mathbb{R}$ *has a* **limit** $l \in \mathbb{R}$ *at* $x_0 \in [a,b]$ *if for all* $\epsilon > 0$ *there exists* $\delta > 0$ *such that whenever* $x \in [a,b]$ *and* $|x - x_0| < \delta$*, then* $|f(x) - l| < \epsilon$*.*

This definition also relies on our ability to express distances between the points as numbers: $|x - x_0|$ and $|f(x) - l|$.
We find that the idea of distance is of crucial importance in analysis. Without it, we wouldn't be able to talk about convergence, limits, and, by implication, continuity.
Real numbers have a natural notion of distance. In particular, if $x, y \in \mathbb{R}$, then the distance between them is given by $|x - y|$. (To make this more concrete, the distance between -3 and 5 is $|-3-5| = 8$.) It turns out that we can generalize the idea of distance. Not only can we introduce this idea for other sets; the same set can have different notions of distance. This generalization enables us to talk about convergence and continuity in many different contexts.
Before we generalize an idea, we need to consider its essential properties. The properties of distance are apparent from a mileage chart (Table 13.1).

	London	Manchester	Sheffield
London	0	190	160
Manchester	190	0	40
Sheffield	160	40	0

Table 13.1: A mileage chart.

The first property is that zeros appear down the diagonal and nowhere else; the second is the symmetry of the mileage chart (the distance from London to Manchester is equal to the distance from Manchester to London). The third property is inherent in the chart but is not so immediately seen. The distance from London to Manchester is 190 miles, whereas the distance from London to Sheffield to Manchester is

$160 + 40 = 200$ miles. If we go via some other point we must travel as far or further than we would on the direct route. With our usual straight-line distances this property can be illustrated by means of the three vertices of a triangle (Figure 13.1): for this reason it is known as the **triangle inequality**:

$$\text{distance } x \text{ to } z \leq (\text{distance } x \text{ to } y) + (\text{distance } y \text{ to } z).$$

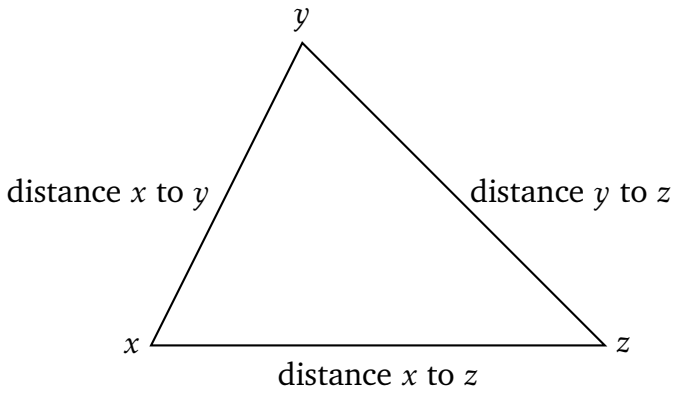


Figure 13.1: The triangle inequality

13.2 Definition of a metric space

Now that we have considered the properties of distance informally, we can, with some additional work, formalize them into a definition:

Definition 34
A pair of objects (X, d) consisting of a nonempty set X and a function $d : X \times X \to \mathbb{R}$ is called a **metric space** *provided that,*

- $d(x, y) = 0$ *iff* $x = y$ *for all* $x, y \in X$ *(***identity of indiscernibles***);*
- $d(x, y) = d(y, x)$ *for all* $x, y \in X$ *(***symmetry***);*
- $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$ *(***subadditivity** *or* **triangle inequality***).*

The function d is called a **distance function** *or* **metric** *on X and the set X is called the* **underlyingset**.

The metric d provides a quantitative measure of the degree of closeness of two points. In particular, the triangle inequality

$$d(x, z) \leq d(x, y) + d(y, z)$$

asserts the transitivity of closeness: if x is close to y and y is close to z, then x is close to z.

We have just performed a standard mathematical ruse of **axiomatization**: the formulation of a system of statements (i.e. **axioms**) that relate a number of primitive terms—in order that a consistent body of propositions (lemmas, theorems) may be derived deductively from these statements. The three axioms above capture the essential nature of "distance". Axioms often sidestep the question of what a particular concept **is** and instead describe how it **behaves**. This is important when axiomatizing a philosophically difficult concept, such as "probability"[1].

13.3 Examples of metric spaces

Example 53
Let X be any nonempty set. For $x, y \in X$ define

$$d(x,y) = \begin{cases} 0, & \text{if } x = y; \\ 1, & \text{if } x \neq y. \end{cases}$$

Then (X,d) is a metric space, called a **discrete space** *or a* **space of isolated points**. *In this metric space no distinct pair of points are "close".*

Example 54
If (X,d) is a metric space, and S a subset of X, we may use in S the very same distance function d, except that it be restricted to S. Obviously S becomes in this way a metric space (S,d)—a **metric subspace** *of (X,d).*

Example 55
We have already seen that $(\mathbb{R},d)$ with $d(x,y) = |x-y|$ is a metric space. Since $\mathbb{Q} \subseteq \mathbb{R}$, $\mathbb{Q}$ being, as usual, the set of rational numbers, $(\mathbb{Q},d)$ is also a metric space—a metric subspace of $(\mathbb{R},d)$.

Example 56
Consider the set of all ordered pairs of real numbers $\mathbb{R}^2$. The **Euclidean distance** *between the points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ is given by $d_2(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$. You can convince yourself that the Euclidean distance is a metric.*

For the sake of brevity we may abuse the notation somewhat and refer to the metric space (X,d) as the metric space X. We will do this when the metric d is clear from the context. In particular, the Euclidean distance is the **usual metric** on $\mathbb{R}^2$. So when we say "the metric space $\mathbb{R}^2$" we imply that the metric is the Euclidean distance.

Example 57
However, the Euclidean distance is not the only metric that can be defined on $\mathbb{R}^2$. For example, here are two other possibilities—the **maximum metric**:

$$d_\infty(x,y) = \max(|x_1-y_1|, |x_2-y_2|)$$

and the **taxi cab metric**:

$$d_1(x,y) = |x_1-y_1| + |x_2-y_2|.$$

[1] Probability was first axiomatized by Andrey Nikolaevich Kolmogorov in *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Kol33].

Some idea of what these metrics are like can be obtained by plotting in the plane the points at distance 1 from the origin (Figure 13.2).

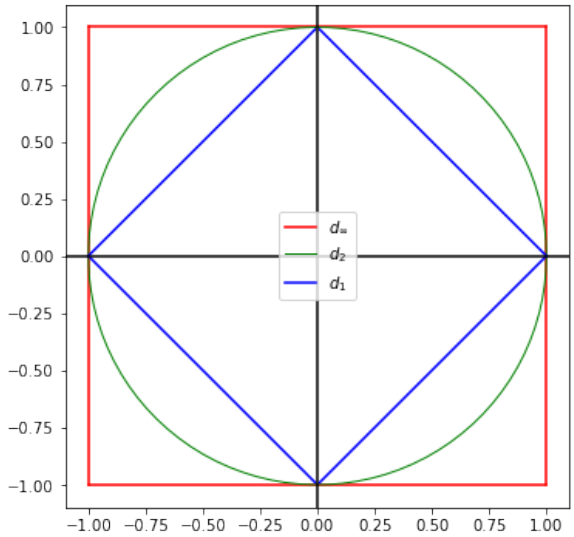


Figure 13.2: Unit circles in $\mathbb{R}^2$ for different metrics.

The above metrics can be extended to $\mathbb{R}^n$ by defining, for $x, y \in \mathbb{R}^n$,

$$d_2(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$d_1(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d_\infty(x,y) = \max_{1 \le i \le n} |x_i - y_i|$$

Note that these three distance functions are *translation invariant* i.e., for all three of them we have $d(x+z, y+z) = d(x,y)$. They are examples of a *norm* on a vector space, such as $\mathbb{R}^n$. A *norm* $\|\cdot\|$ on vectors in $\mathbb{R}^n$ is defined as map $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ satisfying:

- $\|x\| \ge 0$ for all $x \in \mathbb{R}^n$.
- $\|rx\| = |r|\|x\|$ all $x \in \mathbb{R}^n$ and $r \in \mathbb{R}$.
- $\|x+y\| \le \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

Exercise 35
Show that we have $\|x\|_\infty \le \|x\|_2 \le \|x\|_1$ for all $x \in \mathbb{R}^n$, by proving that the squares of these expressions satisfy the above inequalities.

Example 58
We define a metric on the space $\{0,1\}^{\mathbb{N}}$ of infinite sequences of bits of the form

$$b = b_0 b_1 b_2 \ldots$$

where $b_i \in \{0,1\}$ for $i \in \mathbb{N}$. For example

$$00001001100111111\ldots \in \{0,1\}^{\mathbb{N}}$$

Define

$$d(b,b') = \begin{cases} 0 & \text{if } b = b' \\ 1/2^n & \text{if } n \in \mathbb{N} \text{ is the smallest integer with } b_n \neq b'_n \end{cases}$$

Then d clearly satisfies $d(b,b') = 0$ iff $b = b'$, and $d(b,b') = d(b',b)$. Check that d satisfies the triangular inequality and is hence a metric on $\{0,1\}^{\mathbb{N}}$.

Definition 35
We say two metrics d_1 and d_2 on a space X are **equivalent** *if there exist $p > 0$ and $q > 0$ such that*

$$\forall x, y \in X.\ d_1(x,y) \leq p d_2(x,y)\ \&\ d_2(x,y) \leq q d_1(x,y).$$

We say two norms on $\mathbb{R}^n$ are **equivalent** *if they are equivalent as metrics.*

Exercise 36
Show that for all $x \in \mathbb{R}^n$

$$\|x\|_1 \leq \sqrt{n}\|x\|_2 \leq n\|x\|_\infty$$

Then show that the three norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ in $\mathbb{R}^n$ are pairwise equivalent by using Exercise 35.

13.4 Additional material: Continuity

We have already encountered continuity in a very specific context—the Euclidean real line,—it is now time to generalize this definition to metric spaces, of which the Euclidean real line is but one special case.

Definition 36
Let (X,d) and (Y,d') be metric spaces, and let $a \in X$. A function $f : X \to Y$ is said to be **continuous at the point** *$a \in X$ if given $\epsilon > 0$, there is a $\delta > 0$, such that $d'(f(x), f(a)) < \epsilon$ whenever $x \in X$ and $d(x,a) < \delta$.*
The function $f : X \to Y$ is said to be **continuous** *if it is continuous at each point of X.*

Let us consider some trivial examples of continuous functions.

Proposition 44
Let (X,d) and (Y,d') be metric spaces. The **constant function** *$f : X \to Y$ which maps all $x \in X$ to a particular $c \in Y$, $f : x \mapsto c$, is continuous.*

Proof *Let a point $a \in X$ and $\epsilon > 0$ be given. Choose any $\delta > 0$, say $\delta = 1$. Then whenever $d(x,a) < \delta$, we have $d'(f(x), f(a)) = d'(c,c) = 0 < \epsilon$.*

Proposition 45
Let (X,d) be a metric space. Then the **identity function** *$i : X \to X$ which maps each $x \in X$ to itself, $i : x \mapsto x$, is continuous.*

Proof *Let $\epsilon > 0$ be given. Choose $\delta = \epsilon$, then whenever $d(x,a) < \delta$ we have $d(i(x), i(a)) = d(x,a) < \epsilon$.*

In the above proof we could have chosen δ to be any positive number such that $\delta \leq \epsilon$, and the proof would still be valid. The choice of δ need not be a very efficient choice; all that is required is that it "do the job."
The **composition** of two continuous functions is again a continuous function:

Theorem 46
Let (X,d), (Y,d'), and (Z,d'') be metric spaces. Let $f : X \to Y$ be continuous at the point $a \in X$ and let $g : Y \to Z$ be continuous at the point $f(a) \in Y$. Then $gf = g \circ f : X \to Z$, defined by $gf : x \mapsto g(f(x))$, is continuous at the point $a \in X$.

Proof *Let $\epsilon > 0$ be given. We must find a $\delta > 0$ such that whenever $x \in X$ and $d(x,a) < \delta$, then $d''(g(f(x)), g(f(a))) < \epsilon$. Since g is continuous at $f(a)$, there is an $\eta > 0$, such that whenever $y \in Y$ and $d'(y, f(a)) < \eta$, then $d''(g(y), g(f(a))) < \epsilon$. Using the fact that f is continuous at a, we know that given $\eta > 0$, there is a $\delta > 0$, such that $x \in X$ and $d(x,a) < \delta$ imply that $d'(f(x), f(a)) < \eta$ and hence $d''(g(f(x)), g(f(a))) < \epsilon$.*

Corollary 4
Let (X,d), (Y,d'), and (Z,d'') be metric spaces. Let $f : X \to Y$ and $g : Y \to Z$ be continuous. Then $gf : X \to Z$ is continuous.

Exercise 37
Assume that d_1 and d_2 are equivalent metrics on X and d_1' and d_2' are equivalent metrics on X'. Show that $f : (X, d_1) \to (X, d_1')$ is continuous if and only if $f : (X, d_2) \to (X, d_2')$ is continuous.

Exercise 38
Using Exercise 36 and Exercise 37, show that if a function $f : (X^n, \|\cdot\|_a) \to (X^m, \|\cdot\|_b)$, where $a, b \in \{1, 2, \infty\}$, is continuous, then $f : (X^n, \|\cdot\|_c) \to (X^m, \|\cdot\|_d)$ is continuous for any $c, d \in \{1, 2, \infty\}$.

13.5 Additional material: Open balls and neighbourhoods

In the definition of continuity, we were concerned with how the function f transforms the sets of points $x \in X$ with $d(x,a) < \delta$. Let's take a closer look at these objects, which deserve to be named:

Definition 37
Let (X, d) be a metric space. Let $a \in X$ and $\delta > 0$ be given. The subset of X consisting of those points $x \in X$ such that $d(a,x) < \delta$ is called the **open ball about a of radius δ** *and is denoted by $B(a;\delta)$.*

We can now rewrite the definition of continuity in terms of open balls:

Proposition 47
A function $f : (X,d) \to (Y,d')$ is continuous at a point $a \in X$ iff given $\epsilon > 0$ there is a $\delta > 0$ such that

$$f(B(a;\delta)) \subseteq B(f(a);\epsilon).$$

For a function $f : X \to Y$ we have $f(U) \subseteq V$ iff $U \subseteq f^{-1}(V)$, where U and V are subsets of X and Y respectively. Therefore:

Corollary 5
A function $f : (X,d) \to (Y,d')$ is continuous at a point $a \in X$ iff given $\epsilon > 0$ there is a $\delta > 0$ such that

$$B(a;\delta) \subseteq f^{-1}(B(f(a);\epsilon)).$$

For a given point $a \in X$, the open ball $B(a,\delta)$, for each $\delta > 0$, is an example of the type of subset of X that is called a neighbourhood of a:

Definition 38
Let (X,d) be a metric space and $a \in X$. A subset N of X is called a **neighbourhood of** *a if there is a $\delta > 0$ such that*

$$B(a;\delta) \subseteq N.$$

The collection $\mathcal{N}_a$ of all neighbourhoods of a point $a \in X$ is called a **complete system of neighbourhoods** *of the point a.*

A neighbourhood of a point $a \in X$ may be thought of as containing all the points of X that are sufficiently close to a or as "enclosing" a by virtue of the fact that it contains some open ball about a. In particular, for each $\delta > 0$, $B(a;\delta)$ is a neighbourhood of a.

Theorem 48
Let (X,d) be a metric space and $a \in X$. For each $\delta > 0$, the open ball $B(a;\delta)$ is a neighbourhood of each of its points.

Proof *Let $b \in B(a;\delta)$. In order to show that $B(a;\delta)$ is a neighbourhood of b, we must show that there is an $\eta > 0$ such that $B(b;\eta) \subseteq B(a;\delta)$. Since $b \in B(a;\delta)$, $d(a,b) < \delta$. Choose $\eta < \delta - d(a,b)$. If $x \in B(b;\eta)$ then*

$$d(a,x) \leq d(a,b) + d(b,x) < d(a,b) + \eta < d(a,b) + \delta - d(a,b) = \delta,$$

and therefore $x \in B(a;\delta)$. Thus $B(b;\eta) \subseteq B(a;\delta)$ and $B(a;\delta)$ is a neighbourhood of b.

Let us visualize this proof. We start with an open ball $B(a;\delta)$ about a and choose a point $b \in B(a;\delta)$. Then the minimum distance from b to points not in $B(a;\delta)$ is at least $\delta - d(a,b)$. A ball about b of radius $\eta < \delta - d(a,b)$ is contained in $B(a;\delta)$.

13.6 Limits

Definition 39
Let (X,d) be a metric space. Let $x_1, x_2, \ldots$ be a sequence of points of X. A point $x \in X$ is said to be the **limit of the sequence** *$x_1, x_2, \ldots$ if $\lim_{n\to\infty} d(x,x_n) = 0$. In this event, we shall say that the sequence $x_1, x_2, \ldots$* **converges to** *x and write $\lim_{n\to\infty} x_n = x$.*

13.7 Topology and levels of abstractions

A famous aphorism of Butler Lampson goes: "All problems in computer science can be solved by another level of indirection" (the "fundamental theorem of software engineering") [Spi07]. This is often deliberately mis-quoted with "abstraction layer" substituted for "level of indirection". This seems to also be the case in mathematics. We obtained the concept of a metric space when we axiomatized the idea of distance. We can raise the level of abstraction and introduce another notion—that of a **topological space**. A metric space has a notion of distance, while a topological space only has a notion of closeness. If we have a notion of distance then we can say when things are close to each other. However, distance is not necessary to determine when things are close to each other. Therefore every metric space is a topological space, but not every topological space can be thought of as a metric space.
Unfortunately, topological spaces are outside the scope of this course, but it's good to be aware of their existence.

13.8 History

The French mathematician Maurice Fréchet (1878–1973) initiated the study of metric spaces in 1905 with his doctoral dissertation *Sur quelques points du calcul fonctionnel* submitted on 2 April 1906. While he introduced the concept of a metric space, he did not invent the name, which is due to Felix Hausdorff (1868–1942).
Fréchet's thesis concerns "functional operations" and "functional calculus" and is developed from ideas due to Jacques Hadamard (1865–1963) and Vito Volterra (1860–1940). The thesis followed the lead of group theory, which axiomatized algebraic systems, and axiomatized analysis systems. Fréchet required such axiomatization in order to study limits and continuity in a fairly abstract setting. Fréchet's "functional operation" is a numerically valued function defined on arbitrary objects such as points, lines, functions, numbers, surfaces, etc. The "functional calculus" of his thesis is then the systematic study of functional operations.
The theory of metric spaces was vigorously pursued by several Polish mathematicians in the 1920s. A general survey of the results obtained is contained in Wacław Sierpiński's *General Topology* [Sie52] and Kazimierz Kuratowski's *Topology* [Kur66].

13.9 Further reading

Chapters on metric spaces can be found in [KF57, KF70, Bin81, Men90]. Works dedicated specifically to metric spaces include [Cop68, Kap72, Bry85].
The study of topological spaces—called **general topology**—is an interesting and useful subject in its own right. A deeper understanding of topology can help one better appreciate mathematical analysis. The reader can learn general topology from [Dug66, Wil70, Kel55, Mun17], starting perhaps from [Men90].
Metric (and, more generally, topological spaces) can often be quite counterintuitive. It is good to consult *Counterexamples in Topology* [SS95] and inspect some corner

cases.

13.10 Exercises

Exercise 1

Use the fact that

$$d(x,x) \leq d(x,y) + d(y,x)$$

for any x, y in a metric space (X,d) to deduce that $d(x,y) \geq 0$.

Solution

Notice that $d(x,x) = 0$ and $d(y,x) = d(x,y)$. Thus we have

$$0 \leq 2d(x,y).$$

It therefore follows that $d(x,y) \geq 0$.

Exercise 2

Let (X,d) be a metric space. Let k be a positive real number and set $d_k(x,y) = k \cdot d(x,y)$. Prove that (X,d_k) is a metric space.

Solution

TODO.

14 Vector Derivatives

14.1 Introduction

Up to now we have focussed on functions of a single variable. In data science and machine learning we often encounter functions of multiple variables, such as the loss function of a linear regression and neural network.
The extension from one to multiple variables is not difficult, but requires some diligence and care. The easiest way to manage the complexity of multiple variables is to encapsulate it within vectors and matrices—objects studied in the mathematical field of **linear algebra**.
Linear algebra is no less important than calculus. While we have tried to make our notes on vectors and matrices self-sufficient, they do not replace a full linear algebra course.

14.2 Vectors

The set $\mathbb{R}^n$, we recall, consists of all objects of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

in which $x_1, x_2, \ldots, x_n$ are real numbers called the **coordinates** or **components** of $\mathbf{x}$. We refer to such objects as **vectors** as opposed to ordinary real numbers, which we call **scalars**.
If

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

are vectors and $\alpha \in \mathbb{R}$ a scalar, then we define **vector addition** and **scalar multiplication** by

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad \alpha\mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}.$$

These definitions provide the linear structure of $\mathbb{R}^n$ which have a simple geometric interpretation: We also recall that a matrix $M \in \mathbb{R}^{m \times n}$ consists of n vectors in $\mathbb{R}^m$

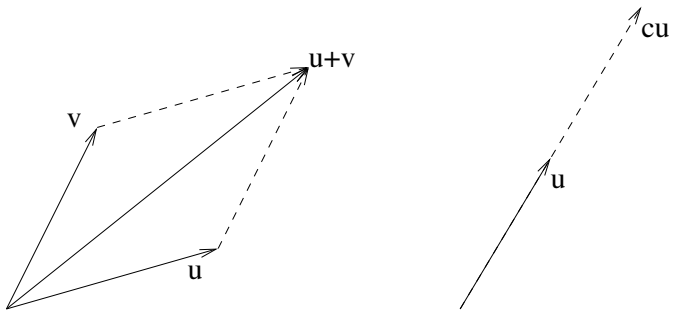


Figure 14.1: Geometric interpretation of addition and scalar multiplication of vectors

giving s rectangular array of mn elements M_{ij} with $1 \leq i \leq m$ and $1 \leq j \leq n$. The *transpose* of M is the matrix $M^T \in \mathbb{R}^{n\times m}$ with $(M)_{ij} = M^T_{ji}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. Given a vector $u \in \mathbb{R}^n$, the action of M on u, written as Mu is defined to be the vector $Mu \in \mathbb{R}^m$ with components $(Mu)_i = \sum_{j=1}^n M_{ij}u_j$ for $1 \leq i \leq m$. If $M \in \mathbb{R}^{m\times n}$ and $N \in \mathbb{R}^{n\times p}$ then the product $MN \in \mathbb{R}^{m\times p}$ with $(MN)_{it} = \sum_{j=1}^n M_{ij}N_{jt}$. We always have $(MN)P = M(NP)$ whenever the matrices M, N and P can be multiplied, i.e., they have pairwise the correct type. For $x, y \in \mathbb{R}^n$, their scalar or dot product is defined by

$$x \cdot y = x^T y = y^T x = \sum_{i=1}^{n} x_i y_i$$

and we always have $x \cdot y = \|x\|_2\|y\| \cos\theta$ where θ is the angle between x and y. (There is a 2d plane that passes through any two vectors x and y and the angle between these vectors lies on this 2d plane.)
We can write partial derivatives in a more concise and elegant manner as follows. For a function $f : \mathbb{R}^n \to \mathbb{R}$ with value $f(\mathbf{x})$ for the vector $\mathbf{x} \in \mathbb{R}^n$, we put:

$$\nabla f = \frac{\partial f}{\partial \mathbf{x}}.$$

If f is an affine map then it is given by $f(\mathbf{x}) = r + \mathbf{x} \cdot \mathbf{c} = r + \mathbf{c}^T\mathbf{x} = r + \mathbf{x}^T \cdot \mathbf{c}$ and we have:

$$\frac{\partial f}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) = \mathbf{c}.$$

On the other hand, suppose $f : \mathbb{R}^n \to \mathbb{R}$ is quadratic function of the form

$$f(\mathbf{x}) = \mathbf{x}^T M\mathbf{x}$$

with $\mathbf{M} \in \mathbb{R}^{n\times n}$ and symmetric $\mathbf{M}^T = \mathbf{M}$. Then we have

$$\frac{\partial f}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) = 2\mathbf{M}\mathbf{x}$$

Exercise 39
Assume $n = 2$. Show that

$$\frac{\partial f}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) = 2\mathbf{M}\mathbf{x}$$

for any symmetric 2×2 matrix $\mathbf{M}$.

14.3 Application: Linear regression

Linear regression [MPV12] is one of the simplest and most useful statistical techniques.
Consider first *simple linear regression*, which consist of finding the line of best fit to a set of points $(x_i, y_i) \in \mathbb{R}^2$. In other words, we would like to model our data points by a straight line. Assume we seek to find the best linear map of the form $y = mx$ to model these points. Our job is to determine m to optimise our model. Thus, we have an error $e_i := mx_i - y_i$ for $i = 1, \ldots, n$, giving us a vector $e = (e_1, \ldots, e_n)^T \in \mathbb{R}^n$. The job is to find m that minimises these errors. We need to compute the total error by combining these n errors, one for each pair of points. Three obvious ways to attempt this are to use one of the three norms for e:

(i) $\|e\|_1 = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |mx_i - y_i|$.

(ii) $\|e\|_\infty = \max_{1\leq i\leq n} |e_i| = \max_{1\leq i\leq n} |mx_i - y_i|$.

(iii) $\|e\|_2 = \sqrt{\sum_{i=1}^n e_i^2} = \sqrt{\sum_{i=1}^n (mx_i - y_i)^2}$.

The first choice (i) would give us a huge problem in minimisation as the absolute value function has a point of non-differnetiability and we need to find the sign of $mx_i - y_i$ for each i, which for large n is inefficient. The second choice would lead to a huge bias for the line of best fit as it would allow outliers in the data to determine the value of m. Thus, the best choice is (iii). We can work with $\|e\|_2^2$ instead of $\|e\|_2$ as these two functions will have the same minimum. We have thus the following loss function $L(D, m)$ with respect to the dataset $D = \{(x_i, y_i) :\leq i \leq n\}$ and parameter m:

$$L(D, m) = \|e\|^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (mx_i - y_i)^2$$

It is now easy to determine the optimum value of m as

$$\frac{dL}{dm} = 2\sum_{i=1}^n x_i(mx_i - y_i) = 0$$

when $m = x \cdot y / x \cdot x$. In addition, we have

$$\frac{d^2L}{dm^2} = 2x \cdot x > 0$$

so our critical point $m = x \cdot y / x \cdot x$ is a minimum as expected.
Next consider the problem of finding the line of best fit of the form $y = c + mx$ i.e., an affine map. So we now have two unknowns m and c and we need partial differentiation to solve the problem. For the loss function we now have:

$$L(D, m, c) = \|e\|^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (c + mx_i - y_i)^2$$

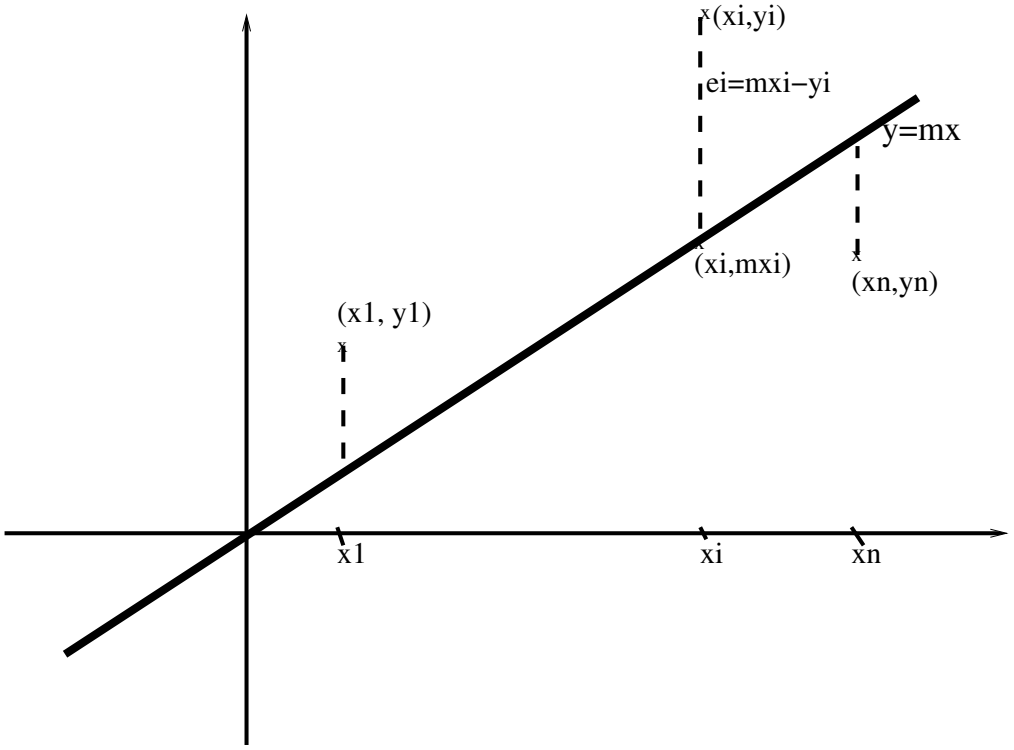


Figure 14.2: Simple linear regression

In finding the critical point, we obtain:

$$\frac{\partial L}{\partial m} = 2\sum_{i=1}^{n} x_i(c + mx_i - y_i) = 0$$

$$\frac{\partial L}{\partial c} = 2\sum_{i=1}^{n} c + mx_i - y_i = 0 \Rightarrow nc = \sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i \Rightarrow c = \overline{y} - m\overline{x}$$

where $\overline{y} = \sum_{i=1}^{n} y_i/n$ and $\overline{x} = \sum_{i=1}^{n} x_i/n$ are the mean of y_i's and x_i's respectively. This gives two equations with two unknowns m and c. A simple calculation shows that

$$\frac{\partial^2 L}{\partial m^2} = 2\sum_{i=1}^{n} x_i^2, \qquad \frac{\partial^2 L}{\partial c^2} = 2n, \qquad \frac{\partial^2 L}{\partial m \partial c} = 2\sum_{i=1}^{n} x_i$$

Exercise 40
Find the critical value of m and show that at the critical point the loss function is a minimum.

Hint.

$$m = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} = \frac{(\sum_i x_i y_i) - n\overline{xy}}{(\sum_i x_i^2) - n(\overline{x})^2}$$

In addition,

$$\frac{1}{4}\left(\left(\frac{\partial^2 L}{\partial c \partial m}\right)^2 - \frac{\partial^2 L}{\partial m^2}\frac{\partial^2 L}{\partial c^2}\right) = \left(\sum_{i=1}^{n} x_i\right)^2 - n\left(\sum_{i=1}^{n} x_i^2\right) = -\sum_{i<j}(x_i - x_j)^2$$

Suppose the data consists of n observations, $\{\mathbf{x}_i, y_i\}_{i=1}^n$. Each observation i includes a scalar label (target) y_i and a column vector $\mathbf{x}_i$ of p features (independent variables). In a **linear regression model**, the target, y_i, is regarded as a linear function of the independent variables:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

or, in vector form,

$$y_i = \beta_0 + \mathbf{x}_i \cdot (\beta_1, \ldots, \beta_p)^T + \epsilon_i,$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a $(p+1) \times 1$ vector of unknown parameters and the random variable ϵ_i represents the error of the ith observation; it accounts for the influences on y_i from sources other than the independent variables $\mathbf{x}_i$.
This model can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{14.1}$$

where $\mathbf{y}$ and ϵ are $n \times 1$ of the labels and errors, and called the **data matrix**, is an $n \times (p+1)$ matrix whose first column has all entries 1. Its ith row starts with 1 followed by $\mathbf{x}_i^\intercal$, which contains the ith observation of all p features:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The **ordinary least squares (OLS)** approach consists in finding such a β in (14.1)—we'll refer to it as $\hat{\beta}$—that will give us the estimate of $\mathbf{y}$,

$$\hat{\mathbf{y}} := \mathbf{X}\hat{\beta},$$

that will minimize a certain loss function of the **residual**

$$\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}},$$

namely the squared error, which we'll now call the **residual sum of squares (RSS)** or the **loss function**,

$$L(D, \hat{\beta}) = \hat{\epsilon}^\intercal \hat{\epsilon} = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)^\intercal \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right).$$

Theorem 49
$\hat{\beta} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$.

Proof *We can rewrite the loss function as*

$$L(D, \hat{\beta}) = \mathbf{y}^\intercal\mathbf{y} - 2\mathbf{y}^\intercal\mathbf{X}\hat{\beta} + \hat{\beta}^\intercal\mathbf{X}^\intercal\mathbf{X}\hat{\beta}.$$

Using the vector notation for vector-by-vector and scalar-by-vector derivatives,

$$\frac{\partial L(D, \hat{\beta})}{\partial \hat{\beta}} = -2\mathbf{X}^\intercal\mathbf{y} + 2\mathbf{X}^\intercal\mathbf{X}\hat{\beta}.$$

When $\frac{\partial L(D,\hat{\beta})}{\partial \beta} = 0$*, we have:*

$$\mathbf{X}^\intercal\mathbf{X}\beta = \mathbf{X}^\intercal\mathbf{y} \tag{14.2}$$

(this is the so-called **normal equation***), hence*

$$\hat{\beta} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}. \tag{14.3}$$

To check that this critical point is a minimum, we examine the Hessian matrix

$$\frac{\partial}{\partial \hat{\beta}}\left(\frac{\partial L(D, \hat{\beta})}{\partial \hat{\beta}}\right) = 2\mathbf{X}^\intercal\mathbf{X}.$$

If $\hat{\beta}'$ *is close to* $\hat{\beta}$*, then the Hessian in the second term in the Taylor series gives*

$$2(\hat{\beta}' - \hat{\beta})^T\mathbf{X}^\intercal\mathbf{X}(\hat{\beta}' - \hat{\beta}) = 2(\mathbf{X}(\hat{\beta}' - \hat{\beta}))^T\mathbf{X}(\hat{\beta}' - \hat{\beta}) \geq 0,$$

since for the vector $\mathbf{z} := \mathbf{X}(\hat{\beta}' - \hat{\beta})$ *we have* $\mathbf{z}^T\mathbf{z} = \mathbf{z} \cdot \mathbf{z} \geq 0$*. Thus,* $\hat{\beta}$ *in Equation 14.3 yields a minimum.*

14.4 Additional material: Backpropagation

Suppose we have two nested functions,

$$z = \sin(\underbrace{\underbrace{3x}_{y=g(x)}}_{z=f(y)}).$$

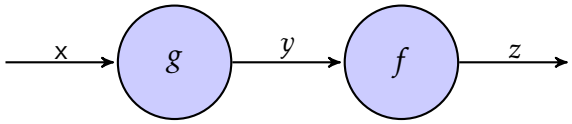
The outer function, $z = f(y) = \sin(y)$, is evaluated at the value of the inner function, $y = g(x) = 3x$. Thus $z = f(y) = f(g(x))$ or $z = (f \circ g)(x)$.
How quickly does z change as x changes?
The answer is given by the **chain rule**:

$$\boxed{\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.}$$

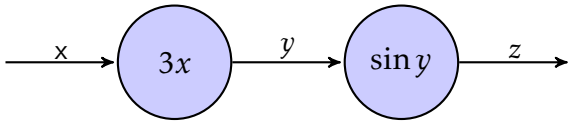
In our example, $\frac{dz}{dy} = \cos y$ and $\frac{dy}{dx} = 3$, and so

$$\frac{dz}{dx} = \underbrace{(\cos y)}_{\frac{dz}{dy}} \underbrace{(3)}_{\frac{dy}{dx}} = (\cos \underbrace{3x}_{y})(3) = 3\cos 3x.$$

We can visualize this computation by means of a **computational graph**, whose nodes correspond to mathematical operations:



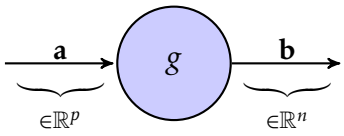
Here x is the input, z the output, y an intermediate result, and g and f are computations.
We can be more specific:



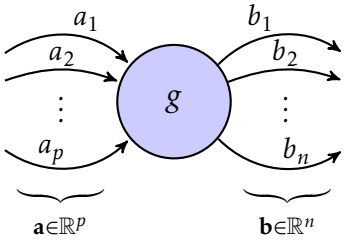
The chain rule proceeds backwards along the graph:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} = (\cos y)(3) = (\cos 3x)(3) = 3\cos 3x.$$

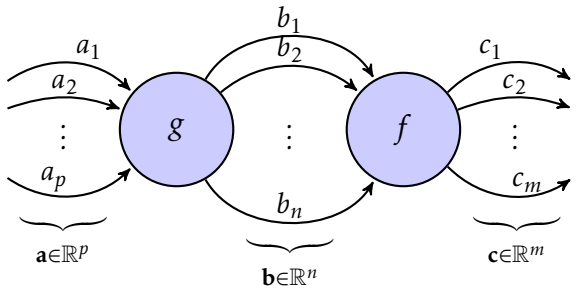
A function $g : \mathbb{R}^p \to \mathbb{R}^n$ can be represented on a computational graph as either



or, more verbosely,



The **partial derivative** $\frac{\partial b_i}{\partial a_j}$, $i \in \{1, 2, \ldots, n\}$, $j \in \{1, 2, \ldots, p\}$, is the instantaneous rate of change of b_i as we change a_j.
If we change a_j slightly to $a_j + \delta$, then (for a small δ), b_i changes to approximately $b_i + \frac{\partial b_i}{\partial a_j}\delta$.
Now suppose that we have two functions, $g : \mathbb{R}^p \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^m$.
Let $\mathbf{b} = g(\mathbf{a})$ and $\mathbf{c} = g(\mathbf{b})$.



A change in a_j may affect each of $b_1,\ldots,b_n$.
Changes in $b_1,\ldots,b_n$ may each affect c_i.
The **multivariable chain rule** tells us that the net change in c_i is the sum of the changes induced along each path from a_j to c_i. The multivariable chain rule becomes simply:

$$\boxed{\frac{\partial c_i}{\partial a_j} = \sum_{k=1}^{n} \frac{\partial c_i}{\partial b_k} \cdot \frac{\partial b_k}{\partial a_j}.}$$

Proof Since we are only dealing with a_j and c_i components, we can simplify the notation by suppressing all other components and write $a := a_j$ and $c := c_i$. Then by the Taylor's theorem we have:

$$c(a+\Delta) = c(a) + \Delta\frac{dc}{da} + O(|\Delta|^2)$$

On the other hand, applying Taylor's theorem to f we obtain:

$$\begin{aligned} c(a+\Delta) &= f(g(a+\Delta)) = f(b_1(a+\Delta),\ldots,(b_n(a+\Delta))) \\ &= f\left(b_1(a) + \Delta\frac{db_1}{da},\ldots b_n(a) + \Delta\frac{db_n}{da}\right) + O(|\Delta|^2) \\ = f(b_1(a),\ldots,b_n(a)) &+ \Delta\left[\nabla f(b_1(a),\ldots,b_n(a)) \cdot \left(\frac{db_1}{da},\ldots,\frac{db_n}{da}\right)\right] + O(|\Delta|^2) \\ &= c(a) + \Delta\left[\sum_{k=1}^{n} \frac{\partial c}{\partial b_k} \cdot \frac{db_k}{da}\right] + O(|\Delta|^2) \end{aligned}$$

Comparing the two above values for $c(a+\Delta)$, we obtain:

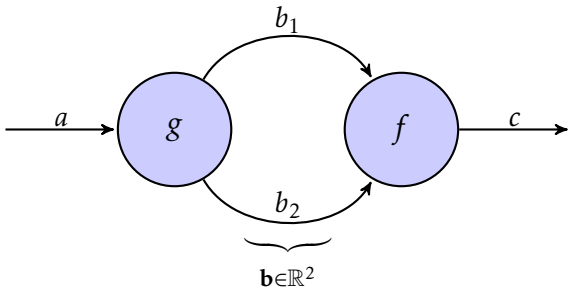
$$\frac{dc}{da} = \sum_{k=1}^{n} \frac{\partial c}{\partial b_k} \cdot \frac{db_k}{da}$$

Switching back to a_j and c_i, we finally have: $\frac{\partial c_i}{\partial a_j} = \sum_{k=1}^{n} \frac{\partial c_i}{\partial b_k} \cdot \frac{\partial b_k}{\partial a_j}$ and we are done □.

Here is an example. Let

$$g : a \mapsto \begin{pmatrix} \sin a \\ \cos a \end{pmatrix} =: \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad f : \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \mapsto 4b_1^2 + 3b_2^2 =: c.$$

What is $\frac{dc}{da}$?
Our computational graph is



$$\frac{dc}{da} = \frac{\partial c}{\partial b_1}\cdot\frac{db_1}{da} + \frac{\partial c}{\partial b_2}\cdot\frac{db_2}{da} = (8\underbrace{\quad b_1 \quad}_{\sin a})(\cos a) + (6\underbrace{\quad b_2 \quad}_{\cos a})(-\sin a) = 2\sin a\cos a.$$

Let us consider a practical example: the **linear regression** model, which we solved in the previous section using multivariate calculus and matrix algebra. This time we consider a solution using the **steepest descent** algorithm.

We have **observations** of the form $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$, where $\mathbf{x}_i \in \mathbb{R}^p$ are the **inputs** and $y_i \in \mathbb{R}$ are the corresponding **outputs** ($i \in \{1, 2, \ldots, n\}$).

We approximate each y_i with

$$\hat{y}_i := \mathbf{w}^\mathsf{T}\mathbf{x}_i + b,$$

where $\mathbf{w} \in \mathbb{R}^n$ is the vector of **weights** and $b \in \mathbb{R}$ a scalar **bias**.
How do we find suitable (optimal?) $\mathbf{w}$ and b?
As in the previous section, using the procedure called **linear least squares**, we express the quality (really, the badness) of our approximation by the **loss function** defined for a single observation as

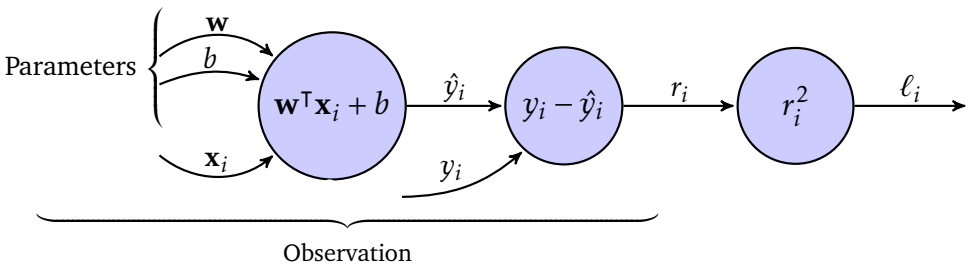
$$\ell_i(\mathbf{w}, b) = \left[(\mathbf{w}^\mathsf{T}\mathbf{x}_i + b) - y_i\right]^2.$$

Our objective, then, is to find such $\mathbf{w}$ and b that the loss function $\ell_i(\mathbf{w}, b)$ is minimized using the steepest descent algorithm:

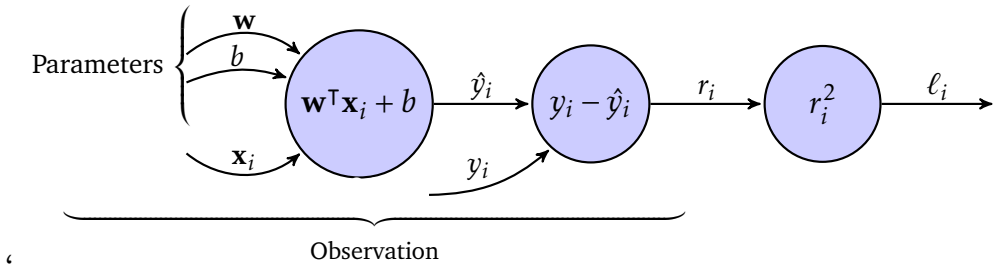
$$w_j \leftarrow w_j - \eta\frac{\partial \ell_i}{\partial w_j}$$

$$b \leftarrow b - \eta\frac{\partial \ell_i}{\partial b}$$

To this end, we require $\frac{\partial \ell_i}{\partial w_j}$ ($j \in \{1, 2, \ldots, p\}$) and $\frac{\partial \ell_i}{\partial b}$. As ever, we represent the computation as a computational graph:



We'll work our way backwards along the computational graph from the output ℓ_i to the parameters $\mathbf{w}$ and b:



Thus we obtain the following partial derivatives:

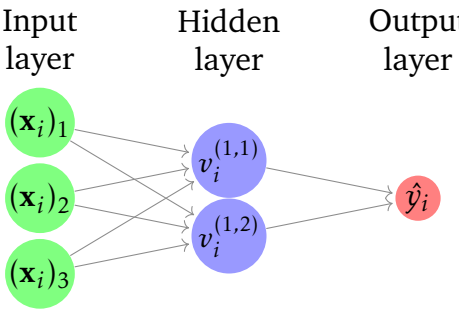
- $\frac{\partial \ell_i}{\partial r_i} = 2r_i$;
- $\frac{\partial \ell_i}{\partial \hat{y}_i} = \frac{\partial \ell_i}{\partial r_i} \cdot \frac{\partial r_i}{\partial \hat{y}_i} = (2r_i)(-1) = -2r_i$;
- $\frac{\partial \ell_i}{\partial w_j} = \frac{\partial \ell_i}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial w_j} = (-2r_i)(\mathbf{x}_i)_j = -2r_i(\mathbf{x}_i)_j$;
- $\frac{\partial \ell_i}{\partial b} = \frac{\partial \ell_i}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial b} = (-2r_i)(1) = -2r_i$.

Notice that there is no summation because there is only one path through the graph. Of course, this is a simple case, and we could have easily computed the partial derivatives directly from the formula:

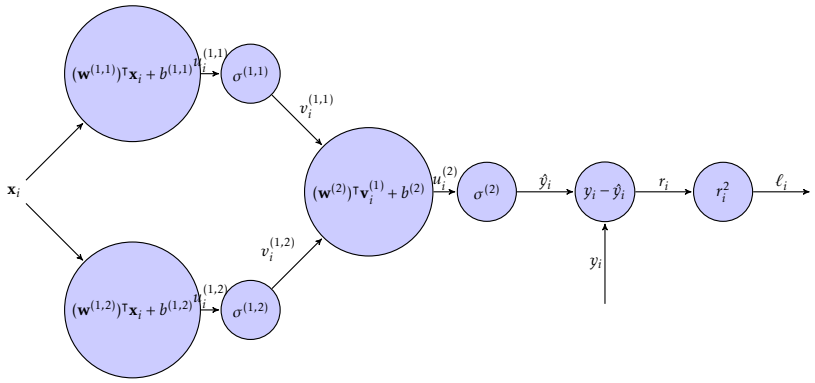
$$\ell_i = r_i^2 = (y_i - \hat{y}_i)^2 = (\mathbf{w_i}^T\mathbf{x_i} + b)^2$$

The example is given to indicate how backpropagation works in a simple case.

Backpropagation is the basis of training a deep neural network for classification of data, such as recognising handwritten digits using the MNIST library. Consider a neural network with three units in the input layer, two units in the single hidden layer, and one unit in the output layer:



The corresponding computational graph:



14.5 Further reading

There is much more to say about vectors and vector spaces. Fortunately, the Department of Computing offers a self-contained Linear Algebra course that you should be able to take.
The reader may also wish to consult [FB94, Str16, Str19] for introductory information on linear algebra.
There is a related mathematical field called **computational linear algebra**, which is concerned with the safe and efficient implementation of linear algebra routines in software [GL13].
Vectors and vector calculus are discussed in Chapters 18 and 19 of [Bin82] and Chapters 3–5 of [BD01]. The entire [Lan87] is dedicated to vector calculus.
We have also briefly touched on **optimization**—we minimized the loss function of a linear regression and of a neural network. Like linear algebra, optimization is a separate mathematical field, and a prerequisite for a thorough understanding of machine learning. An ambitious reader may wish to consult [GMW82, BV04, SNW12].

14.6 Exercises

Exercise 1

Let

$$g : \begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} =: \begin{pmatrix} x \\ y \end{pmatrix}, \quad f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^2 - y^2 =: c.$$

What are $\frac{\partial c}{\partial r}$ and $\frac{\partial c}{\partial \theta}$?

15 Epilogue

... ...The game's fortunes have turned. The less confident Sedol makes questionable moves at stones 119 and 123, followed by a losing move at 129.
AlphaGo wins.
Out of the five games played between Sedol and AlphaGo on 9th–15th March, Sedol would win only one.
He was beaten by a machine.
On 27th November, 2019, a curious article was published on BBC News:

> A master player of Chinese strategy game Go has decided to retire, due to the rise of artificial intelligence that "cannot be defeated."

This article was about Lee Sedol, whom we have met at the beginning of this course. He remained the only human to ever beat the AlphaGo software.

> The South Korean said he had decided to retire after realising: "I'm not at the top even if I become the number one."

... ...The Navy was wrong. In the 1960s, Marvin Minsky published a book proving the theoretical limitations of the perceptron (namely, that it could only learn classifiers on data that was **linearly separable**).
However, when lots and lots of perceptrons are combined such that the outputs of some become the inputs to others, this creates a **neural network**, which does not have the same limitations, and is indeed resulting in machines that can do all the things that the Navy promised, just way after the one-year deadline, and at a much higher cost than predicted.
The principles underlying Frank Rosenblatt's perceptron help spark the modern artificial intelligence revolution. Deep learning and neural networks—which can classify online images, for example, or enable language translation—are transforming society today.

> ... ..."Why should that apple always descend perpendicularly to the ground," thought he to him self: occasion'd by the fall of an apple, as he sat in a contemplative mood: "Why should it not go sideways, or upwards? but constantly to the earths centre? Assuredly, the reason is, that the earth draws it. There must be a drawing power in matter. & the sum of the drawing power in the matter of the earth must be in the earths center, not in any side of the earth. Therefore dos this apple fall perpendicularly, or toward the center. If matter thus draws matter; it must be in proportion of its quantity. Therefore the apple draws the earth, as well as the earth draws the apple."

Thus Newton discovered the law of universal gravitation, classical mechanics, and, indeed, calculus.
Thank you ever so much for your attention!

Bibliography

[Alc14] Lara Alcock. *How to Think about Analysis*. Oxford University Press, 2014. pages 65

[Alm96] Dennis Almeida. Variation in proof standards: implications for mathematics education. *International Journal of Mathematical Education in Science and Technology*, 27(5):659–665, sep 1996. pages 32

[App12] David Applebaum. *Limits, Limits Everywhere: The Tools of Mathematical Analysis*. Oxford University Press, 2012. pages 65

[Ash02] Mark H. Ashcraft. Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11(5):181–185, oct 2002. pages 4

[BD01] Ken Binmore and Joan Davies. *Calculus: Concepts and Methods*. Cambridge University Press, 2001. pages 161

[BD13] David C. Bramlett and Carl T. Drake. A history of mathematical proof: Ancient greece to the computer age. *Journal of Mathematical Sciences & Mathematics Education*, 8(2):20–33, 2013. pages 40

[Bin80] K. G. Binmore. *Foundations of Analysis: A Straightforward Introduction. Book 1: Logic, Sets and Numbers*. Cambridge University Press, 1980. pages 31, 40

[Bin81] K. G. Binmore. *Foundations of Analysis: A Straightforward Introduction. Book 2: Topological Ideas*. Cambridge University Press, 1981. pages 149

[Bin82] K. G. Binmore. *Mathematical Analysis: A Straightforward Approach*. Cambridge University Press, 2nd edition, 1982. pages 65, 161

[BL04] Cliff T. Bekar and Richard G. Lipsey. Science, institutions, and the Industrial Revolution. *Jorunal of European Economic History*, 2004. pages 15

[Bry85] Victor Bryant. *Metric Spaces: Iteration and Application*. Cambridge University Press, 1985. pages 149

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. pages 134, 161

[Che12] Karine Chemla, editor. *The History of Mathematical Proof in Ancient Traditions*. Cambridge University Press, 2012. pages 40

[Cop68] E. T. Copson. *Metric Spaces*, volume 57 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1968. pages 149

[Dug66] James Dugundji. *Topology*. Allyn and Bacon, Inc., 1966. pages 149

[FB94] John B. Fraleigh and Raymond A. Beauregard. *Linear Algebra*. Pearson, 3 edition, 1994. pages 161

[GL13] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013. pages 161

[GMW82] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Emerald Group Publishing Limited, 1982. pages 134, 161

[Goo65] I. J. Good. The mystery of go. *New Scientist*, January 1965. pages 8

[Gra74] Judith V. Grabiner. Is mathematical truth time-dependent. *The American Mathematical Monthly*, 81(4):354–365, April 1974. pages 40

[Hal60] Paul Richard Halmos. *Naïve Set Theory*. University series in undergraduate mathematics. Van Nostrand, 1960. pages 30

[HK68] Seymour Hayden and John F. Kennison. *Zermelo-Fraenkel Set Theory*. C. E. Merrill, 1968. pages 30

[Jac97] Margaret C. Jacob. *Scientific Culture and the Making of the Industrial West*. Oxford University Press, 1997. pages 15

[Kap72] Irving Kaplansky. *Set Theory and Metric Spaces*. Allyn and Bacon, Inc., 1972. pages 31, 149

[Kel55] John L. Kelley. *General Topology*. Springer, 1955. pages 20, 31, 149

[KF57] Andrey Nikolaevich Kolmogorov and Sergei Vasilyevich Fomin. *Elements of the Theory of Functions and Functional Analysis*. Graylock Press, 1957. pages 31, 149

[KF70] Andrey Nikolaevich Kolmogorov and Sergei Vasilyevich Fomin. *Introductory Real Analysis*. Dover Publications, Inc., 1970. pages 31, 149

[Kol33] Andrey Nikolaevich Kolmogorov. Grundbegriffe der Wahrscheinlichkeitrechnung. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 2(3):1–62, 1933. pages 144

[Kur66] Kazimierz Kuratowski. *Topology*. Academic Press, 1966. pages 149

[Lan87] Serge Lang. *Calculus of Several Variables*. Undergraduate Texts in Mathematics. Springer, 1987. pages 161

[Lie15] Martin Liebeck. *A Concise Introduction to Pure Mathematics*. CRC Press, 4th edition, 2015. pages 31, 40, 65

[Men90] Bert Mendelson. *Introduction to Topology*. Dover Publications, Inc., 3 edition, 1990. pages 149

[Mol11] Mieke Molthof. The industrial revolution and a Newtonian culture. *E-International Relations*, 2011. https://www.e-ir.info/2011/08/24/the-industrial-revolution-and-a-newtonian-culture/. pages 15

[MPV12] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 5 edition, 2012. pages 153

[Mun17] James R. Munkres. *Topology*. Pearson, 2nd edition, 2017. pages 149

[New94] Isaac Newton. A quote from a letter dated 25 may, 1694. May 1694. pages 11

[new58] New navy device learns by doing: Psychologist shows embryo of computer designed to read and grow wiser. *New York Times*, 1958. pages 6

[PTVF07] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edition, 2007. pages 133

[Sie52] Wacław Sierpiński. *General Topology*. Dover Publications, 1952. pages 149

[Smi07] George Smith. Newton's Philosophiae Naturalis Principia Mathematica. *Stanford Encyclopedia of Philosophy*, 2007. https://plato.stanford.edu/entries/newton-principia/. pages 12

[SNW12] Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors. *Optimization for Machine Learning*. MIT Press, 2012. pages 134, 161

[Spi07] Diomidis Spinellis. *Beautiful Code: Leading Programmers Explain How They Think*, chapter Another level of indirection, pages 279–291. O'Reilly and Associates, 2007. pages 149

[SS95] Lynn Arthur Steen and J. Arthur Seebach. *Counterexamples in Topology*. Courier Dover Publications, 1995. pages 149

[Ste15] Ian Stewart. *Galois Theory*. CRC Press, 4th edition, 2015. pages 123

[Str01] Gilbert Strang. Too much calculus. http://web.mit.edu/18.06/www/Essays/too-much-calculus.pdf, 2001. pages 4

[Str16] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley–Cambridge Press, 5 edition, 2016. pages 161

[Str17] Gilbert Strang. *Calculus*. Wellesley–Cambridge Press, 3 edition, 2017. pages 4

[Str19] Gilbert Strang. *Linear Algebra and Learning from Data*. Wellesley–Cambridge Press, 2019. pages 161

[Stu52] William Stukeley. *Memoirs of Sir Isaac Newton's Life*. 1752. pages 5

[Tal07] Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007. pages 40

[Vel19] Daniel J. Velleman. *How to Prove It: A Structured Approach*. Cambridge University Press, 3rd edition, 2019. pages 40

[Wil70] Stephen Willard. *General Topology*. Addison-Wesley Publishing Company, 1970. pages 149

[WS02] Robert C. Wrede and Murray Spiegel. *Advanced Calculus*. McGraw–Hill, second edition, 2002. pages 65