

Basic Analysis

Introduction to Real Analysis

by **Jiří Lebl**

December 16, 2012

Typeset in L^AT_EX.

Copyright ©2009–2012 Jiří Lebl



This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

You can use, print, duplicate, share these notes as much as you want. You can base your own notes on these and reuse parts if you keep the license the same. If you plan to use these commercially (sell them for more than just duplicating cost), then you need to contact me and we will work something out. If you are printing a course pack for your students, then it is fine if the duplication service is charging a fee for printing and selling the printed copy. I consider that duplicating cost.

During the writing of these notes, the author was in part supported by NSF grant DMS-0900885.

See <http://www.jirka.org/ra/> for more information (including contact information).

Contents

Introduction	5
0.1 Notes about these notes	5
0.2 About analysis	7
0.3 Basic set theory	8
1 Real Numbers	21
1.1 Basic properties	21
1.2 The set of real numbers	25
1.3 Absolute value	31
1.4 Intervals and the size of \mathbb{R}	35
2 Sequences and Series	39
2.1 Sequences and limits	39
2.2 Facts about limits of sequences	47
2.3 Limit superior, limit inferior, and Bolzano-Weierstrass	57
2.4 Cauchy sequences	65
2.5 Series	68
3 Continuous Functions	79
3.1 Limits of functions	79
3.2 Continuous functions	86
3.3 Min-max and intermediate value theorems	92
3.4 Uniform continuity	98
4 The Derivative	103
4.1 The derivative	103
4.2 Mean value theorem	109
4.3 Taylor's theorem	115

5	The Riemann Integral	117
5.1	The Riemann integral	117
5.2	Properties of the integral	125
5.3	Fundamental theorem of calculus	133
6	Sequences of Functions	139
6.1	Pointwise and uniform convergence	139
6.2	Interchange of limits	145
6.3	Picard's theorem	151
7	Metric Spaces	157
7.1	Metric spaces	157
7.2	Open and closed sets	164
7.3	Sequences and convergence	171
7.4	Completeness and compactness	175
7.5	Continuous functions	180
7.6	Fixed point theorem and Picard's theorem again	184
	Further Reading	187
	Index	189

Introduction

0.1 Notes about these notes

This book is a one semester course in basic analysis. It started its life as my lecture notes for teaching Math 444 at the University of Illinois at Urbana-Champaign (UIUC) in Fall semester 2009. Later I added the metric space chapter to teach Math 521 at University of Wisconsin–Madison. A prerequisite for the course should be a basic proof course, for example one using the (unfortunately rather pricey) book [DW]. It should be possible to use the book for both a basic course for students who do not necessarily wish to go to graduate school, but also as a more advanced one-semester course that also covers topics such as metric spaces.

The book normally used for the class at UIUC is Bartle and Sherbert, *Introduction to Real Analysis* third edition [BS]. The structure of the beginning of the book mostly follows the syllabus of UIUC Math 444 and therefore has some similarities with [BS]. A major difference is that we define the Riemann integral using Darboux sums and not tagged partitions. The Darboux approach is far more appropriate for a course of this level.

As the integral is treated more lightly, we can spend some extra time on the interchange of limits and in particular on a section on Picard's theorem on the existence and uniqueness of solutions of ordinary differential equations if time allows. This theorem is a wonderful example that uses many results proved in the book. For more advanced students, material may be covered faster so that we arrive at metric spaces and prove Picard's theorem using the fixed point theorem as is usual.

Other excellent books exist. My favorite is Rudin's excellent *Principles of Mathematical Analysis* [R2] or as it is commonly and lovingly called *baby Rudin* (to distinguish it from his other great analysis textbook). I have taken a lot of inspiration and ideas from Rudin. However, Rudin is a bit more advanced and ambitious than this present course. For those that wish to continue mathematics, Rudin is a fine investment. An inexpensive and somewhat simpler alternative to Rudin is Rosenlicht's *Introduction to Analysis* [R1]. There is also the freely downloadable *Introduction to Real Analysis* by William Trench [T] for those that do not wish to invest much money.

A note about the style of some of the proofs: Many proofs that are traditionally done by contradiction, I prefer to do by a direct proof or by contrapositive. While the book does include proofs by contradiction, I only do so when the contrapositive statement seemed too awkward, or when the contradiction follows rather quickly. In my opinion, contradiction is more likely to get beginning students into trouble, as we are talking about objects that do not exist.

I try to avoid unnecessary formalism where it is unhelpful. Furthermore, the proofs and the language get slightly less formal as we progress through the book, as more and more details are left out to avoid clutter.

As a general rule, I use $:=$ instead of $=$ to define an object rather than to simply show equality. I use this symbol rather more liberally than is usual for emphasis. I use it even when the context is “local,” that is, I may simply define a function $f(x) := x^2$ for a single exercise or example.

It is possible to skip or skim some material in the book as it is not used later on. Some material that can safely be skipped is marked as such in the notes that appear below every section title. Section §0.3 can usually be covered lightly, or left as reading, depending on the prerequisites of the course.

It is possible to fill a semester with a course that ends with Picards theorem in §6.3 by skipping little. On the other hand, if metric spaces are covered, one might have to go a little faster and skip some topics in the first six chapters.

If you are teaching (or being taught) with [BS], here is an approximate correspondence of the sections. Note that the material in this book and in [BS] differs. The correspondence to other books is more complicated.

Section	Section in [BS]	Section	Section in [BS]
§0.3	§1.1–§1.3	§3.3	§5.3
§1.1	§2.1 and §2.3	§3.4	§5.4
§1.2	§2.3 and §2.4	§4.1	§6.1
§1.3	§2.2	§4.2	§6.2
§1.4	§2.5	§4.3	§6.3
§2.1	parts of §3.1, §3.2, §3.3, §3.4	§5.1	§7.1, §7.2
§2.2	§3.2	§5.2	§7.2
§2.3	§3.3 and §3.4	§5.3	§7.3
§2.4	§3.5	§6.1	§8.1
§2.5	§3.7	§6.2	§8.2
§3.1	§4.1–§4.2	§6.3	Not in [BS]
§3.2	§5.1 and §5.2	chapter 7	partially §11

Finally, I would like to acknowledge Jana Maříková, Glen Pugh, Paul Vojta, Frank Beatrous, and Sönmez Şahutoğlu for teaching with the book and giving me useful feedback. I would also like to thank Dan Stoneham, Jeremy Sutter, Eliya Gwetta, Daniel Alarcon, Steve Hoerning, Yi Zhang, an anonymous reader, and in general all the students in my classes for suggestions and finding errors and typos.

0.2 About analysis

Analysis is the branch of mathematics that deals with inequalities and limits. The present course deals with the most basic concepts in analysis. The goal of the course is to acquaint the reader with rigorous proofs in analysis and also to set a firm foundation for calculus of one variable.

Calculus has prepared you, the student, for using mathematics without telling you why what you have learned is true. To use, or teach, mathematics effectively, you cannot simply know *what* is true, you must know *why* it is true. This course shows you *why* calculus is true. It is here to give you a good understanding of the concept of a limit, the derivative, and the integral.

Let us use an analogy. An auto mechanic that has learned to change the oil, fix broken headlights, and charge the battery, will only be able to do those simple tasks. He will be unable to work independently to diagnose and fix problems. A high school teacher that does not understand the definition of the Riemann integral will not be able to properly answer all the student's questions that could come up. To this day I remember several nonsensical statements I heard from my calculus teacher in high school, who simply did not understand the concept of the limit, though he could "do" all problems in calculus.

We will start with discussion of the real number system, most importantly its completeness property, which is the basis for all that comes after. We will then discuss the simplest form of a limit, the limit of a sequence. Afterwards, we will study functions of one variable, continuity, and the derivative. Next, we will define the Riemann integral and prove the fundamental theorem of calculus. We will discuss sequences of functions and the interchange of limits. Finally, we will give a short introduction to metric spaces.

Let us give the most important difference between analysis and algebra. In algebra, we prove equalities directly; we prove that an object, a number perhaps, is equal to another object. In analysis, we usually prove inequalities. To illustrate the point, consider the following statement.

Let x be a real number. If $0 \leq x < \varepsilon$ is true for all real numbers $\varepsilon > 0$, then $x = 0$.

This statement is the general idea of what we do in analysis. If we wish to show that $x = 0$, we will show that $0 \leq x < \varepsilon$ for all positive ε .

The term *real analysis* is a little bit of a misnomer. I prefer to use simply *analysis*. The other type of analysis, *complex analysis*, really builds up on the present material, rather than being distinct. Furthermore, a more advanced course on real analysis would talk about complex numbers often. I suspect the nomenclature is historical baggage.

Let us get on with the show...

0.3 Basic set theory

Note: 1–3 lectures (some material can be skipped or covered lightly)

Before we start talking about analysis we need to fix some language. Modern* analysis uses the language of sets, and therefore that is where we start. We talk about sets in a rather informal way, using the so-called “naïve set theory.” Do not worry, that is what majority of mathematicians use, and it is hard to get into trouble.

We assume that the reader has seen basic set theory and has had a course in basic proof writing. This section should be thought of as a refresher.

0.3.1 Sets

Definition 0.3.1. A *set* is a collection of objects called *elements* or *members*. A set with no objects is called the *empty set* and is denoted by \emptyset (or sometimes by $\{\}$).

Think of a set as a club with a certain membership. For example, the students who play chess are members of the chess club. However, do not take the analogy too far. A set is only defined by the members that form the set; two sets that have the same members are the same set.

Most of the time we will consider sets of numbers. For example, the set

$$S := \{0, 1, 2\}$$

is the set containing the three elements 0, 1, and 2. We write

$$1 \in S$$

to denote that the number 1 belongs to the set S . That is, 1 is a member of S . Similarly we write

$$7 \notin S$$

to denote that the number 7 is not in S . That is, 7 is not a member of S . The elements of all sets under consideration come from some set we call the *universe*. For simplicity, we often consider the universe to be the set that contains only the elements we are interested in. The universe is generally understood from context and is not explicitly mentioned. In this course, our universe will most often be the set of real numbers.

While the elements of a set are often numbers, other object, such as other sets, can be elements of a set.

A set may contain some of the same elements as another set. For example,

$$T := \{0, 2\}$$

contains the numbers 0 and 2. In this case all elements of T also belong to S . We write $T \subset S$. More formally we have the following definition.

*The term “modern” refers to late 19th century up to the present.

Definition 0.3.2.

- (i) A set A is a *subset* of a set B if $x \in A$ implies that $x \in B$, and we write $A \subset B$. That is, all members of A are also members of B .
- (ii) Two sets A and B are *equal* if $A \subset B$ and $B \subset A$. We write $A = B$. That is, A and B contain the exactly the same elements. If it is not true that A and B are equal, then we write $A \neq B$.
- (iii) A set A is a *proper subset* of B if $A \subset B$ and $A \neq B$. We write $A \subsetneq B$.

When $A = B$, we consider A and B to just be two names for the same exact set. For example, for S and T defined above we have $T \subset S$, but $T \neq S$. So T is a proper subset of S . At this juncture, we also mention the *set building notation*,

$$\{x \in A : P(x)\}.$$

This notation refers to a subset of the set A containing all elements of A that satisfy the property $P(x)$. The notation is sometimes abbreviated (A is not mentioned) when understood from context. Furthermore, $x \in A$ is sometimes replaced with a formula to make the notation easier to read.

Example 0.3.3: The following are sets including the standard notations.

- (i) The set of *natural numbers*, $\mathbb{N} := \{1, 2, 3, \dots\}$.
- (ii) The set of *integers*, $\mathbb{Z} := \{0, -1, 1, -2, 2, \dots\}$.
- (iii) The set of *rational numbers*, $\mathbb{Q} := \{\frac{m}{n} : m, n \in \mathbb{Z} \text{ and } n \neq 0\}$.
- (iv) The set of even natural numbers, $\{2m : m \in \mathbb{N}\}$.
- (v) The set of real numbers, \mathbb{R} .

Note that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

There are many operations we want to do with sets.

Definition 0.3.4.

- (i) A *union* of two sets A and B is defined as

$$A \cup B := \{x : x \in A \text{ or } x \in B\}.$$

- (ii) An *intersection* of two sets A and B is defined as

$$A \cap B := \{x : x \in A \text{ and } x \in B\}.$$

- (iii) A *complement of B relative to A* (or *set-theoretic difference* of A and B) is defined as

$$A \setminus B := \{x : x \in A \text{ and } x \notin B\}.$$

(iv) We say *complement* of B and write B^c if A is understood from context. The set A is either the entire universe or is the obvious set containing B .

(v) We say that sets A and B are *disjoint* if $A \cap B = \emptyset$.

The notation B^c may be a little vague at this point. If the set B is a subset of the real numbers \mathbb{R} , then B^c will mean $\mathbb{R} \setminus B$. If B is naturally a subset of the natural numbers, then B^c is $\mathbb{N} \setminus B$. If ambiguity would ever arise, we will use the set difference notation $A \setminus B$.

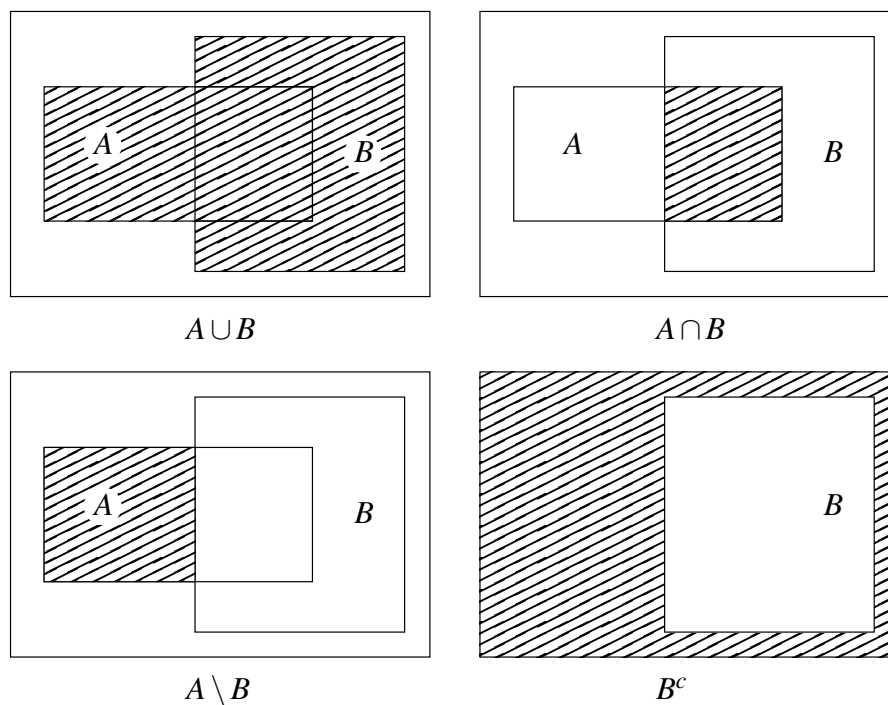


Figure 1: Venn diagrams of set operations.

We illustrate the operations on the *Venn diagrams* in Figure 1. Let us now establish one of most basic theorems about sets and logic.

Theorem 0.3.5 (DeMorgan). *Let A, B, C be sets. Then*

$$\begin{aligned}(B \cup C)^c &= B^c \cap C^c, \\ (B \cap C)^c &= B^c \cup C^c,\end{aligned}$$

or, more generally,

$$\begin{aligned}A \setminus (B \cup C) &= (A \setminus B) \cap (A \setminus C), \\ A \setminus (B \cap C) &= (A \setminus B) \cup (A \setminus C).\end{aligned}$$

Proof. The first statement is proved by the second statement if we assume that set A is our “universe.”

Let us prove $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$. Remember the definition of equality of sets. First, we must show that if $x \in A \setminus (B \cup C)$, then $x \in (A \setminus B) \cap (A \setminus C)$. Second, we must also show that if $x \in (A \setminus B) \cap (A \setminus C)$, then $x \in A \setminus (B \cup C)$.

So let us assume that $x \in A \setminus (B \cup C)$. Then x is in A , but not in B nor C . Hence x is in A and not in B , that is, $x \in A \setminus B$. Similarly $x \in A \setminus C$. Thus $x \in (A \setminus B) \cap (A \setminus C)$.

On the other hand suppose that $x \in (A \setminus B) \cap (A \setminus C)$. In particular $x \in (A \setminus B)$ and so $x \in A$ and $x \notin B$. Also as $x \in (A \setminus C)$, then $x \notin C$. Hence $x \in A \setminus (B \cup C)$.

The proof of the other equality is left as an exercise. \square

We will also need to intersect or union several sets at once. If there are only finitely many, then we simply apply the union or intersection operation several times. However, suppose that we have an infinite collection of sets (a set of sets) $\{A_1, A_2, A_3, \dots\}$. We define

$$\bigcup_{n=1}^{\infty} A_n := \{x : x \in A_n \text{ for some } n \in \mathbb{N}\},$$

$$\bigcap_{n=1}^{\infty} A_n := \{x : x \in A_n \text{ for all } n \in \mathbb{N}\}.$$

We can also have sets indexed by two integers. For example, we can have the set of sets $\{A_{1,1}, A_{1,2}, A_{2,1}, A_{1,3}, A_{2,2}, A_{3,1}, \dots\}$. Then we write

$$\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} A_{n,m} = \bigcup_{n=1}^{\infty} \left(\bigcup_{m=1}^{\infty} A_{n,m} \right).$$

And similarly with intersections.

It is not hard to see that we can take the unions in any order. However, switching unions and intersections is not generally permitted without proof. For example:

$$\bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \{k \in \mathbb{N} : mk < n\} = \bigcup_{n=1}^{\infty} \emptyset = \emptyset.$$

However,

$$\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{k \in \mathbb{N} : mk < n\} = \bigcap_{m=1}^{\infty} \mathbb{N} = \mathbb{N}.$$

0.3.2 Induction

A common method of proof is the principle of induction. We start with the set of natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$. The natural ordering on \mathbb{N} (that is, $1 < 2 < 3 < 4 < \dots$) has a wonderful property.

The natural numbers \mathbb{N} ordered in the natural way possess the *well ordering property*. We take this property as an axiom.

Well ordering property of \mathbb{N} . *Every nonempty subset of \mathbb{N} has a least (smallest) element.*

By $S \subset \mathbb{N}$ having a least element we mean that there exist an $x \in S$, such that for every $y \in S$, we have $x \leq y$.

The *principle of induction* is the following theorem, which is equivalent to the well ordering property of the natural numbers.

Theorem 0.3.6 (Principle of induction). *Let $P(n)$ be a statement depending on a natural number n . Suppose that*

- (i) (basis statement) $P(1)$ is true,
- (ii) (induction step) if $P(n)$ is true, then $P(n+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. Suppose that S is the set of natural numbers m for which $P(m)$ is not true. Suppose that S is nonempty. Then S has a least element by the well ordering property. Let us call m the least element of S . We know that $1 \notin S$ by assumption. Therefore $m > 1$ and $m - 1$ is a natural number as well. Since m was the least element of S , we know that $P(m - 1)$ is true. But by the induction step we see that $P(m - 1 + 1) = P(m)$ is true, contradicting the statement that $m \in S$. Therefore S is empty and $P(n)$ is true for all $n \in \mathbb{N}$. \square

Sometimes it is convenient to start at a different number than 1, but all that changes is the labeling. The assumption that $P(n)$ is true in “if $P(n)$ is true, then $P(n+1)$ is true” is usually called the *induction hypothesis*.

Example 0.3.7: Let us prove that for all $n \in \mathbb{N}$ we have

$$2^{n-1} \leq n!.$$

We let $P(n)$ be the statement that $2^{n-1} \leq n!$ is true. By plugging in $n = 1$, we see that $P(1)$ is true.

Suppose that $P(n)$ is true. That is, suppose that $2^{n-1} \leq n!$ holds. Multiply both sides by 2 to obtain

$$2^n \leq 2(n!).$$

As $2 \leq (n+1)$ when $n \in \mathbb{N}$, we have $2(n!) \leq (n+1)(n!) = (n+1)!$. That is,

$$2^n \leq 2(n!) \leq (n+1!),$$

and hence $P(n+1)$ is true. By the principle of induction, we see that $P(n)$ is true for all n , and hence $2^{n-1} \leq n!$ is true for all $n \in \mathbb{N}$.

Example 0.3.8: We claim that for all $c \neq 1$, we have that

$$1 + c + c^2 + \cdots + c^n = \frac{1 - c^{n+1}}{1 - c}.$$

Proof: It is easy to check that the equation holds with $n = 1$. Suppose that it is true for n . Then

$$\begin{aligned} 1 + c + c^2 + \cdots + c^n + c^{n+1} &= (1 + c + c^2 + \cdots + c^n) + c^{n+1} \\ &= \frac{1 - c^{n+1}}{1 - c} + c^{n+1} \\ &= \frac{1 - c^{n+1} + (1 - c)c^{n+1}}{1 - c} \\ &= \frac{1 - c^{n+2}}{1 - c}. \end{aligned}$$

There is an equivalent principle called strong induction. The proof that strong induction is equivalent to induction is left as an exercise.

Theorem 0.3.9 (Principle of strong induction). *Let $P(n)$ be a statement depending on a natural number n . Suppose that*

- (i) (basis statement) $P(1)$ is true,
- (ii) (induction step) if $P(k)$ is true for all $k = 1, 2, \dots, n$, then $P(n+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

0.3.3 Functions

Informally, a *set-theoretic function* f taking a set A to a set B is a mapping that to each $x \in A$ assigns a unique $y \in B$. We write $f: A \rightarrow B$. For example, we define a function $f: S \rightarrow T$ taking $S = \{0, 1, 2\}$ to $T = \{0, 2\}$ by assigning $f(0) := 2$, $f(1) := 2$, and $f(2) := 0$. That is, a function $f: A \rightarrow B$ is a black box, into which we stick an element of A and the function spits out an element of B . Sometimes f is called a *mapping* and we say that f maps A to B .

Often, functions are defined by some sort of formula, however, you should really think of a function as just a very big table of values. The subtle issue here is that a single function can have several different formulas, all giving the same function. Also, for many functions, there is no formula that expresses its values.

To define a function rigorously first let us define the Cartesian product.

Definition 0.3.10. Let A and B be sets. Then the *Cartesian product* is the set of tuples defined as follows.

$$A \times B := \{(x, y) : x \in A, y \in B\}.$$

For example, the set $[0, 1] \times [0, 1]$ is a set in the plane bounded by a square with vertices $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$. When A and B are the same set we sometimes use a superscript 2 to denote such a product. For example $[0, 1]^2 = [0, 1] \times [0, 1]$, or $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ (the Cartesian plane).

Definition 0.3.11. A function $f: A \rightarrow B$ is a subset f of $A \times B$ such that for each $x \in A$, there is a unique $(x, y) \in f$. Sometimes the set f is called the *graph* of the function rather than the function itself.

The set A is called the *domain* of f (and sometimes confusingly denoted $D(f)$). The set

$$R(f) := \{y \in B : \text{there exists an } x \text{ such that } (x, y) \in f\}$$

is called the *range* of f .

Note that $R(f)$ can possibly be a proper subset of B , while the domain of f is always equal to A . We usually assume that the domain of f is nonempty.

Example 0.3.12: From calculus, you are most familiar with functions taking real numbers to real numbers. However, you have seen some other types of functions as well. For example the derivative is a function mapping the set of differentiable functions to the set of all functions. Another example is the Laplace transform, which also takes functions to functions. Yet another example is the function that takes a continuous function g defined on the interval $[0, 1]$ and returns the number $\int_0^1 g(x)dx$.

Definition 0.3.13. Let $f: A \rightarrow B$ be a function, and $C \subset A$. Define the *image* (or *direct image*) of C as

$$f(C) := \{f(x) \in B : x \in C\}.$$

Let $D \subset B$. Define the *inverse image* as

$$f^{-1}(D) := \{x \in A : f(x) \in D\}.$$

Example 0.3.14: Define the function $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) := \sin(\pi x)$. Then $f([0, 1/2]) = [0, 1]$, $f^{-1}(\{0\}) = \mathbb{Z}$, etc. . . .

Proposition 0.3.15. Let $f: A \rightarrow B$. Let C, D be subsets of B . Then

$$\begin{aligned} f^{-1}(C \cup D) &= f^{-1}(C) \cup f^{-1}(D), \\ f^{-1}(C \cap D) &= f^{-1}(C) \cap f^{-1}(D), \\ f^{-1}(C^c) &= (f^{-1}(C))^c. \end{aligned}$$

Read the last line as $f^{-1}(B \setminus C) = A \setminus f^{-1}(C)$.

Proof. Let us start with the union. Suppose that $x \in f^{-1}(C \cup D)$. That means that x maps to C or D . Thus $f^{-1}(C \cup D) \subset f^{-1}(C) \cup f^{-1}(D)$. Conversely if $x \in f^{-1}(C)$, then $x \in f^{-1}(C \cup D)$. Similarly for $x \in f^{-1}(D)$. Hence $f^{-1}(C \cup D) \supset f^{-1}(C) \cup f^{-1}(D)$, and we have equality.

The rest of the proof is left as an exercise. □

The proposition does not hold for direct images. We do have the following weaker result.

Proposition 0.3.16. *Let $f: A \rightarrow B$. Let C, D be subsets of A . Then*

$$\begin{aligned} f(C \cup D) &= f(C) \cup f(D), \\ f(C \cap D) &\subset f(C) \cap f(D). \end{aligned}$$

The proof is left as an exercise.

Definition 0.3.17. Let $f: A \rightarrow B$ be a function. The function f is said to be *injective* or *one-to-one* if $f(x_1) = f(x_2)$ implies $x_1 = x_2$. In other words, $f^{-1}(\{y\})$ is empty or consists of a single element for all $y \in B$. We call such an f an *injection*.

The function f is said to be *surjective* or *onto* if $f(A) = B$. We call such an f a *surjection*.

A function f that is both an injection and a surjection is said to be *bijective*, and we say f is a *bijection*.

When $f: A \rightarrow B$ is a bijection, then $f^{-1}(\{y\})$ is always a unique element of A , and we can consider f^{-1} as a function $f^{-1}: B \rightarrow A$. In this case, we call f^{-1} the *inverse function* of f . For example, for the bijection $f(x) := x^3$ we have $f^{-1}(x) = \sqrt[3]{x}$.

A final piece of notation for functions that we need is the *composition of functions*.

Definition 0.3.18. Let $f: A \rightarrow B$, $g: B \rightarrow C$. The function $g \circ f: A \rightarrow C$ is defined as

$$(g \circ f)(x) := g(f(x)).$$

0.3.4 Cardinality

A subtle issue in set theory and one generating a considerable amount of confusion among students is that of cardinality, or “size” of sets. The concept of cardinality is important in modern mathematics in general and in analysis in particular. In this section, we will see the first really unexpected theorem.

Definition 0.3.19. Let A and B be sets. We say A and B have the same *cardinality* when there exists a bijection $f: A \rightarrow B$. We denote by $|A|$ the equivalence class of all sets with the same cardinality as A and we simply call $|A|$ the *cardinality* of A .

Note that A has the same cardinality as the empty set if and only if A itself is the empty set. We then write $|A| := 0$.

Definition 0.3.20. Suppose that A has the same cardinality as $\{1, 2, 3, \dots, n\}$ for some $n \in \mathbb{N}$. We then write $|A| := n$, and we say that A is *finite*. When A is the empty set, we also call A finite.

We say that A is *infinite* or “of infinite cardinality” if A is not finite.

That the notation $|A| = n$ is justified we leave as an exercise. That is, for each nonempty finite set A , there exists a unique natural number n such that there exists a bijection from A to $\{1, 2, 3, \dots, n\}$.

We can also order sets by size.

Definition 0.3.21. We write

$$|A| \leq |B|$$

if there exists an injection from A to B . We write $|A| = |B|$ if A and B have the same cardinality. We write $|A| < |B|$ if $|A| \leq |B|$, but A and B do not have the same cardinality.

We state without proof that $|A| = |B|$ have the same cardinality if and only if $|A| \leq |B|$ and $|B| \leq |A|$. This is the so-called Cantor-Bernstein-Schroeder theorem. Furthermore, if A and B are any two sets, we can always write $|A| \leq |B|$ or $|B| \leq |A|$. The issues surrounding this last statement are very subtle. As we do not require either of these two statements, we omit proofs.

The truly interesting cases of cardinality are infinite sets. We start with the following definition.

Definition 0.3.22. If $|A| = |\mathbb{N}|$, then A is said to be *countably infinite*. If A is finite or countably infinite, then we say A is *countable*. If A is not countable, then A is said to be *uncountable*.

The cardinality of \mathbb{N} is usually denoted as \aleph_0 (read as aleph-naught)*.

Example 0.3.23: The set of even natural numbers has the same cardinality as \mathbb{N} . Proof: Given an even natural number, write it as $2n$ for some $n \in \mathbb{N}$. Then create a bijection taking $2n$ to n .

In fact, let us mention without proof the following characterization of infinite sets: *A set is infinite if and only if it is in one-to-one correspondence with a proper subset of itself.*

Example 0.3.24: $\mathbb{N} \times \mathbb{N}$ is a countably infinite set. Proof: Arrange the elements of $\mathbb{N} \times \mathbb{N}$ as follows $(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1), \dots$. That is, always write down first all the elements whose two entries sum to k , then write down all the elements whose entries sum to $k + 1$ and so on. Then define a bijection with \mathbb{N} by letting 1 go to $(1, 1)$, 2 go to $(1, 2)$ and so on.

Example 0.3.25: The set of rational numbers is countable. Proof: (informal) Follow the same procedure as in the previous example, writing $1/1, 1/2, 2/1$, etc. . . . However, leave out any fraction (such as $2/2$) that has already appeared.

For completeness we mention the following statement. *If $A \subset B$ and B is countable, then A is countable. Similarly if A is uncountable, then B is uncountable.* As we will not need this statement in the sequel, and as the proof requires the Cantor-Bernstein-Schroeder theorem mentioned above, we will not give it here.

We give the first truly striking result. First, we need a notation for the set of all subsets of a set.

Definition 0.3.26. If A is a set, we define the *power set* of A , denoted by $\mathcal{P}(A)$, to be the set of all subsets of A .

*For the fans of the TV show *Futurama*, there is a movie theater in one episode called an \aleph_0 -plex.

For example, if $A := \{1, 2\}$, then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Note that for a finite set A of cardinality n , the cardinality of $\mathcal{P}(A)$ is 2^n . This fact is left as an exercise. That is, the cardinality of $\mathcal{P}(A)$ is strictly larger than the cardinality of A , at least for finite sets. What is an unexpected and striking fact is that this statement is still true for infinite sets.

Theorem 0.3.27 (Cantor). $|A| < |\mathcal{P}(A)|$. In particular, there exists no surjection from A onto $\mathcal{P}(A)$.

Proof. There of course exists an injection $f: A \rightarrow \mathcal{P}(A)$. For any $x \in A$, define $f(x) := \{x\}$. Therefore $|A| \leq |\mathcal{P}(A)|$.

To finish the proof, we have to show that no function $f: A \rightarrow \mathcal{P}(A)$ is a surjection. Suppose that $f: A \rightarrow \mathcal{P}(A)$ is a function. So for $x \in A$, $f(x)$ is a subset of A . Define the set

$$B := \{x \in A : x \notin f(x)\}.$$

We claim that B is not in the range of f and hence f is not a surjection. Suppose that there exists an x_0 such that $f(x_0) = B$. Either $x_0 \in B$ or $x_0 \notin B$. If $x_0 \in B$, then $x_0 \notin f(x_0) = B$, which is a contradiction. If $x_0 \notin B$, then $x_0 \in f(x_0) = B$, which is again a contradiction. Thus such an x_0 does not exist. Therefore, B is not in the range of f , and f is not a surjection. As f was an arbitrary function, no surjection exists. \square

One particular consequence of this theorem is that there do exist uncountable sets, as $\mathcal{P}(\mathbb{N})$ must be uncountable. This fact is related to the fact that the set of real numbers (which we study in the next chapter) is uncountable. The existence of uncountable sets may seem unintuitive, and the theorem caused quite a controversy at the time it was announced. The theorem not only says that uncountable sets exist, but that there in fact exist progressively larger and larger infinite sets \mathbb{N} , $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathcal{P}(\mathbb{N}))$, $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))$, etc. . . .

0.3.5 Exercises

Exercise 0.3.1: Show $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

Exercise 0.3.2: Prove that the principle of strong induction is equivalent to the standard induction.

Exercise 0.3.3: Finish the proof of Proposition 0.3.15.

Exercise 0.3.4: a) Prove Proposition 0.3.16.

b) Find an example for which equality of sets in $f(C \cap D) \subset f(C) \cap f(D)$ fails. That is, find an f , A , B , C , and D such that $f(C \cap D)$ is a proper subset of $f(C) \cap f(D)$.

Exercise 0.3.5 (Tricky): Prove that if A is finite, then there exists a unique number n such that there exists a bijection between A and $\{1, 2, 3, \dots, n\}$. In other words, the notation $|A| := n$ is justified. Hint: Show that if $n > m$, then there is no injection from $\{1, 2, 3, \dots, n\}$ to $\{1, 2, 3, \dots, m\}$.

Exercise 0.3.6: Prove

- a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Exercise 0.3.7: Let $A \Delta B$ denote the symmetric difference, that is, the set of all elements that belong to either A or B , but not to both A and B .

- a) Draw a Venn diagram for $A \Delta B$.
- b) Show $A \Delta B = (A \setminus B) \cup (B \setminus A)$.
- c) Show $A \Delta B = (A \cup B) \setminus (A \cap B)$.

Exercise 0.3.8: For each $n \in \mathbb{N}$, let $A_n := \{(n+1)k : k \in \mathbb{N}\}$.

- a) Find $A_1 \cap A_2$.
- b) Find $\bigcup_{n=1}^{\infty} A_n$.
- c) Find $\bigcap_{n=1}^{\infty} A_n$.

Exercise 0.3.9: Determine $\mathcal{P}(S)$ (the power set) for each of the following:

- a) $S = \emptyset$,
- b) $S = \{1\}$,
- c) $S = \{1, 2\}$,
- d) $S = \{1, 2, 3, 4\}$.

Exercise 0.3.10: Let $f: A \rightarrow B$ and $g: B \rightarrow C$ be functions.

- a) Prove that if $g \circ f$ is injective, then f is injective.
- b) Prove that if $g \circ f$ is surjective, then g is surjective.
- c) Find an explicit example where $g \circ f$ is bijective, but neither f nor g are bijective.

Exercise 0.3.11: Prove that $n < 2^n$ by induction.

Exercise 0.3.12: Show that for a finite set A of cardinality n , the cardinality of $\mathcal{P}(A)$ is 2^n .

Exercise 0.3.13: Prove $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}$ for all $n \in \mathbb{N}$.

Exercise 0.3.14: Prove $1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$ for all $n \in \mathbb{N}$.

Exercise 0.3.15: Prove that $n^3 + 5n$ is divisible by 6 for all $n \in \mathbb{N}$.

Exercise 0.3.16: Find the smallest $n \in \mathbb{N}$ such that $2(n+5)^2 < n^3$ and call it n_0 . Show that $2(n+5)^2 < n^3$ for all $n \geq n_0$.

Exercise 0.3.17: Find all $n \in \mathbb{N}$ such that $n^2 < 2^n$.

Exercise 0.3.18: *Finish the proof that the principle of induction is equivalent to the well ordering property of \mathbb{N} . That is, prove the well ordering property for \mathbb{N} using the principle of induction.*

Exercise 0.3.19: *Give an example of a countable collection of finite sets A_1, A_2, \dots , whose union is not a finite set.*

Exercise 0.3.20: *Give an example of a countable collection of infinite sets A_1, A_2, \dots , with $A_j \cap A_k$ being infinite for all j and k , such that $\bigcap_{j=1}^{\infty} A_j$ is nonempty and finite.*

Chapter 1

Real Numbers

1.1 Basic properties

Note: 1.5 lectures

The main object we work with in analysis is the set of real numbers. As this set is so fundamental, often much time is spent on formally constructing the set of real numbers. However, we will take an easier approach here and just assume that a set with the correct properties exists. We need to start with some basic definitions.

Definition 1.1.1. An *ordered set* is a set A , together with a relation $<$ such that

- (i) For any $x, y \in A$, exactly one of $x < y$, $x = y$, or $y < x$ holds.
- (ii) If $x < y$ and $y < z$, then $x < z$.

For example, the rational numbers \mathbb{Q} are an ordered set by letting $x < y$ if and only if $y - x$ is a positive rational number. Similarly, \mathbb{N} and \mathbb{Z} are also ordered sets.

We write $x \leq y$ if $x < y$ or $x = y$. We define $>$ and \geq in the obvious way.

Definition 1.1.2. Let $E \subset A$, where A is an ordered set.

- (i) If there exists a $b \in A$ such that $x \leq b$ for all $x \in E$, then we say E is *bounded above* and b is an *upper bound* of E .
- (ii) If there exists a $b \in A$ such that $x \geq b$ for all $x \in E$, then we say E is *bounded below* and b is a *lower bound* of E .
- (iii) If there exists an upper bound b_0 of E such that whenever b is any upper bound for E we have $b_0 \leq b$, then b_0 is called the *least upper bound* or the *supremum* of E . We write

$$\sup E := b_0.$$

- (iv) Similarly, if there exists a lower bound b_0 of E such that whenever b is any lower bound for E we have $b_0 \geq b$, then b_0 is called the *greatest lower bound* or the *infimum* of E . We write

$$\inf E := b_0.$$

When a set E is both bounded above and bounded below, we will say simply that E is *bounded*.

Note that a supremum or infimum for E (even if they exist) need not be in E . For example the set $\{x \in \mathbb{Q} : x < 1\}$ has a least upper bound of 1, but 1 is not in the set itself.

Definition 1.1.3. An ordered set A has the *least-upper-bound property* if every nonempty subset $E \subset A$ that is bounded above has a least upper bound, that is $\sup E$ exists in A .

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

Example 1.1.4: For example \mathbb{Q} does not have the least-upper-bound property. The set $\{x \in \mathbb{Q} : x^2 < 2\}$ does not have a supremum. The obvious supremum $\sqrt{2}$ is not rational. Suppose that $x^2 = 2$ for some $x \in \mathbb{Q}$. Write $x = m/n$ in lowest terms. So $(m/n)^2 = 2$ or $m^2 = 2n^2$. Hence m^2 is divisible by 2 and so m is divisible by 2. We write $m = 2k$ and so we have $(2k)^2 = 2n^2$. We divide by 2 and note that $2k^2 = n^2$ and hence n is divisible by 2. But that is a contradiction as we said m/n was in lowest terms.

That \mathbb{Q} does not have the least-upper-bound property is one of the most important reasons why we work with \mathbb{R} in analysis. The set \mathbb{Q} is just fine for algebraists. But analysts require the least-upper-bound property to do any work. We also require our real numbers to have many algebraic properties. In particular, we require that they are a field.

Definition 1.1.5. A set F is called a *field* if it has two operations defined on it, addition $x + y$ and multiplication xy , and if it satisfies the following axioms.

- (A1) If $x \in F$ and $y \in F$, then $x + y \in F$.
- (A2) (*commutativity of addition*) If $x + y = y + x$ for all $x, y \in F$.
- (A3) (*associativity of addition*) If $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.
- (A4) There exists an element $0 \in F$ such that $0 + x = x$ for all $x \in F$.
- (A5) For every element $x \in F$ there exists an element $-x \in F$ such that $x + (-x) = 0$.
- (M1) If $x \in F$ and $y \in F$, then $xy \in F$.
- (M2) (*commutativity of multiplication*) If $xy = yx$ for all $x, y \in F$.
- (M3) (*associativity of multiplication*) If $(xy)z = x(yz)$ for all $x, y, z \in F$.
- (M4) There exists an element 1 (and $1 \neq 0$) such that $1x = x$ for all $x \in F$.
- (M5) For every $x \in F$ such that $x \neq 0$ there exists an element $1/x \in F$ such that $x(1/x) = 1$.
- (D) (*distributive law*) $x(y + z) = xy + xz$ for all $x, y, z \in F$.

Example 1.1.6: The set \mathbb{Q} of rational numbers is a field. On the other hand \mathbb{Z} is not a field, as it does not contain multiplicative inverses.

We will assume the basic facts about fields that can easily be proved from the axioms. For example, $0x = 0$ can be easily proved by noting that $xx = (0 + x)x = 0x + xx$, using (A4), (D), and (M2). Then using (A5) on xx we obtain $0 = 0x$.

Definition 1.1.7. A field F is said to be an *ordered field* if F is also an ordered set such that:

- (i) For $x, y, z \in F$, $x < y$ implies $x + z < y + z$.
- (ii) For $x, y \in F$, $x > 0$ and $y > 0$ implies $xy > 0$.

If $x > 0$, we say x is *positive*. If $x < 0$, we say x is *negative*. We also say x is *nonnegative* if $x \geq 0$, and x is *nonpositive* if $x \leq 0$.

Proposition 1.1.8. Let F be an ordered field and $x, y, z \in F$. Then:

- (i) If $x > 0$, then $-x < 0$ (and vice-versa).
- (ii) If $x > 0$ and $y < z$, then $xy < xz$.
- (iii) If $x < 0$ and $y < z$, then $xy > xz$.
- (iv) If $x \neq 0$, then $x^2 > 0$.
- (v) If $0 < x < y$, then $0 < 1/y < 1/x$.

Note that (iv) implies in particular that $1 > 0$.

Proof. Let us prove (i). The inequality $x > 0$ implies by item (i) of definition of ordered field that $x + (-x) > 0 + (-x)$. Now apply the algebraic properties of fields to obtain $0 > -x$. The “vice-versa” follows by similar calculation.

For (ii), first notice that $y < z$ implies $0 < z - y$ by applying item (i) of the definition of ordered fields. Now apply item (ii) of the definition of ordered fields to obtain $0 < x(z - y)$. By algebraic properties we get $0 < xz - xy$, and again applying item (i) of the definition we obtain $xy < xz$.

Part (iii) is left as an exercise.

To prove part (iv) first suppose that $x > 0$. Then by item (ii) of the definition of ordered fields we obtain that $x^2 > 0$ (use $y = x$). If $x < 0$, we use part (iii) of this proposition. Plug in $y = x$ and $z = 0$.

Finally to prove part (v), notice that $1/x$ cannot be equal to zero (why?). Suppose $1/x < 0$, then $-1/x > 0$ by (i). Then apply part (ii) (as $x > 0$) to obtain $x(-1/x) > 0x$ or $-1 > 0$, which contradicts $1 > 0$ by using part (i) again. Hence $1/x > 0$. Similarly $1/y > 0$. Thus $(1/x)(1/y) > 0$ by definition of ordered field and by part (ii)

$$(1/x)(1/y)x < (1/x)(1/y)y.$$

By algebraic properties we get $1/y < 1/x$. □

Product of two positive numbers (elements of an ordered field) is positive. However, it is not true that if the product is positive, then each of the two factors must be positive.

Proposition 1.1.9. *Let $x, y \in F$ where F is an ordered field. Suppose that $xy > 0$. Then either both x and y are positive, or both are negative.*

Proof. It is clear that both possibilities can happen. If either x and y are zero, then xy is zero and hence not positive. Hence we can assume that x and y are nonzero, and we simply need to show that if they have opposite signs, then $xy < 0$. Without loss of generality suppose that $x > 0$ and $y < 0$. Multiply $y < 0$ by x to get $xy < 0x = 0$. The result follows by contrapositive. \square

1.1.1 Exercises

Exercise 1.1.1: *Prove part (iii) of Proposition 1.1.8.*

Exercise 1.1.2: *Let S be an ordered set. Let $A \subset S$ be a nonempty finite subset. Then A is bounded. Furthermore, $\inf A$ exists and is in A and $\sup A$ exists and is in A . Hint: Use induction.*

Exercise 1.1.3: *Let $x, y \in F$, where F is an ordered field. Suppose that $0 < x < y$. Show that $x^2 < y^2$.*

Exercise 1.1.4: *Let S be an ordered set. Let $B \subset S$ be bounded (above and below). Let $A \subset B$ be a nonempty subset. Suppose that all the inf's and sup's exist. Show that*

$$\inf B \leq \inf A \leq \sup A \leq \sup B.$$

Exercise 1.1.5: *Let S be an ordered set. Let $A \subset S$ and suppose that b is an upper bound for A . Suppose that $b \in A$. Show that $b = \sup A$.*

Exercise 1.1.6: *Let S be an ordered set. Let $A \subset S$ be a nonempty subset that is bounded above. Suppose that $\sup A$ exists and that $\sup A \notin A$. Show that A contains a countably infinite subset. In particular, A is infinite.*

Exercise 1.1.7: *Find a (nonstandard) ordering of the set of natural numbers \mathbb{N} such that there exists a nonempty proper subset $A \subsetneq \mathbb{N}$ and such that $\sup A$ exists in \mathbb{N} , but $\sup A \notin A$.*

Exercise 1.1.8: *Let $F = \{0, 1, 2\}$. a) Prove that there is exactly one way to define addition and multiplication so that F is a field if 0 and 1 have their usual meaning of (A4) and (M4). b) Show that F cannot be an ordered field.*

Exercise 1.1.9: *Let S be an ordered set and A is a nonempty subset such that $\sup A$ exists. Suppose that there is a $B \subset A$ such that whenever $x \in A$ there is a $y \in B$ such that $x \leq y$. Show that $\sup B$ exists and $\sup B = \sup A$.*

1.2 The set of real numbers

Note: 2 lectures, the extended real numbers are optional

1.2.1 The set of real numbers

We finally get to the real number system. Instead of constructing the real number set from the rational numbers, we simply state their existence as a theorem without proof. Notice that \mathbb{Q} is an ordered field.

Theorem 1.2.1. *There exists a unique* ordered field \mathbb{R} with the least-upper-bound property such that $\mathbb{Q} \subset \mathbb{R}$.*

Note that also $\mathbb{N} \subset \mathbb{Q}$. As we have seen, $1 > 0$. By induction (exercise) we can prove that $n > 0$ for all $n \in \mathbb{N}$. Similarly we can easily verify all the statements we know about rational numbers and their natural ordering.

Let us prove one of the most basic but useful results about the real numbers. The following proposition is essentially how an analyst proves that a number is zero.

Proposition 1.2.2. *If $x \in \mathbb{R}$ is such that $x \geq 0$ and $x \leq \varepsilon$ for all $\varepsilon \in \mathbb{R}$ where $\varepsilon > 0$, then $x = 0$.*

Proof. If $x > 0$, then $0 < x/2 < x$ (why?). Taking $\varepsilon = x/2$ obtains a contradiction. Thus $x = 0$. \square

A related simple fact is that any time we have two real numbers $a < b$, then there is another real number c such that $a < c < b$. Just take for example $c = \frac{a+b}{2}$ (why?). In fact, there are infinitely many real numbers between a and b .

The most useful property of \mathbb{R} for analysts is not just that it is an ordered field, but that it has the least-upper-bound property. Essentially we want \mathbb{Q} , but we also want to take suprema (and infima) willy-nilly. So what we do is to throw in enough numbers to obtain \mathbb{R} .

We have already seen that \mathbb{R} must contain elements that are not in \mathbb{Q} because of the least-upper-bound property. We have seen that there is no rational square root of two. The set $\{x \in \mathbb{Q} : x^2 < 2\}$ implies the existence of the real number $\sqrt{2}$ that is not rational, although this fact requires a bit of work.

Example 1.2.3: Claim: There exists a unique positive real number r such that $r^2 = 2$. We denote r by $\sqrt{2}$.

Proof. Take the set $A := \{x \in \mathbb{R} : x^2 < 2\}$. First if $x^2 < 2$, then $x < 2$. To see this fact, note that $x \geq 2$ implies $x^2 \geq 4$ (use Proposition 1.1.8, we will not explicitly mention its use from now on), hence any number x such that $x \geq 2$ is not in A . Thus A is bounded above. On the other hand, $1 \in A$, so A is nonempty.

*Uniqueness is up to isomorphism, but we wish to avoid excessive use of algebra. For us, it is simply enough to assume that a set of real numbers exists. See Rudin [R2] for the construction and more details.

Let us define $r := \sup A$. We will show that $r^2 = 2$ by showing that $r^2 \geq 2$ and $r^2 \leq 2$. This is the way analysts show equality, by showing two inequalities. We already know that $r \geq 1 > 0$.

In the following, it may seem that we are pulling certain expressions out of a hat. When writing a proof such as this we would of course come up with the expressions only after playing around with what we wish to prove. The order in which we write the proof is not necessarily the order in which we come up with the proof.

Let us first show that $r^2 \geq 2$. Take a positive number s such that $s^2 < 2$. We wish to find an $h > 0$ such that $(s+h)^2 < 2$. As $2 - s^2 > 0$, we have $\frac{2-s^2}{2s+1} > 0$. We choose an $h \in \mathbb{R}$ such that $0 < h < \frac{2-s^2}{2s+1}$. Furthermore, we assume that $h < 1$.

$$\begin{aligned} (s+h)^2 - s^2 &= h(2s+h) \\ &< h(2s+1) && \text{(since } h < 1\text{)} \\ &< 2 - s^2 && \left(\text{since } h < \frac{2-s^2}{2s+1} \right). \end{aligned}$$

Therefore, $(s+h)^2 < 2$. Hence $s+h \in A$, but as $h > 0$ we have $s+h > s$. So $s < r = \sup A$. As s was an arbitrary positive number such that $s^2 < 2$, it follows that $r^2 \geq 2$.

Now take a positive number s such that $s^2 > 2$. We wish to find an $h > 0$ such that $(s-h)^2 > 2$. As $s^2 - 2 > 0$ we have $\frac{s^2-2}{2s} > 0$. We choose an $h \in \mathbb{R}$ such that $0 < h < \frac{s^2-2}{2s}$ and $h < s$.

$$\begin{aligned} s^2 - (s-h)^2 &= 2sh - h^2 \\ &< 2sh \\ &< s^2 - 2 && \left(\text{since } h < \frac{s^2-2}{2s} \right). \end{aligned}$$

By subtracting s^2 from both sides and multiplying by -1 , we find $(s-h)^2 > 2$. Therefore $s-h \notin A$.

Furthermore, if $x \geq s-h$, then $x^2 \geq (s-h)^2 > 2$ (as $x > 0$ and $s-h > 0$) and so $x \notin A$. Thus $s-h$ is an upper bound for A . However, $s-h < s$, or in other words $s > r = \sup A$. Thus $r^2 \leq 2$.

Together, $r^2 \geq 2$ and $r^2 \leq 2$ imply $r^2 = 2$. The existence part is finished. We still need to handle uniqueness. Suppose that $s \in \mathbb{R}$ such that $s^2 = 2$ and $s > 0$. Thus $s^2 = r^2$. However, if $0 < s < r$, then $s^2 < r^2$. Similarly $0 < r < s$ implies $r^2 < s^2$. Hence $s = r$. \square

The number $\sqrt{2} \notin \mathbb{Q}$. The set $\mathbb{R} \setminus \mathbb{Q}$ is called the set of *irrational* numbers. We have just seen that $\mathbb{R} \setminus \mathbb{Q}$ is nonempty. Not only is it nonempty, we will see later that it is very large indeed.

Using the same technique as above, we can show that a positive real number $x^{1/n}$ exists for all $n \in \mathbb{N}$ and all $x > 0$. That is, for each $x > 0$, there exists a unique positive real number r such that $r^n = x$. The proof is left as an exercise.

1.2.2 Archimedean property

As we have seen, there are plenty of real numbers in any interval. But there are also infinitely many rational numbers in any interval. The following is one of the fundamental facts about the real numbers. The two parts of the next theorem are actually equivalent, even though it may not seem like that at first sight.

Theorem 1.2.4.

(i) (Archimedean property) *If $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that*

$$nx > y.$$

(ii) (\mathbb{Q} is dense in \mathbb{R}) *If $x, y \in \mathbb{R}$ and $x < y$, then there exists an $r \in \mathbb{Q}$ such that $x < r < y$.*

Proof. Let us prove (i). We divide through by x and then (i) says that for any real number $t := y/x$, we can find natural number n such that $n > t$. In other words, (i) says that $\mathbb{N} \subset \mathbb{R}$ is not bounded above. Suppose for contradiction that \mathbb{N} is bounded above. Let $b := \sup \mathbb{N}$. The number $b - 1$ cannot possibly be an upper bound for \mathbb{N} as it is strictly less than b (the supremum). Thus there exists an $m \in \mathbb{N}$ such that $m > b - 1$. We add one to obtain $m + 1 > b$, which contradicts b being an upper bound.

Let us tackle (ii). First assume that $x \geq 0$. Note that $y - x > 0$. By (i), there exists an $n \in \mathbb{N}$ such that

$$n(y - x) > 1.$$

Also by (i) the set $A := \{k \in \mathbb{N} : k > nx\}$ is nonempty. By the well ordering property of \mathbb{N} , A has a least element m . As $m \in A$, then $m > nx$. As m is the least element of A , $m - 1 \notin A$. If $m > 1$, then $m - 1 \in \mathbb{N}$, but $m - 1 \notin A$ and so $m - 1 \leq nx$. If $m = 1$, then $m - 1 = 0$, and $m - 1 \leq nx$ still holds as $x \geq 0$. In other words,

$$m - 1 \leq nx < m.$$

We divide through by n to get $x < m/n$. On the other hand from $n(y - x) > 1$ we obtain $ny > 1 + nx$. As $nx \geq m - 1$ we get that $1 + nx \geq m$ and hence $ny > m$ and therefore $y > m/n$. Putting everything together we obtain $x < m/n < y$. So let $r = m/n$.

Now assume that $x < 0$. If $y > 0$, then we just take $r = 0$. If $y < 0$, then note that $0 < -y < -x$ and find a rational q such that $-y < q < -x$. Then take $r = -q$. \square

Let us state and prove a simple but useful corollary of the Archimedean property.

Corollary 1.2.5. $\inf\{1/n : n \in \mathbb{N}\} = 0$.

Proof. Let $A := \{1/n : n \in \mathbb{N}\}$. Obviously A is not empty. Furthermore, $1/n > 0$ and so 0 is a lower bound, and $b := \inf A$ exists. As 0 is a lower bound, then $b \geq 0$. Now take an arbitrary $a > 0$. By the Archimedean property there exists an n such that $na > 1$, or in other words $a > 1/n \in A$. Therefore a cannot be a lower bound for A . Hence $b = 0$. \square

1.2.3 Using supremum and infimum

We want to make sure that suprema and infima are compatible with algebraic operations. For a set $A \subset \mathbb{R}$ and a number x define

$$\begin{aligned}x + A &:= \{x + y \in \mathbb{R} : y \in A\}, \\xA &:= \{xy \in \mathbb{R} : y \in A\}.\end{aligned}$$

Proposition 1.2.6. *Let $A \subset \mathbb{R}$ be bounded and nonempty.*

- (i) *If $x \in \mathbb{R}$, then $\sup(x + A) = x + \sup A$.*
- (ii) *If $x \in \mathbb{R}$, then $\inf(x + A) = x + \inf A$.*
- (iii) *If $x > 0$, then $\sup(xA) = x(\sup A)$.*
- (iv) *If $x > 0$, then $\inf(xA) = x(\inf A)$.*
- (v) *If $x < 0$, then $\sup(xA) = x(\inf A)$.*
- (vi) *If $x < 0$, then $\inf(xA) = x(\sup A)$.*

Do note that multiplying a set by a negative number switches supremum for an infimum and vice-versa. The proposition also implies that if A is nonempty and bounded then xA and $x + A$ are nonempty and bounded.

Proof. Let us only prove the first statement. The rest are left as exercises.

Suppose that b is an upper bound for A . That is, $y \leq b$ for all $y \in A$. Then $x + y \leq x + b$ for all $y \in A$, and so $x + b$ is an upper bound for $x + A$. In particular, if $b = \sup A$, then

$$\sup(x + A) \leq x + b = x + \sup A.$$

The other direction is similar. If b is an upper bound for $x + A$, then $x + y \leq b$ for all $y \in A$ and so $y \leq b - x$ for all $y \in A$. So $b - x$ is an upper bound for A . If $b = \sup(x + A)$, then

$$\sup A \leq b - x = \sup(x + A) - x.$$

And the result follows. □

Sometimes we will need to apply supremum twice. Here is an example.

Proposition 1.2.7. *Let $A, B \subset \mathbb{R}$ be nonempty sets such that $x \leq y$ whenever $x \in A$ and $y \in B$. Then A is bounded above, B is bounded below, and $\sup A \leq \inf B$.*

Proof. First note that any $x \in A$ is a lower bound for B . Therefore $x \leq \inf B$. Now $\inf B$ is an upper bound for A and therefore $\sup A \leq \inf B$. □

We have to be careful about strict inequalities and taking suprema and infima. Note that $x < y$ whenever $x \in A$ and $y \in B$ still only implies $\sup A \leq \inf B$, and not a strict inequality. This is an important subtle point that comes up often. For example, take $A := \{0\}$ and take $B := \{1/n : n \in \mathbb{N}\}$. Then $0 < 1/n$ for all $n \in \mathbb{N}$. However, $\sup A = 0$ and $\inf B = 0$ as we have seen.

The proof of the following often used elementary fact is left to the reader. A similar statement holds for infima.

Proposition 1.2.8. *If $S \subset \mathbb{R}$ is a nonempty bounded set, then for every $\varepsilon > 0$ there exists $x \in S$ such that $\sup S - \varepsilon < x \leq \sup S$.*

To make using suprema and infima even easier, we may want to write $\sup A$ and $\inf A$ without worrying about A being bounded and nonempty. We make the following natural definitions.

Definition 1.2.9. Let $A \subset \mathbb{R}$ be a set.

- (i) If A is empty, then $\sup A := -\infty$.
- (ii) If A is not bounded above, then $\sup A := \infty$.
- (iii) If A is empty, then $\inf A := \infty$.
- (iv) If A is not bounded below, then $\inf A := -\infty$.

For convenience, ∞ and $-\infty$ are sometimes treated as if they were numbers, except we do not allow arbitrary arithmetic with them. We make $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, \infty\}$ into an ordered set by letting

$$-\infty < \infty \quad \text{and} \quad -\infty < x \quad \text{and} \quad x < \infty \quad \text{for all } x \in \mathbb{R}.$$

The set \mathbb{R}^* is called the set of *extended real numbers*. It is possible to define some arithmetic on \mathbb{R}^* . Most operations are extended in an obvious way, but we must leave $\infty - \infty$, $0 \cdot (\pm\infty)$, and $\frac{\pm\infty}{\pm\infty}$ undefined. We refrain from using this arithmetic so as it leads to easy mistakes as \mathbb{R}^* is not a field. Now we can take suprema and infima without fear of emptiness or unboundedness. In this book we mostly avoid using \mathbb{R}^* outside of exercises, and leave such generalizations to the interested reader.

1.2.4 Maxima and minima

By Exercise 1.1.2 we know that a finite set of numbers always has a supremum or an infimum that is contained in the set itself. In this case we usually do not use the words supremum or infimum.

When we have a set A of real numbers bounded above, such that $\sup A \in A$, then we can use the word *maximum* and the notation $\max A$ to denote the supremum. Similarly for infimum; when a set A is bounded below and $\inf A \in A$, then we can use the word *minimum* and the notation $\min A$. For example,

$$\begin{aligned} \max\{1, 2.4, \pi, 100\} &= 100, \\ \min\{1, 2.4, \pi, 100\} &= 1. \end{aligned}$$

While writing \sup and \inf may be technically correct in this situation, \max and \min are generally used to emphasize that the supremum or infimum is in the set itself.

1.2.5 Exercises

Exercise 1.2.1: Prove that if $t > 0$ ($t \in \mathbb{R}$), then there exists an $n \in \mathbb{N}$ such that $\frac{1}{n^2} < t$.

Exercise 1.2.2: Prove that if $t \geq 0$ ($t \in \mathbb{R}$), then there exists an $n \in \mathbb{N}$ such that $n - 1 \leq t < n$.

Exercise 1.2.3: Finish the proof of Proposition 1.2.6.

Exercise 1.2.4: Let $x, y \in \mathbb{R}$. Suppose that $x^2 + y^2 = 0$. Prove that $x = 0$ and $y = 0$.

Exercise 1.2.5: Show that $\sqrt{3}$ is irrational.

Exercise 1.2.6: Let $n \in \mathbb{N}$. Show that either \sqrt{n} is either an integer or it is irrational.

Exercise 1.2.7: Prove the arithmetic-geometric mean inequality. That is, for two positive real numbers x, y we have

$$\sqrt{xy} \leq \frac{x+y}{2}.$$

Furthermore, equality occurs if and only if $x = y$.

Exercise 1.2.8: Show that for any two real numbers such that $x < y$, we have an irrational number s such that $x < s < y$. Hint: Apply the density of \mathbb{Q} to $\frac{x}{\sqrt{2}}$ and $\frac{y}{\sqrt{2}}$.

Exercise 1.2.9: Let A and B be two nonempty bounded sets of real numbers. Let $C := \{a + b : a \in A, b \in B\}$. Show that C is a bounded set and that

$$\sup C = \sup A + \sup B \quad \text{and} \quad \inf C = \inf A + \inf B.$$

Exercise 1.2.10: Let A and B be two nonempty bounded sets of nonnegative real numbers. Define the set $C := \{ab : a \in A, b \in B\}$. Show that C is a bounded set and that

$$\sup C = (\sup A)(\sup B) \quad \text{and} \quad \inf C = (\inf A)(\inf B).$$

Exercise 1.2.11 (Hard): Given $x > 0$ and $n \in \mathbb{N}$, show that there exists a unique positive real number r such that $x = r^n$. Usually r is denoted by $x^{1/n}$.

1.3 Absolute value

Note: 0.5–1 lecture

A concept we will encounter over and over is the concept of *absolute value*. You want to think of the absolute value as the “size” of a real number. Let us give a formal definition.

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

Let us give the main features of the absolute value as a proposition.

Proposition 1.3.1.

- (i) $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$.
- (ii) $|-x| = |x|$ for all $x \in \mathbb{R}$.
- (iii) $|xy| = |x||y|$ for all $x, y \in \mathbb{R}$.
- (iv) $|x|^2 = x^2$ for all $x \in \mathbb{R}$.
- (v) $|x| \leq y$ if and only if $-y \leq x \leq y$.
- (vi) $-|x| \leq x \leq |x|$ for all $x \in \mathbb{R}$.

Proof. (i): This statement is obvious from the definition.

(ii): Suppose that $x > 0$, then $|-x| = -(-x) = x = |x|$. Similarly when $x < 0$, or $x = 0$.

(iii): If x or y is zero, then the result is obvious. When x and y are both positive, then $|x||y| = xy$. xy is also positive and hence $xy = |xy|$. If x and y are both negative then xy is still positive and $xy = |xy|$, and $|x||y| = (-x)(-y) = xy$. Next assume that $x > 0$ and $y < 0$. Then $|x||y| = x(-y) = -(xy)$. Now xy is negative and hence $|xy| = -(xy)$. Similarly if $x < 0$ and $y > 0$.

(iv): Obvious if $x \geq 0$. If $x < 0$, then $|x|^2 = (-x)^2 = x^2$.

(v): Suppose that $|x| \leq y$. If $x \geq 0$, then $x \leq y$. Obviously $y \geq 0$ and hence $-y \leq 0 \leq x$ so $-y \leq x \leq y$ holds. If $x < 0$, then $|x| \leq y$ means $-x \leq y$. Negating both sides we get $x \geq -y$. Again $y \geq 0$ and so $y \geq 0 > x$. Hence, $-y \leq x \leq y$.

On the other hand, suppose that $-y \leq x \leq y$ is true. If $x \geq 0$, then $x \leq y$ is equivalent to $|x| \leq y$. If $x < 0$, then $-y \leq x$ implies $(-x) \leq y$, which is equivalent to $|x| \leq y$.

(vi): Apply (v) with $y = |x|$. □

A property used frequently enough to give it a name is the so-called *triangle inequality*.

Proposition 1.3.2 (Triangle Inequality). $|x + y| \leq |x| + |y|$ for all $x, y \in \mathbb{R}$.

Proof. From Proposition 1.3.1 we have $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. We add these two inequalities to obtain

$$-(|x| + |y|) \leq x + y \leq |x| + |y|.$$

Again by Proposition 1.3.1 we have that $|x + y| \leq |x| + |y|$. □

There are other versions of the triangle inequality that are applied often.

Corollary 1.3.3. *Let $x, y \in \mathbb{R}$*

(i) (reverse triangle inequality) $|(|x| - |y|)| \leq |x - y|$.

(ii) $|x - y| \leq |x| + |y|$.

Proof. Let us plug in $x = a - b$ and $y = b$ into the standard triangle inequality to obtain

$$|a| = |a - b + b| \leq |a - b| + |b|,$$

or $|a| - |b| \leq |a - b|$. Switching the roles of a and b we obtain or $|b| - |a| \leq |b - a| = |a - b|$. Now applying Proposition 1.3.1 again we obtain the reverse triangle inequality.

The second version of the triangle inequality is obtained from the standard one by just replacing y with $-y$ and noting again that $|-y| = |y|$. □

Corollary 1.3.4. *Let $x_1, x_2, \dots, x_n \in \mathbb{R}$. Then*

$$|x_1 + x_2 + \dots + x_n| \leq |x_1| + |x_2| + \dots + |x_n|.$$

Proof. We proceed by induction. The conclusion holds trivially for $n = 1$, and for $n = 2$ it is the standard triangle inequality. Suppose that the corollary holds for n . Take $n + 1$ numbers x_1, x_2, \dots, x_{n+1} and first use the standard triangle inequality, then the induction hypothesis

$$\begin{aligned} |x_1 + x_2 + \dots + x_n + x_{n+1}| &\leq |x_1 + x_2 + \dots + x_n| + |x_{n+1}| \\ &\leq |x_1| + |x_2| + \dots + |x_n| + |x_{n+1}|. \end{aligned}$$

□

Let us see an example of the use of the triangle inequality.

Example 1.3.5: Find a number M such that $|x^2 - 9x + 1| \leq M$ for all $-1 \leq x \leq 5$.

Using the triangle inequality, write

$$|x^2 - 9x + 1| \leq |x^2| + |9x| + |1| = |x|^2 + 9|x| + 1.$$

It is obvious that $|x|^2 + 9|x| + 1$ is largest when $|x|$ is largest. In the interval provided, $|x|$ is largest when $x = 5$ and so $|x| = 5$. One possibility for M is

$$M = 5^2 + 9(5) + 1 = 71.$$

There are, of course, other M that work. The bound of 71 is much higher than it need be, but we didn't ask for the best possible M , just one that works.

The last example leads us to the concept of bounded functions.

Definition 1.3.6. Suppose $f: D \rightarrow \mathbb{R}$ is a function. We say f is *bounded* if there exists a number M such that $|f(x)| \leq M$ for all $x \in D$.

In the example we have shown that $x^2 - 9x + 1$ is bounded when considered as a function on $D = \{x: -1 \leq x \leq 5\}$. On the other hand, if we consider the same polynomial as a function on the whole real line \mathbb{R} , then it is not bounded.

For a function $f: D \rightarrow \mathbb{R}$ we write

$$\begin{aligned}\sup_{x \in D} f(x) &:= \sup f(D), \\ \inf_{x \in D} f(x) &:= \inf f(D).\end{aligned}$$

To illustrate some common issues, let us prove the following proposition.

Proposition 1.3.7. If $f: D \rightarrow \mathbb{R}$ and $g: D \rightarrow \mathbb{R}$ (D nonempty) are bounded* functions and

$$f(x) \leq g(x) \quad \text{for all } x \in D,$$

then

$$\sup_{x \in D} f(x) \leq \sup_{x \in D} g(x) \quad \text{and} \quad \inf_{x \in D} f(x) \leq \inf_{x \in D} g(x). \quad (1.1)$$

You should be careful with the variables. The x on the left side of the inequality in 1.1 is different from the x on the right. You should really think of the first inequality as

$$\sup_{x \in D} f(x) \leq \sup_{y \in D} g(y).$$

Let us prove this inequality. If b is an upper bound for $g(D)$, then $f(x) \leq g(x) \leq b$ for all $x \in D$, and hence b is an upper bound for $f(D)$. Taking the least upper bound we get that for all $x \in D$

$$f(x) \leq \sup_{y \in D} g(y).$$

Therefore $\sup_{y \in D} g(y)$ is an upper bound for $f(D)$ and thus greater than or equal to the least upper bound of $f(D)$.

$$\sup_{x \in D} f(x) \leq \sup_{y \in D} g(y).$$

The second inequality (the statement about the inf) is left as an exercise.

Do note that a common mistake is to conclude that

$$\sup_{x \in D} f(x) \leq \inf_{y \in D} g(y). \quad (1.2)$$

*The boundedness hypothesis is for simplicity, it can be dropped if we allow for the extended real numbers.

The inequality (1.2) is not true given the hypothesis of the claim above. For this stronger inequality we need the stronger hypothesis

$$f(x) \leq g(y) \quad \text{for all } x \in D \text{ and } y \in D.$$

The proof is left as an exercise.

1.3.1 Exercises

Exercise 1.3.1: Show that $|x - y| < \varepsilon$ if and only if $x - \varepsilon < y < x + \varepsilon$.

Exercise 1.3.2: Show that

$$a) \max\{x, y\} = \frac{x+y+|x-y|}{2}$$

$$b) \min\{x, y\} = \frac{x+y-|x-y|}{2}$$

Exercise 1.3.3: Find a number M such that $|x^3 - x^2 + 8x| \leq M$ for all $-2 \leq x \leq 10$.

Exercise 1.3.4: Finish the proof of Proposition 1.3.7. That is, prove that given any set D , and two bounded functions $f: D \rightarrow \mathbb{R}$ and $g: D \rightarrow \mathbb{R}$ such that $f(x) \leq g(x)$ for all $x \in D$, then

$$\inf_{x \in D} f(x) \leq \inf_{x \in D} g(x).$$

Exercise 1.3.5: Let $f: D \rightarrow \mathbb{R}$ and $g: D \rightarrow \mathbb{R}$ be functions (D nonempty).

a) Suppose that $f(x) \leq g(y)$ for all $x \in D$ and $y \in D$. Show that

$$\sup_{x \in D} f(x) \leq \inf_{x \in D} g(x).$$

b) Find a specific D , f , and g , such that $f(x) \leq g(x)$ for all $x \in D$, but

$$\sup_{x \in D} f(x) > \inf_{x \in D} g(x).$$

Exercise 1.3.6: Prove Proposition 1.3.7 without the assumption that the functions are bounded. Hint: You will need to use the extended real numbers.

1.4 Intervals and the size of \mathbb{R}

Note: 0.5–1 lecture (proof of uncountability of \mathbb{R} can be optional)

You have seen the notation for intervals before, but let us give a formal definition here. For $a, b \in \mathbb{R}$ such that $a < b$ we define

$$\begin{aligned} [a, b] &:= \{x \in \mathbb{R} : a \leq x \leq b\}, \\ (a, b) &:= \{x \in \mathbb{R} : a < x < b\}, \\ (a, b] &:= \{x \in \mathbb{R} : a < x \leq b\}, \\ [a, b) &:= \{x \in \mathbb{R} : a \leq x < b\}. \end{aligned}$$

The interval $[a, b]$ is called a *closed interval* and (a, b) is called an *open interval*. The intervals of the form $(a, b]$ and $[a, b)$ are called *half-open intervals*.

The above intervals were all *bounded intervals*, since both a and b were real numbers. We define *unbounded intervals*,

$$\begin{aligned} [a, \infty) &:= \{x \in \mathbb{R} : a \leq x\}, \\ (a, \infty) &:= \{x \in \mathbb{R} : a < x\}, \\ (-\infty, b] &:= \{x \in \mathbb{R} : x \leq b\}, \\ (-\infty, b) &:= \{x \in \mathbb{R} : x < b\}. \end{aligned}$$

For completeness we define $(-\infty, \infty) := \mathbb{R}$.

We have already seen that any open interval (a, b) (where $a < b$ of course) must be nonempty. For example, it contains the number $\frac{a+b}{2}$. An unexpected fact is that from a set-theoretic perspective, all intervals have the same “size,” that is, they all have the same cardinality. For example the map $f(x) := 2x$ takes the interval $[0, 1]$ bijectively to the interval $[0, 2]$.

Maybe more interestingly, the function $f(x) := \tan(x)$ is a bijective map from $(-\pi, \pi)$ to \mathbb{R} . Hence the bounded interval $(-\pi, \pi)$ has the same cardinality as \mathbb{R} . It is not completely straightforward to construct a bijective map from $[0, 1]$ to say $(0, 1)$, but it is possible.

And do not worry, there does exist a way to measure the “size” of subsets of real numbers that “sees” the difference between $[0, 1]$ and $[0, 2]$. However, its proper definition requires much more machinery than we have right now.

Let us say more about the cardinality of intervals and hence about the cardinality of \mathbb{R} . We have seen that there exist irrational numbers, that is $\mathbb{R} \setminus \mathbb{Q}$ is nonempty. The question is: How many irrational numbers are there? It turns out there are a lot more irrational numbers than rational numbers. We have seen that \mathbb{Q} is countable, and we will show that \mathbb{R} is uncountable. In fact, the cardinality of \mathbb{R} is the same as the cardinality of $\mathcal{P}(\mathbb{N})$, although we will not prove this claim.

Theorem 1.4.1 (Cantor). \mathbb{R} is uncountable.

We give a modified version of Cantor's original proof from 1874 as this proof requires the least setup. Normally this proof is stated as a contradiction proof, but a proof by contrapositive is easier to understand.

Proof. Let $X \subset \mathbb{R}$ be a countably infinite subset such that for any two real numbers $a < b$, there is an $x \in X$ such that $a < x < b$. Were \mathbb{R} countable, then we could take $X = \mathbb{R}$. If we show that X is necessarily a proper subset, then X cannot equal \mathbb{R} , and \mathbb{R} must be uncountable.

As X is countably infinite, there is a bijection from \mathbb{N} to X . Consequently, we write X as a sequence of real numbers x_1, x_2, x_3, \dots , such that each number in X is given by x_j for some $j \in \mathbb{N}$.

Let us inductively construct two sequences of real numbers a_1, a_2, a_3, \dots and b_1, b_2, b_3, \dots . Let $a_1 := x_1$ and $b_1 := x_1 + 1$. Note that $a_1 < b_1$ and $x_1 \notin (a_1, b_1)$. For $k > 1$, suppose that a_{k-1} and b_{k-1} has been defined. Let us also suppose that (a_{k-1}, b_{k-1}) does not contain any x_j for any $j = 1, \dots, k-1$.

- (i) Define $a_k := x_j$, where j is the smallest $j \in \mathbb{N}$ such that $x_j \in (a_{k-1}, b_{k-1})$. Such an x_j exists by our assumption on X .
- (ii) Next, define $b_k := x_j$ where j is the smallest $j \in \mathbb{N}$ such that $x_j \in (a_k, b_{k-1})$.

Notice that $a_k < b_k$ and $a_{k-1} < a_k < b_k < b_{k-1}$. Also notice that (a_k, b_k) does not contain x_k and hence does not contain any x_j for $j = 1, \dots, k$.

Claim: $a_j < b_k$ for all j and k in \mathbb{N} . Let us first assume that $j < k$. Then $a_j < a_{j+1} < \dots < a_{k-1} < a_k < b_k$. Similarly for $j > k$. The claim follows.

Let $A = \{a_j : j \in \mathbb{N}\}$ and $B = \{b_j : j \in \mathbb{N}\}$. By Proposition 1.2.7 and the claim above we have

$$\sup A \leq \inf B.$$

Define $y := \sup A$. The number y cannot be a member of A . If $y = a_j$ for some j , then $y < a_{j+1}$, which is impossible. Similarly y cannot be a member of B . Therefore, $a_j < y$ for all $j \in \mathbb{N}$ and $y < b_j$ for all $j \in \mathbb{N}$. In other words $y \in (a_j, b_j)$ for all $j \in \mathbb{N}$.

Finally we have to show that $y \notin X$. If we do so, then we have constructed a real number not in X showing that X must have been a proper subset. Take any $x_k \in X$. By the above construction $x_k \notin (a_k, b_k)$, so $x_k \neq y$ as $y \in (a_k, b_k)$.

Therefore, the sequence x_1, x_2, \dots cannot contain all elements of \mathbb{R} and thus \mathbb{R} is uncountable. \square

1.4.1 Exercises

Exercise 1.4.1: For $a < b$, construct an explicit bijection from $(a, b]$ to $(0, 1]$.

Exercise 1.4.2: Suppose $f: [0, 1] \rightarrow (0, 1)$ is a bijection. Using f , construct a bijection from $[-1, 1]$ to \mathbb{R} .

Exercise 1.4.3 (Hard): Show that the cardinality of \mathbb{R} is the same as the cardinality of $\mathcal{P}(\mathbb{N})$. Hint: If you have a binary representation of a real number in the interval $[0, 1]$, then you have a sequence of 1's and 0's. Use the sequence to construct a subset of \mathbb{N} . The tricky part is to notice that some numbers have more than one binary representation.

Exercise 1.4.4 (Hard): Construct an explicit bijection from $(0, 1]$ to $(0, 1)$. *Hint: One approach is as follows: First map $(1/2, 1]$ to $(0, 1/2]$, then map $(1/4, 1/2]$ to $(1/2, 3/4]$, etc. . . . Write down the map explicitly, that is, write down an algorithm that tells you exactly what number goes where. Then prove that the map is a bijection.*

Exercise 1.4.5 (Hard): Construct an explicit bijection from $[0, 1]$ to $(0, 1)$.

Chapter 2

Sequences and Series

2.1 Sequences and limits

Note: 2.5 lectures

Analysis is essentially about taking limits. The most basic type of a limit is a limit of a sequence of real numbers. We have already seen sequences used informally. Let us give the formal definition.

Definition 2.1.1. A *sequence* (of real numbers) is a function $x: \mathbb{N} \rightarrow \mathbb{R}$. Instead of $x(n)$ we usually denote the n th element in the sequence by x_n . We use the notation $\{x_n\}$, or more precisely

$$\{x_n\}_{n=1}^{\infty},$$

to denote a sequence.

A sequence $\{x_n\}$ is *bounded* if there exists a $B \in \mathbb{R}$ such that

$$|x_n| \leq B \quad \text{for all } n \in \mathbb{N}.$$

In other words, the sequence $\{x_n\}$ is bounded whenever the set $\{x_n : n \in \mathbb{N}\}$ is bounded.

When we need to give a concrete sequence we often give each term as a formula in terms of n . For example, $\{1/n\}_{n=1}^{\infty}$, or simply $\{1/n\}$, stands for the sequence $1, 1/2, 1/3, 1/4, 1/5, \dots$. The sequence $\{1/n\}$ is a bounded sequence ($B = 1$ will suffice). On the other hand the sequence $\{n\}$ stands for $1, 2, 3, 4, \dots$, and this sequence is not bounded (why?).

While the notation for a sequence is similar* to that of a set, the notions are distinct. For example, the sequence $\{(-1)^n\}$ is the sequence $-1, 1, -1, 1, -1, 1, \dots$, whereas the set of values, the *range of the sequence*, is just the set $\{-1, 1\}$. We can write this set as $\{(-1)^n : n \in \mathbb{N}\}$. When ambiguity can arise, we use the words *sequence* or *set* to distinguish the two concepts.

Another example of a sequence is the so-called *constant sequence*. That is a sequence $\{c\} = c, c, c, c, \dots$ consisting of a single constant $c \in \mathbb{R}$ repeating indefinitely.

*[BS] use the notation (x_n) to denote a sequence instead of $\{x_n\}$, which is what [R2] uses. Both are common.

We now get to the idea of a *limit of a sequence*. We will see in Proposition 2.1.6 that the notation below is well defined. That is, if a limit exists, then it is unique. So it makes sense to talk about *the* limit of a sequence.

Definition 2.1.2. A sequence $\{x_n\}$ is said to *converge* to a number $x \in \mathbb{R}$, if for every $\varepsilon > 0$, there exists an $M \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for all $n \geq M$. The number x is said to be the *limit* of $\{x_n\}$. We write

$$\lim_{n \rightarrow \infty} x_n := x.$$

A sequence that converges is said to be *convergent*. Otherwise, the sequence is said to be *divergent*.

It is good to know intuitively what a limit means. It means that eventually every number in the sequence is close to the number x . More precisely, we can be arbitrarily close to the limit, provided we go far enough in the sequence. It does not mean we can ever reach the limit. It is possible, and quite common, that there is no x_n in the sequence that equals the limit x .

When we write $\lim x_n = x$ for some real number x , we are saying two things. First, that $\{x_n\}$ is convergent, and second that the limit is x .

The above definition is one of the most important definitions in analysis, and it is necessary to understand it perfectly. The key point in the definition is that given *any* $\varepsilon > 0$, we can find an M . The M can depend on ε , so we only pick an M once we know ε . Let us illustrate this concept on a few examples.

Example 2.1.3: The constant sequence $1, 1, 1, 1, \dots$ is convergent and the limit is 1. For every $\varepsilon > 0$, we pick $M = 1$.

Example 2.1.4: Claim: The sequence $\{1/n\}$ is convergent and

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

Proof: Given an $\varepsilon > 0$, we find an $M \in \mathbb{N}$ such that $0 < 1/M < \varepsilon$ (Archimedean property at work). Then for all $n \geq M$ we have that

$$|x_n - 0| = \left| \frac{1}{n} \right| = \frac{1}{n} \leq \frac{1}{M} < \varepsilon.$$

Example 2.1.5: The sequence $\{(-1)^n\}$ is divergent. Proof: If there were a limit x , then for $\varepsilon = \frac{1}{2}$ we expect an M that satisfies the definition. Suppose such an M exists, then for an even $n \geq M$ we compute

$$\frac{1}{2} > |x_n - x| = |1 - x| \quad \text{and} \quad \frac{1}{2} > |x_{n+1} - x| = |-1 - x|.$$

But

$$2 = |1 - x - (-1 - x)| \leq |1 - x| + |-1 - x| < \frac{1}{2} + \frac{1}{2} = 1,$$

and that is a contradiction.

Proposition 2.1.6. *A convergent sequence has a unique limit.*

The proof of this proposition exhibits a useful technique in analysis. Many proofs follow the same general scheme. We want to show a certain quantity is zero. We write the quantity using the triangle inequality as two quantities, and we estimate each one by arbitrarily small numbers.

Proof. Suppose that the sequence $\{x_n\}$ has the limit x and the limit y . Take an arbitrary $\varepsilon > 0$. From the definition we find an M_1 such that for all $n \geq M_1$, $|x_n - x| < \varepsilon/2$. Similarly we find an M_2 such that for all $n \geq M_2$ we have $|x_n - y| < \varepsilon/2$. Take $M := \max\{M_1, M_2\}$. For $n \geq M$ (so that both $n \geq M_1$ and $n \geq M_2$) we have

$$\begin{aligned} |y - x| &= |x_n - x - (x_n - y)| \\ &\leq |x_n - x| + |x_n - y| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

As $|y - x| < \varepsilon$ for all $\varepsilon > 0$, then $|y - x| = 0$ and $y = x$. Hence the limit (if it exists) is unique. \square

Proposition 2.1.7. *A convergent sequence $\{x_n\}$ is bounded.*

Proof. Suppose that $\{x_n\}$ converges to x . Thus there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $|x_n - x| < 1$. Let $B_1 := |x| + 1$ and note that for $n \geq M$ we have

$$\begin{aligned} |x_n| &= |x_n - x + x| \\ &\leq |x_n - x| + |x| \\ &< 1 + |x| = B_1. \end{aligned}$$

The set $\{|x_1|, |x_2|, \dots, |x_{M-1}|\}$ is a finite set and hence let

$$B_2 := \max\{|x_1|, |x_2|, \dots, |x_{M-1}|\}.$$

Let $B := \max\{B_1, B_2\}$. Then for all $n \in \mathbb{N}$ we have

$$|x_n| \leq B.$$

\square

The sequence $\{(-1)^n\}$ shows that the converse does not hold. A bounded sequence is not necessarily convergent.

Example 2.1.8: Let us show the sequence $\left\{\frac{n^2+1}{n^2+n}\right\}$ converges and

$$\lim_{n \rightarrow \infty} \frac{n^2+1}{n^2+n} = 1.$$

Given $\varepsilon > 0$, find $M \in \mathbb{N}$ such that $\frac{1}{M+1} < \varepsilon$. Then for any $n \geq M$ we have

$$\begin{aligned} \left| \frac{n^2+1}{n^2+n} - 1 \right| &= \left| \frac{n^2+1 - (n^2+n)}{n^2+n} \right| \\ &= \left| \frac{1-n}{n^2+n} \right| \\ &= \frac{n-1}{n^2+n} \\ &\leq \frac{n}{n^2+n} = \frac{1}{n+1} \\ &\leq \frac{1}{M+1} < \varepsilon. \end{aligned}$$

Therefore, $\lim_{n \rightarrow \infty} \frac{n^2+1}{n^2+n} = 1$.

2.1.1 Monotone sequences

The simplest type of a sequence is a monotone sequence. Checking that a monotone sequence converges is as easy as checking that it is bounded. It is also easy to find the limit for a convergent monotone sequence, provided we can find the supremum or infimum of a countable set of numbers.

Definition 2.1.9. A sequence $\{x_n\}$ is *monotone increasing* if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$. A sequence $\{x_n\}$ is *monotone decreasing* if $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$. If a sequence is either monotone increasing or monotone decreasing, we simply say the sequence is *monotone*. Some authors also use the word *monotonic*.

Theorem 2.1.10. A monotone sequence $\{x_n\}$ is bounded if and only if it is convergent.

Furthermore, if $\{x_n\}$ is monotone increasing and bounded, then

$$\lim_{n \rightarrow \infty} x_n = \sup\{x_n : n \in \mathbb{N}\}.$$

If $\{x_n\}$ is monotone decreasing and bounded, then

$$\lim_{n \rightarrow \infty} x_n = \inf\{x_n : n \in \mathbb{N}\}.$$

Proof. Let us suppose that the sequence is monotone increasing. Suppose that the sequence is bounded. That means that there exists a B such that $x_n \leq B$ for all n , that is the set $\{x_n : n \in \mathbb{N}\}$ is bounded from above. Let

$$x := \sup\{x_n : n \in \mathbb{N}\}.$$

Let $\varepsilon > 0$ be arbitrary. As x is the supremum, then there must be at least one $M \in \mathbb{N}$ such that $x_M > x - \varepsilon$ (because x is the supremum). As $\{x_n\}$ is monotone increasing, then it is easy to see (by induction) that $x_n \geq x_M$ for all $n \geq M$. Hence

$$|x_n - x| = x - x_n \leq x - x_M < \varepsilon.$$

Therefore the sequence converges to x . We already know that a convergent sequence is bounded, which completes the other direction of the implication.

The proof for monotone decreasing sequences is left as an exercise. \square

Example 2.1.11: Take the sequence $\{\frac{1}{\sqrt{n}}\}$.

First we note that $\frac{1}{\sqrt{n}} > 0$ and hence the sequence is bounded from below. Let us show that it is monotone decreasing. We start with $\sqrt{n+1} \geq \sqrt{n}$ (why is that true?). From this inequality we obtain

$$\frac{1}{\sqrt{n+1}} \leq \frac{1}{\sqrt{n}}.$$

So the sequence is monotone decreasing and bounded from below (hence bounded). We apply the theorem to note that the sequence is convergent and that in fact

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = \inf \left\{ \frac{1}{\sqrt{n}} : n \in \mathbb{N} \right\}.$$

We already know that the infimum is greater than or equal to 0, as 0 is a lower bound. Take a number $b \geq 0$ such that $b \leq \frac{1}{\sqrt{n}}$ for all n . We square both sides to obtain

$$b^2 \leq \frac{1}{n} \quad \text{for all } n \in \mathbb{N}.$$

We have seen before that this implies that $b^2 \leq 0$ (a consequence of the Archimedean property). As we also have $b^2 \geq 0$, then $b^2 = 0$ and so $b = 0$. Hence $b = 0$ is the greatest lower bound, and $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0$.

Example 2.1.12: A word of caution: We have to show that a monotone sequence is bounded in order to use Theorem 2.1.10. For example, the sequence $\{1 + 1/2 + \cdots + 1/n\}$ is a monotone increasing sequence that grows very slowly. We will see, once we get to series, that this sequence has no upper bound and so does not converge. It is not at all obvious that this sequence has no upper bound.

A common example of where monotone sequences arise is the following proposition. The proof is left as an exercise.

Proposition 2.1.13. *Let $S \subset \mathbb{R}$ be a nonempty bounded set. Then there exist monotone sequences $\{x_n\}$ and $\{y_n\}$ such that $x_n, y_n \in S$ and*

$$\sup S = \lim_{n \rightarrow \infty} x_n \quad \text{and} \quad \inf S = \lim_{n \rightarrow \infty} y_n.$$

2.1.2 Tail of a sequence

Definition 2.1.14. For a sequence $\{x_n\}$, the K -tail (where $K \in \mathbb{N}$) or just the *tail* of the sequence is the sequence starting at $K + 1$, usually written as

$$\{x_{n+K}\}_{n=1}^{\infty} \quad \text{or} \quad \{x_n\}_{n=K+1}^{\infty}.$$

The main result about the tail of a sequence is the following proposition.

Proposition 2.1.15. For any $K \in \mathbb{N}$, the sequence $\{x_n\}_{n=1}^{\infty}$ converges if and only if the K -tail $\{x_{n+K}\}_{n=1}^{\infty}$ converges. Furthermore, if the limit exists, then

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_{n+K}.$$

Proof. Define $y_n := x_{n+K}$. We wish to show that $\{x_n\}$ converges if and only if $\{y_n\}$ converges. Furthermore we wish to show that the limits are equal.

Suppose that $\{x_n\}$ converges to some $x \in \mathbb{R}$. That is, given an $\varepsilon > 0$, there exists an $M \in \mathbb{N}$ such that $|x - x_n| < \varepsilon$ for all $n \geq M$. Note that $n \geq M$ implies $n + K \geq M$. Therefore, it is true that for all $n \geq M$ we have that

$$|x - y_n| = |x - x_{n+K}| < \varepsilon.$$

Therefore $\{y_n\}$ converges to x .

Now suppose that $\{y_n\}$ converges to $x \in \mathbb{R}$. That is, given an $\varepsilon > 0$, there exists an $M' \in \mathbb{N}$ such that $|x - y_n| < \varepsilon$ for all $n \geq M'$. Let $M := M' + K$. Then $n \geq M$ implies that $n - K \geq M'$. Thus, whenever $n \geq M$ we have

$$|x - x_n| = |x - y_{n-K}| < \varepsilon.$$

Therefore $\{x_n\}$ converges to x . □

Essentially, the limit does not care about how the sequence begins, it only cares about the tail of the sequence. That is, the beginning of the sequence may be arbitrary.

2.1.3 Subsequences

A very useful concept related to sequences is that of a subsequence. A subsequence of $\{x_n\}$ is a sequence that contains only some of the numbers from $\{x_n\}$ in the same order.

Definition 2.1.16. Let $\{x_n\}$ be a sequence. Let $\{n_i\}$ be a strictly increasing sequence of natural numbers (that is $n_1 < n_2 < n_3 < \dots$). The sequence

$$\{x_{n_i}\}_{i=1}^{\infty}$$

is called a *subsequence* of $\{x_n\}$.

For example, take the sequence $\{1/n\}$. The sequence $\{1/3n\}$ is a subsequence. To see how these two sequences fit in the definition, take $n_i := 3i$. Note that the numbers in the subsequence must come from the original sequence, so $1, 0, 1/3, 0, 1/5, \dots$ is not a subsequence of $\{1/n\}$. Similarly order must be preserved, so the sequence $1, 1/3, 1/2, 1/5, \dots$ is not a subsequence of $\{1/n\}$.

Note that a tail of a sequence is one type of a subsequence. For an arbitrary subsequence, we have the following proposition.

Proposition 2.1.17. *If $\{x_n\}$ is a convergent sequence, then any subsequence $\{x_{n_i}\}$ is also convergent and*

$$\lim_{n \rightarrow \infty} x_n = \lim_{i \rightarrow \infty} x_{n_i}.$$

Proof. Suppose that $\lim_{n \rightarrow \infty} x_n = x$. That means that for every $\varepsilon > 0$ we have an $M \in \mathbb{N}$ such that for all $n \geq M$

$$|x_n - x| < \varepsilon.$$

It is not hard to prove (do it!) by induction that $n_i \geq i$. Hence $i \geq M$ implies that $n_i \geq M$. Thus, for all $i \geq M$ we have

$$|x_{n_i} - x| < \varepsilon,$$

and we are done. □

Example 2.1.18: Existence of a convergent subsequence does not imply convergence of the sequence itself. Take the sequence $0, 1, 0, 1, 0, 1, \dots$. That is $x_n = 0$ if n is odd, and $x_n = 1$ if n is even. It is not hard to see that $\{x_n\}$ is divergent, however, the subsequence $\{x_{2n}\}$ converges to 1 and the subsequence $\{x_{2n+1}\}$ converges to 0. Compare Theorem 2.3.7.

2.1.4 Exercises

In the following exercises, feel free to use what you know from calculus to find the limit, if it exists. But you must *prove* that you have found the correct limit, or prove that the series is divergent.

Exercise 2.1.1: *Is the sequence $\{3n\}$ bounded? Prove or disprove.*

Exercise 2.1.2: *Is the sequence $\{n\}$ convergent? If so, what is the limit.*

Exercise 2.1.3: *Is the sequence $\left\{\frac{(-1)^n}{2n}\right\}$ convergent? If so, what is the limit.*

Exercise 2.1.4: *Is the sequence $\{2^{-n}\}$ convergent? If so, what is the limit.*

Exercise 2.1.5: *Is the sequence $\left\{\frac{n}{n+1}\right\}$ convergent? If so, what is the limit.*

Exercise 2.1.6: *Is the sequence $\left\{\frac{n}{n^2+1}\right\}$ convergent? If so, what is the limit.*

Exercise 2.1.7: Let $\{x_n\}$ be a sequence.

- a) Show that $\lim x_n = 0$ (that is, the limit exists and is zero) if and only if $\lim |x_n| = 0$.
 b) Find an example such that $\{|x_n|\}$ converges and $\{x_n\}$ diverges.

Exercise 2.1.8: Is the sequence $\left\{\frac{2^n}{n!}\right\}$ convergent? If so, what is the limit.

Exercise 2.1.9: Show that the sequence $\left\{\frac{1}{\sqrt[3]{n}}\right\}$ is monotone, bounded, and use Theorem 2.1.10 to find the limit.

Exercise 2.1.10: Show that the sequence $\left\{\frac{n+1}{n}\right\}$ is monotone, bounded, and use Theorem 2.1.10 to find the limit.

Exercise 2.1.11: Finish the proof of Theorem 2.1.10 for monotone decreasing sequences.

Exercise 2.1.12: Prove Proposition 2.1.13.

Exercise 2.1.13: Let $\{x_n\}$ be a convergent monotone sequence. Suppose that there exists a $k \in \mathbb{N}$ such that

$$\lim_{n \rightarrow \infty} x_n = x_k.$$

Show that $x_n = x_k$ for all $n \geq k$.

Exercise 2.1.14: Find a convergent subsequence of the sequence $\{(-1)^n\}$.

Exercise 2.1.15: Let $\{x_n\}$ be a sequence defined by

$$x_n := \begin{cases} n & \text{if } n \text{ is odd,} \\ 1/n & \text{if } n \text{ is even.} \end{cases}$$

- a) Is the sequence bounded? (prove or disprove)
 b) Is there a convergent subsequence? If so, find it.

Exercise 2.1.16: Let $\{x_n\}$ be a sequence. Suppose that there are two convergent subsequences $\{x_{n_i}\}$ and $\{x_{m_i}\}$. Suppose that

$$\lim_{i \rightarrow \infty} x_{n_i} = a \quad \text{and} \quad \lim_{i \rightarrow \infty} x_{m_i} = b,$$

where $a \neq b$. Prove that $\{x_n\}$ is not convergent, without using Proposition 2.1.17.

Exercise 2.1.17: Find a sequence $\{x_n\}$ such that for any $y \in \mathbb{R}$, there exists a subsequence $\{x_{n_i}\}$ converging to y .

Exercise 2.1.18 (Easy): Let $\{x_n\}$ be a sequence and $x \in \mathbb{R}$. Suppose that for any $\varepsilon > 0$, there is an M such that for all $n \geq M$, $|x_n - x| \leq \varepsilon$. Show that $\lim x_n = x$.

2.2 Facts about limits of sequences

Note: 2–2.5 lectures, recursively defined sequences can safely be skipped

In this section we will go over some basic results about the limits of sequences. We start with looking at how sequences interact with inequalities.

2.2.1 Limits and inequalities

A basic lemma about limits and inequalities is the so-called squeeze lemma. It allows us to show convergence of sequences in difficult cases if we find two other simpler convergent sequences that “squeeze” the original sequence.

Lemma 2.2.1 (Squeeze lemma). *Let $\{a_n\}$, $\{b_n\}$, and $\{x_n\}$ be sequences such that*

$$a_n \leq x_n \leq b_n \quad \text{for all } n \in \mathbb{N}.$$

Suppose that $\{a_n\}$ and $\{b_n\}$ converge and

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n.$$

Then $\{x_n\}$ converges and

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n.$$

The intuitive idea of the proof is illustrated in Figure 2.1. If x is the limit of a_n and b_n , then if they are both within $\varepsilon/3$ of x , then the distance between a_n and b_n is at most $2\varepsilon/3$. As x_n is between a_n and b_n it is at most $2\varepsilon/3$ from a_n . Since a_n is at most $\varepsilon/3$ away from x , then x_n must be at most ε away from x . Let us follow through on this intuition rigorously.

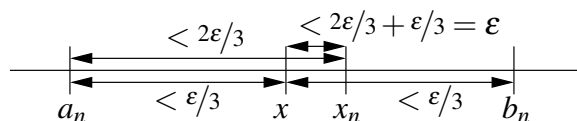


Figure 2.1: Squeeze lemma proof in picture.

Proof. Let $x := \lim a_n = \lim b_n$. Let $\varepsilon > 0$ be given.

Find an M_1 such that for all $n \geq M_1$ we have that $|a_n - x| < \varepsilon/3$, and an M_2 such that for all $n \geq M_2$ we have $|b_n - x| < \varepsilon/3$. Set $M := \max\{M_1, M_2\}$. Suppose that $n \geq M$. We compute

$$\begin{aligned} |x_n - a_n| &= x_n - a_n \leq b_n - a_n \\ &= |b_n - x + x - a_n| \\ &\leq |b_n - x| + |x - a_n| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \frac{2\varepsilon}{3}. \end{aligned}$$

Armed with this information we estimate

$$\begin{aligned} |x_n - x| &= |x_n - x + a_n - a_n| \\ &\leq |x_n - a_n| + |a_n - x| \\ &< \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

And we are done. □

Example 2.2.2: One application of the squeeze lemma is to compute limits of sequences using limits that are already known. For example, suppose that we have the sequence $\{\frac{1}{n\sqrt{n}}\}$. Since $\sqrt{n} \geq 1$ for all $n \in \mathbb{N}$ we have

$$0 \leq \frac{1}{n\sqrt{n}} \leq \frac{1}{n}$$

for all $n \in \mathbb{N}$. We already know that $\lim 1/n = 0$. Hence, using the constant sequence $\{0\}$ and the sequence $\{1/n\}$ in the squeeze lemma, we conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n\sqrt{n}} = 0.$$

Limits also preserve inequalities.

Lemma 2.2.3. Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences and

$$x_n \leq y_n,$$

for all $n \in \mathbb{N}$. Then

$$\lim_{n \rightarrow \infty} x_n \leq \lim_{n \rightarrow \infty} y_n.$$

Proof. Let $x := \lim x_n$ and $y := \lim y_n$. Let $\varepsilon > 0$ be given. Find an M_1 such that for all $n \geq M_1$ we have $|x_n - x| < \varepsilon/2$. Find an M_2 such that for all $n \geq M_2$ we have $|y_n - y| < \varepsilon/2$. In particular, for some $n \geq \max\{M_1, M_2\}$ we have $x - x_n < \varepsilon/2$ and $y_n - y < \varepsilon/2$. We add these inequalities to obtain

$$y_n - x_n + x - y < \varepsilon, \quad \text{or} \quad y_n - x_n < y - x + \varepsilon.$$

Since $x_n \leq y_n$ we have $0 \leq y_n - x_n$ and hence $0 < y - x + \varepsilon$. In other words

$$x - y < \varepsilon.$$

Because $\varepsilon > 0$ was arbitrary we obtain $x - y \leq 0$, as we have seen that a nonnegative number less than any positive ε is zero. Therefore $x \leq y$. □

We give an easy corollary that can be proved using constant sequences and an application of Lemma 2.2.3. The proof is left as an exercise.

Corollary 2.2.4.

(i) Let $\{x_n\}$ be a convergent sequence such that $x_n \geq 0$, then

$$\lim_{n \rightarrow \infty} x_n \geq 0.$$

(ii) Let $a, b \in \mathbb{R}$ and let $\{x_n\}$ be a convergent sequence such that

$$a \leq x_n \leq b,$$

for all $n \in \mathbb{N}$. Then

$$a \leq \lim_{n \rightarrow \infty} x_n \leq b.$$

In Lemma 2.2.3 and Corollary 2.2.4 we cannot simply replace all the non-strict inequalities with strict inequalities. For example, let $x_n := -1/n$ and $y_n := 1/n$. Then $x_n < y_n$, $x_n < 0$, and $y_n > 0$ for all n . However, these inequalities are not preserved by the limit operation as we have $\lim x_n = \lim y_n = 0$. The moral of this example is that strict inequalities may become non-strict inequalities when limits are applied; if we know that $x_n < y_n$ for all n , we can only conclude that

$$\lim_{n \rightarrow \infty} x_n \leq \lim_{n \rightarrow \infty} y_n.$$

This issue is a common source of errors.

2.2.2 Continuity of algebraic operations

Limits interact nicely with algebraic operations.

Proposition 2.2.5. Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences.

(i) The sequence $\{z_n\}$, where $z_n := x_n + y_n$, converges and

$$\lim_{n \rightarrow \infty} (x_n + y_n) = \lim_{n \rightarrow \infty} z_n = \lim_{n \rightarrow \infty} x_n + \lim_{n \rightarrow \infty} y_n.$$

(ii) The sequence $\{z_n\}$, where $z_n := x_n - y_n$, converges and

$$\lim_{n \rightarrow \infty} (x_n - y_n) = \lim_{n \rightarrow \infty} z_n = \lim_{n \rightarrow \infty} x_n - \lim_{n \rightarrow \infty} y_n.$$

(iii) The sequence $\{z_n\}$, where $z_n := x_n y_n$, converges and

$$\lim_{n \rightarrow \infty} (x_n y_n) = \lim_{n \rightarrow \infty} z_n = \left(\lim_{n \rightarrow \infty} x_n \right) \left(\lim_{n \rightarrow \infty} y_n \right).$$

(iv) If $\lim y_n \neq 0$ and $y_n \neq 0$ for all $n \in \mathbb{N}$, then the sequence $\{z_n\}$, where $z_n := \frac{x_n}{y_n}$, converges and

$$\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \lim_{n \rightarrow \infty} z_n = \frac{\lim_{n \rightarrow \infty} x_n}{\lim_{n \rightarrow \infty} y_n}.$$

Proof. Let us start with (i). Suppose that $\{x_n\}$ and $\{y_n\}$ are convergent sequences and write $z_n := x_n + y_n$. Let $x := \lim x_n$, $y := \lim y_n$, and $z := x + y$.

Let $\varepsilon > 0$ be given. Find an M_1 such that for all $n \geq M_1$ we have $|x_n - x| < \varepsilon/2$. Find an M_2 such that for all $n \geq M_2$ we have $|y_n - y| < \varepsilon/2$. Take $M := \max\{M_1, M_2\}$. For all $n \geq M$ we have

$$\begin{aligned} |z_n - z| &= |(x_n + y_n) - (x + y)| = |x_n - x + y_n - y| \\ &\leq |x_n - x| + |y_n - y| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore (i) is proved. Proof of (ii) is almost identical and is left as an exercise.

Let us tackle (iii). Suppose again that $\{x_n\}$ and $\{y_n\}$ are convergent sequences and write $z_n := x_n y_n$. Let $x := \lim x_n$, $y := \lim y_n$, and $z := xy$.

Let $\varepsilon > 0$ be given. As $\{x_n\}$ is convergent, it is bounded. Therefore, find a $B > 0$ such that $|x_n| \leq B$ for all $n \in \mathbb{N}$. Find an M_1 such that for all $n \geq M_1$ we have $|x_n - x| < \frac{\varepsilon}{2(|y|+1)}$. Find an M_2 such that for all $n \geq M_2$ we have $|y_n - y| < \frac{\varepsilon}{2B}$. Take $M := \max\{M_1, M_2\}$. For all $n \geq M$ we have

$$\begin{aligned} |z_n - z| &= |(x_n y_n) - (xy)| \\ &= |x_n y_n - (x + x_n - x_n)y| \\ &= |x_n(y_n - y) + (x_n - x)y| \\ &\leq |x_n(y_n - y)| + |(x_n - x)y| \\ &= |x_n| |y_n - y| + |x_n - x| |y| \\ &\leq B |y_n - y| + |x_n - x| |y| \\ &< B \frac{\varepsilon}{2B} + \frac{\varepsilon}{2(|y|+1)} |y| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Finally let us tackle (iv). Instead of proving (iv) directly, we prove the following simpler claim:

Claim: If $\{y_n\}$ is a convergent sequence such that $\lim y_n \neq 0$ and $y_n \neq 0$ for all $n \in \mathbb{N}$, then

$$\lim_{n \rightarrow \infty} \frac{1}{y_n} = \frac{1}{\lim y_n}.$$

Once the claim is proved, we take the sequence $\{1/y_n\}$, multiply it by the sequence $\{x_n\}$ and apply item (iii).

Proof of claim: Let $\varepsilon > 0$ be given. Let $y := \lim y_n$. Find an M such that for all $n \geq M$ we have

$$|y_n - y| < \min \left\{ |y|^2 \frac{\varepsilon}{2}, \frac{|y|}{2} \right\}.$$

Note that

$$|y| = |y - y_n + y_n| \leq |y - y_n| + |y_n|,$$

or in other words $|y_n| \geq |y| - |y - y_n|$. Now $|y_n - y| < \frac{|y|}{2}$ implies that

$$|y| - |y_n - y| > \frac{|y|}{2}.$$

Therefore

$$|y_n| \geq |y| - |y - y_n| > \frac{|y|}{2},$$

and consequently

$$\frac{1}{|y_n|} < \frac{2}{|y|}.$$

Now we can finish the proof of the claim.

$$\begin{aligned} \left| \frac{1}{y_n} - \frac{1}{y} \right| &= \left| \frac{y - y_n}{yy_n} \right| \\ &= \frac{|y - y_n|}{|y| |y_n|} \\ &< \frac{|y - y_n|}{|y|} \frac{2}{|y|} \\ &< \frac{|y|^2 \frac{\varepsilon}{2}}{|y|} \frac{2}{|y|} = \varepsilon. \end{aligned}$$

And we are done. □

By plugging in constant sequences, we get several easy corollaries. If $c \in \mathbb{R}$ and $\{x_n\}$ is a convergent sequence, then for example

$$\lim_{n \rightarrow \infty} cx_n = c \left(\lim_{n \rightarrow \infty} x_n \right) \quad \text{and} \quad \lim_{n \rightarrow \infty} (c + x_n) = c + \lim_{n \rightarrow \infty} x_n.$$

Similarly with constant subtraction and division.

As we can take limits past multiplication we can show (exercise) that $\lim x_n^k = (\lim x_n)^k$ for all $k \in \mathbb{N}$. That is, we can take limits past powers. Let us see if we can do the same with roots.

Proposition 2.2.6. *Let $\{x_n\}$ be a convergent sequence such that $x_n \geq 0$. Then*

$$\lim_{n \rightarrow \infty} \sqrt{x_n} = \sqrt{\lim_{n \rightarrow \infty} x_n}.$$

Of course to even make this statement, we need to apply Corollary 2.2.4 to show that $\lim x_n \geq 0$, so that we can take the square root without worry.

Proof. Let $\{x_n\}$ be a convergent sequence and let $x := \lim x_n$.

First suppose that $x = 0$. Let $\varepsilon > 0$ be given. Then there is an M such that for all $n \geq M$ we have $x_n = |x_n| < \varepsilon^2$, or in other words $\sqrt{x_n} < \varepsilon$. Hence

$$|\sqrt{x_n} - \sqrt{x}| = \sqrt{x_n} < \varepsilon.$$

Now suppose that $x > 0$ (and hence $\sqrt{x} > 0$).

$$\begin{aligned} |\sqrt{x_n} - \sqrt{x}| &= \left| \frac{x_n - x}{\sqrt{x_n} + \sqrt{x}} \right| \\ &= \frac{1}{\sqrt{x_n} + \sqrt{x}} |x_n - x| \\ &\leq \frac{1}{\sqrt{x}} |x_n - x|. \end{aligned}$$

We leave the rest of the proof to the reader. □

A similar proof works for the k th root. That is, we also obtain $\lim x_n^{1/k} = (\lim x_n)^{1/k}$. We leave this to the reader as a challenging exercise.

We may also want to take the limit past the absolute value sign.

Proposition 2.2.7. *If $\{x_n\}$ is a convergent sequence, then $\{|x_n|\}$ is convergent and*

$$\lim_{n \rightarrow \infty} |x_n| = \left| \lim_{n \rightarrow \infty} x_n \right|.$$

Proof. We simply note the reverse triangle inequality

$$||x_n| - |x|| \leq |x_n - x|.$$

Hence if $|x_n - x|$ can be made arbitrarily small, so can $||x_n| - |x||$. Details are left to the reader. □

2.2.3 Recursively defined sequences

Now that we know we can interchange limits and algebraic operations, we can compute the limits of many sequences. One such class are recursively defined sequences, that is, sequences where the next number in the sequence computed using a formula from a fixed number of preceding elements in the sequence.

Example 2.2.8: Let $\{x_n\}$ be defined by $x_1 := 2$ and

$$x_{n+1} := x_n - \frac{x_n^2 - 2}{2x_n}.$$

We must first find out if this sequence is well defined; we must show we never divide by zero. Then we must find out if the sequence converges. Only then can we attempt to find the limit.

First let us prove that $x_n > 0$ for all n (and the sequence is well defined). Let us show this by induction. We know that $x_1 = 2 > 0$. For the induction step, suppose that $x_n > 0$. Then

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{2x_n^2 - x_n^2 + 2}{2x_n} = \frac{x_n^2 + 2}{2x_n}.$$

If $x_n > 0$, then $x_n^2 + 2 > 0$ and hence $x_{n+1} > 0$.

Next let us show that the sequence is monotone decreasing. If we show that $x_n^2 - 2 \geq 0$ for all n , then $x_{n+1} \leq x_n$ for all n . Obviously $x_1^2 - 2 = 4 - 2 = 2 > 0$. For an arbitrary n we have that

$$x_{n+1}^2 - 2 = \left(\frac{x_n^2 + 2}{2x_n} \right)^2 - 2 = \frac{x_n^4 + 4x_n^2 + 4 - 8x_n^2}{4x_n^2} = \frac{x_n^4 - 4x_n^2 + 4}{4x_n^2} = \frac{(x_n^2 - 2)^2}{4x_n^2}.$$

Since $x_n > 0$ and any number squared is nonnegative, we have that $x_{n+1}^2 - 2 \geq 0$ for all n . Therefore, $\{x_n\}$ is monotone decreasing and bounded, and the limit exists. It remains to find the limit.

Let us write

$$2x_n x_{n+1} = x_n^2 + 2.$$

Since $\{x_{n+1}\}$ is the 1-tail of $\{x_n\}$, it converges to the same limit. Let us define $x := \lim x_n$. We take the limit of both sides to obtain

$$2x^2 = x^2 + 2,$$

or $x^2 = 2$. As $x \geq 0$, we know that $x = \sqrt{2}$.

You should, however, be careful. Before taking any limits, you must make sure the sequence converges. Let us see an example.

Example 2.2.9: Suppose $x_1 := 1$ and $x_{n+1} := x_n^2 + x_n$. If we blindly assumed that the limit exists (call it x), then we would get the equation $x = x^2 + x$, from which we might conclude that $x = 0$. However, it is not hard to show that $\{x_n\}$ is unbounded and therefore does not converge.

The thing to notice in this example is that the method still works, but it depends on the initial value x_1 . If we set $x_1 := 0$, then the sequence converges and the limit really is 0. An entire branch of mathematics, called dynamics, deals precisely with these issues.

2.2.4 Some convergence tests

It is not always necessary to go back to the definition of convergence to prove that a sequence is convergent. We first give a simple convergence test. The main idea is that $\{x_n\}$ converges to x if and only if $\{|x_n - x|\}$ converges to zero.

Proposition 2.2.10. Let $\{x_n\}$ be a sequence. Suppose that there is an $x \in \mathbb{R}$ and a convergent sequence $\{a_n\}$ such that

$$\lim_{n \rightarrow \infty} a_n = 0$$

and

$$|x_n - x| \leq a_n$$

for all n . Then $\{x_n\}$ converges and $\lim x_n = x$.

Proof. Let $\varepsilon > 0$ be given. Note that $a_n \geq 0$ for all n . Find an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $a_n = |a_n - 0| < \varepsilon$. Then, for all $n \geq M$ we have

$$|x_n - x| \leq a_n < \varepsilon.$$

□

As the proposition shows, to study when a sequence has a limit is the same as studying when another sequence goes to zero. For some special sequences we can test the convergence easily. First let us compute the limit of a very specific sequence.

Proposition 2.2.11. Let $c > 0$.

(i) If $c < 1$, then

$$\lim_{n \rightarrow \infty} c^n = 0.$$

(ii) If $c > 1$, then $\{c^n\}$ is unbounded.

Proof. First let us suppose that $c > 1$. We write $c = 1 + r$ for some $r > 0$. By induction (or using the binomial theorem if you know it) we see that

$$c^n = (1 + r)^n \geq 1 + nr.$$

By the Archimedean property of the real numbers, the sequence $\{1 + nr\}$ is unbounded (for any number B , we can find an n such that $nr \geq B - 1$). Therefore c^n is unbounded.

Now let $c < 1$. Write $c = \frac{1}{1+r}$, where $r > 0$. Then

$$c^n = \frac{1}{(1+r)^n} \leq \frac{1}{1+nr} \leq \frac{1}{r} \frac{1}{n}.$$

As $\{\frac{1}{n}\}$ converges to zero, so does $\{\frac{1}{r} \frac{1}{n}\}$. Hence, $\{c^n\}$ converges to zero. □

If we look at the above proposition, we note that the ratio of the $(n+1)$ th term and the n th term is c . We can generalize this simple result to a larger class of sequences. The following lemma will come up again once we get to series.

Lemma 2.2.12 (Ratio test for sequences). *Let $\{x_n\}$ be a sequence such that $x_n \neq 0$ for all n and such that the limit*

$$L := \lim_{n \rightarrow \infty} \frac{|x_{n+1}|}{|x_n|}$$

exists.

(i) *If $L < 1$, then $\{x_n\}$ converges and $\lim x_n = 0$.*

(ii) *If $L > 1$, then $\{x_n\}$ is unbounded (hence diverges).*

If L exists, but $L = 1$, the lemma says nothing. We cannot make any conclusion based on that information alone. For example, consider the sequences $1, 1, 1, 1, \dots$ and $1, -1, 1, -1, 1, \dots$

Proof. Suppose $L < 1$. As $\frac{|x_{n+1}|}{|x_n|} \geq 0$, we have that $L \geq 0$. Pick r such that $L < r < 1$. As $r - L > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ we have

$$\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < r - L.$$

Therefore,

$$\frac{|x_{n+1}|}{|x_n|} < r.$$

For $n > M$ (that is for $n \geq M + 1$) we write

$$|x_n| = |x_M| \frac{|x_{M+1}|}{|x_M|} \frac{|x_{M+2}|}{|x_{M+1}|} \dots \frac{|x_n|}{|x_{n-1}|} < |x_M| r r \dots r = |x_M| r^{n-M} = (|x_M| r^{-M}) r^n.$$

The sequence $\{r^n\}$ converges to zero and hence $|x_M| r^{-M} r^n$ converges to zero. By Proposition 2.2.10, the M -tail of $\{x_n\}$ converges to zero and therefore $\{x_n\}$ converges to zero.

Now suppose $L > 1$. Pick r such that $1 < r < L$. As $L - r > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ we have

$$\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < L - r.$$

Therefore,

$$\frac{|x_{n+1}|}{|x_n|} > r.$$

Again for $n > M$ we write

$$|x_n| = |x_M| \frac{|x_{M+1}|}{|x_M|} \frac{|x_{M+2}|}{|x_{M+1}|} \dots \frac{|x_n|}{|x_{n-1}|} > |x_M| r r \dots r = |x_M| r^{n-M} = (|x_M| r^{-M}) r^n.$$

The sequence $\{r^n\}$ is unbounded (since $r > 1$), and therefore $\{x_n\}$ cannot be bounded (if $|x_n| \leq B$ for all n , then $r^n < \frac{B}{|x_M|} r^M$ for all n , which is impossible). Consequently, $\{x_n\}$ cannot converge. \square

Example 2.2.13: A simple application of the above lemma is to prove that

$$\lim_{n \rightarrow \infty} \frac{2^n}{n!} = 0.$$

Proof: We find that

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

It is not hard to see that $\{\frac{2}{n+1}\}$ converges to zero. The conclusion follows by the lemma.

2.2.5 Exercises

Exercise 2.2.1: Prove Corollary 2.2.4. Hint: Use constant sequences and Lemma 2.2.3.

Exercise 2.2.2: Prove part (ii) of Proposition 2.2.5.

Exercise 2.2.3: Prove that if $\{x_n\}$ is a convergent sequence, $k \in \mathbb{N}$, then

$$\lim_{n \rightarrow \infty} x_n^k = \left(\lim_{n \rightarrow \infty} x_n \right)^k.$$

Hint: Use induction.

Exercise 2.2.4: Suppose that $x_1 := \frac{1}{2}$ and $x_{n+1} := x_n^2$. Show that $\{x_n\}$ converges and find $\lim x_n$. Hint: You cannot divide by zero!

Exercise 2.2.5: Let $x_n := \frac{n - \cos(n)}{n}$. Use the squeeze lemma to show that $\{x_n\}$ converges and find the limit.

Exercise 2.2.6: Let $x_n := \frac{1}{n^2}$ and $y_n := \frac{1}{n}$. Define $z_n := \frac{x_n}{y_n}$ and $w_n := \frac{y_n}{x_n}$. Do $\{z_n\}$ and $\{w_n\}$ converge? What are the limits? Can you apply Proposition 2.2.5? Why or why not?

Exercise 2.2.7: True or false, prove or find a counterexample. If $\{x_n\}$ is a sequence such that $\{x_n^2\}$ converges, then $\{x_n\}$ converges.

Exercise 2.2.8: Show that

$$\lim_{n \rightarrow \infty} \frac{n^2}{2^n} = 0.$$

Exercise 2.2.9: Suppose that $\{x_n\}$ is a sequence and suppose that for some $x \in \mathbb{R}$, the limit

$$L := \lim_{n \rightarrow \infty} \frac{|x_{n+1} - x|}{|x_n - x|}$$

exists and $L < 1$. Show that $\{x_n\}$ converges to x .

Exercise 2.2.10 (Challenging): Let $\{x_n\}$ be a convergent sequence such that $x_n \geq 0$ and $k \in \mathbb{N}$. Then

$$\lim_{n \rightarrow \infty} x_n^{1/k} = \left(\lim_{n \rightarrow \infty} x_n \right)^{1/k}.$$

Hint: Find an expression q such that $\frac{x_n^{1/k} - x^{1/k}}{x_n - x} = \frac{1}{q}$.

2.3 Limit superior, limit inferior, and Bolzano-Weierstrass

Note: 1–2 lectures, alternative proof of BW optional

In this section we study bounded sequences and their subsequences. In particular we define the so-called limit superior and limit inferior of a bounded sequence and talk about limits of subsequences. Furthermore, we prove the so-called Bolzano-Weierstrass theorem*, which is an indispensable tool in analysis.

We have seen that every convergent sequence is bounded, but there exist many bounded divergent sequences. For example, the sequence $\{(-1)^n\}$ is bounded, but we have seen it is divergent. All is not lost however and we can still compute certain limits with a bounded divergent sequence.

2.3.1 Upper and lower limits

There are ways of creating monotone sequences out of any sequence, and in this way we get the so-called *limit superior* and *limit inferior*. These limits always exist for bounded sequences.

If a sequence $\{x_n\}$ is bounded, then the set $\{x_k : k \in \mathbb{N}\}$ is bounded. Then for every n the set $\{x_k : k \geq n\}$ is also bounded (as it is a subset).

Definition 2.3.1. Let $\{x_n\}$ be a bounded sequence. Let $a_n := \sup\{x_k : k \geq n\}$ and $b_n := \inf\{x_k : k \geq n\}$. We note that the sequence $\{a_n\}$ is bounded monotone decreasing and $\{b_n\}$ is bounded monotone increasing (more on this point below). We define

$$\begin{aligned}\limsup_{n \rightarrow \infty} x_n &:= \lim_{n \rightarrow \infty} a_n, \\ \liminf_{n \rightarrow \infty} x_n &:= \lim_{n \rightarrow \infty} b_n.\end{aligned}$$

For a bounded sequence, \liminf and \limsup always exist. It is possible to define \liminf and \limsup for unbounded sequences if we allow ∞ and $-\infty$. It is not hard to generalize the following results to include unbounded sequences, however, we restrict our attention to bounded ones.

Let us see why $\{a_n\}$ is a decreasing sequence. As a_n is the least upper bound for $\{x_k : k \geq n\}$, it is also an upper bound for the subset $\{x_k : k \geq (n+1)\}$. Therefore a_{n+1} , the least upper bound for $\{x_k : k \geq (n+1)\}$, has to be less than or equal to a_n , that is, $a_n \geq a_{n+1}$. Similarly, b_n is an increasing sequence. It is left as an exercise to show that if x_n is bounded, then a_n and b_n must be bounded.

Proposition 2.3.2. Let $\{x_n\}$ be a bounded sequence. Define a_n and b_n as in the definition above.

- (i) $\limsup_{n \rightarrow \infty} x_n = \inf\{a_n : n \in \mathbb{N}\}$ and $\liminf_{n \rightarrow \infty} x_n = \sup\{b_n : n \in \mathbb{N}\}$.
- (ii) $\liminf_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} x_n$.

*Named after the Czech mathematician Bernhard Placidus Johann Nepomuk Bolzano (1781 – 1848), and the German mathematician Karl Theodor Wilhelm Weierstrass (1815 – 1897).

Proof. The first item in the proposition follows as the sequences $\{a_n\}$ and $\{b_n\}$ are monotone.

For the second item, we note that $b_n \leq a_n$, as the inf of a set is less than or equal to its sup. We know that $\{a_n\}$ and $\{b_n\}$ converge to the limsup and the liminf (respectively). We apply Lemma 2.2.3 to obtain

$$\lim_{n \rightarrow \infty} b_n \leq \lim_{n \rightarrow \infty} a_n.$$

□

Example 2.3.3: Let $\{x_n\}$ be defined by

$$x_n := \begin{cases} \frac{n+1}{n} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

Let us compute the lim inf and lim sup of this sequence. First the limit inferior:

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} (\inf\{x_k : k \geq n\}) = \lim_{n \rightarrow \infty} 0 = 0.$$

For the limit superior we write

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} (\sup\{x_k : k \geq n\}).$$

It is not hard to see that

$$\sup\{x_k : k \geq n\} = \begin{cases} \frac{n+1}{n} & \text{if } n \text{ is odd,} \\ \frac{n+2}{n+1} & \text{if } n \text{ is even.} \end{cases}$$

We leave it to the reader to show that the limit is 1. That is,

$$\limsup_{n \rightarrow \infty} x_n = 1.$$

Do note that the sequence $\{x_n\}$ is not a convergent sequence.

We can associate with lim sup and lim inf certain subsequences.

Theorem 2.3.4. *If $\{x_n\}$ is a bounded sequence, then there exists a subsequence $\{x_{n_k}\}$ such that*

$$\lim_{k \rightarrow \infty} x_{n_k} = \limsup_{n \rightarrow \infty} x_n.$$

Similarly, there exists a (perhaps different) subsequence $\{x_{m_k}\}$ such that

$$\lim_{k \rightarrow \infty} x_{m_k} = \liminf_{n \rightarrow \infty} x_n.$$

Proof. Define $a_n := \sup\{x_k : k \geq n\}$. Write $x := \limsup x_n = \lim a_n$. Define the subsequence as follows. Pick $n_1 := 1$ and work inductively. Suppose we have defined the subsequence until n_k for some k . Now pick some $m > n_k$ such that

$$a_{(n_k+1)} - x_m < \frac{1}{k+1}.$$

We can do this as $a_{(n_k+1)}$ is a supremum of the set $\{x_n : n \geq n_k + 1\}$ and hence there are elements of the sequence arbitrarily close (or even possibly equal) to the supremum. Set $n_{k+1} := m$. The subsequence $\{x_{n_k}\}$ is defined. Next we need to prove that it converges and has the right limit.

Note that $a_{(n_{k-1}+1)} \geq a_{n_k}$ (why?) and that $a_{n_k} \geq x_{n_k}$. Therefore, for every $k > 1$ we have

$$\begin{aligned} |a_{n_k} - x_{n_k}| &= a_{n_k} - x_{n_k} \\ &\leq a_{(n_{k-1}+1)} - x_{n_k} \\ &< \frac{1}{k}. \end{aligned}$$

Let us show that $\{x_{n_k}\}$ converges x . Note that the subsequence need not be monotone. Let $\varepsilon > 0$ be given. As $\{a_n\}$ converges to x , then the subsequence $\{a_{n_k}\}$ converges to x . Thus there exists an $M_1 \in \mathbb{N}$ such that for all $k \geq M_1$ we have

$$|a_{n_k} - x| < \frac{\varepsilon}{2}.$$

Find an $M_2 \in \mathbb{N}$ such that

$$\frac{1}{M_2} \leq \frac{\varepsilon}{2}.$$

Take $M := \max\{M_1, M_2\}$ and compute. For all $k \geq M$ we have

$$\begin{aligned} |x - x_{n_k}| &= |a_{n_k} - x_{n_k} + x - a_{n_k}| \\ &\leq |a_{n_k} - x_{n_k}| + |x - a_{n_k}| \\ &< \frac{1}{k} + \frac{\varepsilon}{2} \\ &\leq \frac{1}{M_2} + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

We leave the statement for \liminf as an exercise. □

2.3.2 Using limit inferior and limit superior

The advantage of \liminf and \limsup is that we can always write them down for any (bounded) sequence. Working with \liminf and \limsup is a little bit like working with limits, although there are subtle differences. If we could somehow compute them, we could also compute the limit of the sequence if it exists, or show that the sequence diverges.

Theorem 2.3.5. Let $\{x_n\}$ be a bounded sequence. Then $\{x_n\}$ converges if and only if

$$\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n.$$

Furthermore, if $\{x_n\}$ converges, then

$$\lim_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n.$$

Proof. Define a_n and b_n as in Definition 2.3.1. Note that

$$b_n \leq x_n \leq a_n.$$

If $\liminf x_n = \limsup x_n$, then we know that $\{a_n\}$ and $\{b_n\}$ have limits and that these two limits are the same. By the squeeze lemma (Lemma 2.2.1), $\{x_n\}$ converges and

$$\lim_{n \rightarrow \infty} b_n = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} a_n.$$

Now suppose that $\{x_n\}$ converges to x . We know by Theorem 2.3.4 that there exists a subsequence $\{x_{n_k}\}$ that converges to $\limsup x_n$. As $\{x_n\}$ converges to x , we know that every subsequence converges to x and therefore $\limsup x_n = \lim x_{n_k} = x$. Similarly $\liminf x_n = x$. \square

Limit superior and limit inferior behave nicely with subsequences.

Proposition 2.3.6. Suppose that $\{x_n\}$ is a bounded sequence and $\{x_{n_k}\}$ is a subsequence. Then

$$\liminf_{n \rightarrow \infty} x_n \leq \liminf_{k \rightarrow \infty} x_{n_k} \leq \limsup_{k \rightarrow \infty} x_{n_k} \leq \limsup_{n \rightarrow \infty} x_n.$$

Proof. The middle inequality has been proved already. We will prove the third inequality, and leave the first inequality as an exercise.

That is, we want to prove that $\limsup x_{n_k} \leq \limsup x_n$. Define $a_j := \sup\{x_k : k \geq j\}$ as usual. Also define $c_j := \sup\{x_{n_k} : k \geq j\}$. It is not true that c_j is necessarily a subsequence of a_j . However, as $n_k \geq k$ for all k , we have that $\{x_{n_k} : k \geq j\} \subset \{x_k : k \geq j\}$. A supremum of a subset is less than or equal to the supremum of the set and therefore

$$c_j \leq a_j.$$

We apply Lemma 2.2.3 to conclude that

$$\lim_{j \rightarrow \infty} c_j \leq \lim_{j \rightarrow \infty} a_j,$$

which is the desired conclusion. \square

Limit superior and limit inferior are the largest and smallest subsequential limits. If the subsequence in the previous proposition is convergent, then of course we have that $\liminf x_{n_k} = \lim x_{n_k} = \limsup x_{n_k}$. Therefore,

$$\liminf_{n \rightarrow \infty} x_n \leq \lim_{k \rightarrow \infty} x_{n_k} \leq \limsup_{n \rightarrow \infty} x_n.$$

Similarly we get the following useful test for convergence of a bounded sequence. We leave the proof as an exercise.

Theorem 2.3.7. *A bounded sequence $\{x_n\}$ is convergent and converges to x if and only if every convergent subsequence $\{x_{n_k}\}$ converges to x .*

2.3.3 Bolzano-Weierstrass theorem

While it is not true that a bounded sequence is convergent, the Bolzano-Weierstrass theorem tells us that we can at least find a convergent subsequence. The version of Bolzano-Weierstrass that we present in this section is the Bolzano-Weierstrass for sequences.

Theorem 2.3.8 (Bolzano-Weierstrass). *Suppose that a sequence $\{x_n\}$ of real numbers is bounded. Then there exists a convergent subsequence $\{x_{n_i}\}$.*

Proof. We use Theorem 2.3.4. It says that there exists a subsequence whose limit is $\limsup x_n$. \square

The reader might complain right now that Theorem 2.3.4 is strictly stronger than the Bolzano-Weierstrass theorem as presented above. That is true. However, Theorem 2.3.4 only applies to the real line, but Bolzano-Weierstrass applies in more general contexts (that is, in \mathbb{R}^n) with pretty much the exact same statement.

As the theorem is so important to analysis, we present an explicit proof. The following proof generalizes more easily to different contexts.

Alternate proof of Bolzano-Weierstrass. As the sequence is bounded, then there exist two numbers $a_1 < b_1$ such that $a_1 \leq x_n \leq b_1$ for all $n \in \mathbb{N}$.

We will define a subsequence $\{x_{n_i}\}$ and two sequences $\{a_i\}$ and $\{b_i\}$, such that $\{a_i\}$ is monotone increasing, $\{b_i\}$ is monotone decreasing, $a_i \leq x_{n_i} \leq b_i$ and such that $\lim a_i = \lim b_i$. That x_{n_i} converges follows by the squeeze lemma.

We define the sequence inductively. We will always assume that $a_i < b_i$. Further we will always have that $x_n \in [a_i, b_i]$ for infinitely many $n \in \mathbb{N}$. We have already defined a_1 and b_1 . We take $n_1 := 1$, that is $x_{n_1} = x_1$.

Now suppose we have defined the subsequence $x_{n_1}, x_{n_2}, \dots, x_{n_k}$, and the sequences $\{a_i\}$ and $\{b_i\}$ up to some $k \in \mathbb{N}$. We find $y = \frac{a_k + b_k}{2}$. It is clear that $a_k < y < b_k$. If there exist infinitely many $j \in \mathbb{N}$ such that $x_j \in [a_k, y]$, then set $a_{k+1} := a_k$, $b_{k+1} := y$, and pick $n_{k+1} > n_k$ such that $x_{n_{k+1}} \in [a_k, y]$. If there are not infinitely many j such that $x_j \in [a_k, y]$, then it must be true that there are infinitely

many $j \in \mathbb{N}$ such that $x_j \in [y, b_k]$. In this case pick $a_{k+1} := y$, $b_{k+1} := b_k$, and pick $n_{k+1} > n_k$ such that $x_{n_{k+1}} \in [y, b_k]$.

Now we have the sequences defined. What is left to prove is that $\lim a_i = \lim b_i$. Obviously the limits exist as the sequences are monotone. From the construction, it is obvious that $b_i - a_i$ is cut in half in each step. Therefore $b_{i+1} - a_{i+1} = \frac{b_i - a_i}{2}$. By induction, we obtain that

$$b_i - a_i = \frac{b_1 - a_1}{2^{i-1}}.$$

Let $x := \lim a_i$. As $\{a_i\}$ is monotone we have that

$$x = \sup\{a_i : i \in \mathbb{N}\}$$

Now let $y := \lim b_i = \inf\{b_i : i \in \mathbb{N}\}$. Obviously $y \leq x$ as $a_i < b_i$ for all i . As the sequences are monotone, then for any i we have (why?)

$$y - x \leq b_i - a_i = \frac{b_1 - a_1}{2^{i-1}}.$$

As $\frac{b_1 - a_1}{2^{i-1}}$ is arbitrarily small and $y - x \geq 0$, we have that $y - x = 0$. We finish by the squeeze lemma. \square

Yet another proof of the Bolzano-Weierstrass theorem proves the following claim, which is left as a challenging exercise. *Claim: Every sequence has a monotone subsequence.*

2.3.4 Exercises

Exercise 2.3.1: Suppose that $\{x_n\}$ is a bounded sequence. Define a_n and b_n as in Definition 2.3.1. Show that $\{a_n\}$ and $\{b_n\}$ are bounded.

Exercise 2.3.2: Suppose that $\{x_n\}$ is a bounded sequence. Define b_n as in Definition 2.3.1. Show that $\{b_n\}$ is an increasing sequence.

Exercise 2.3.3: Finish the proof of Proposition 2.3.6. That is, suppose that $\{x_n\}$ is a bounded sequence and $\{x_{n_k}\}$ is a subsequence. Prove $\liminf_{n \rightarrow \infty} x_n \leq \liminf_{k \rightarrow \infty} x_{n_k}$.

Exercise 2.3.4: Prove Theorem 2.3.7.

Exercise 2.3.5: a) Let $x_n := \frac{(-1)^n}{n}$, find $\limsup x_n$ and $\liminf x_n$.

b) Let $x_n := \frac{(n-1)(-1)^n}{n}$, find $\limsup x_n$ and $\liminf x_n$.

Exercise 2.3.6: Let $\{x_n\}$ and $\{y_n\}$ be bounded sequences such that $x_n \leq y_n$ for all n . Then show that

$$\limsup_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} y_n$$

and

$$\liminf_{n \rightarrow \infty} x_n \leq \liminf_{n \rightarrow \infty} y_n.$$

Exercise 2.3.7: Let $\{x_n\}$ and $\{y_n\}$ be bounded sequences.

a) Show that $\{x_n + y_n\}$ is bounded.

b) Show that

$$(\liminf_{n \rightarrow \infty} x_n) + (\liminf_{n \rightarrow \infty} y_n) \leq \liminf_{n \rightarrow \infty} (x_n + y_n).$$

Hint: Find a subsequence $\{x_{n_i} + y_{n_i}\}$ of $\{x_n + y_n\}$ that converges. Then find a subsequence $\{x_{n_{m_i}}\}$ of $\{x_{n_i}\}$ that converges. Then apply what you know about limits.

c) Find an explicit $\{x_n\}$ and $\{y_n\}$ such that

$$(\liminf_{n \rightarrow \infty} x_n) + (\liminf_{n \rightarrow \infty} y_n) < \liminf_{n \rightarrow \infty} (x_n + y_n).$$

Hint: Look for examples that do not have a limit.

Exercise 2.3.8: Let $\{x_n\}$ and $\{y_n\}$ be bounded sequences (from the previous exercise we know that $\{x_n + y_n\}$ is bounded).

a) Show that

$$(\limsup_{n \rightarrow \infty} x_n) + (\limsup_{n \rightarrow \infty} y_n) \geq \limsup_{n \rightarrow \infty} (x_n + y_n).$$

Hint: See previous exercise.

b) Find an explicit $\{x_n\}$ and $\{y_n\}$ such that

$$(\limsup_{n \rightarrow \infty} x_n) + (\limsup_{n \rightarrow \infty} y_n) > \limsup_{n \rightarrow \infty} (x_n + y_n).$$

Hint: See previous exercise.

Exercise 2.3.9: If $S \subset \mathbb{R}$ is a set, then $x \in \mathbb{R}$ is a cluster point if for every $\varepsilon > 0$, the set $(x - \varepsilon, x + \varepsilon) \cap S \setminus \{x\}$ is not empty. That is, if there are points of S arbitrarily close to x . For example, $S := \{1/n : n \in \mathbb{N}\}$ has a unique (only one) cluster point 0, but $0 \notin S$. Prove the following version of the Bolzano-Weierstrass theorem:

Theorem. Let $S \subset \mathbb{R}$ be a bounded infinite set, then there exists at least one cluster point of S .

Hint: If S is infinite, then S contains a countably infinite subset. That is, there is a sequence $\{x_n\}$ of distinct numbers in S .

Exercise 2.3.10 (Challenging): a) Prove that any sequence contains a monotone subsequence. *Hint:* Call $n \in \mathbb{N}$ a peak if $a_m \leq a_n$ for all $m \geq n$. Now there are two possibilities: either the sequence has at most finitely many peaks, or it has infinitely many peaks.

b) Now conclude the Bolzano-Weierstrass theorem.

Exercise 2.3.11: Let us prove a stronger version of Theorem 2.3.7. Suppose that $\{x_n\}$ is a sequence such that every subsequence $\{x_{n_i}\}$ has a subsequence $\{x_{n_{m_i}}\}$ that converges to x . a) First show that $\{x_n\}$ is bounded. b) Now show that $\{x_n\}$ converges to x .

Exercise 2.3.12: Let $\{x_n\}$ be a bounded sequence.

a) Prove that there exists an s such that for any $r > s$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $x_n < r$.

b) If s is a number as in a), then prove $\limsup x_n \leq s$.

c) Show that if S is the set of all s as in a), then $\limsup x_n = \inf S$.

2.4 Cauchy sequences

Note: 0.5–1 lecture

Often we wish to describe a certain number by a sequence that converges to it. In this case, it is impossible to use the number itself in the proof that the sequence converges. It would be nice if we could check for convergence without being able to find the limit.

Definition 2.4.1. A sequence $\{x_n\}$ is a *Cauchy sequence** if for every $\varepsilon > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $k \geq M$ we have

$$|x_n - x_k| < \varepsilon.$$

Intuitively what it means is that the terms of the sequence are eventually arbitrarily close to each other. We would expect such a sequence to be convergent. It turns out that is true because \mathbb{R} has the least-upper-bound property. First, let us look at some examples.

Example 2.4.2: The sequence $\{1/n\}$ is a Cauchy sequence.

Proof: Given $\varepsilon > 0$, find M such that $M > 2/\varepsilon$. Then for $n, k \geq M$ we have that $1/n < \varepsilon/2$ and $1/k < \varepsilon/2$. Therefore for $n, k \geq M$ we have

$$\left| \frac{1}{n} - \frac{1}{k} \right| \leq \left| \frac{1}{n} \right| + \left| \frac{1}{k} \right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Example 2.4.3: The sequence $\{\frac{n+1}{n}\}$ is a Cauchy sequence.

Proof: Given $\varepsilon > 0$, find M such that $M > 2/\varepsilon$. Then for $n, k \geq M$ we have that $1/n < \varepsilon/2$ and $1/k < \varepsilon/2$. Therefore for $n, k \geq M$ we have

$$\begin{aligned} \left| \frac{n+1}{n} - \frac{k+1}{k} \right| &= \left| \frac{k(n+1) - n(k+1)}{nk} \right| \\ &= \left| \frac{kn + k - nk - n}{nk} \right| \\ &= \left| \frac{k - n}{nk} \right| \\ &\leq \left| \frac{k}{nk} \right| + \left| \frac{-n}{nk} \right| \\ &= \frac{1}{n} + \frac{1}{k} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Proposition 2.4.4. *A Cauchy sequence is bounded.*

* Named after the French mathematician Augustin-Louis Cauchy (1789–1857).

Proof. Suppose that $\{x_n\}$ is Cauchy. Pick M such that for all $n, k \geq M$ we have $|x_n - x_k| < 1$. In particular, we have that for all $n \geq M$

$$|x_n - x_M| < 1.$$

Or by the reverse triangle inequality, $|x_n| - |x_M| \leq |x_n - x_M| < 1$. Hence for $n \geq M$ we have

$$|x_n| < 1 + |x_M|.$$

Let

$$B := \max\{|x_1|, |x_2|, \dots, |x_{M-1}|, 1 + |x_M|\}.$$

Then $|x_n| \leq B$ for all $n \in \mathbb{N}$. □

Theorem 2.4.5. *A sequence of real numbers is Cauchy if and only if it converges.*

Proof. Let $\varepsilon > 0$ be given and suppose that $\{x_n\}$ converges to x . Then there exists an M such that for $n \geq M$ we have

$$|x_n - x| < \frac{\varepsilon}{2}.$$

Hence for $n \geq M$ and $k \geq M$ we have

$$|x_n - x_k| = |x_n - x + x - x_k| \leq |x_n - x| + |x - x_k| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Alright, that direction was easy. Now suppose that $\{x_n\}$ is Cauchy. We have shown that $\{x_n\}$ is bounded. If we show that

$$\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n,$$

then $\{x_n\}$ must be convergent by Theorem 2.3.5. Assuming that \liminf and \limsup exist is where we use the least-upper-bound property.

Define $a := \limsup x_n$ and $b := \liminf x_n$. By Theorem 2.3.7, there exist subsequences $\{x_{n_i}\}$ and $\{x_{m_i}\}$, such that

$$\lim_{i \rightarrow \infty} x_{n_i} = a \quad \text{and} \quad \lim_{i \rightarrow \infty} x_{m_i} = b.$$

Given an $\varepsilon > 0$, there exists an M_1 such that for all $i \geq M_1$ we have $|x_{n_i} - a| < \varepsilon/3$ and an M_2 such that for all $i \geq M_2$ we have $|x_{m_i} - b| < \varepsilon/3$. There also exists an M_3 such that for all $n, k \geq M_3$ we have $|x_n - x_k| < \varepsilon/3$. Let $M := \max\{M_1, M_2, M_3\}$. Note that if $i \geq M$, then $n_i \geq M$ and $m_i \geq M$. Hence

$$\begin{aligned} |a - b| &= |a - x_{n_i} + x_{n_i} - x_{m_i} + x_{m_i} - b| \\ &\leq |a - x_{n_i}| + |x_{n_i} - x_{m_i}| + |x_{m_i} - b| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

As $|a - b| < \varepsilon$ for all $\varepsilon > 0$, then $a = b$ and the sequence converges. □

Remark 2.4.6. The statement of this proposition is sometimes used to define the completeness property of the real numbers. That is, we can say that \mathbb{R} is *Cauchy-complete* (or sometimes just *complete*). We have proved above that as \mathbb{R} has the least-upper-bound property, then \mathbb{R} is Cauchy-complete. We can “complete” \mathbb{Q} by “throwing in” just enough points to make all Cauchy sequences converge (we omit the details). The resulting field has the least-upper-bound property. The advantage of using Cauchy sequences to define completeness is that this idea generalizes to more abstract settings.

2.4.1 Exercises

Exercise 2.4.1: Prove that $\{\frac{n^2-1}{n^2}\}$ is Cauchy using directly the definition of Cauchy sequences.

Exercise 2.4.2: Let $\{x_n\}$ be a sequence such that there exists a $0 < C < 1$ such that

$$|x_{n+1} - x_n| \leq C|x_n - x_{n-1}|.$$

Prove that $\{x_n\}$ is Cauchy. Hint: You can freely use the formula (for $C \neq 1$)

$$1 + C + C^2 + \cdots + C^n = \frac{1 - C^{n+1}}{1 - C}.$$

Exercise 2.4.3 (Challenging): Suppose that F is an ordered field that contains the rational numbers \mathbb{Q} , such that \mathbb{Q} is dense, that is: whenever $x, y \in F$ are such that $x < y$, then there exists a $q \in \mathbb{Q}$ such that $x < q < y$. Say a sequence $\{x_n\}_{n=1}^{\infty}$ of rational numbers is Cauchy if given any $\varepsilon \in \mathbb{Q}$ with $\varepsilon > 0$, there exists an M such that for all $n, k \geq M$ we have $|x_n - x_k| < \varepsilon$. Suppose that any Cauchy sequence of rational numbers has a limit in F . Prove that F has the least-upper-bound property.

Exercise 2.4.4: Let $\{x_n\}$ and $\{y_n\}$ be sequences such that $\lim y_n = 0$. Suppose that for all $k \in \mathbb{N}$ and for all $m \geq k$ we have

$$|x_m - x_k| \leq y_k.$$

Show that $\{x_n\}$ is Cauchy.

Exercise 2.4.5: Suppose that a Cauchy sequence $\{x_n\}$ is such that for every $M \in \mathbb{N}$, there exists a $k \geq M$ and an $n \geq M$ such that $x_k < 0$ and $x_n > 0$. Using simply the definition of a Cauchy sequence and of a convergent sequence, show that the sequence converges to 0.

2.5 Series

Note: 2 lectures

A fundamental object in mathematics is that of a series. In fact, when foundations of analysis were being developed, the motivation was to understand series. Understanding series is very important in applications of analysis. For example, solving differential equations often includes series, and differential equations are the basis for understanding almost all of modern science.

2.5.1 Definition

Definition 2.5.1. Given a sequence $\{x_n\}$, we write the formal object

$$\sum_{n=1}^{\infty} x_n \quad \text{or sometimes just} \quad \sum x_n$$

and call it a *series*. A series *converges*, if the sequence $\{s_n\}$ defined by

$$s_n := \sum_{k=1}^n x_k = x_1 + x_2 + \cdots + x_n,$$

converges. The numbers s_n are called *partial sums*. If $x := \lim s_n$, we write

$$\sum_{n=1}^{\infty} x_n = x.$$

In this case, we cheat a little and treat $\sum_{n=1}^{\infty} x_n$ as a number.

On the other hand, if the sequence $\{s_n\}$ diverges, we say that the series is *divergent*. In this case, $\sum x_n$ is simply a formal object and not a number.

In other words, for a convergent series we have

$$\sum_{n=1}^{\infty} x_n = \lim_{n \rightarrow \infty} \sum_{k=1}^n x_k.$$

We should be careful to only use this equality if the limit on the right actually exists. That is, the right-hand side does not make sense (the limit does not exist) if the series does not converge.

Remark 2.5.2. Before going further, let us remark that it is sometimes convenient to start the series at an index different from 1. That is, for example we can write

$$\sum_{n=0}^{\infty} r^n := \sum_{n=1}^{\infty} r^{n-1}.$$

The left-hand side is more convenient to write. The idea is the same as the notation for the tail of a sequence.

Remark 2.5.3. It is common to write the series $\sum x_n$ as

$$x_1 + x_2 + x_3 + \cdots$$

with the understanding that the ellipsis indicates a series and not a simple sum. We do not use this notation as it often leads to mistakes in proofs.

Example 2.5.4: The series

$$\sum_{n=1}^{\infty} \frac{1}{2^n}$$

converges and the limit is 1. That is,

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{2^k} = 1.$$

Proof: First we prove the following equality

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

The equality is easy to see when $n = 1$. The proof for general n follows by induction, which we leave to the reader. Let s_n be the partial sum. We write

$$|1 - s_n| = \left| 1 - \sum_{k=1}^n \frac{1}{2^k} \right| = \left| \frac{1}{2^n} \right| = \frac{1}{2^n}.$$

The sequence $\{\frac{1}{2^n}\}$ and therefore $\{|1 - s_n|\}$ converges to zero. So, $\{s_n\}$ converges to 1.

For $-1 < r < 1$, the geometric series

$$\sum_{n=0}^{\infty} r^n$$

converges. In fact, $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$. The proof is left as an exercise to the reader. The proof consists of showing that

$$\sum_{k=0}^{n-1} r^k = \frac{1 - r^n}{1 - r},$$

and then taking the limit.

A fact we often use is the following analogue of looking at the tail of a sequence.

Proposition 2.5.5. Let $\sum x_n$ be a series. Let $M \in \mathbb{N}$. Then

$$\sum_{n=1}^{\infty} x_n \text{ converges if and only if } \sum_{n=M}^{\infty} x_n \text{ converges.}$$

Proof. We look at partial sums of the two series (for $k \geq M$)

$$\sum_{n=1}^k x_n = \left(\sum_{n=1}^{M-1} x_n \right) + \sum_{n=M}^k x_n.$$

Note that $\sum_{n=1}^{M-1} x_n$ is a fixed number. Now use Proposition 2.2.5 to finish the proof. \square

2.5.2 Cauchy series

Definition 2.5.6. A series $\sum x_n$ is said to be *Cauchy* or a *Cauchy series*, if the sequence of partial sums $\{s_n\}$ is a Cauchy sequence.

A sequence of real numbers converges if and only if it is Cauchy. Therefore a series is convergent if and only if it is Cauchy.

The series $\sum x_n$ is Cauchy if for every $\varepsilon > 0$, there exists an $M \in \mathbb{N}$, such that for every $n \geq M$ and $k \geq M$ we have

$$\left| \left(\sum_{j=1}^k x_j \right) - \left(\sum_{j=1}^n x_j \right) \right| < \varepsilon.$$

Without loss of generality we assume that $n < k$. Then we write

$$\left| \left(\sum_{j=1}^k x_j \right) - \left(\sum_{j=1}^n x_j \right) \right| = \left| \sum_{j=n+1}^k x_j \right| < \varepsilon.$$

We have proved the following simple proposition.

Proposition 2.5.7. *The series $\sum x_n$ is Cauchy if for every $\varepsilon > 0$, there exists an $M \in \mathbb{N}$ such that for every $n \geq M$ and every $k > n$ we have*

$$\left| \sum_{j=n+1}^k x_j \right| < \varepsilon.$$

2.5.3 Basic properties

Proposition 2.5.8. *Let $\sum x_n$ be a convergent series. Then the sequence $\{x_n\}$ is convergent and*

$$\lim_{n \rightarrow \infty} x_n = 0.$$

Proof. Let $\varepsilon > 0$ be given. As $\sum x_n$ is convergent, it is Cauchy. Thus we find an M such that for every $n \geq M$ we have

$$\varepsilon > \left| \sum_{j=n+1}^{n+1} x_j \right| = |x_{n+1}|.$$

Hence for every $n \geq M + 1$ we have that $|x_n| < \varepsilon$. \square

So if a series converges, the terms of the series go to zero. The implication, however, goes only one way. Let us give an example.

Example 2.5.9: The series $\sum \frac{1}{n}$ diverges (despite the fact that $\lim \frac{1}{n} = 0$). This is the famous *harmonic series**.

Proof: We will show that the sequence of partial sums is unbounded, and hence cannot converge. Write the partial sums s_n for $n = 2^k$ as:

$$\begin{aligned} s_1 &= 1, \\ s_2 &= (1) + \left(\frac{1}{2}\right), \\ s_4 &= (1) + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right), \\ s_8 &= (1) + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right), \\ &\vdots \\ s_{2^k} &= 1 + \sum_{j=1}^k \left(\sum_{m=2^{j-1}+1}^{2^j} \frac{1}{m} \right). \end{aligned}$$

We note that $1/3 + 1/4 \geq 1/4 + 1/4 = 1/2$ and $1/5 + 1/6 + 1/7 + 1/8 \geq 1/8 + 1/8 + 1/8 + 1/8 = 1/2$. More generally

$$\sum_{m=2^{k-1}+1}^{2^k} \frac{1}{m} \geq \sum_{m=2^{k-1}+1}^{2^k} \frac{1}{2^k} = (2^{k-1}) \frac{1}{2^k} = \frac{1}{2}.$$

Therefore

$$s_{2^k} = 1 + \sum_{j=1}^k \left(\sum_{m=2^{j-1}+1}^{2^j} \frac{1}{m} \right) \geq 1 + \sum_{j=1}^k \frac{1}{2} = 1 + \frac{k}{2}.$$

As $\{\frac{k}{2}\}$ is unbounded by the Archimedean property, that means that $\{s_{2^k}\}$ is unbounded, and therefore $\{s_n\}$ is unbounded. Hence $\{s_n\}$ diverges, and consequently $\sum \frac{1}{n}$ diverges.

Convergent series are linear. That is, we can multiply them by constants and add them and these operations are done term by term.

Proposition 2.5.10 (Linearity of series). *Let $\alpha \in \mathbb{R}$ and $\sum x_n$ and $\sum y_n$ be convergent series. Then*

(i) $\sum \alpha x_n$ is a convergent series and

$$\sum_{n=1}^{\infty} \alpha x_n = \alpha \sum_{n=1}^{\infty} x_n.$$

*The divergence of the harmonic series was known before the theory of series was made rigorous. In fact the proof we give is the earliest proof and was given by Nicole Oresme (1323–1382).

(ii) $\sum(x_n + y_n)$ is a convergent series and

$$\sum_{n=1}^{\infty} (x_n + y_n) = \left(\sum_{n=1}^{\infty} x_n \right) + \left(\sum_{n=1}^{\infty} y_n \right).$$

Proof. For the first item, we simply write the n th partial sum

$$\sum_{k=1}^n \alpha x_k = \alpha \left(\sum_{k=1}^n x_k \right).$$

We look at the right-hand side and note that the constant multiple of a convergent sequence is convergent. Hence, we simply take the limit of both sides to obtain the result.

For the second item we also look at the n th partial sum

$$\sum_{k=1}^n (x_k + y_k) = \left(\sum_{k=1}^n x_k \right) + \left(\sum_{k=1}^n y_k \right).$$

We look at the right-hand side and note that the sum of convergent sequences is convergent. Hence, we simply take the limit of both sides to obtain the proposition. \square

Note that multiplying series is not as simple as adding, and we will not cover this topic here. It is not true, of course, that we can multiply term by term, since that strategy does not work even for finite sums. For example, $(a + b)(c + d) \neq ac + bd$.

2.5.4 Absolute convergence

Since monotone sequences are easier to work with than arbitrary sequences, it is generally easier to work with series $\sum x_n$ where $x_n \geq 0$ for all n . Then the sequence of partial sums is monotone increasing and converges if it is bounded from above. Let us formalize this statement as a proposition.

Proposition 2.5.11. *If $x_n \geq 0$ for all n , then $\sum x_n$ converges if and only if the sequence of partial sums is bounded from above.*

The following criterion often gives a convenient way to test for convergence of a series.

Definition 2.5.12. A series $\sum x_n$ converges absolutely if the series $\sum |x_n|$ converges. If a series converges, but does not converge absolutely, we say it is *conditionally convergent*.

Proposition 2.5.13. *If the series $\sum x_n$ converges absolutely, then it converges.*

Proof. We know that a series is convergent if and only if it is Cauchy. Hence suppose that $\sum |x_n|$ is Cauchy. That is for every $\varepsilon > 0$, there exists an M such that for all $k \geq M$ and $n > k$ we have that

$$\sum_{j=k+1}^n |x_j| = \left| \sum_{j=k+1}^n |x_j| \right| < \varepsilon.$$

We apply the triangle inequality for a finite sum to obtain

$$\left| \sum_{j=k+1}^n x_j \right| \leq \sum_{j=k+1}^n |x_j| < \varepsilon.$$

Hence $\sum x_n$ is Cauchy and therefore it converges. \square

Of course, if $\sum x_n$ converges absolutely, the limits of $\sum x_n$ and $\sum |x_n|$ are different. Computing one will not help us compute the other.

Absolutely convergent series have many wonderful properties for which we do not have space in this book. For example, absolutely convergent series can be rearranged arbitrarily.

We leave as an exercise to show that

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$$

converges. On the other hand we have already seen that

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges. Therefore $\sum \frac{(-1)^n}{n}$ is a conditionally convergent subsequence.

2.5.5 Comparison test and the p -series

We have noted above that for a series to converge the terms not only have to go to zero, but they have to go to zero “fast enough.” If we know about convergence of a certain series we can use the following comparison test to see if the terms of another series go to zero “fast enough.”

Proposition 2.5.14 (Comparison test). *Let $\sum x_n$ and $\sum y_n$ be series such that $0 \leq x_n \leq y_n$ for all $n \in \mathbb{N}$.*

- (i) *If $\sum y_n$ converges, then so does $\sum x_n$.*
- (ii) *If $\sum x_n$ diverges, then so does $\sum y_n$.*

Proof. Since the terms of the series are all nonnegative, the sequences of partial sums are both monotone increasing. Since $x_n \leq y_n$ for all n , the partial sums satisfy

$$\sum_{k=1}^n x_k \leq \sum_{k=1}^n y_k. \quad (2.1)$$

If the series $\sum y_n$ converges the partial sums for the series are bounded. Therefore the right-hand side of (2.1) is bounded for all n . Hence the partial sums for $\sum x_n$ are also bounded. Since the partial sums are a monotone increasing sequence they are convergent. The first item is thus proved.

On the other hand if $\sum x_n$ diverges, the sequence of partial sums must be unbounded since it is monotone increasing. That is, the partial sums for $\sum x_n$ are bigger than any real number. Putting this together with (2.1) we see that for any $B \in \mathbb{R}$, there is an n such that

$$B \leq \sum_{k=1}^n x_k \leq \sum_{k=1}^n y_k.$$

Hence the partial sums for $\sum y_n$ are also unbounded, and $\sum y_n$ also diverges. □

A useful series to use with the comparison test is the p -series.

Proposition 2.5.15 (p -series or the p -test). *For $p \in \mathbb{R}$, the series*

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if and only if $p > 1$.

Proof. First suppose $p \leq 1$. As $n \geq 1$, we have $\frac{1}{n^p} \geq \frac{1}{n}$. Since $\sum \frac{1}{n}$ diverges, we see that the $\sum \frac{1}{n^p}$ must diverge for all $p \leq 1$ by the comparison test.

Now suppose that $p > 1$. We proceed in a similar fashion as we did in the case of the harmonic series, but instead of showing that the sequence of partial sums is unbounded we show that it is bounded. Since the terms of the series are positive, the sequence of partial sums is monotone increasing and will converge if we show that it is bounded above. Let s_k denote the k th partial sum.

$$\begin{aligned} s_1 &= 1, \\ s_3 &= (1) + \left(\frac{1}{2^p} + \frac{1}{3^p} \right), \\ s_7 &= (1) + \left(\frac{1}{2^p} + \frac{1}{3^p} \right) + \left(\frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p} \right), \\ &\vdots \\ s_{2^k-1} &= 1 + \sum_{j=1}^{k-1} \left(\sum_{m=2^j}^{2^{j+1}-1} \frac{1}{m^p} \right). \end{aligned}$$

Instead of estimating from below, we estimate from above. In particular, as $p > 1$, then $2^p < 3^p$, and hence $\frac{1}{2^p} + \frac{1}{3^p} < \frac{1}{2^p} + \frac{1}{2^p}$. Similarly $\frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p} < \frac{1}{4^p} + \frac{1}{4^p} + \frac{1}{4^p} + \frac{1}{4^p}$. Therefore

$$\begin{aligned} s_{2^k-1} &= 1 + \sum_{j=1}^k \left(\sum_{m=2^j}^{2^{j+1}-1} \frac{1}{m^p} \right) \\ &< 1 + \sum_{j=1}^k \left(\sum_{m=2^j}^{2^{j+1}-1} \frac{1}{(2^j)^p} \right) \\ &= 1 + \sum_{j=1}^k \left(\frac{2^j}{(2^j)^p} \right) \\ &= 1 + \sum_{j=1}^k \left(\frac{1}{2^{p-1}} \right)^j. \end{aligned}$$

As $p > 1$, then $\frac{1}{2^{p-1}} < 1$. Then by using the result of Exercise 2.5.2, we note that

$$\sum_{j=1}^{\infty} \left(\frac{1}{2^{p-1}} \right)^j$$

converges. Therefore

$$s_{2^k-1} < 1 + \sum_{j=1}^k \left(\frac{1}{2^{p-1}} \right)^j \leq 1 + \sum_{j=1}^{\infty} \left(\frac{1}{2^{p-1}} \right)^j.$$

As $\{s_n\}$ is a monotone sequence, then all $s_n \leq s_{2^k-1}$ for all $n \leq 2^k - 1$. Thus for all n ,

$$s_n < 1 + \sum_{j=1}^{\infty} \left(\frac{1}{2^{p-1}} \right)^j.$$

The sequence of partial sums is bounded and hence converges. \square

Note that neither the p -series test nor the comparison test will tell us what the sum converges to. They only tell us that a limit of the partial sums exists. For example, while we know that $\sum 1/n^2$ converges it is far harder to find* that the limit is $\pi^2/6$. If we treat $\sum 1/n^p$ as a function of p , we get the so-called Riemann ζ function. Understanding the behavior of this function contains one of the most famous unsolved problems in mathematics today and has applications in seemingly unrelated areas such as modern cryptography.

Example 2.5.16: The series $\sum \frac{1}{n^2+1}$ converges.

Proof: First note that $\frac{1}{n^2+1} < \frac{1}{n^2}$ for all $n \in \mathbb{N}$. Note that $\sum \frac{1}{n^2}$ converges by the p -series test. Therefore, by the comparison test, $\sum \frac{1}{n^2+1}$ converges.

*Demonstration of this fact is what made the Swiss mathematician Leonhard Paul Euler (1707 – 1783) famous.

2.5.6 Ratio test

Proposition 2.5.17 (Ratio test). *Let $\sum x_n$ be a series such that*

$$L := \lim_{n \rightarrow \infty} \frac{|x_{n+1}|}{|x_n|}$$

exists. Then

(i) *If $L < 1$, then $\sum x_n$ converges absolutely.*

(ii) *If $L > 1$, then $\sum x_n$ diverges.*

Proof. From Lemma 2.2.12 we note that if $L > 1$, then x_n diverges. Since it is a necessary condition for the convergence of series that the terms go to zero, we know that $\sum x_n$ must diverge.

Thus suppose that $L < 1$. We will argue that $\sum |x_n|$ must converge. The proof is similar to that of Lemma 2.2.12. Of course $L \geq 0$. Pick r such that $L < r < 1$. As $r - L > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$

$$\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < r - L.$$

Therefore,

$$\frac{|x_{n+1}|}{|x_n|} < r.$$

For $n > M$ (that is for $n \geq M + 1$) write

$$|x_n| = |x_M| \frac{|x_{M+1}|}{|x_M|} \frac{|x_{M+2}|}{|x_{M+1}|} \cdots \frac{|x_n|}{|x_{n-1}|} < |x_M| r r \cdots r = |x_M| r^{n-M} = (|x_M| r^{-M}) r^n.$$

For $n > M$ we write the partial sum as

$$\begin{aligned} \sum_{k=1}^n |x_k| &= \left(\sum_{k=1}^M |x_k| \right) + \left(\sum_{k=M+1}^n |x_k| \right) \\ &\leq \left(\sum_{k=1}^M |x_k| \right) + \left(\sum_{k=M+1}^n (|x_M| r^{-M}) r^k \right) \\ &\leq \left(\sum_{k=1}^M |x_k| \right) + (|x_M| r^{-M}) \left(\sum_{k=M+1}^n r^k \right). \end{aligned}$$

As $0 < r < 1$ the geometric series $\sum_{k=0}^{\infty} r^k$ converges, and of course $\sum_{k=M+1}^{\infty} r^k$ converges as well (why?). We take the limit as n goes to infinity on the right-hand side to obtain

$$\begin{aligned} \sum_{k=1}^n |x_k| &\leq \left(\sum_{k=1}^M |x_k| \right) + (|x_M| r^{-M}) \left(\sum_{k=M+1}^n r^k \right) \\ &\leq \left(\sum_{k=1}^M |x_k| \right) + (|x_M| r^{-M}) \left(\sum_{k=M+1}^{\infty} r^k \right). \end{aligned}$$

The right-hand side is a number that does not depend on n . Hence the sequence of partial sums of $\sum |x_n|$ is bounded and $\sum |x_n|$ is convergent. Thus $\sum x_n$ is absolutely convergent. \square

Example 2.5.18: The series

$$\sum_{n=1}^{\infty} \frac{2^n}{n!}$$

converges absolutely.

Proof: We write

$$\lim_{n \rightarrow \infty} \frac{2^{(n+1)}/(n+1)!}{2^n/n!} = \lim_{n \rightarrow \infty} \frac{2}{n+1} = 0.$$

Therefore, the series converges absolutely by the ratio test.

2.5.7 Exercises

Exercise 2.5.1: For $r \neq 1$, prove

$$\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}.$$

Hint: Let $s := \sum_{k=0}^{n-1} r^k$, then compute $s(1-r) = s - rs$, and solve for s .

Exercise 2.5.2: Prove that for $-1 < r < 1$ we have

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

Hint: Use the previous exercise.

Exercise 2.5.3: Decide the convergence or divergence of the following series.

$$\begin{array}{lll} a) \sum_{n=1}^{\infty} \frac{3}{9n+1} & b) \sum_{n=1}^{\infty} \frac{1}{2n-1} & c) \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \\ d) \sum_{n=1}^{\infty} \frac{1}{n(n+1)} & e) \sum_{n=1}^{\infty} ne^{-n^2} & \end{array}$$

Exercise 2.5.4:

a) Prove that if $\sum_{n=1}^{\infty} x_n$ converges, then $\sum_{n=1}^{\infty} (x_{2n} + x_{2n+1})$ also converges.

b) Find an explicit example where the converse does not hold.

Exercise 2.5.5: For $j = 1, 2, \dots, n$, let $\{x_{j,k}\}_{k=1}^{\infty}$ denote n sequences. Suppose that for each j

$$\sum_{k=1}^{\infty} x_{j,k}$$

is convergent. Then show

$$\sum_{j=1}^n \left(\sum_{k=1}^{\infty} x_{j,k} \right) = \sum_{k=1}^{\infty} \left(\sum_{j=1}^n x_{j,k} \right).$$

Exercise 2.5.6: Prove the following stronger version of the ratio test: Let $\sum x_n$ be a series.

- a) If there is an N and a $\rho < 1$ such that for all $n \geq N$ we have $\frac{|x_{n+1}|}{|x_n|} < \rho$, then the series converges absolutely.
- b) If there is an N such that for all $n \geq N$ we have $\frac{|x_{n+1}|}{|x_n|} \geq 1$, then the series diverges.

Exercise 2.5.7 (Challenging): Let $\{x_n\}$ be a decreasing sequence such that $\sum x_n$ converges. Show that $\lim_{n \rightarrow \infty} nx_n = 0$.

Exercise 2.5.8: Show that $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges. Hint: consider the sum of two subsequent entries.

Exercise 2.5.9:

- a) Prove that if $\sum x_n$ and $\sum y_n$ converge absolutely, then $\sum x_n y_n$ converges absolutely.
- b) Find an explicit example where the converse does not hold.
- c) Find an explicit example where all three series are absolutely convergent, are not just finite sums, and $(\sum x_n)(\sum y_n) \neq \sum x_n y_n$. That is, show that series are not multiplied term-by-term.

Exercise 2.5.10: Prove the triangle inequality for series, that is if $\sum x_n$ converges absolutely then

$$\left| \sum_{n=1}^{\infty} x_n \right| \leq \sum_{n=1}^{\infty} |x_n|.$$

Chapter 3

Continuous Functions

3.1 Limits of functions

Note: 2–3 lectures

Before we define continuity of functions, we need to visit a somewhat more general notion of a limit. That is, given a function $f: S \rightarrow \mathbb{R}$, we want to see how $f(x)$ behaves as x tends to a certain point.

3.1.1 Cluster points

First, let us return to a concept we have previously seen in an exercise.

Definition 3.1.1. Let $S \subset \mathbb{R}$ be a set. A number $x \in \mathbb{R}$ is called a *cluster point* of S if for every $\varepsilon > 0$, the set $(x - \varepsilon, x + \varepsilon) \cap S \setminus \{x\}$ is not empty.

That is, x is a cluster point of S if there are points of S arbitrarily close to x . Another way of phrasing the definition is to say that x is a cluster point of S if for every $\varepsilon > 0$, there exists a $y \in S$ such that $y \neq x$ and $|x - y| < \varepsilon$. Note that a cluster point of S need not lie in S .

Let us see some examples.

- (i) The set $\{1/n : n \in \mathbb{N}\}$ has a unique cluster point zero.
- (ii) The cluster points of the open interval $(0, 1)$ are all points in the closed interval $[0, 1]$.
- (iii) For the set \mathbb{Q} , the set of cluster points is the whole real line \mathbb{R} .
- (iv) For the set $[0, 1) \cup \{2\}$, the set of cluster points is the interval $[0, 1]$.
- (v) The set \mathbb{N} has no cluster points in \mathbb{R} .

Proposition 3.1.2. Let $S \subset \mathbb{R}$. Then $x \in \mathbb{R}$ is a cluster point of S if and only if there exists a convergent sequence of numbers $\{x_n\}$ such that $x_n \neq x$, $x_n \in S$, and $\lim x_n = x$.

Proof. First suppose that x is a cluster point of S . For any $n \in \mathbb{N}$, we pick x_n to be an arbitrary point of $(x - 1/n, x + 1/n) \cap S \setminus \{x\}$, which we know is nonempty because x is a cluster point of S . Then x_n is within $1/n$ of x , that is,

$$|x - x_n| < 1/n.$$

As $\{1/n\}$ converges to zero, $\{x_n\}$ converges to x .

On the other hand, if we start with a sequence of numbers $\{x_n\}$ in S converging to x such that $x_n \neq x$ for all n , then for every $\varepsilon > 0$ there is an M such that in particular $|x_M - x| < \varepsilon$. That is, $x_M \in (x - \varepsilon, x + \varepsilon) \cap S \setminus \{x\}$. \square

3.1.2 Limits of functions

If a function f is defined on a set S and c is a cluster point of S , then we can define the limit of $f(x)$ as x gets close to c . Do note that it is irrelevant for the definition if f is defined at c or not. Furthermore, even if the function is defined at c , the limit of the function as x goes to c could very well be different from $f(c)$.

Definition 3.1.3. Let $f: S \rightarrow \mathbb{R}$ be a function and c a cluster point of S . Suppose that there exists an $L \in \mathbb{R}$ and for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in S \setminus \{c\}$ and $|x - c| < \delta$, then

$$|f(x) - L| < \varepsilon.$$

In this case we say that $f(x)$ converges to L as x goes to c . We also say that L is the *limit* of $f(x)$ as x goes to c . We write

$$\lim_{x \rightarrow c} f(x) := L,$$

or

$$f(x) \rightarrow L \quad \text{as } x \rightarrow c.$$

If no such L exists, then we say that the limit does not exist or that f *diverges* at c .

Again the notation and language we are using above assumes that the limit is unique even though we have not yet proved that. Let us do that now.

Proposition 3.1.4. Let c be a cluster point of $S \subset \mathbb{R}$ and let $f: S \rightarrow \mathbb{R}$ be a function such that $f(x)$ converges as x goes to c . Then the limit of $f(x)$ as x goes to c is unique.

Proof. Let L_1 and L_2 be two numbers that both satisfy the definition. Take an $\varepsilon > 0$ and find a $\delta_1 > 0$ such that $|f(x) - L_1| < \varepsilon/2$ for all $x \in S \setminus \{c\}$ with $|x - c| < \delta_1$. Also find $\delta_2 > 0$ such that $|f(x) - L_2| < \varepsilon/2$ for all $x \in S \setminus \{c\}$ with $|x - c| < \delta_2$. Put $\delta := \min\{\delta_1, \delta_2\}$. Suppose that $x \in S$, $|x - c| < \delta$, and $x \neq c$. Then

$$|L_1 - L_2| = |L_1 - f(x) + f(x) - L_2| \leq |L_1 - f(x)| + |f(x) - L_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

As $|L_1 - L_2| < \varepsilon$ for arbitrary $\varepsilon > 0$, then $L_1 = L_2$. \square

Example 3.1.5: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be defined as $f(x) := x^2$. Then

$$\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} x^2 = c^2.$$

Proof: First let c be fixed. Let $\varepsilon > 0$ be given. Take

$$\delta := \min \left\{ 1, \frac{\varepsilon}{2|c|+1} \right\}.$$

Take $x \neq c$ such that $|x - c| < \delta$. In particular, $|x - c| < 1$. Then by reverse triangle inequality we get

$$|x| - |c| \leq |x - c| < 1.$$

Adding $2|c|$ to both sides we obtain $|x| + |c| < 2|c| + 1$. We compute

$$\begin{aligned} |f(x) - c^2| &= |x^2 - c^2| \\ &= |(x+c)(x-c)| \\ &= |x+c| |x-c| \\ &\leq (|x| + |c|) |x-c| \\ &< (2|c| + 1) |x-c| \\ &< (2|c| + 1) \frac{\varepsilon}{2|c| + 1} = \varepsilon. \end{aligned}$$

Example 3.1.6: Define $f: [0, 1) \rightarrow \mathbb{R}$ by

$$f(x) := \begin{cases} x & \text{if } x > 0, \\ 1 & \text{if } x = 0. \end{cases}$$

Then

$$\lim_{x \rightarrow 0^+} f(x) = 0,$$

even though $f(0) = 1$.

Proof: Let $\varepsilon > 0$ be given. Let $\delta := \varepsilon$. Then for $x \in [0, 1)$, $x \neq 0$, and $|x - 0| < \delta$ we get

$$|f(x) - 0| = |x| < \delta = \varepsilon.$$

3.1.3 Sequential limits

Let us connect the limit as defined above with limits of sequences.

Lemma 3.1.7. Let $S \subset \mathbb{R}$ and c be a cluster point of S . Let $f: S \rightarrow \mathbb{R}$ be a function.

Then $f(x) \rightarrow L$ as $x \rightarrow c$, if and only if for every sequence $\{x_n\}$ of numbers such that $x_n \in S \setminus \{c\}$ for all n , and such that $\lim x_n = c$, we have that the sequence $\{f(x_n)\}$ converges to L .

Proof. Suppose that $f(x) \rightarrow L$ as $x \rightarrow c$, and $\{x_n\}$ is a sequence such that $x_n \in S \setminus \{c\}$ and $\lim x_n = c$. We wish to show that $\{f(x_n)\}$ converges to L . Let $\varepsilon > 0$ be given. Find a $\delta > 0$ such that if $x \in S \setminus \{c\}$ and $|x - c| < \delta$, then $|f(x) - L| < \varepsilon$. As $\{x_n\}$ converges to c , find an M such that for $n \geq M$ we have that $|x_n - c| < \delta$. Therefore

$$|f(x_n) - L| < \varepsilon.$$

Thus $\{f(x_n)\}$ converges to L .

For the other direction, we will use proof by contrapositive. Suppose that it is not true that $f(x) \rightarrow L$ as $x \rightarrow c$. The negation of the definition is that there exists an $\varepsilon > 0$ such that for every $\delta > 0$ there exists an $x \in S \setminus \{c\}$, where $|x - c| < \delta$ and $|f(x) - L| \geq \varepsilon$.

Let us use $1/n$ for δ in the above statement to construct a sequence $\{x_n\}$. We have that there exists an $\varepsilon > 0$ such that for every n , there exists a point $x_n \in S \setminus \{c\}$, where $|x_n - c| < 1/n$ and $|f(x_n) - L| \geq \varepsilon$. The sequence $\{x_n\}$ just constructed converges to c , but the sequence $\{f(x_n)\}$ does not converge to L . And we are done. \square

It is possible to strengthen the reverse direction of the lemma by simply stating that $\{f(x_n)\}$ converges without requiring a specific limit. See Exercise 3.1.11.

Example 3.1.8: $\lim_{x \rightarrow 0} \sin(1/x)$ does not exist, but $\lim_{x \rightarrow 0} x \sin(1/x) = 0$. See Figure 3.1.

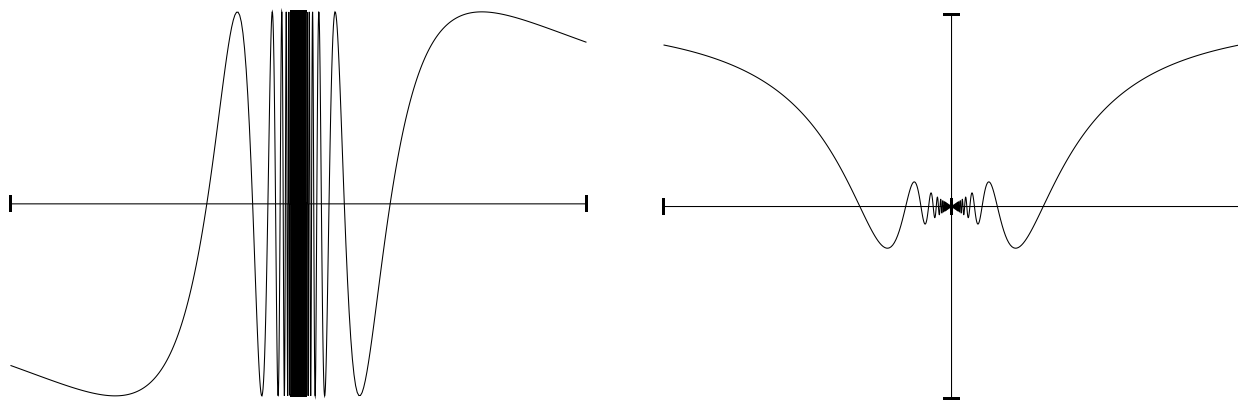


Figure 3.1: Graphs of $\sin(1/x)$ and $x \sin(1/x)$. Note that the computer cannot properly graph $\sin(1/x)$ near zero as it oscillates too fast.

Proof: Let us work with $\sin(1/x)$ first. Let us define the sequence $x_n := \frac{1}{\pi n + \pi/2}$. It is not hard to see that $\lim x_n = 0$. Furthermore,

$$\sin(1/x_n) = \sin(\pi n + \pi/2) = (-1)^n.$$

Therefore, $\{\sin(1/x_n)\}$ does not converge. Thus, by Lemma 3.1.7, $\lim_{x \rightarrow 0} \sin(1/x)$ does not exist.

Now let us look at $x \sin(1/x)$. Let x_n be a sequence such that $x_n \neq 0$ for all n and such that $\lim x_n = 0$. Notice that $|\sin(t)| \leq 1$ for any $t \in \mathbb{R}$. Therefore,

$$|x_n \sin(1/x_n) - 0| = |x_n| |\sin(1/x_n)| \leq |x_n|.$$

As x_n goes to 0, then $|x_n|$ goes to zero, and hence $\{x_n \sin(1/x_n)\}$ converges to zero. By Lemma 3.1.7, $\lim_{x \rightarrow 0} x \sin(1/x) = 0$.

Using Lemma 3.1.7, we can start applying everything we know about sequential limits to limits of functions. Let us give a few important examples.

Corollary 3.1.9. *Let $S \subset \mathbb{R}$ and c be a cluster point of S . Let $f: S \rightarrow \mathbb{R}$ and $g: S \rightarrow \mathbb{R}$ be functions. Suppose that the limits of $f(x)$ and $g(x)$ as x goes to c both exist, and that*

$$f(x) \leq g(x) \quad \text{for all } x \in S.$$

Then

$$\lim_{x \rightarrow c} f(x) \leq \lim_{x \rightarrow c} g(x).$$

Proof. Take $\{x_n\}$ be a sequence of numbers in $S \setminus \{c\}$ that converges to c . Let

$$L_1 := \lim_{x \rightarrow c} f(x), \quad \text{and} \quad L_2 := \lim_{x \rightarrow c} g(x).$$

By Lemma 3.1.7 we know $\{f(x_n)\}$ converges to L_1 and $\{g(x_n)\}$ converges to L_2 . We also have $f(x_n) \leq g(x_n)$. We obtain $L_1 \leq L_2$ using Lemma 2.2.3. \square

By applying constant functions, we get the following corollary. The proof is left as an exercise.

Corollary 3.1.10. *Let $S \subset \mathbb{R}$ and c be a cluster point of S . Let $f: S \rightarrow \mathbb{R}$ be a function. And suppose that the limit of $f(x)$ as x goes to c exists. Suppose that there are two real numbers a and b such that*

$$a \leq f(x) \leq b \quad \text{for all } x \in S.$$

Then

$$a \leq \lim_{x \rightarrow c} f(x) \leq b.$$

Using Lemma 3.1.7 in the same way as above we also get the following corollaries, whose proofs are again left as an exercise.

Corollary 3.1.11. *Let $S \subset \mathbb{R}$ and c be a cluster point of S . Let $f: S \rightarrow \mathbb{R}$, $g: S \rightarrow \mathbb{R}$, and $h: S \rightarrow \mathbb{R}$ be functions. Suppose that*

$$f(x) \leq g(x) \leq h(x) \quad \text{for all } x \in S,$$

and the limits of $f(x)$ and $h(x)$ as x goes to c both exist, and

$$\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} h(x).$$

Then the limit of $g(x)$ as x goes to c exists and

$$\lim_{x \rightarrow c} g(x) = \lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} h(x).$$

Corollary 3.1.12. Let $S \subset \mathbb{R}$ and c be a cluster point of S . Let $f: S \rightarrow \mathbb{R}$ and $g: S \rightarrow \mathbb{R}$ be functions. Suppose that limits of $f(x)$ and $g(x)$ as x goes to c both exist. Then

$$(i) \lim_{x \rightarrow c} (f(x) + g(x)) = \left(\lim_{x \rightarrow c} f(x) \right) + \left(\lim_{x \rightarrow c} g(x) \right).$$

$$(ii) \lim_{x \rightarrow c} (f(x) - g(x)) = \left(\lim_{x \rightarrow c} f(x) \right) - \left(\lim_{x \rightarrow c} g(x) \right).$$

$$(iii) \lim_{x \rightarrow c} (f(x)g(x)) = \left(\lim_{x \rightarrow c} f(x) \right) \left(\lim_{x \rightarrow c} g(x) \right).$$

(iv) If $\lim_{x \rightarrow c} g(x) \neq 0$, and $g(x) \neq 0$ for all $x \in S \setminus \{c\}$, then

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow c} f(x)}{\lim_{x \rightarrow c} g(x)}.$$

3.1.4 Restrictions and limits

It is not necessary to always consider all of S . Sometimes we may be able to work with the function defined on a smaller set.

Definition 3.1.13. Let $f: S \rightarrow \mathbb{R}$ be a function. Let $A \subset S$. Define the function $f|_A: A \rightarrow \mathbb{R}$ by

$$f|_A(x) := f(x) \quad \text{for } x \in A.$$

The function $f|_A$ is called the *restriction* of f to A .

The function $f|_A$ is simply the function f taken on a smaller domain. The following proposition is the analogue of taking a tail of a sequence.

Proposition 3.1.14. Let $S \subset \mathbb{R}$, $c \in \mathbb{R}$, and let $f: S \rightarrow \mathbb{R}$ be a function. Suppose $A \subset S$ is such that there is some $\alpha > 0$ such that $A \cap (c - \alpha, c + \alpha) = S \cap (c - \alpha, c + \alpha)$.

(i) The point c is a cluster point of A if and only if c is a cluster point of S .

(ii) Supposing c is a cluster point of S , then $f(x) \rightarrow L$ as $x \rightarrow c$ if and only if $f|_A(x) \rightarrow L$ as $x \rightarrow c$.

Proof. First, let c be a cluster point of A . Since $A \subset S$, then if $(A \setminus \{c\}) \cap (c - \varepsilon, c + \varepsilon)$ is nonempty for every $\varepsilon > 0$, then $(S \setminus \{c\}) \cap (c - \varepsilon, c + \varepsilon)$ is nonempty for every $\varepsilon > 0$. Thus c is a cluster point of S . Second, suppose that c is a cluster point of S . Then for $\varepsilon > 0$ such that $\varepsilon < \alpha$ we get that $(A \setminus \{c\}) \cap (c - \varepsilon, c + \varepsilon) = (S \setminus \{c\}) \cap (c - \varepsilon, c + \varepsilon)$, which is nonempty. This is true for all $\varepsilon < \alpha$ and hence $(A \setminus \{c\}) \cap (c - \varepsilon, c + \varepsilon)$ must be nonempty for all $\varepsilon > 0$. Thus c is a cluster point of A .

Now suppose that $f(x) \rightarrow L$ as $x \rightarrow c$. That is, for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $x \in S \setminus \{c\}$ and $|x - c| < \delta$, then $|f(x) - L| < \varepsilon$. As $A \subset S$, then if x is in $A \setminus \{c\}$, then x is in $S \setminus \{c\}$, and hence $f|_A(x) \rightarrow L$ as $x \rightarrow c$.

Now suppose that $f|_A(x) \rightarrow L$ as $x \rightarrow c$. Hence for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $x \in A \setminus \{c\}$ and $|x - c| < \delta$, then $|f|_A(x) - L| < \varepsilon$. Without loss of generality assume $\delta \leq \alpha$. If $|x - c| < \delta$, then $x \in S \setminus \{c\}$ if and only if $x \in A \setminus \{c\}$. Thus $|f(x) - L| = |f|_A(x) - L| < \varepsilon$. \square

3.1.5 Exercises

Exercise 3.1.1: Find the limit or prove that the limit does not exist

$$\begin{array}{lll} a) \lim_{x \rightarrow c} \sqrt{x}, \text{ for } c \geq 0 & b) \lim_{x \rightarrow c} x^2 + x + 1, \text{ for any } c \in \mathbb{R} & c) \lim_{x \rightarrow 0} x^2 \cos(1/x) \\ d) \lim_{x \rightarrow 0} \sin(1/x) \cos(1/x) & e) \lim_{x \rightarrow 0} \sin(x) \cos(1/x) & \end{array}$$

Exercise 3.1.2: Prove Corollary 3.1.10.

Exercise 3.1.3: Prove Corollary 3.1.11.

Exercise 3.1.4: Prove Corollary 3.1.12.

Exercise 3.1.5: Let $A \subset S$. Show that if c is a cluster point of A , then c is a cluster point of S . Note the difference from Proposition 3.1.14.

Exercise 3.1.6: Let $A \subset S$. Suppose that c is a cluster point of A and it is also a cluster point of S . Let $f: S \rightarrow \mathbb{R}$ be a function. Show that if $f(x) \rightarrow L$ as $x \rightarrow c$, then $f|_A(x) \rightarrow L$ as $x \rightarrow c$. Note the difference from Proposition 3.1.14.

Exercise 3.1.7: Find an example of a function $f: [-1, 1] \rightarrow \mathbb{R}$ such that for $A := [0, 1]$, the restriction $f|_A(x) \rightarrow 0$ as $x \rightarrow 0$, but the limit of $f(x)$ as $x \rightarrow 0$ does not exist. Note why you cannot apply Proposition 3.1.14.

Exercise 3.1.8: Find example functions f and g such that the limit of neither $f(x)$ nor $g(x)$ exists as $x \rightarrow 0$, but such that the limit of $f(x) + g(x)$ exists as $x \rightarrow 0$.

Exercise 3.1.9: Let c_1 be a cluster point of $A \subset \mathbb{R}$ and c_2 be a cluster point of $B \subset \mathbb{R}$. Suppose that $f: A \rightarrow B$ and $g: B \rightarrow \mathbb{R}$ are functions such that $f(x) \rightarrow c_2$ as $x \rightarrow c_1$ and $g(y) \rightarrow L$ as $y \rightarrow c_2$. Let $h(x) := g(f(x))$ and show $h(x) \rightarrow L$ as $x \rightarrow c_1$.

Exercise 3.1.10: Let c be a cluster point of $A \subset \mathbb{R}$, and $f: A \rightarrow \mathbb{R}$ be a function. Suppose that for every sequence $\{x_n\}$ in A , such that $\lim x_n = c$, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy. Prove that $\lim_{x \rightarrow c} f(x)$ exists.

Exercise 3.1.11: Prove the following stronger version of one direction of Lemma 3.1.7: Let $S \subset \mathbb{R}$, c be a cluster point of S , and $f: S \rightarrow \mathbb{R}$ be a function. Suppose that for every sequence $\{x_n\}$ in $S \setminus \{c\}$ such that $\lim x_n = c$ the sequence $\{f(x_n)\}$ is convergent. Then show $f(x) \rightarrow L$ as $x \rightarrow c$ for some $L \in \mathbb{R}$.

3.2 Continuous functions

Note: 2–2.5 lectures

You have undoubtedly heard of continuous functions in your schooling. A high school criterion for this concept is that a function is continuous if we can draw its graph without lifting the pen from the paper. While that intuitive concept may be useful in simple situations, we require a rigorous concept. The following definition took three great mathematicians (Bolzano, Cauchy, and finally Weierstrass) to get correctly and its final form dates only to the late 1800s.

3.2.1 Definition and basic properties

Definition 3.2.1. Let $S \subset \mathbb{R}$, $c \in S$, and let $f: S \rightarrow \mathbb{R}$ be a function. We say that f is *continuous at c* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that whenever $x \in S$ and $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$.

When $f: S \rightarrow \mathbb{R}$ is continuous at all $c \in S$, then we simply say that f is a *continuous function*.

Sometimes we say that f is continuous on $A \subset S$. We then mean that f is continuous at all $c \in A$. It is left as an exercise to prove that if f is continuous on A , then $f|_A$ is continuous.

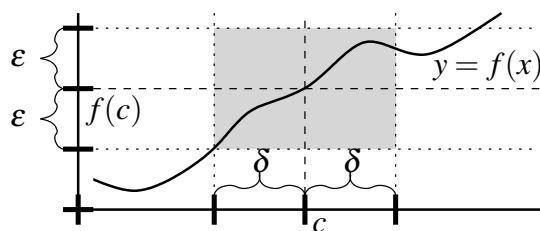


Figure 3.2: For $|x - c| < \delta$, $f(x)$ should be within the gray region.

Continuity may be the most important definition to understand in analysis, and it is not an easy one. See Figure 3.2. Note that δ not only depends on ε , but also on c ; we need not have to pick one δ for all $c \in S$. It is no accident that the definition of continuity is similar to the definition of a limit of a function. The main feature of continuous functions is that these are precisely the functions that behave nicely with limits.

Proposition 3.2.2. *Suppose that $f: S \rightarrow \mathbb{R}$ is a function and $c \in S$. Then*

- (i) *If c is not a cluster point of S , then f is continuous at c .*
- (ii) *If c is a cluster point of S , then f is continuous at c if and only if the limit of $f(x)$ as $x \rightarrow c$ exists and*

$$\lim_{x \rightarrow c} f(x) = f(c).$$

- (iii) *f is continuous at c if and only if for every sequence $\{x_n\}$ where $x_n \in S$ and $\lim x_n = c$, the sequence $\{f(x_n)\}$ converges to $f(c)$.*

Proof. Let us start with the first item. Suppose that c is not a cluster point of S . Then there exists a $\delta > 0$ such that $S \cap (c - \delta, c + \delta) = \{c\}$. Therefore, for any $\varepsilon > 0$, simply pick this given delta. The only $x \in S$ such that $|x - c| < \delta$ is $x = c$. Then $|f(x) - f(c)| = |f(c) - f(c)| = 0 < \varepsilon$.

Let us move to the second item. Suppose that c is a cluster point of S . Let us first suppose that $\lim_{x \rightarrow c} f(x) = f(c)$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $x \in S \setminus \{c\}$ and $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$. As $|f(c) - f(c)| = 0 < \varepsilon$, then the definition of continuity at c is satisfied. On the other hand, suppose that f is continuous at c . For every $\varepsilon > 0$, there exists a $\delta > 0$ such that for $x \in S$ where $|x - c| < \delta$ we have $|f(x) - f(c)| < \varepsilon$. Then the statement is, of course, still true if $x \in S \setminus \{c\} \subset S$. Therefore $\lim_{x \rightarrow c} f(x) = f(c)$.

For the third item, suppose that f is continuous at c . Let $\{x_n\}$ be a sequence such that $x_n \in S$ and $\lim x_n = c$. Let $\varepsilon > 0$ be given. Find a $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ for all $x \in S$ where $|x - c| < \delta$. Find an $M \in \mathbb{N}$ such that for $n \geq M$ we have $|x_n - c| < \delta$. Then for $n \geq M$ we have that $|f(x_n) - f(c)| < \varepsilon$, so $\{f(x_n)\}$ converges to $f(c)$.

Let us prove the converse of the third item by contrapositive. Suppose that f is not continuous at c . Then there exists an $\varepsilon > 0$ such that for all $\delta > 0$, there exists an $x \in S$ such that $|x - c| < \delta$ and $|f(x) - f(c)| \geq \varepsilon$. Let us define a sequence $\{x_n\}$ as follows. Let $x_n \in S$ be such that $|x_n - c| < 1/n$ and $|f(x_n) - f(c)| \geq \varepsilon$. Now $\{x_n\}$ is a sequence of numbers in S such that $\lim x_n = c$ and such that $|f(x_n) - f(c)| \geq \varepsilon$ for all $n \in \mathbb{N}$. Thus $\{f(x_n)\}$ does not converge to $f(c)$. It may or may not converge, but it definitely does not converge to $f(c)$. \square

The last item in the proposition is particularly powerful. It allows us to quickly apply what we know about limits of sequences to continuous functions and even to prove that certain functions are continuous. It can also be strengthened, see Exercise 3.2.13.

Example 3.2.3: $f: (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := 1/x$ is continuous.

Proof: Fix $c \in (0, \infty)$. Let $\{x_n\}$ be a sequence in $(0, \infty)$ such that $\lim x_n = c$. Then we know that

$$\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} \frac{1}{x_n} = \frac{1}{\lim x_n} = \frac{1}{c} = f(c).$$

Thus f is continuous at c . As f is continuous at all $c \in (0, \infty)$, f is continuous.

We have previously shown that $\lim_{x \rightarrow c} x^2 = c^2$ directly. Therefore the function x^2 is continuous. We can use the continuity of algebraic operations with respect to limits of sequences, which we proved in the previous chapter, to prove a much more general result.

Proposition 3.2.4. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a polynomial. That is

$$f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0,$$

for some constants a_0, a_1, \dots, a_d . Then f is continuous.

Proof. Fix $c \in \mathbb{R}$. Let $\{x_n\}$ be a sequence such that $\lim x_n = c$. Then

$$\begin{aligned}\lim_{n \rightarrow \infty} f(x_n) &= \lim_{n \rightarrow \infty} \left(a_d x_n^d + a_{d-1} x_n^{d-1} + \cdots + a_1 x_n + a_0 \right) \\ &= a_d (\lim x_n)^d + a_{d-1} (\lim x_n)^{d-1} + \cdots + a_1 (\lim x_n) + a_0 \\ &= a_d c^d + a_{d-1} c^{d-1} + \cdots + a_1 c + a_0 = f(c).\end{aligned}$$

Thus f is continuous at c . As f is continuous at all $c \in \mathbb{R}$, f is continuous. \square

By similar reasoning, or by appealing to Corollary 3.1.12, we can prove the following. The details of the proof are left as an exercise.

Proposition 3.2.5. *Let $f: S \rightarrow \mathbb{R}$ and $g: S \rightarrow \mathbb{R}$ be functions continuous at $c \in S$.*

- (i) *The function $h: S \rightarrow \mathbb{R}$ defined by $h(x) := f(x) + g(x)$ is continuous at c .*
- (ii) *The function $h: S \rightarrow \mathbb{R}$ defined by $h(x) := f(x) - g(x)$ is continuous at c .*
- (iii) *The function $h: S \rightarrow \mathbb{R}$ defined by $h(x) := f(x)g(x)$ is continuous at c .*
- (iv) *If $g(x) \neq 0$ for all $x \in S$, the function $h: S \rightarrow \mathbb{R}$ defined by $h(x) := \frac{f(x)}{g(x)}$ is continuous at c .*

Example 3.2.6: The functions $\sin(x)$ and $\cos(x)$ are continuous. In the following computations we use the sum-to-product trigonometric identities. We also use the simple facts that $|\sin(x)| \leq |x|$, $|\cos(x)| \leq 1$, and $|\sin(x)| \leq 1$.

$$\begin{aligned}|\sin(x) - \sin(c)| &= \left| 2 \sin\left(\frac{x-c}{2}\right) \cos\left(\frac{x+c}{2}\right) \right| \\ &= 2 \left| \sin\left(\frac{x-c}{2}\right) \right| \left| \cos\left(\frac{x+c}{2}\right) \right| \\ &\leq 2 \left| \sin\left(\frac{x-c}{2}\right) \right| \\ &\leq 2 \left| \frac{x-c}{2} \right| = |x-c|\end{aligned}$$

$$\begin{aligned}|\cos(x) - \cos(c)| &= \left| -2 \sin\left(\frac{x-c}{2}\right) \sin\left(\frac{x+c}{2}\right) \right| \\ &= 2 \left| \sin\left(\frac{x-c}{2}\right) \right| \left| \sin\left(\frac{x+c}{2}\right) \right| \\ &\leq 2 \left| \sin\left(\frac{x-c}{2}\right) \right| \\ &\leq 2 \left| \frac{x-c}{2} \right| = |x-c|\end{aligned}$$

The claim that \sin and \cos are continuous follows by taking an arbitrary sequence $\{x_n\}$ converging to c . Details are left to the reader.

3.2.2 Composition of continuous functions

You have probably already realized that one of the basic tools in constructing complicated functions out of simple ones is composition. A useful property of continuous functions is that compositions of continuous functions are again continuous. Recall that for two functions f and g , the composition $f \circ g$ is defined by $(f \circ g)(x) := f(g(x))$.

Proposition 3.2.7. *Let $A, B \subset \mathbb{R}$ and $f: B \rightarrow \mathbb{R}$ and $g: A \rightarrow B$ be functions. If g is continuous at $c \in A$ and f is continuous at $g(c)$, then $f \circ g: A \rightarrow \mathbb{R}$ is continuous at c .*

Proof. Let $\{x_n\}$ be a sequence in A such that $\lim x_n = c$. Then as g is continuous at c , then $\{g(x_n)\}$ converges to $g(c)$. As f is continuous at $g(c)$, then $\{f(g(x_n))\}$ converges to $f(g(c))$. Thus $f \circ g$ is continuous at c . \square

Example 3.2.8: Claim: $(\sin(1/x))^2$ is a continuous function on $(0, \infty)$.

Proof: First note that $1/x$ is a continuous function on $(0, \infty)$ and $\sin(x)$ is a continuous function on $(0, \infty)$ (actually on all of \mathbb{R} , but $(0, \infty)$ is the range for $1/x$). Hence the composition $\sin(1/x)$ is continuous. We also know that x^2 is continuous on the interval $(-1, 1)$ (the range of \sin). Thus the composition $(\sin(1/x))^2$ is also continuous on $(0, \infty)$.

3.2.3 Discontinuous functions

When f is not continuous at c , we say that f is *discontinuous* at c , or that it has a *discontinuity* at c . If we state the contrapositive of the third item of Proposition 3.2.2 as a separate claim we get an easy to use test for discontinuities.

Proposition 3.2.9. *Let $f: S \rightarrow \mathbb{R}$ be a function. Suppose that for some $c \in S$, there exists a sequence $\{x_n\}$, $x_n \in S$, and $\lim x_n = c$ such that $\{f(x_n)\}$ does not converge to $f(c)$ (or does not converge at all), then f is not continuous at c .*

Example 3.2.10: The function $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by

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

is not continuous at 0.

Proof: Take the sequence $\{-1/n\}$. Then $f(-1/n) = -1$ and so $\lim f(-1/n) = -1$, but $f(0) = 1$.

Example 3.2.11: For an extreme example we take the so-called *Dirichlet function*.

$$f(x) := \begin{cases} 1 & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

The function f is discontinuous at all $c \in \mathbb{R}$.

Proof: Suppose that c is rational. Take a sequence $\{x_n\}$ of irrational numbers such that $\lim x_n = c$ (why can we?). Then $f(x_n) = 0$ and so $\lim f(x_n) = 0$, but $f(c) = 1$. If c is irrational, take a sequence of rational numbers $\{x_n\}$ that converges to c (why can we?). Then $\lim f(x_n) = 1$, but $f(c) = 0$.

As a final example, let us yet again test the limits of your intuition. Can there exist a function that is continuous on all irrational numbers, but discontinuous at all rational numbers? There are rational numbers arbitrarily close to any irrational number. Perhaps strangely, the answer is yes. The following example is called the *Thomae function** or the *popcorn function*.

Example 3.2.12: Let $f: (0, 1) \rightarrow \mathbb{R}$ be defined by

$$f(x) := \begin{cases} 1/k & \text{if } x = m/k \text{ where } m, k \in \mathbb{N} \text{ and } m \text{ and } k \text{ have no common divisors,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Then f is continuous at all irrational $c \in (0, 1)$ and discontinuous at all rational c . See the graph of f in Figure 3.3.

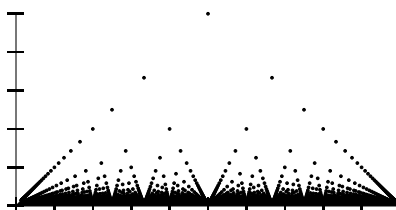


Figure 3.3: Graph of the “popcorn function.”

Proof: Suppose that $c = m/k$ is rational. Take a sequence of irrational numbers $\{x_n\}$ such that $\lim x_n = c$. Then $\lim f(x_n) = \lim 0 = 0$, but $f(c) = 1/k \neq 0$. So f is discontinuous at c .

Now let c be irrational, so $f(c) = 0$. Take a sequence $\{x_n\}$ of rational numbers in $(0, 1)$ such that $\lim x_n = c$. Given $\varepsilon > 0$, find $K \in \mathbb{N}$ such that $1/K < \varepsilon$ by the Archimedean property. If $m/k \in (0, 1)$ is lowest terms (no common divisors), then $m < k$. So there are only finitely many rational numbers in $(0, 1)$ whose denominator k in lowest terms is less than K . Hence there is an M such that for $n \geq M$, all the rational numbers x_n have a denominator larger than or equal to K . Thus for $n \geq M$

$$|f(x_n) - 0| = f(x_n) \leq 1/K < \varepsilon.$$

Therefore f is continuous at irrational c .

*Named after the German mathematician Johannes Karl Thomae (1840 – 1921).

3.2.4 Exercises

Exercise 3.2.1: Using the definition of continuity directly prove that $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) := x^2$ is continuous.

Exercise 3.2.2: Using the definition of continuity directly prove that $f: (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := 1/x$ is continuous.

Exercise 3.2.3: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(x) := \begin{cases} x & \text{if } x \text{ is rational,} \\ x^2 & \text{if } x \text{ is irrational.} \end{cases}$$

Using the definition of continuity directly prove that f is continuous at 1 and discontinuous at 2.

Exercise 3.2.4: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(x) := \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Is f continuous? Prove your assertion.

Exercise 3.2.5: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(x) := \begin{cases} x \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Is f continuous? Prove your assertion.

Exercise 3.2.6: Prove Proposition 3.2.5.

Exercise 3.2.7: Prove the following statement. Let $S \subset \mathbb{R}$ and $A \subset S$. Let $f: S \rightarrow \mathbb{R}$ be a continuous function. Then the restriction $f|_A$ is continuous.

Exercise 3.2.8: Suppose that $S \subset \mathbb{R}$. Suppose that for some $c \in \mathbb{R}$ and $\alpha > 0$, we have $A = (c - \alpha, c + \alpha) \subset S$. Let $f: S \rightarrow \mathbb{R}$ be a function. Prove that if $f|_A$ is continuous at c , then f is continuous at c .

Exercise 3.2.9: Give an example of functions $f: \mathbb{R} \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}$ such that the function h defined by $h(x) := f(x) + g(x)$ is continuous, but f and g are not continuous. Can you find f and g that are nowhere continuous, but h is a continuous function?

Exercise 3.2.10: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}$ be continuous functions. Suppose that for all rational numbers r , $f(r) = g(r)$. Show that $f(x) = g(x)$ for all x .

Exercise 3.2.11: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be continuous. Suppose that $f(c) > 0$. Show that there exists an $\alpha > 0$ such that for all $x \in (c - \alpha, c + \alpha)$ we have $f(x) > 0$.

Exercise 3.2.12: Let $f: \mathbb{Z} \rightarrow \mathbb{R}$ be a function. Show that f is continuous.

Exercise 3.2.13: Let $f: S \rightarrow \mathbb{R}$ be a function and $c \in S$, such that for every sequence $\{x_n\}$ in S with $\lim x_n = c$, the sequence $\{f(x_n)\}$ converges. Show that f is continuous at c .

3.3 Min-max and intermediate value theorems

Note: 1.5 lectures

Let us now state and prove some very important results about continuous functions defined on the real line. In particular, on closed bounded intervals of the real line.

3.3.1 Min-max theorem

Recall that a function $f: [a, b] \rightarrow \mathbb{R}$ is *bounded* if there exists a $B \in \mathbb{R}$ such that $|f(x)| \leq B$ for all $x \in [a, b]$. We have the following lemma.

Lemma 3.3.1. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then f is bounded.*

Proof. Let us prove this claim by contrapositive. Suppose that f is not bounded, then for each $n \in \mathbb{N}$, there is an $x_n \in [a, b]$, such that

$$|f(x_n)| \geq n.$$

Now $\{x_n\}$ is a bounded sequence as $a \leq x_n \leq b$. By the Bolzano-Weierstrass theorem, there is a convergent subsequence $\{x_{n_i}\}$. Let $x := \lim x_{n_i}$. Since $a \leq x_{n_i} \leq b$ for all i , then $a \leq x \leq b$. The limit $\lim f(x_{n_i})$ does not exist as the sequence is not bounded as $|f(x_{n_i})| \geq n_i \geq i$. On the other hand $f(x)$ is a finite number and

$$f(x) = f\left(\lim_{i \rightarrow \infty} x_{n_i}\right).$$

Thus f is not continuous at x . □

In fact, for a continuous f , we will see that the minimum and the maximum are actually achieved. Recall from calculus that $f: S \rightarrow \mathbb{R}$ achieves an *absolute minimum* at $c \in S$ if

$$f(x) \geq f(c) \quad \text{for all } x \in S.$$

On the other hand, f achieves an *absolute maximum* at $c \in S$ if

$$f(x) \leq f(c) \quad \text{for all } x \in S.$$

We say that f achieves an absolute minimum or an absolute maximum on S if such a $c \in S$ exists. If S is a closed and bounded interval, then a continuous f must have an absolute minimum and an absolute maximum on S .

Theorem 3.3.2 (Minimum-maximum theorem). *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then f achieves both an absolute minimum and an absolute maximum on $[a, b]$.*

Proof. We have shown that f is bounded by the lemma. Therefore, the set $f([a, b]) = \{f(x) : x \in [a, b]\}$ has a supremum and an infimum. From what we know about suprema and infima, there exist sequences in the set $f([a, b])$ that approach them. That is, there are sequences $\{f(x_n)\}$ and $\{f(y_n)\}$, where x_n, y_n are in $[a, b]$, such that

$$\lim_{n \rightarrow \infty} f(x_n) = \inf f([a, b]) \quad \text{and} \quad \lim_{n \rightarrow \infty} f(y_n) = \sup f([a, b]).$$

We are not done yet, we need to find where the minimum and the maxima are. The problem is that the sequences $\{x_n\}$ and $\{y_n\}$ need not converge. We know that $\{x_n\}$ and $\{y_n\}$ are bounded (their elements belong to a bounded interval $[a, b]$). We apply the Bolzano-Weierstrass theorem. Hence there exist convergent subsequences $\{x_{n_i}\}$ and $\{y_{m_i}\}$. Let

$$x := \lim_{i \rightarrow \infty} x_{n_i} \quad \text{and} \quad y := \lim_{i \rightarrow \infty} y_{m_i}.$$

Then as $a \leq x_{n_i} \leq b$, we have that $a \leq x \leq b$. Similarly $a \leq y \leq b$, so x and y are in $[a, b]$. We apply that a limit of a subsequence is the same as the limit of the sequence, and we apply the continuity of f to obtain

$$\inf f([a, b]) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{i \rightarrow \infty} f(x_{n_i}) = f\left(\lim_{i \rightarrow \infty} x_{n_i}\right) = f(x).$$

Similarly,

$$\sup f([a, b]) = \lim_{n \rightarrow \infty} f(m_n) = \lim_{i \rightarrow \infty} f(y_{m_i}) = f\left(\lim_{i \rightarrow \infty} y_{m_i}\right) = f(y).$$

Therefore, f achieves an absolute minimum at x and f achieves an absolute maximum at y . \square

Example 3.3.3: The function $f(x) := x^2 + 1$ defined on the interval $[-1, 2]$ achieves a minimum at $x = 0$ when $f(0) = 1$. It achieves a maximum at $x = 2$ where $f(2) = 5$. Do note that the domain of definition matters. If we instead took the domain to be $[-10, 10]$, then $x = 2$ would no longer be a maximum of f . Instead the maximum would be achieved at either $x = 10$ or $x = -10$.

Let us show by examples that the different hypotheses of the theorem are truly needed.

Example 3.3.4: The function $f(x) := x$, defined on the whole real line, achieves neither a minimum, nor a maximum. So it is important that we are looking at a bounded interval.

Example 3.3.5: The function $f(x) := 1/x$, defined on $(0, 1)$ achieves neither a minimum, nor a maximum. The values of the function are unbounded as we approach 0. Also as we approach $x = 1$, the values of the function approach 1, but $f(x) > 1$ for all $x \in (0, 1)$. There is no $x \in (0, 1)$ such that $f(x) = 1$. So it is important that we are looking at a closed interval.

Example 3.3.6: Continuity is important. Define $f: [0, 1] \rightarrow \mathbb{R}$ by $f(x) := 1/x$ for $x > 0$ and let $f(0) := 0$. Then the function does not achieve a maximum. The problem is that the function is not continuous at 0.

3.3.2 Bolzano's intermediate value theorem

Bolzano's intermediate value theorem is one of the cornerstones of analysis. It is sometimes called only intermediate value theorem, or just Bolzano's theorem. To prove Bolzano's theorem we prove the following simpler lemma.

Lemma 3.3.7. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Suppose that $f(a) < 0$ and $f(b) > 0$. Then there exists a number $c \in (a, b)$ such that $f(c) = 0$.*

Proof. We define two sequences $\{a_n\}$ and $\{b_n\}$ inductively:

- (i) Let $a_1 := a$ and $b_1 := b$.
- (ii) If $f\left(\frac{a_n + b_n}{2}\right) \geq 0$, let $a_{n+1} := a_n$ and $b_{n+1} := \frac{a_n + b_n}{2}$.
- (iii) If $f\left(\frac{a_n + b_n}{2}\right) < 0$, let $a_{n+1} := \frac{a_n + b_n}{2}$ and $b_{n+1} := b_n$.

From the definition of the two sequences it is obvious that if $a_n < b_n$, then $a_{n+1} < b_{n+1}$. Thus by induction $a_n < b_n$ for all n . Furthermore, $a_n \leq a_{n+1}$ and $b_n \geq b_{n+1}$ for all n . Finally we notice that

$$b_{n+1} - a_{n+1} = \frac{b_n - a_n}{2}.$$

By induction we see that

$$b_n - a_n = \frac{b_1 - a_1}{2^{n-1}} = 2^{1-n}(b - a).$$

As $\{a_n\}$ and $\{b_n\}$ are monotone and bounded, they converge. Let $c := \lim a_n$ and $d := \lim b_n$. As $a_n < b_n$ for all n , then $c \leq d$. Furthermore, as a_n is increasing and b_n is decreasing, c is the supremum of a_n and d is the infimum of the b_n . Thus $d - c \leq b_n - a_n$ for all n . So

$$|d - c| = d - c \leq b_n - a_n = 2^{1-n}(b - a)$$

for all n . As $2^{1-n}(b - a) \rightarrow 0$ as $n \rightarrow \infty$, we see that $c = d$. By construction, for all n we have

$$f(a_n) < 0 \quad \text{and} \quad f(b_n) \geq 0.$$

We use the fact that $\lim a_n = \lim b_n = c$ and the continuity of f to take limits in those inequalities to get

$$f(c) = \lim f(a_n) \leq 0 \quad \text{and} \quad f(c) = \lim f(b_n) \geq 0.$$

As $f(c) \geq 0$ and $f(c) \leq 0$, we conclude $f(c) = 0$. Obviously, $a < c < b$. □

Notice that the proof tells us how to find the c . The proof is not only useful for us pure mathematicians, but it is a useful idea in applied mathematics.

Theorem 3.3.8 (Bolzano's intermediate value theorem). *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Suppose that there exists a y such that $f(a) < y < f(b)$ or $f(a) > y > f(b)$. Then there exists a $c \in (a, b)$ such that $f(c) = y$.*

The theorem says that a continuous function on a closed interval achieves all the values between the values at the endpoints.

Proof. If $f(a) < y < f(b)$, then define $g(x) := f(x) - y$. Then we see that $g(a) < 0$ and $g(b) > 0$ and we can apply Lemma 3.3.7 to g . If $g(c) = 0$, then $f(c) = y$.

Similarly if $f(a) > y > f(b)$, then define $g(x) := y - f(x)$. Then again $g(a) < 0$ and $g(b) > 0$ and we can apply Lemma 3.3.7. Again if $g(c) = 0$, then $f(c) = y$. \square

If a function is continuous, then the restriction to a subset is continuous. So if $f: S \rightarrow \mathbb{R}$ is continuous and $[a, b] \subset S$, then $f|_{[a, b]}$ is also continuous. Hence, we generally apply the theorem to a function continuous on some large set S , but we restrict attention to an interval.

Example 3.3.9: The polynomial $f(x) := x^3 - 2x^2 + x - 1$ has a real root in $(1, 2)$. We simply notice that $f(1) = -1$ and $f(2) = 1$. Hence there must exist a point $c \in (1, 2)$ such that $f(c) = 0$. To find a better approximation of the root we could follow the proof of Lemma 3.3.7. For example, next we would look at 1.5 and find that $f(1.5) = -0.625$. Therefore, there is a root of the equation in $(1.5, 2)$. Next we look at 1.75 and note that $f(1.75) \approx -0.016$. Hence there is a root of f in $(1.75, 2)$. Next we look at 1.875 and find that $f(1.875) \approx 0.44$, thus there is a root in $(1.75, 1.875)$. We follow this procedure until we gain sufficient precision.

The technique above is the simplest method of finding roots of polynomials. Finding roots of polynomials is perhaps the most common problem in applied mathematics. In general it is hard to do quickly, precisely and automatically. We can use the intermediate value theorem to find roots for any continuous function, not just a polynomial.

There are better and faster methods of finding roots of equations, such as Newton's method. One advantage of the above method is its simplicity. Another advantage is that the moment we find an initial interval where the intermediate value theorem applies, we are guaranteed that we will find a root up to a desired precision in finitely many steps.

The theorem guarantees a single c such that $f(c) = y$, but there may be many different roots of the equation $f(c) = y$. If we follow the procedure of the proof, we are guaranteed to find approximations to one such root. We need to work harder to find any other roots.

Polynomials of even degree may not have any real roots. For example, there is no real number x such that $x^2 + 1 = 0$. Odd polynomials, on the other hand, always have at least one real root.

Proposition 3.3.10. *Let $f(x)$ be a polynomial of odd degree. Then f has a real root.*

Proof. Suppose f is a polynomial of odd degree d . We write

$$f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0,$$

where $a_d \neq 0$. We divide by a_d to obtain a polynomial

$$g(x) := x^d + b_{d-1} x^{d-1} + \cdots + b_1 x + b_0,$$

where $b_k = a_k/a_d$. Let us show that $g(n)$ is positive for some large $n \in \mathbb{N}$. So

$$\begin{aligned} \left| \frac{b_{d-1}n^{d-1} + \cdots + b_1n + b_0}{n^d} \right| &= \frac{|b_{d-1}n^{d-1} + \cdots + b_1n + b_0|}{n^d} \\ &\leq \frac{|b_{d-1}|n^{d-1} + \cdots + |b_1|n + |b_0|}{n^d} \\ &\leq \frac{|b_{d-1}|n^{d-1} + \cdots + |b_1|n^{d-1} + |b_0|n^{d-1}}{n^d} \\ &= \frac{n^{d-1}(|b_{d-1}| + \cdots + |b_1| + |b_0|)}{n^d} \\ &= \frac{1}{n}(|b_{d-1}| + \cdots + |b_1| + |b_0|). \end{aligned}$$

Therefore

$$\lim_{n \rightarrow \infty} \frac{b_{d-1}n^{d-1} + \cdots + b_1n + b_0}{n^d} = 0.$$

Thus there exists an $M \in \mathbb{N}$ such that

$$\left| \frac{b_{d-1}M^{d-1} + \cdots + b_1M + b_0}{M^d} \right| < 1,$$

which implies

$$-(b_{d-1}M^{d-1} + \cdots + b_1M + b_0) < M^d.$$

Therefore $g(M) > 0$.

Next we look at $g(-n)$ for $n \in \mathbb{N}$. By a similar argument (exercise) we find that there exists some $K \in \mathbb{N}$ such that $b_{d-1}(-K)^{d-1} + \cdots + b_1(-K) + b_0 < K^d$ and therefore $g(-K) < 0$ (why?). In the proof make sure you use the fact that d is odd. In particular, if d is odd then $(-n)^d = -(n^d)$.

We appeal to the intermediate value theorem, to find a $c \in [-K, M]$ such that $g(c) = 0$. As $g(x) = \frac{f(x)}{a_d}$, we see that $f(c) = 0$, and the proof is done. \square

Example 3.3.11: An interesting fact is that there do exist discontinuous functions that have the intermediate value property. The function

$$f(x) := \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is not continuous at 0, however, it has the intermediate value property. That is, for any $a < b$, and any y such that $f(a) < y < f(b)$ or $f(a) > y > f(b)$, there exists a c such that $f(y) = c$. Proof is left as an exercise.

3.3.3 Exercises

Exercise 3.3.1: Find an example of a discontinuous function $f: [0, 1] \rightarrow \mathbb{R}$ where the intermediate value theorem fails.

Exercise 3.3.2: Find an example of a bounded discontinuous function $f: [0, 1] \rightarrow \mathbb{R}$ that has neither an absolute minimum nor an absolute maximum.

Exercise 3.3.3: Let $f: (0, 1) \rightarrow \mathbb{R}$ be a continuous function such that $\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 1} f(x) = 0$. Show that f achieves either an absolute minimum or an absolute maximum on $(0, 1)$ (but perhaps not both).

Exercise 3.3.4: Let

$$f(x) := \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Show that f has the intermediate value property. That is, for any $a < b$, if there exists a y such that $f(a) < y < f(b)$ or $f(a) > y > f(b)$, then there exists a $c \in (a, b)$ such that $f(c) = y$.

Exercise 3.3.5: Suppose that $g(x)$ is a polynomial of odd degree d such that

$$g(x) = x^d + b_{d-1}x^{d-1} + \cdots + b_1x + b_0,$$

for some real numbers b_0, b_1, \dots, b_{d-1} . Show that there exists a $K \in \mathbb{N}$ such that $g(-K) < 0$. Hint: Make sure to use the fact that $(-n)^d = -(n^d)$.

Exercise 3.3.6: Suppose that $g(x)$ is a polynomial of even degree d such that

$$g(x) = x^d + b_{d-1}x^{d-1} + \cdots + b_1x + b_0,$$

for some real numbers b_0, b_1, \dots, b_{d-1} . Suppose that $g(0) < 0$. Show that g has at least two distinct real roots.

Exercise 3.3.7: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is a continuous function. Prove that the direct image $f([a, b])$ is a closed and bounded interval or a single number.

Exercise 3.3.8: Suppose that $f: \mathbb{R} \rightarrow \mathbb{R}$ is continuous and periodic with period $P > 0$. That is, $f(x + P) = f(x)$ for all $x \in \mathbb{R}$. Show that f achieves an absolute minimum and an absolute maximum.

Exercise 3.3.9 (Challenging): Suppose that $f(x)$ is a bounded polynomial, that is, there is an M such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$. Prove that f must be a constant.

3.4 Uniform continuity

Note: 1.5–2 lectures (Continuous extension and Lipschitz can be optional)

3.4.1 Uniform continuity

We made a fuss of saying that the δ in the definition of continuity depended on the point c . There are situations when it is advantageous to have a δ independent of any point. Let us give a name to this concept.

Definition 3.4.1. Let $S \subset \mathbb{R}$, and let $f: S \rightarrow \mathbb{R}$ be a function. Suppose that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $x, c \in S$ and $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$. Then we say f is *uniformly continuous*.

It is not hard to see that a uniformly continuous function must be continuous. The only difference in the definitions is that for a given $\varepsilon > 0$ we pick a $\delta > 0$ that works for all $c \in S$. That is, δ can no longer depend on c , it only depends on ε . The domain of definition of the function makes a difference now. A function that is not uniformly continuous on a larger set, may be uniformly continuous when restricted to a smaller set.

Example 3.4.2: The function $f: (0, 1) \rightarrow \mathbb{R}$, defined by $f(x) := 1/x$ is not uniformly continuous, but it is continuous.

Proof: Given $\varepsilon > 0$, then for $\varepsilon > |1/x - 1/y|$ to hold we must have

$$\varepsilon > |1/x - 1/y| = \frac{|y - x|}{|xy|} = \frac{|y - x|}{xy},$$

or

$$|x - y| < xy\varepsilon.$$

Therefore, to satisfy the definition of uniform continuity we would have to have $\delta \leq xy\varepsilon$ for all x, y in $(0, 1)$, but that would mean that $\delta \leq 0$. Therefore there is no single $\delta > 0$.

Example 3.4.3: $f: [0, 1] \rightarrow \mathbb{R}$, defined by $f(x) := x^2$ is uniformly continuous.

Proof: Note that $0 \leq x, c \leq 1$. Then

$$|x^2 - c^2| = |x + c||x - c| \leq (|x| + |c|)|x - c| \leq (1 + 1)|x - c|.$$

Therefore given $\varepsilon > 0$, let $\delta := \varepsilon/2$. If $|x - c| < \delta$, then $|x^2 - c^2| < \varepsilon$.

On the other hand, $f: \mathbb{R} \rightarrow \mathbb{R}$, defined by $f(x) := x^2$ is not uniformly continuous.

Proof: Suppose it is, then for all $\varepsilon > 0$, there would exist a $\delta > 0$ such that if $|x - c| < \delta$, then $|x^2 - c^2| < \varepsilon$. Take $x > 0$ and let $c := x + \delta/2$. Write

$$\varepsilon \geq |x^2 - c^2| = |x + c||x - c| = (2x + \delta/2)\delta/2 \geq \delta x.$$

Therefore $x \leq \varepsilon/\delta$ for all $x > 0$, which is a contradiction.

We have seen that if f is defined on an interval that is either not closed or not bounded, then f can be continuous, but not uniformly continuous. For a closed and bounded interval $[a, b]$, we can, however, make the following statement.

Theorem 3.4.4. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then f is uniformly continuous.*

Proof. We prove the statement by contrapositive. Suppose that f is not uniformly continuous. We will prove that there is some $c \in [a, b]$ where f is not continuous. Let us negate the definition of uniformly continuous. There exists an $\varepsilon > 0$ such that for every $\delta > 0$, there exist points x, y in S with $|x - y| < \delta$ and $|f(x) - f(y)| \geq \varepsilon$.

So for the $\varepsilon > 0$ above, we can find sequences $\{x_n\}$ and $\{y_n\}$ such that $|x_n - y_n| < 1/n$ and such that $|f(x_n) - f(y_n)| \geq \varepsilon$. By Bolzano-Weierstrass, there exists a convergent subsequence $\{x_{n_k}\}$. Let $c := \lim x_{n_k}$. Note that as $a \leq x_{n_k} \leq b$, then $a \leq c \leq b$. Write

$$|c - y_{n_k}| = |c - x_{n_k} + x_{n_k} - y_{n_k}| \leq |c - x_{n_k}| + |x_{n_k} - y_{n_k}| < |c - x_{n_k}| + 1/n_k.$$

As $|c - x_{n_k}|$ and $1/n_k$ go to zero when k goes to infinity, we see that $\{y_{n_k}\}$ converges and the limit is c . We now show that f is not continuous at c . We estimate

$$\begin{aligned} |f(c) - f(x_{n_k})| &= |f(c) - f(y_{n_k}) + f(y_{n_k}) - f(x_{n_k})| \\ &\geq |f(y_{n_k}) - f(x_{n_k})| - |f(c) - f(y_{n_k})| \\ &\geq \varepsilon - |f(c) - f(y_{n_k})|. \end{aligned}$$

Or in other words

$$|f(c) - f(x_{n_k})| + |f(c) - f(y_{n_k})| \geq \varepsilon.$$

At least one of the sequences $\{f(x_{n_k})\}$ or $\{f(y_{n_k})\}$ cannot converge to $f(c)$, otherwise the left hand side of the inequality would go to zero while the right-hand side is positive. Thus f cannot be continuous at c . \square

3.4.2 Continuous extension

Before we get to continuous extension, we show the following useful lemma. It says that uniformly continuous functions behave nicely with respect to Cauchy sequences. The new issue here is that for a Cauchy sequence we no longer know where the limit ends up; it may not end up in the domain of the function.

Lemma 3.4.5. *Let $f: S \rightarrow \mathbb{R}$ be a uniformly continuous function. Let $\{x_n\}$ be a Cauchy sequence in S . Then $\{f(x_n)\}$ is Cauchy.*

Proof. Let $\varepsilon > 0$ be given. Then there is a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. Now find an $M \in \mathbb{N}$ such that for all $n, k \geq M$ we have $|x_n - x_k| < \delta$. Then for all $n, k \geq M$ we have $|f(x_n) - f(x_k)| < \varepsilon$. \square

An application of the above lemma is the following theorem. It says that a function on an open interval is uniformly continuous if and only if it can be extended to a continuous function on the closed interval.

Theorem 3.4.6. *A function $f: (a, b) \rightarrow \mathbb{R}$ is uniformly continuous if and only if the limits*

$$L_a := \lim_{x \rightarrow a} f(x) \quad \text{and} \quad L_b := \lim_{x \rightarrow b} f(x)$$

exist and the function $\tilde{f}: [a, b] \rightarrow \mathbb{R}$ defined by

$$\tilde{f}(x) := \begin{cases} f(x) & \text{if } x \in (a, b), \\ L_a & \text{if } x = a, \\ L_b & \text{if } x = b, \end{cases}$$

is continuous.

Proof. One direction is not hard to prove. If \tilde{f} is continuous, then it is uniformly continuous by Theorem 3.4.4. As f is the restriction of \tilde{f} to (a, b) , then f is also uniformly continuous (easy exercise).

Now suppose that f is uniformly continuous. We must first show that the limits L_a and L_b exist. Let us concentrate on L_a . Take a sequence $\{x_n\}$ in (a, b) such that $\lim x_n = a$. The sequence is a Cauchy sequence and hence by Lemma 3.4.5, the sequence $\{f(x_n)\}$ is Cauchy and therefore convergent. We have some number $L_1 := \lim f(x_n)$. Take another sequence $\{y_n\}$ in (a, b) such that $\lim y_n = a$. By the same reasoning we get $L_2 := \lim f(y_n)$. If we show that $L_1 = L_2$, then the limit $L_a = \lim_{x \rightarrow a} f(x)$ exists. Let $\varepsilon > 0$ be given, find $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon/3$. Find $M \in \mathbb{N}$ such that for $n \geq M$ we have $|a - x_n| < \delta/2$, $|a - y_n| < \delta/2$, $|f(x_n) - L_1| < \varepsilon/3$, and $|f(y_n) - L_2| < \varepsilon/3$. Then for $n \geq M$ we have

$$|x_n - y_n| = |x_n - a + a - y_n| \leq |x_n - a| + |a - y_n| < \delta/2 + \delta/2 = \delta.$$

So

$$\begin{aligned} |L_1 - L_2| &= |L_1 - f(x_n) + f(x_n) - f(y_n) + f(y_n) - L_2| \\ &\leq |L_1 - f(x_n)| + |f(x_n) - f(y_n)| + |f(y_n) - L_2| \\ &\leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

Therefore $L_1 = L_2$. Thus L_a exists. To show that L_b exists is left as an exercise.

Now that we know that the limits L_a and L_b exist, we are done. If $\lim_{x \rightarrow a} f(x)$ exists, then $\lim_{x \rightarrow a} \tilde{f}(x)$ exists (See Proposition 3.1.14). Similarly with L_b . Hence \tilde{f} is continuous at a and b . And since f is continuous at $c \in (a, b)$, then \tilde{f} is continuous at $c \in (a, b)$. \square

3.4.3 Lipschitz continuous functions

Definition 3.4.7. Let $f: S \rightarrow \mathbb{R}$ be a function such that there exists a number K such that for all x and y in S we have

$$|f(x) - f(y)| \leq K|x - y|.$$

Then f is said to be *Lipschitz continuous**.

A large class of functions is Lipschitz continuous. Be careful, just as for uniformly continuous functions, the domain of definition of the function is important. See the examples below and the exercises. First we justify the use of the word *continuous*.

Proposition 3.4.8. A Lipschitz continuous function is uniformly continuous.

Proof. Let $f: S \rightarrow \mathbb{R}$ be a function and let K be a constant such that for all x, y in S we have $|f(x) - f(y)| \leq K|x - y|$.

Let $\varepsilon > 0$ be given. Take $\delta := \varepsilon/K$. For any x and y in S such that $|x - y| < \delta$ we have that

$$|f(x) - f(y)| \leq K|x - y| < K\delta = K \frac{\varepsilon}{K} = \varepsilon.$$

Therefore f is uniformly continuous. □

We can interpret Lipschitz continuity geometrically. If f is a Lipschitz continuous function with some constant K . We rewrite the inequality to say that for $x \neq y$ we have

$$\left| \frac{f(x) - f(y)}{x - y} \right| \leq K.$$

The quantity $\frac{f(x) - f(y)}{x - y}$ is the slope of the line between the points $(x, f(x))$ and $(y, f(y))$. Therefore, f is Lipschitz continuous if and only if every line that intersects the graph of f in at least two distinct points has slope less than or equal to K .

Example 3.4.9: The functions $\sin(x)$ and $\cos(x)$ are Lipschitz continuous. We have seen (Example 3.2.6) the following two inequalities.

$$|\sin(x) - \sin(y)| \leq |x - y| \quad \text{and} \quad |\cos(x) - \cos(y)| \leq |x - y|.$$

Hence \sin and \cos are Lipschitz continuous with $K = 1$.

Example 3.4.10: The function $f: [1, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := \sqrt{x}$ is Lipschitz continuous. *Proof:*

$$|\sqrt{x} - \sqrt{y}| = \left| \frac{x - y}{\sqrt{x} + \sqrt{y}} \right| = \frac{|x - y|}{\sqrt{x} + \sqrt{y}}.$$

*Named after the German mathematician Rudolf Otto Sigismund Lipschitz (1832–1903).

As $x \geq 1$ and $y \geq 1$, we see that $\frac{1}{\sqrt{x} + \sqrt{y}} \leq \frac{1}{2}$. Therefore

$$|\sqrt{x} - \sqrt{y}| = \left| \frac{x - y}{\sqrt{x} + \sqrt{y}} \right| \leq \frac{1}{2} |x - y|.$$

On the other hand $f: [0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := \sqrt{x}$ is not Lipschitz continuous. Let us see why: Suppose that we have

$$|\sqrt{x} - \sqrt{y}| \leq K|x - y|,$$

for some K . Let $y = 0$ to obtain $\sqrt{x} \leq Kx$. If $K > 0$, then for $x > 0$ we then get $1/K \leq \sqrt{x}$. This cannot possibly be true for all $x > 0$. Thus no such $K > 0$ can exist and f is not Lipschitz continuous.

The last example is a function that is uniformly continuous but not Lipschitz continuous. To see that \sqrt{x} is uniformly continuous on $[0, \infty)$ note that it is uniformly continuous on $[0, 1]$ by Theorem 3.4.4. It is also Lipschitz (and therefore uniformly continuous) on $[1, \infty)$. It is not hard (exercise) to show that this means that \sqrt{x} is uniformly continuous on $[0, \infty)$.

3.4.4 Exercises

Exercise 3.4.1: Let $f: S \rightarrow \mathbb{R}$ be uniformly continuous. Let $A \subset S$. Then the restriction $f|_A$ is uniformly continuous.

Exercise 3.4.2: Let $f: (a, b) \rightarrow \mathbb{R}$ be a uniformly continuous function. Finish the proof of Theorem 3.4.6 by showing that the limit $\lim_{x \rightarrow b} f(x)$ exists.

Exercise 3.4.3: Show that $f: (c, \infty) \rightarrow \mathbb{R}$ for some $c > 0$ and defined by $f(x) := 1/x$ is Lipschitz continuous.

Exercise 3.4.4: Show that $f: (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := 1/x$ is not Lipschitz continuous.

Exercise 3.4.5: Let A, B be intervals. Let $f: A \rightarrow \mathbb{R}$ and $g: B \rightarrow \mathbb{R}$ be uniformly continuous functions such that $f(x) = g(x)$ for $x \in A \cap B$. Define the function $h: A \cup B \rightarrow \mathbb{R}$ by $h(x) := f(x)$ if $x \in A$ and $h(x) := g(x)$ if $x \in B \setminus A$. a) Prove that if $A \cap B \neq \emptyset$, then h is uniformly continuous. b) Find an example where $A \cap B = \emptyset$ and h is not even continuous.

Exercise 3.4.6: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a polynomial of degree $d \geq 2$. Show that f is not Lipschitz continuous.

Exercise 3.4.7: Let $f: (0, 1) \rightarrow \mathbb{R}$ be a bounded continuous function. Show that the function $g(x) := x(1-x)f(x)$ is uniformly continuous.

Exercise 3.4.8: Show that $f: (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) := \sin(1/x)$ is not uniformly continuous.

Exercise 3.4.9 (Challenging): Let $f: \mathbb{Q} \rightarrow \mathbb{R}$ be a uniformly continuous function. Show that there exists a uniformly continuous function $\tilde{f}: \mathbb{R} \rightarrow \mathbb{R}$ such that $f(x) = \tilde{f}(x)$ for all $x \in \mathbb{Q}$.

Chapter 4

The Derivative

4.1 The derivative

Note: 1 lecture

The idea of a derivative is the following. Let us suppose that a graph of a function looks locally like a straight line. We can then talk about the slope of this line. The slope tells us how fast is the value of the function changing at the particular point. Of course, we are leaving out any function that has corners or discontinuities. Let us be precise.

4.1.1 Definition and basic properties

Definition 4.1.1. Let I be an interval, let $f: I \rightarrow \mathbb{R}$ be a function, and let $c \in I$. If the limit

$$L := \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}$$

exists, then we say f is *differentiable* at c , that L is the *derivative* of f at c , and write $f'(c) := L$.

If f is differentiable at all $c \in I$, then we simply say that f is *differentiable*, and then we obtain a function $f': I \rightarrow \mathbb{R}$.

The expression $\frac{f(x) - f(c)}{x - c}$ is called the *difference quotient*.

The graphical interpretation of the derivative is depicted in Figure 4.1. The left-hand plot gives the line through $(c, f(c))$ and $(x, f(x))$ with slope $\frac{f(x) - f(c)}{x - c}$. When we take the limit as x goes to c , we get the right-hand plot, where we see that the derivative of the function at the point c is the slope of the line tangent to the graph of f at the point $(c, f(c))$.

Note that we allow I to be a closed interval and we allow c to be an endpoint of I . Some calculus books do not allow c to be an endpoint of an interval, but all the theory still works by allowing it, and it makes our work easier.

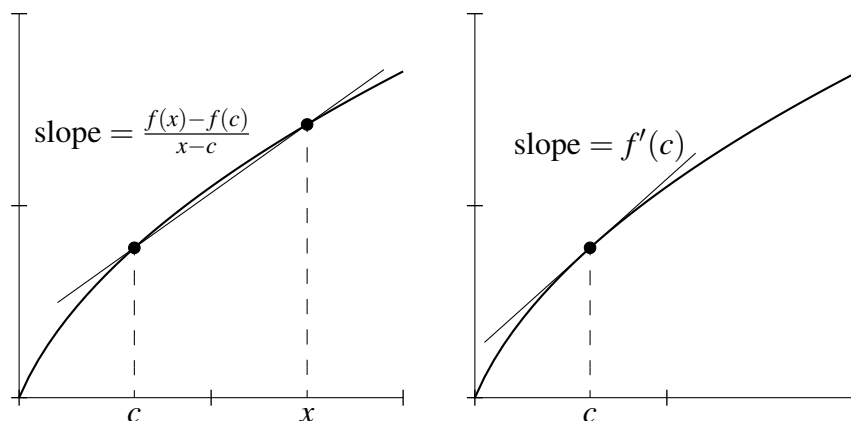


Figure 4.1: Graphical interpretation of the derivative.

Example 4.1.2: Let $f(x) := x^2$ defined on the whole real line. We find that

$$\lim_{x \rightarrow c} \frac{x^2 - c^2}{x - c} = \lim_{x \rightarrow c} \frac{(x + c)(x - c)}{x - c} = \lim_{x \rightarrow c} (x + c) = 2c.$$

Therefore $f'(c) = 2c$.

Example 4.1.3: The function $f(x) := |x|$ is not differentiable at the origin. When $x > 0$, then

$$\frac{|x| - |0|}{x - 0} = \frac{x - 0}{x - 0} = 1,$$

and when $x < 0$ we have

$$\frac{|x| - |0|}{x - 0} = \frac{-x - 0}{x - 0} = -1.$$

A famous example of Weierstrass shows that there exists a continuous function that is not differentiable at *any* point. The construction of this function is beyond the scope of this book. On the other hand, a differentiable function is always continuous.

Proposition 4.1.4. Let $f: I \rightarrow \mathbb{R}$ be differentiable at $c \in I$, then it is continuous at c .

Proof. We know that the limits

$$\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} = f'(c) \quad \text{and} \quad \lim_{x \rightarrow c} (x - c) = 0$$

exist. Furthermore,

$$f(x) - f(c) = \left(\frac{f(x) - f(c)}{x - c} \right) (x - c).$$

Therefore the limit of $f(x) - f(c)$ exists and

$$\lim_{x \rightarrow c} (f(x) - f(c)) = \left(\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} \right) \left(\lim_{x \rightarrow c} (x - c) \right) = f'(c) \cdot 0 = 0.$$

Hence $\lim_{x \rightarrow c} f(x) = f(c)$, and f is continuous at c . \square

An important property of the derivative is linearity. The derivative is the approximation of a function by a straight line. The slope of a line through two points changes linearly when the y -coordinates are changed linearly. By taking the limit, it makes sense that the derivative is linear.

Proposition 4.1.5. *Let I be an interval, let $f: I \rightarrow \mathbb{R}$ and $g: I \rightarrow \mathbb{R}$ be differentiable at $c \in I$, and let $\alpha \in \mathbb{R}$.*

(i) *Define $h: I \rightarrow \mathbb{R}$ by $h(x) := \alpha f(x)$. Then h is differentiable at c and $h'(c) = \alpha f'(c)$.*

(ii) *Define $h: I \rightarrow \mathbb{R}$ by $h(x) := f(x) + g(x)$. Then h is differentiable at c and $h'(c) = f'(c) + g'(c)$.*

Proof. First, let $h(x) = \alpha f(x)$. For $x \in I$, $x \neq c$ we have

$$\frac{h(x) - h(c)}{x - c} = \frac{\alpha f(x) - \alpha f(c)}{x - c} = \alpha \frac{f(x) - f(c)}{x - c}.$$

The limit as x goes to c exists on the right by Corollary 3.1.12. We get

$$\lim_{x \rightarrow c} \frac{h(x) - h(c)}{x - c} = \alpha \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}.$$

Therefore h is differentiable at c , and the derivative is computed as given.

Next, define $h(x) := f(x) + g(x)$. For $x \in I$, $x \neq c$ we have

$$\frac{h(x) - h(c)}{x - c} = \frac{(f(x) + g(x)) - (f(c) + g(c))}{x - c} = \frac{f(x) - f(c)}{x - c} + \frac{g(x) - g(c)}{x - c}.$$

The limit as x goes to c exists on the right by Corollary 3.1.12. We get

$$\lim_{x \rightarrow c} \frac{h(x) - h(c)}{x - c} = \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} + \lim_{x \rightarrow c} \frac{g(x) - g(c)}{x - c}.$$

Therefore h is differentiable at c and the derivative is computed as given. \square

It is not true that the derivative of a multiple of two functions is the multiple of the derivatives. Instead we get the so-called *product rule* or the *Leibniz rule**.

*Named for the German mathematician Gottfried Wilhelm Leibniz (1646–1716).

Proposition 4.1.6 (Product rule). *Let I be an interval, let $f: I \rightarrow \mathbb{R}$ and $g: I \rightarrow \mathbb{R}$ be functions differentiable at c . If $h: I \rightarrow \mathbb{R}$ is defined by*

$$h(x) := f(x)g(x),$$

then h is differentiable at c and

$$h'(c) = f(c)g'(c) + f'(c)g(c).$$

The proof of the product rule is left as an exercise. The key is to use the identity $f(x)g(x) - f(c)g(c) = f(x)(g(x) - g(c)) + g(c)(f(x) - f(c))$.

Proposition 4.1.7 (Quotient Rule). *Let I be an interval, let $f: I \rightarrow \mathbb{R}$ and $g: I \rightarrow \mathbb{R}$ be differentiable at c and $g(x) \neq 0$ for all $x \in I$. If $h: I \rightarrow \mathbb{R}$ is defined by*

$$h(x) := \frac{f(x)}{g(x)},$$

then h is differentiable at c and

$$h'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{(g(c))^2}.$$

Again the proof is left as an exercise.

4.1.2 Chain rule

A useful rule for computing derivatives is the chain rule.

Proposition 4.1.8 (Chain Rule). *Let I_1, I_2 be intervals, let $g: I_1 \rightarrow I_2$ be differentiable at $c \in I_1$, and $f: I_2 \rightarrow \mathbb{R}$ be differentiable at $g(c)$. If $h: I_1 \rightarrow \mathbb{R}$ is defined by*

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

then h is differentiable at c and

$$h'(c) = f'(g(c))g'(c).$$

Proof. Let $d := g(c)$. Define $u: I_2 \rightarrow \mathbb{R}$ and $v: I_1 \rightarrow \mathbb{R}$ by

$$u(y) := \begin{cases} \frac{f(y)-f(d)}{y-d} & \text{if } y \neq d, \\ f'(d) & \text{if } y = d, \end{cases}$$

$$v(x) := \begin{cases} \frac{g(x)-g(c)}{x-c} & \text{if } x \neq c, \\ g'(c) & \text{if } x = c. \end{cases}$$

We note that

$$f(y) - f(d) = u(y)(y - d) \quad \text{and} \quad g(x) - g(c) = v(x)(x - c).$$

We plug in to obtain

$$h(x) - h(c) = f(g(x)) - f(g(c)) = u(g(x))(g(x) - g(c)) = u(g(x))(v(x)(x - c)).$$

Therefore,

$$\frac{h(x) - h(c)}{x - c} = u(g(x))v(x). \quad (4.1)$$

We compute the limits $\lim_{y \rightarrow d} u(y) = f'(d) = f'(g(c))$ and $\lim_{x \rightarrow c} v(x) = g'(c)$. That is, the functions u and v are continuous at $d = g(c)$ and c respectively. Furthermore the function g is continuous at c . Hence the limit of the right-hand side of (4.1) as x goes to c exists and is equal to $f'(g(c))g'(c)$. Thus h is differentiable at c and the limit is $f'(g(c))g'(c)$. \square

4.1.3 Exercises

Exercise 4.1.1: Prove the product rule. Hint: Use $f(x)g(x) - f(c)g(c) = f(x)(g(x) - g(c)) + g(c)(f(x) - f(c))$.

Exercise 4.1.2: Prove the quotient rule. Hint: You can do this directly, but it may be easier to find the derivative of $1/x$ and then use the chain rule and the product rule.

Exercise 4.1.3: For $n \in \mathbb{Z}$, prove that x^n is differentiable and find the derivative, unless, of course, $n < 0$ and $x = 0$. Hint: Use the product rule.

Exercise 4.1.4: Prove that a polynomial is differentiable and find the derivative. Hint: Use the previous exercise.

Exercise 4.1.5: Let

$$f(x) := \begin{cases} x^2 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases}$$

Prove that f is differentiable at 0, but discontinuous at all points except 0.

Exercise 4.1.6: Assume the inequality $|x - \sin(x)| \leq x^2$. Prove that \sin is differentiable at 0, and find the derivative at 0.

Exercise 4.1.7: Using the previous exercise, prove that \sin is differentiable at all x and that the derivative is $\cos(x)$. Hint: Use the sum-to-product trigonometric identity as we did before.

Exercise 4.1.8: Let $f: I \rightarrow \mathbb{R}$ be differentiable. Given $n \in \mathbb{Z}$, define f^n be the function defined by $f^n(x) := (f(x))^n$. If $n < 0$ assume $f(x) \neq 0$. Prove that $(f^n)'(x) = n(f(x))^{n-1}f'(x)$.

Exercise 4.1.9: Suppose that $f: \mathbb{R} \rightarrow \mathbb{R}$ is a differentiable Lipschitz continuous function. Prove that f' is a bounded function.

Exercise 4.1.10: Let I_1, I_2 be intervals. Let $f: I_1 \rightarrow I_2$ be a bijective function and $g: I_2 \rightarrow I_1$ be the inverse. Suppose that both f is differentiable at $c \in I_1$ and $f'(c) \neq 0$ and g is differentiable at $f(c)$. Use the chain rule to find a formula for $g'(f(c))$ (in terms of $f'(c)$).

Exercise 4.1.11: Suppose that $f: I \rightarrow \mathbb{R}$ is a bounded function and $g: I \rightarrow \mathbb{R}$ is a function differentiable at $c \in I$ and $g(c) = g'(c) = 0$. Show that $h(x) := f(x)g(x)$ is differentiable at c . Hint: Note that you cannot apply the product rule.

4.2 Mean value theorem

Note: 2 lectures (some applications may be skipped)

4.2.1 Relative minima and maxima

We talked about absolute maxima and minima. These are the tallest peaks and lowest valleys in the whole mountain range. We might also want to talk about peaks of individual mountains and valleys.

Definition 4.2.1. Let $S \subset \mathbb{R}$ be a set and let $f: S \rightarrow \mathbb{R}$ be a function. The function f is said to have a *relative maximum* at $c \in S$ if there exists a $\delta > 0$ such that for all $x \in S$ such that $|x - c| < \delta$ we have $f(x) \leq f(c)$. The definition of *relative minimum* is analogous.

Theorem 4.2.2. Let $f: [a, b] \rightarrow \mathbb{R}$ be a function differentiable at $c \in (a, b)$, and c is a relative minimum or a relative maximum of f . Then $f'(c) = 0$.

Proof. We prove the statement for a maximum. For a minimum the statement follows by considering the function $-f$.

Let c be a relative maximum of f . In particular as long as $|x - c| < \delta$ we have $f(x) - f(c) \leq 0$. Then we look at the difference quotient. If $x > c$ we note that

$$\frac{f(x) - f(c)}{x - c} \leq 0,$$

and if $x < c$ we have

$$\frac{f(x) - f(c)}{x - c} \geq 0.$$

We now take sequences $\{x_n\}$ and $\{y_n\}$, such that $x_n > c$, and $y_n < c$ for all $n \in \mathbb{N}$, and such that $\lim x_n = \lim y_n = c$. Since f is differentiable at c we know that

$$0 \geq \lim_{n \rightarrow \infty} \frac{f(x_n) - f(c)}{x_n - c} = f'(c) = \lim_{n \rightarrow \infty} \frac{f(y_n) - f(c)}{y_n - c} \geq 0.$$

□

4.2.2 Rolle's theorem

Suppose that a function is zero at both endpoints of an interval. Intuitively it ought to attain a minimum or a maximum in the interior of the interval, then at such a minimum or a maximum, the derivative should be zero. See Figure 4.2 for the geometric idea. This is the content of the so-called Rolle's theorem.

Theorem 4.2.3 (Rolle). Let $f: [a, b] \rightarrow \mathbb{R}$ be continuous function differentiable on (a, b) such that $f(a) = f(b) = 0$. Then there exists a $c \in (a, b)$ such that $f'(c) = 0$.

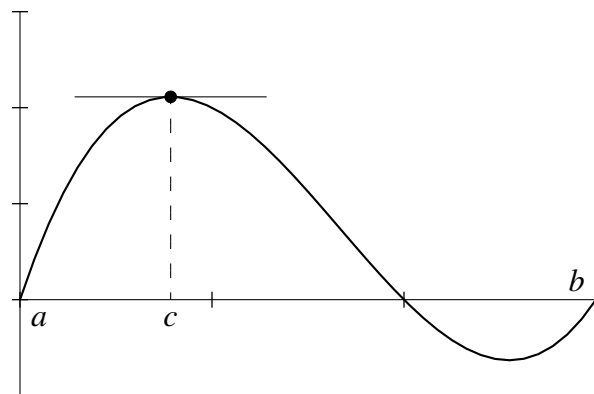


Figure 4.2: Point where tangent line is horizontal, that is $f'(c) = 0$.

Proof. As f is continuous on $[a, b]$ it attains an absolute minimum and an absolute maximum in $[a, b]$. If it attains an absolute maximum at $c \in (a, b)$, then c is also a relative maximum and we apply Theorem 4.2.2 to find that $f'(c) = 0$. If the absolute maximum is at a or at b , then we look for the absolute minimum. If the absolute minimum is at $c \in (a, b)$, then again we find that $f'(c) = 0$. So suppose that the absolute minimum is also at a or b . Hence the absolute minimum is 0 and the absolute maximum is 0, and the function is identically zero. Thus $f'(x) = 0$ for all $x \in [a, b]$, so pick an arbitrary c . \square

4.2.3 Mean value theorem

We extend Rolle's theorem to functions that attain different values at the endpoints.

Theorem 4.2.4 (Mean value theorem). *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function differentiable on (a, b) . Then there exists a point $c \in (a, b)$ such that*

$$f(b) - f(a) = f'(c)(b - a).$$

Proof. The theorem follows from Rolle's theorem. Define the function $g: [a, b] \rightarrow \mathbb{R}$ by

$$g(x) := f(x) - f(b) + (f(b) - f(a)) \frac{b - x}{b - a}.$$

The function g is a differentiable on (a, b) , continuous on $[a, b]$, such that $g(a) = 0$ and $g(b) = 0$. Thus there exists $c \in (a, b)$ such that $g'(c) = 0$.

$$0 = g'(c) = f'(c) + (f(b) - f(a)) \frac{-1}{b - a}.$$

Or in other words $f'(c)(b - a) = f(b) - f(a)$. \square

For a geometric interpretation of the mean value theorem, see Figure 4.3. The idea is that the value $\frac{f(b)-f(a)}{b-a}$ is the slope of the line between the points $(a, f(a))$ and $(b, f(b))$. Then c is the point such that $f'(c) = \frac{f(b)-f(a)}{b-a}$, that is, the tangent line at the point $(c, f(c))$ has the same slope as the line between $(a, f(a))$ and $(b, f(b))$.

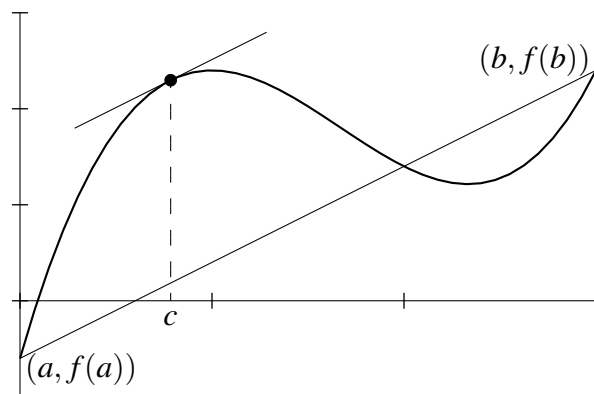


Figure 4.3: Graphical interpretation of the mean value theorem.

4.2.4 Applications

We now solve our very first differential equation.

Proposition 4.2.5. *Let I be an interval and let $f: I \rightarrow \mathbb{R}$ be a differentiable function such that $f'(x) = 0$ for all $x \in I$. Then f is constant.*

Proof. Take arbitrary $x, y \in I$ with $x < y$. Then f restricted to $[x, y]$ satisfies the hypotheses of the mean value theorem. Therefore there is a $c \in (x, y)$ such that

$$f(y) - f(x) = f'(c)(y - x).$$

as $f'(c) = 0$, we have $f(y) = f(x)$. Therefore, the function is constant. \square

Now that we know what it means for the function to stay constant, let us look at increasing and decreasing functions. We say that $f: I \rightarrow \mathbb{R}$ is *increasing* (resp. *strictly increasing*) if $x < y$ implies $f(x) \leq f(y)$ (resp. $f(x) < f(y)$). We define *decreasing* and *strictly decreasing* in the same way by switching the inequalities for f .

Proposition 4.2.6. *Let $f: I \rightarrow \mathbb{R}$ be a differentiable function.*

- (i) f is increasing if and only if $f'(x) \geq 0$ for all $x \in I$.
- (ii) f is decreasing if and only if $f'(x) \leq 0$ for all $x \in I$.

Proof. Let us prove the first item. Suppose that f is increasing, then for all x and c in I we have

$$\frac{f(x) - f(c)}{x - c} \geq 0.$$

Taking a limit as x goes to c we see that $f'(c) \geq 0$.

For the other direction, suppose that $f'(x) \geq 0$ for all $x \in I$. Let $x < y$ in I . Then by the mean value theorem there is some $c \in (x, y)$ such that

$$f(x) - f(y) = f'(c)(x - y).$$

As $f'(c) \geq 0$, and $x - y < 0$, then $f(x) - f(y) \leq 0$ and so f is increasing.

We leave the decreasing part to the reader as exercise. □

Example 4.2.7: We can make a similar but weaker statement about strictly increasing and decreasing functions. If $f'(x) > 0$ for all $x \in I$, then f is strictly increasing. The proof is left as an exercise. The converse is not true. For example, $f(x) := x^3$ is a strictly increasing function, but $f'(0) = 0$.

Another application of the mean value theorem is the following result about location of extrema. The theorem is stated for an absolute minimum and maximum, but the way it is applied to find relative minima and maxima is to restrict f to an interval $(c - \delta, c + \delta)$.

Proposition 4.2.8. *Let $f: (a, b) \rightarrow \mathbb{R}$ be continuous. Let $c \in (a, b)$ and suppose f is differentiable on (a, c) and (c, b) .*

- (i) *If $f'(x) \leq 0$ for $x \in (a, c)$ and $f'(x) \geq 0$ for $x \in (c, b)$, then f has an absolute minimum at c .*
- (ii) *If $f'(x) \geq 0$ for $x \in (a, c)$ and $f'(x) \leq 0$ for $x \in (c, b)$, then f has an absolute maximum at c .*

Proof. Let us prove the first item. The second is left to the reader. Let x be in (a, c) and $\{y_n\}$ a sequence such that $x < y_n < c$ and $\lim y_n = c$. By the previous proposition, the function is decreasing on (a, c) so $f(x) \geq f(y_n)$. The function is continuous at c so we can take the limit to get $f(x) \geq f(c)$ for all $x \in (a, c)$.

Similarly take $x \in (c, b)$ and $\{y_n\}$ a sequence such that $c < y_n < x$ and $\lim y_n = c$. The function is increasing on (c, b) so $f(x) \geq f(y_n)$. By continuity of f we get $f(x) \geq f(c)$ for all $x \in (c, b)$. Thus $f(x) \geq f(c)$ for all $x \in (a, b)$. □

The converse of the proposition does not hold. See Example 4.2.10 below.

4.2.5 Continuity of derivatives and the intermediate value theorem

Derivatives of functions satisfy an intermediate value property. The result is usually called Darboux's theorem.

Theorem 4.2.9 (Darboux). *Let $f: [a, b] \rightarrow \mathbb{R}$ be differentiable. Suppose that there exists a $y \in \mathbb{R}$ such that $f'(a) < y < f'(b)$ or $f'(a) > y > f'(b)$. Then there exists a $c \in (a, b)$ such that $f'(c) = y$.*

Proof. Suppose without loss of generality that $f'(a) < y < f'(b)$. Define

$$g(x) := yx - f(x).$$

As g is continuous on $[a, b]$, then g attains a maximum at some $c \in [a, b]$.

Now compute $g'(x) = y - f'(x)$. Thus $g'(a) > 0$. As the derivative is the limit of difference quotients and is positive, there must be some difference quotient that is positive. That is, there must exist an $x > a$ such that

$$\frac{g(x) - g(a)}{x - a} > 0,$$

or $g(x) > g(a)$. Thus a cannot possibly be a maximum of g . Similarly as $g'(b) < 0$, we find an $x < b$ (a different x) such that $\frac{g(x) - g(b)}{x - b} < 0$ or that $g(x) > g(b)$, thus b cannot possibly be a maximum.

Therefore $c \in (a, b)$. Then as c is a maximum of g we find $g'(c) = 0$ and $f'(c) = y$. \square

We have seen already that there exist discontinuous functions that have the intermediate value property. While it is hard to imagine at first, there also exist functions that are differentiable everywhere and the derivative is not continuous.

Example 4.2.10: Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be the function defined by

$$f(x) := \begin{cases} (x \sin(1/x))^2 & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

We claim that f is differentiable, but $f': \mathbb{R} \rightarrow \mathbb{R}$ is not continuous at the origin. Furthermore, f has a minimum at 0, but the derivative changes sign infinitely often near the origin.

Proof: It is easy to see from the definition that f has an absolute minimum at 0: we know that $f(x) \geq 0$ for all x and $f(0) = 0$.

The function f is differentiable for $x \neq 0$ and the derivative is $2 \sin(1/x)(x \sin(1/x) - \cos(1/x))$. As an exercise show that for $x_n = \frac{4}{(8n+1)\pi}$ we have $\lim f'(x_n) = -1$, and for $y_n = \frac{4}{(8n+3)\pi}$ we have $\lim f'(y_n) = 1$. Hence if f' exists at 0, then it cannot be continuous.

Let us show that f' exists at 0. We claim that the derivative is zero. In other words $\left| \frac{f(x) - f(0)}{x - 0} - 0 \right|$ goes to zero as x goes to zero. For $x \neq 0$ we have

$$\left| \frac{f(x) - f(0)}{x - 0} - 0 \right| = \left| \frac{x^2 \sin^2(1/x)}{x} \right| = |x \sin^2(1/x)| \leq |x|.$$

And, of course, as x tends to zero, then $|x|$ tends to zero and hence $\left| \frac{f(x) - f(0)}{x - 0} - 0 \right|$ goes to zero. Therefore, f is differentiable at 0 and the derivative at 0 is 0.

It is sometimes useful to assume that the derivative of a differentiable function is continuous. If $f: I \rightarrow \mathbb{R}$ is differentiable and the derivative f' is continuous on I , then we say that f is *continuously differentiable*. It is common to write $C^1(I)$ for the set of continuously differentiable functions on I .

4.2.6 Exercises

Exercise 4.2.1: Finish the proof of Proposition 4.2.6.

Exercise 4.2.2: Finish the proof of Proposition 4.2.8.

Exercise 4.2.3: Suppose that $f: \mathbb{R} \rightarrow \mathbb{R}$ is a differentiable function such that f' is a bounded function. Then show that f is a Lipschitz continuous function.

Exercise 4.2.4: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is differentiable and $c \in [a, b]$. Then show that there exists a sequence $\{x_n\}$ converging to c , $x_n \neq c$ for all n , such that

$$f'(c) = \lim_{n \rightarrow \infty} f'(x_n).$$

Do note that this does not imply that f' is continuous (why?).

Exercise 4.2.5: Suppose that $f: \mathbb{R} \rightarrow \mathbb{R}$ is a function such that $|f(x) - f(y)| \leq |x - y|^2$ for all x and y . Show that $f(x) = C$ for some constant C . Hint: Show that f is differentiable at all points and compute the derivative.

Exercise 4.2.6: Suppose that I is an interval and $f: I \rightarrow \mathbb{R}$ is a differentiable function. If $f'(x) > 0$ for all $x \in I$, show that f is strictly increasing.

Exercise 4.2.7: Suppose $f: (a, b) \rightarrow \mathbb{R}$ is a differentiable function such that $f'(x) \neq 0$ for all $x \in (a, b)$. Suppose that there exists a point $c \in (a, b)$ such that $f'(c) > 0$. Prove that $f'(x) > 0$ for all $x \in (a, b)$.

Exercise 4.2.8: Suppose that $f: (a, b) \rightarrow \mathbb{R}$ and $g: (a, b) \rightarrow \mathbb{R}$ are differentiable functions such that $f'(x) = g'(x)$ for all $x \in (a, b)$, then show that there exists a constant C such that $f(x) = g(x) + C$.

Exercise 4.2.9: Prove the following version of L'Hopital's rule. Suppose that $f: (a, b) \rightarrow \mathbb{R}$ and $g: (a, b) \rightarrow \mathbb{R}$ are differentiable functions. Suppose that at $c \in (a, b)$, $f(c) = 0$, $g(c) = 0$, and that the limit of $f'(x)/g'(x)$ as x goes to c exists. Show that

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}.$$

Exercise 4.2.10: Let $f: (a, b) \rightarrow \mathbb{R}$ be an unbounded differentiable function. Show that $f': (a, b) \rightarrow \mathbb{R}$ is unbounded.

4.3 Taylor's theorem

Note: 0.5 lecture (optional section)

4.3.1 Derivatives of higher orders

When $f: I \rightarrow \mathbb{R}$ is differentiable, we obtain a function $f': I \rightarrow \mathbb{R}$. The function f' is called the *first derivative* of f . If f' is differentiable, we denote by $f'': I \rightarrow \mathbb{R}$ the derivative of f' . The function f'' is called the *second derivative* of f . We similarly obtain f''' , f'''' , and so on. With a larger number of derivatives the notation would get out of hand; we denote by $f^{(n)}$ the *n th derivative* of f .

When f possesses n derivatives, we say that f is *n times differentiable*.

4.3.2 Taylor's theorem

Taylor's theorem* is a generalization of the mean value theorem. It tells us that up to a small error, any n times differentiable function can be approximated at a point x_0 by a polynomial. The error of this approximation behaves like $(x - x_0)^n$ near the point x_0 . To see why this is a good approximation notice that for a big n , $(x - x_0)^n$ is very small in a small interval around x_0 .

Definition 4.3.1. For an n times differentiable function f defined near a point $x_0 \in \mathbb{R}$, define the *n th Taylor polynomial* for f at x_0 as

$$\begin{aligned} P_n(x) &:= \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \\ &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \frac{f^{(3)}(x_0)}{6}(x - x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n. \end{aligned}$$

Taylor's theorem tells us that a function behaves like its the n th Taylor polynomial. We can think of the theorem as a generalization of the mean value theorem, which is really Taylor's theorem for the first derivative.

Theorem 4.3.2 (Taylor). Suppose $f: [a, b] \rightarrow \mathbb{R}$ is a function with n continuous derivatives on $[a, b]$ and such that $f^{(n+1)}$ exists on (a, b) . Given distinct points x_0 and x in $[a, b]$, we can find a point c between x_0 and x such that

$$f(x) = P_n(x) + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

The term $R_n(x) := \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}$ is called the *remainder term*. This form of the remainder term is called the *Lagrange form* of the remainder. There are other ways to write the remainder term, but we skip those. Note that c depends on both x and x_0 .

*Named for the English mathematician Brook Taylor (1685–1731).

Proof. Find a number M_{x,x_0} (depending on x and x_0) solving the equation

$$f(x) = P_n(x) + M_{x,x_0}(x - x_0)^{n+1}.$$

Define a function $g(s)$ by

$$g(s) := f(s) - P_n(s) - M_{x,x_0}(s - x_0)^{n+1}.$$

A simple computation shows that $P_n^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 0, 1, 2, \dots, n$ (the zeroth derivative corresponds to the function itself). Therefore,

$$g(x_0) = g'(x_0) = g''(x_0) = \dots = g^{(n)}(x_0) = 0.$$

In particular $g(x_0) = 0$. On the other hand $g(x) = 0$. Thus by the mean value theorem there exists an x_1 between x_0 and x such that $g'(x_1) = 0$. Applying the mean value theorem to g' we obtain that there exists x_2 between x_0 and x_1 (and therefore between x_0 and x) such that $g''(x_2) = 0$. We repeat the argument $n + 1$ times to obtain a number x_{n+1} between x_0 and x_n (and therefore between x_0 and x) such that $g^{(n+1)}(x_{n+1}) = 0$.

Let $c := x_{n+1}$. We compute the $(n + 1)$ th derivative of g to find

$$g^{(n+1)}(s) = f^{(n+1)}(s) - (n + 1)!M_{x,x_0}.$$

Plugging in c for s we obtain that $M_{x,x_0} = \frac{f^{(n+1)}(c)}{(n+1)!}$, and we are done. \square

In the proof we have computed that $P_n^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 0, 1, 2, \dots, n$. Therefore the Taylor polynomial has the same derivatives as f at x_0 up to the n th derivative. That is why the Taylor polynomial is a good approximation to f .

In simple terms, a differentiable function is locally approximated by a line, that's the definition of the derivative. There does exist a converse to Taylor's theorem, which we will not state nor prove, saying that if a function is locally approximated in a certain way by a polynomial of degree d , then it has d derivatives.

4.3.3 Exercises

Exercise 4.3.1: Compute the n th Taylor Polynomial at 0 for the exponential function.

Exercise 4.3.2: Suppose that p is a polynomial of degree d . Given any $x_0 \in \mathbb{R}$, show that the $(d + 1)$ th Taylor polynomial for p at x_0 is equal to p .

Exercise 4.3.3: Let $f(x) := |x|^3$. Compute $f'(x)$ and $f''(x)$ for all x , but show that $f^{(3)}(0)$ does not exist.

Exercise 4.3.4: Suppose $f: \mathbb{R} \rightarrow \mathbb{R}$ has n continuous derivatives. Show that for any $x_0 \in \mathbb{R}$, there exist polynomials P and Q of degree n and an $\varepsilon > 0$ such that $P(x) \leq f(x) \leq Q(x)$ for all $x \in [x_0 - \varepsilon, x_0 + \varepsilon]$ and $Q(x) - P(x) = \lambda(x - x_0)^n$ for some $\lambda \geq 0$.

Chapter 5

The Riemann Integral

5.1 The Riemann integral

Note: 1.5 lectures

We now get to the fundamental concept of an integral. There is often confusion among students of calculus between *integral* and *antiderivative*. The integral is (informally) the area under the curve, nothing else. That we can compute an antiderivative using the integral is a nontrivial result we have to prove. In this chapter we define the *Riemann integral** using the Darboux integral†, which is technically simpler than (but equivalent to) the traditional definition as done by Riemann.

5.1.1 Partitions and lower and upper integrals

We want to integrate a bounded function defined on an interval $[a, b]$. We first define two auxiliary integrals that can be defined for all bounded functions. Only then can we talk about the Riemann integral and the Riemann integrable functions.

Definition 5.1.1. A *partition* P of the interval $[a, b]$ is a finite set of numbers $\{x_0, x_1, x_2, \dots, x_n\}$ such that

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b.$$

We write

$$\Delta x_i := x_i - x_{i-1}.$$

*Named after the German mathematician Georg Friedrich Bernhard Riemann (1826–1866).

†Named after the French mathematician Jean-Gaston Darboux (1842–1917).

Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function. Let P be a partition of $[a, b]$. Define

$$\begin{aligned} m_i &:= \inf\{f(x) : x_{i-1} \leq x \leq x_i\}, \\ M_i &:= \sup\{f(x) : x_{i-1} \leq x \leq x_i\}, \\ L(P, f) &:= \sum_{i=1}^n m_i \Delta x_i, \\ U(P, f) &:= \sum_{i=1}^n M_i \Delta x_i. \end{aligned}$$

We call $L(P, f)$ the *lower Darboux sum* and $U(P, f)$ the *upper Darboux sum*.

The geometric idea of Darboux sums is indicated in Figure 5.1. The lower sum is the area of the shaded rectangles, and the upper sum is the area of the entire rectangles. The width of the i th rectangle is Δx_i , the height of the shaded rectangle is m_i and the height of the entire rectangle is M_i .

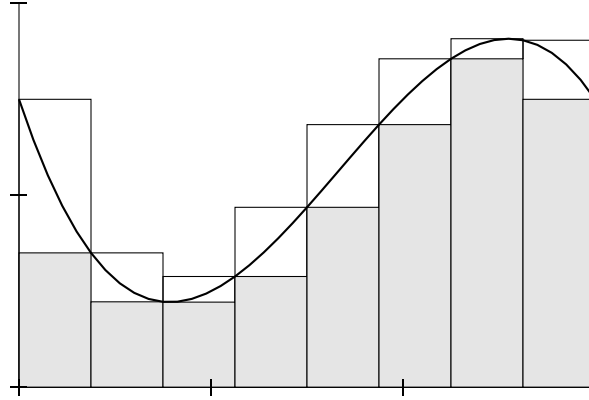


Figure 5.1: Sample Darboux sums.

Proposition 5.1.2. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function. Let $m, M \in \mathbb{R}$ be such that for all x we have $m \leq f(x) \leq M$. For any partition P of $[a, b]$ we have*

$$m(b-a) \leq L(P, f) \leq U(P, f) \leq M(b-a). \quad (5.1)$$

Proof. Let P be a partition. Then note that $m \leq m_i$ for all i and $M_i \leq M$ for all i . Also $m_i \leq M_i$ for all i . Finally $\sum_{i=1}^n \Delta x_i = (b-a)$. Therefore,

$$\begin{aligned} m(b-a) &= m \left(\sum_{i=1}^n \Delta x_i \right) = \sum_{i=1}^n m \Delta x_i \leq \sum_{i=1}^n m_i \Delta x_i \leq \\ &\leq \sum_{i=1}^n M_i \Delta x_i \leq \sum_{i=1}^n M \Delta x_i = M \left(\sum_{i=1}^n \Delta x_i \right) = M(b-a). \end{aligned}$$

Hence we get (5.1). In other words, the set of lower and upper sums are bounded sets. \square

Definition 5.1.3. As the sets of lower and upper Darboux sums are bounded, we define

$$\int_a^b f(x) dx := \sup\{L(P, f) : P \text{ a partition of } [a, b]\},$$

$$\overline{\int_a^b} f(x) dx := \inf\{U(P, f) : P \text{ a partition of } [a, b]\}.$$

We call \int the *lower Darboux integral* and $\overline{\int}$ the *upper Darboux integral*. To avoid worrying about the variable of integration, we often simply write

$$\int_a^b f := \int_a^b f(x) dx \quad \text{and} \quad \overline{\int_a^b} f := \overline{\int_a^b} f(x) dx.$$

If integration is to make sense, then the lower and upper Darboux integrals should be the same number, as we want a single number to call *the integral*. However, these two integrals may in fact differ for some functions.

Example 5.1.4: Take the Dirichlet function $f: [0, 1] \rightarrow \mathbb{R}$, where $f(x) := 1$ if $x \in \mathbb{Q}$ and $f(x) := 0$ if $x \notin \mathbb{Q}$. Then

$$\int_0^1 f = 0 \quad \text{and} \quad \overline{\int_0^1} f = 1.$$

The reason is that for every i we have that $m_i = \inf\{f(x) : x \in [x_{i-1}, x_i]\} = 0$ and $M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\} = 1$. Thus

$$L(P, f) = \sum_{i=1}^n 0 \cdot \Delta x_i = 0,$$

$$U(P, f) = \sum_{i=1}^n 1 \cdot \Delta x_i = \sum_{i=1}^n \Delta x_i = 1.$$

Remark 5.1.5. The same definition of $\int_a^b f$ and $\overline{\int_a^b} f$ is used when f is defined on a larger set S such that $[a, b] \subset S$. In that case, we use the restriction of f to $[a, b]$ and we must ensure that the restriction is bounded on $[a, b]$.

To compute the integral we often take a partition P and make it finer. That is, we cut intervals in the partition into yet smaller pieces.

Definition 5.1.6. Let $P = \{x_0, x_1, \dots, x_n\}$ and $\tilde{P} = \{\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_m\}$ be partitions of $[a, b]$. We say \tilde{P} is a *refinement* of P if as sets $P \subset \tilde{P}$.

That is, \tilde{P} is a refinement of a partition if it contains all the numbers in P and perhaps some other numbers in between. For example, $\{0, 0.5, 1, 2\}$ is a partition of $[0, 2]$ and $\{0, 0.2, 0.5, 1, 1.5, 1.75, 2\}$ is a refinement. The main reason for introducing refinements is the following proposition.

Proposition 5.1.7. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function, and let P be a partition of $[a, b]$. Let \tilde{P} be a refinement of P . Then*

$$L(P, f) \leq L(\tilde{P}, f) \quad \text{and} \quad U(\tilde{P}, f) \leq U(P, f).$$

Proof. The tricky part of this proof is to get the notation correct. Let $\tilde{P} := \{\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_m\}$ be a refinement of $P := \{x_0, x_1, \dots, x_n\}$. Then $x_0 = \tilde{x}_0$ and $x_n = \tilde{x}_m$. In fact, we can find integers $k_0 < k_1 < \dots < k_n$ such that $x_j = \tilde{x}_{k_j}$ for $j = 0, 1, 2, \dots, n$.

Let $\Delta\tilde{x}_j = \tilde{x}_{j-1} - \tilde{x}_j$. We get that

$$\Delta x_j = \sum_{p=k_{j-1}+1}^{k_j} \Delta\tilde{x}_p.$$

Let m_j be as before and correspond to the partition P . Let $\tilde{m}_j := \inf\{f(x) : \tilde{x}_{j-1} \leq x \leq \tilde{x}_j\}$. Now, $m_j \leq \tilde{m}_p$ for $k_{j-1} < p \leq k_j$. Therefore,

$$m_j \Delta x_j = m_j \sum_{p=k_{j-1}+1}^{k_j} \Delta\tilde{x}_p = \sum_{p=k_{j-1}+1}^{k_j} m_j \Delta\tilde{x}_p \leq \sum_{p=k_{j-1}+1}^{k_j} \tilde{m}_p \Delta\tilde{x}_p.$$

So

$$L(P, f) = \sum_{j=1}^n m_j \Delta x_j \leq \sum_{j=1}^n \sum_{p=k_{j-1}+1}^{k_j} \tilde{m}_p \Delta\tilde{x}_p = \sum_{j=1}^m \tilde{m}_j \Delta\tilde{x}_j = L(\tilde{P}, f).$$

The proof of $U(\tilde{P}, f) \leq U(P, f)$ is left as an exercise. \square

Armed with refinements we prove the following. The key point of this next proposition is that the lower Darboux integral is less than or equal to the upper Darboux integral.

Proposition 5.1.8. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function. Let $m, M \in \mathbb{R}$ be such that for all x we have $m \leq f(x) \leq M$. Then*

$$m(b-a) \leq \int_a^b f \leq \int_a^b f \leq M(b-a). \quad (5.2)$$

Proof. By Proposition 5.1.2 we have for any partition P

$$m(b-a) \leq L(P, f) \leq U(P, f) \leq M(b-a).$$

The inequality $m(b-a) \leq L(P, f)$ implies $m(b-a) \leq \underline{\int_a^b} f$. Also $U(P, f) \leq M(b-a)$ implies $\overline{\int_a^b} f \leq M(b-a)$.

The key point of this proposition is the middle inequality in (5.2). Let P_1, P_2 be partitions of $[a, b]$. Define $\tilde{P} := P_1 \cup P_2$. The set \tilde{P} is a partition of $[a, b]$. Furthermore, \tilde{P} is a refinement of P_1 and it is also a refinement of P_2 . By Proposition 5.1.7 we have $L(P_1, f) \leq L(\tilde{P}, f)$ and $U(\tilde{P}, f) \leq U(P_2, f)$. Putting it all together we have

$$L(P_1, f) \leq L(\tilde{P}, f) \leq U(\tilde{P}, f) \leq U(P_2, f).$$

In other words, for two arbitrary partitions P_1 and P_2 we have $L(P_1, f) \leq U(P_2, f)$. Now we recall Proposition 1.2.7. Taking the supremum and infimum over all partitions we get

$$\sup\{L(P, f) : P \text{ a partition}\} \leq \inf\{U(P, f) : P \text{ a partition}\}.$$

In other words $\underline{\int_a^b} f \leq \overline{\int_a^b} f$. □

5.1.2 Riemann integral

We can finally define the Riemann integral. However, the Riemann integral is only defined on a certain class of functions, called the Riemann integrable functions.

Definition 5.1.9. Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function such that

$$\underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx.$$

Then f is said to be *Riemann integrable*. The set of Riemann integrable functions on $[a, b]$ is denoted by $\mathcal{R}[a, b]$. When $f \in \mathcal{R}[a, b]$ we define

$$\int_a^b f(x) dx := \underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx.$$

As before, we often simply write

$$\int_a^b f := \int_a^b f(x) dx.$$

The number $\int_a^b f$ is called the *Riemann integral* of f , or sometimes simply the *integral* of f .

By definition, any Riemann integrable function is bounded. By appealing to Proposition 5.1.8 we immediately obtain the following proposition.

Proposition 5.1.10. Let $f: [a, b] \rightarrow \mathbb{R}$ be a Riemann integrable function. Let $m, M \in \mathbb{R}$ be such that $m \leq f(x) \leq M$. Then

$$m(b-a) \leq \int_a^b f \leq M(b-a).$$

Often we use a weaker form of this proposition. That is, if $|f(x)| \leq M$ for all $x \in [a, b]$, then

$$\left| \int_a^b f \right| \leq M(b-a).$$

Example 5.1.11: We integrate constant functions using Proposition 5.1.8. If $f(x) := c$ for some constant c , then we take $m = M = c$. In inequality (5.2) all the inequalities must be equalities. Thus f is integrable on $[a, b]$ and $\int_a^b f = c(b-a)$.

Example 5.1.12: Let $f: [0, 2] \rightarrow \mathbb{R}$ be defined by

$$f(x) := \begin{cases} 1 & \text{if } x < 1, \\ 1/2 & \text{if } x = 1, \\ 0 & \text{if } x > 1. \end{cases}$$

We claim that f is Riemann integrable and that $\int_0^2 f = 1$.

Proof: Let $0 < \varepsilon < 1$ be arbitrary. Let $P := \{0, 1 - \varepsilon, 1 + \varepsilon, 2\}$ be a partition. We use the notation from the definition of the Darboux sums. Then

$$\begin{aligned} m_1 &= \inf\{f(x) : x \in [0, 1 - \varepsilon]\} = 1, & M_1 &= \sup\{f(x) : x \in [0, 1 - \varepsilon]\} = 1, \\ m_2 &= \inf\{f(x) : x \in [1 - \varepsilon, 1 + \varepsilon]\} = 0, & M_2 &= \sup\{f(x) : x \in [1 - \varepsilon, 1 + \varepsilon]\} = 1, \\ m_3 &= \inf\{f(x) : x \in [1 + \varepsilon, 2]\} = 0, & M_3 &= \sup\{f(x) : x \in [1 + \varepsilon, 2]\} = 0. \end{aligned}$$

Furthermore, $\Delta x_1 = 1 - \varepsilon$, $\Delta x_2 = 2\varepsilon$ and $\Delta x_3 = 1 - \varepsilon$. We compute

$$\begin{aligned} L(P, f) &= \sum_{i=1}^3 m_i \Delta x_i = 1 \cdot (1 - \varepsilon) + 0 \cdot 2\varepsilon + 0 \cdot (1 - \varepsilon) = 1 - \varepsilon, \\ U(P, f) &= \sum_{i=1}^3 M_i \Delta x_i = 1 \cdot (1 - \varepsilon) + 1 \cdot 2\varepsilon + 0 \cdot (1 - \varepsilon) = 1 + \varepsilon. \end{aligned}$$

Thus,

$$\overline{\int_0^2 f} - \underline{\int_0^2 f} \leq U(P, f) - L(P, f) = (1 + \varepsilon) - (1 - \varepsilon) = 2\varepsilon.$$

By Proposition 5.1.8 we have $\underline{\int_0^2 f} \leq \overline{\int_0^2 f}$. As ε was arbitrary we see that $\overline{\int_0^2 f} = \underline{\int_0^2 f}$. So f is Riemann integrable. Finally,

$$1 - \varepsilon = L(P, f) \leq \int_0^2 f \leq U(P, f) = 1 + \varepsilon.$$

Hence, $|\int_0^2 f - 1| \leq \varepsilon$. As ε was arbitrary, we have that $\int_0^2 f = 1$.

5.1.3 More notation

When $f: S \rightarrow \mathbb{R}$ is defined on a larger set S and $[a, b] \subset S$, we say that f is Riemann integrable on $[a, b]$ if the restriction of f to $[a, b]$ is Riemann integrable. In this case, we say $f \in \mathcal{R}[a, b]$, and we write $\int_a^b f$ to mean the Riemann integral of the restriction of f to $[a, b]$.

It is useful to define the integral $\int_a^b f$ even if $a \not\leq b$. Suppose that $b < a$ and that $f \in \mathcal{R}[b, a]$, then define

$$\int_a^b f := - \int_b^a f.$$

For any function f we define

$$\int_a^a f := 0.$$

At times, the variable x may already have some other meaning. When we need to write down the variable of integration, we may simply use a different letter. For example,

$$\int_a^b f(s) ds := \int_a^b f(x) dx.$$

5.1.4 Exercises

Exercise 5.1.1: Let $f: [0, 1] \rightarrow \mathbb{R}$ be defined by $f(x) := x^3$ and let $P := \{0, 0.1, 0.4, 1\}$. Compute $L(P, f)$ and $U(P, f)$.

Exercise 5.1.2: Let $f: [0, 1] \rightarrow \mathbb{R}$ be defined by $f(x) := x$. Show that $f \in \mathcal{R}[0, 1]$ and compute $\int_0^1 f$ using the definition of the integral (but feel free to use the propositions of this section).

Exercise 5.1.3: Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function. Suppose that there exists a sequence of partitions $\{P_k\}$ of $[a, b]$ such that

$$\lim_{k \rightarrow \infty} (U(P_k, f) - L(P_k, f)) = 0.$$

Show that f is Riemann integrable and that

$$\int_a^b f = \lim_{k \rightarrow \infty} U(P_k, f) = \lim_{k \rightarrow \infty} L(P_k, f).$$

Exercise 5.1.4: Finish the proof of Proposition 5.1.7.

Exercise 5.1.5: Suppose that $f: [-1, 1] \rightarrow \mathbb{R}$ is defined as

$$f(x) := \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Prove that $f \in \mathcal{R}[-1, 1]$ and compute $\int_{-1}^1 f$ using the definition of the integral (but feel free to use the propositions of this section).

Exercise 5.1.6: Let $c \in (a, b)$ and let $d \in \mathbb{R}$. Define $f: [a, b] \rightarrow \mathbb{R}$ as

$$f(x) := \begin{cases} d & \text{if } x = c, \\ 0 & \text{if } x \neq c. \end{cases}$$

Prove that $f \in \mathcal{R}[a, b]$ and compute $\int_a^b f$ using the definition of the integral (but feel free to use the propositions of this section).

Exercise 5.1.7: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is Riemann integrable. Let $\varepsilon > 0$ be given. Then show that there exists a partition $P = \{x_0, x_1, \dots, x_n\}$ such that if we pick any set of numbers $\{c_1, c_2, \dots, c_n\}$ with $c_k \in [x_{k-1}, x_k]$ for all k , then

$$\left| \int_a^b f - \sum_{k=1}^n f(c_k) \Delta x_k \right| < \varepsilon.$$

Exercise 5.1.8: Let $f: [a, b] \rightarrow \mathbb{R}$ be a Riemann integrable function. Let $\alpha > 0$ and $\beta \in \mathbb{R}$. Then define $g(x) := f(\alpha x + \beta)$ on the interval $I = [1/\alpha(a - \beta), 1/\alpha(b - \beta)]$. Show that g is Riemann integrable on I .

Exercise 5.1.9: Let $f: [a, b] \rightarrow \mathbb{R}$ be a Riemann integrable function. Let $\alpha > 0$ and $\beta \in \mathbb{R}$. Then define $g(x) := f(\alpha x + \beta)$ on the interval $I = [1/\alpha(a - \beta), 1/\alpha(b - \beta)]$. Show that g is Riemann integrable on I .

Exercise 5.1.10: Let $f: [0, 1] \rightarrow \mathbb{R}$ be a bounded function. Let $P_n = \{x_0, x_1, \dots, x_n\}$ be a uniform partition of $[0, 1]$, that is $x_j := j/n$. Is $\{L(P_n, f)\}_{n=1}^{\infty}$ always monotone? Yes/No: Prove or find a counterexample.

Exercise 5.1.11: For a bounded function $f: [0, 1] \rightarrow \mathbb{R}$ let $R_n := (1/n) \sum_{j=1}^n f(j/n)$ (the uniform right hand rule). a) If f is Riemann integrable show $\int_0^1 f = \lim R_n$. b) Find an f that is not Riemann integrable, but $\lim R_n$ exists.

5.2 Properties of the integral

Note: 2 lectures, integrability of functions with discontinuities can safely be skipped

5.2.1 Additivity

The next result we prove is usually referred to as the additive property of the integral. First we prove the additivity property for the lower and upper Darboux integrals.

Lemma 5.2.1. *Suppose $a < b < c$ and $f: [a, c] \rightarrow \mathbb{R}$ is a bounded function. Then*

$$\int_a^c f = \int_a^b f + \int_b^c f$$

and

$$\overline{\int}_a^c f = \overline{\int}_a^b f + \overline{\int}_b^c f.$$

Proof. If we have partitions $P_1 = \{x_0, x_1, \dots, x_k\}$ of $[a, b]$ and $P_2 = \{x_k, x_{k+1}, \dots, x_n\}$ of $[b, c]$, then the set $P := P_1 \cup P_2 = \{x_0, x_1, \dots, x_n\}$ is a partition of $[a, c]$. Then

$$L(P, f) = \sum_{j=1}^n m_j \Delta x_j = \sum_{j=1}^k m_j \Delta x_j + \sum_{j=k+1}^n m_j \Delta x_j = L(P_1, f) + L(P_2, f).$$

When we take the supremum of the right hand side over all P_1 and P_2 , we are taking a supremum of the left hand side over all partitions P of $[a, c]$ that contain b . If Q is any partition of $[a, c]$ and $P = Q \cup \{b\}$, then P is a refinement of Q and so $L(Q, f) \leq L(P, f)$. Therefore, taking a supremum only over the P that contain b is sufficient to find the supremum of $L(P, f)$ over all partitions P , see Exercise 1.1.9. Finally recall Exercise 1.2.9 to compute

$$\begin{aligned} \int_a^c f &= \sup\{L(P, f) : P \text{ a partition of } [a, c]\} \\ &= \sup\{L(P, f) : P \text{ a partition of } [a, c], b \in P\} \\ &= \sup\{L(P_1, f) + L(P_2, f) : P_1 \text{ a partition of } [a, b], P_2 \text{ a partition of } [b, c]\} \\ &= \sup\{L(P_1, f) : P_1 \text{ a partition of } [a, b]\} + \sup\{L(P_2, f) : P_2 \text{ a partition of } [b, c]\} \\ &= \int_a^b f + \int_b^c f. \end{aligned}$$

Similarly, for $P, P_1,$ and P_2 as above we obtain

$$U(P, f) = \sum_{j=1}^n M_j \Delta x_j = \sum_{j=1}^k M_j \Delta x_j + \sum_{j=k+1}^n M_j \Delta x_j = U(P_1, f) + U(P_2, f).$$

We wish to take the infimum on the right over all P_1 and P_2 , and so we are taking the infimum over all partitions P of $[a, c]$ that contain b . If Q is any partition of $[a, c]$ and $P = Q \cup \{b\}$, then P is a refinement of Q and so $U(Q, f) \geq U(P, f)$. Therefore, taking an infimum only over the P that contain b is sufficient to find the infimum of $U(P, f)$ for all P . We obtain

$$\overline{\int_a^c} f = \overline{\int_a^b} f + \overline{\int_b^c} f.$$

□

Theorem 5.2.2. *Let $a < b < c$. A function $f: [a, c] \rightarrow \mathbb{R}$ is Riemann integrable if and only if f is Riemann integrable on $[a, b]$ and $[b, c]$. If f is Riemann integrable, then*

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Proof. Suppose that $f \in \mathcal{R}[a, c]$, then $\overline{\int_a^c} f = \underline{\int_a^c} f = \int_a^c f$. We apply the lemma to get

$$\int_a^c f = \underline{\int_a^c} f = \underline{\int_a^b} f + \underline{\int_b^c} f \leq \overline{\int_a^b} f + \overline{\int_b^c} f = \overline{\int_a^c} f = \int_a^c f.$$

Thus the inequality is an equality and

$$\underline{\int_a^b} f + \underline{\int_b^c} f = \overline{\int_a^b} f + \overline{\int_b^c} f.$$

As we also know that $\underline{\int_a^b} f \leq \overline{\int_a^b} f$ and $\underline{\int_b^c} f \leq \overline{\int_b^c} f$, we conclude that

$$\underline{\int_a^b} f = \overline{\int_a^b} f \quad \text{and} \quad \underline{\int_b^c} f = \overline{\int_b^c} f.$$

Thus f is Riemann integrable on $[a, b]$ and $[b, c]$ and the desired formula holds.

Now assume that the restrictions of f to $[a, b]$ and to $[b, c]$ are Riemann integrable. We again apply the lemma to get

$$\underline{\int_a^c} f = \underline{\int_a^b} f + \underline{\int_b^c} f = \int_a^b f + \int_b^c f = \overline{\int_a^b} f + \overline{\int_b^c} f = \overline{\int_a^c} f.$$

Therefore f is Riemann integrable on $[a, c]$, and the integral is computed as indicated. □

An easy consequence of the additivity is the following corollary. We leave the details to the reader as an exercise.

Corollary 5.2.3. *If $f \in \mathcal{R}[a, b]$ and $[c, d] \subset [a, b]$, then the restriction $f|_{[c, d]}$ is in $\mathcal{R}[c, d]$.*

5.2.2 Linearity and monotonicity

Proposition 5.2.4 (Linearity). *Let f and g be in $\mathcal{R}[a, b]$ and $\alpha \in \mathbb{R}$.*

(i) αf is in $\mathcal{R}[a, b]$ and

$$\int_a^b \alpha f(x) dx = \alpha \int_a^b f(x) dx.$$

(ii) $f + g$ is in $\mathcal{R}[a, b]$ and

$$\int_a^b f(x) + g(x) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$$

Proof. Let us prove the first item. First suppose that $\alpha \geq 0$. For a partition P we notice that (details are left to the reader)

$$L(P, \alpha f) = \alpha L(P, f) \quad \text{and} \quad U(P, \alpha f) = \alpha U(P, f).$$

For a bounded set of real numbers we can move multiplication by a positive number α past the supremum. Hence,

$$\begin{aligned} \underline{\int_a^b \alpha f(x) dx} &= \sup\{L(P, \alpha f) : P \text{ a partition}\} \\ &= \sup\{\alpha L(P, f) : P \text{ a partition}\} \\ &= \alpha \sup\{L(P, f) : P \text{ a partition}\} \\ &= \alpha \underline{\int_a^b f(x) dx}. \end{aligned}$$

Similarly we show that

$$\overline{\int_a^b \alpha f(x) dx} = \alpha \overline{\int_a^b f(x) dx}.$$

The conclusion now follows for $\alpha \geq 0$.

To finish the proof of the first item, we need to show that $\int_a^b -f(x) dx = -\int_a^b f(x) dx$. The proof of this fact is left as an exercise.

The proof of the second item in the proposition is also left as an exercise. It is not as trivial as it may appear at first glance. \square

Proposition 5.2.5 (Monotonicity). *Let f and g be in $\mathcal{R}[a, b]$ and let $f(x) \leq g(x)$ for all $x \in [a, b]$. Then*

$$\int_a^b f \leq \int_a^b g.$$

Proof. Let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$. Then let

$$m_i := \inf\{f(x) : x \in [x_{i-1}, x_i]\} \quad \text{and} \quad \tilde{m}_i := \inf\{g(x) : x \in [x_{i-1}, x_i]\}.$$

As $f(x) \leq g(x)$, then $m_i \leq \tilde{m}_i$. Therefore,

$$L(P, f) = \sum_{i=1}^n m_i \Delta x_i \leq \sum_{i=1}^n \tilde{m}_i \Delta x_i = L(P, g).$$

We take the supremum over all P (see Proposition 1.3.7) to obtain that

$$\int_a^b f \leq \int_a^b g.$$

As f and g are Riemann integrable, the conclusion follows. \square

5.2.3 Continuous functions

We say that a function $f: [a, b] \rightarrow \mathbb{R}$ has *finitely many discontinuities* if there exists a finite set $S := \{x_1, x_2, \dots, x_n\} \subset [a, b]$, and f is continuous at all points of $[a, b] \setminus S$. Before we prove that bounded functions with finitely many discontinuities are Riemann integrable, we need some lemmas. The first lemma says that bounded continuous functions are Riemann integrable.

Lemma 5.2.6. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then $f \in \mathcal{R}[a, b]$.*

Proof. As f is continuous on a closed bounded interval, it is uniformly continuous. Let $\varepsilon > 0$ be given. Find a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \frac{\varepsilon}{b-a}$.

Let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$ such that $\Delta x_i < \delta$ for all $i = 1, 2, \dots, n$. For example, take n such that $\frac{b-a}{n} < \delta$ and let $x_i := \frac{i}{n}(b-a) + a$. Then for all $x, y \in [x_{i-1}, x_i]$ we have that $|x - y| \leq \Delta x_i < \delta$ and hence

$$f(x) - f(y) < \frac{\varepsilon}{b-a}.$$

As f is continuous on $[x_{i-1}, x_i]$ it attains a maximum and a minimum. Let x be a point where f attains the maximum and y be a point where f attains the minimum. Then $f(x) = M_i$ and $f(y) = m_i$ in the notation from the definition of the integral. Therefore,

$$M_i - m_i = f(x) - f(y) < \frac{\varepsilon}{b-a}.$$

And so

$$\begin{aligned}
 \overline{\int_a^b f} - \underline{\int_a^b f} &\leq U(P, f) - L(P, f) \\
 &= \left(\sum_{i=1}^n M_i \Delta x_i \right) - \left(\sum_{i=1}^n m_i \Delta x_i \right) \\
 &= \sum_{i=1}^n (M_i - m_i) \Delta x_i \\
 &< \frac{\varepsilon}{b-a} \sum_{i=1}^n \Delta x_i \\
 &= \frac{\varepsilon}{b-a} (b-a) = \varepsilon.
 \end{aligned}$$

As $\varepsilon > 0$ was arbitrary,

$$\overline{\int_a^b f} = \underline{\int_a^b f},$$

and f is Riemann integrable on $[a, b]$. □

The second lemma says that we need the function to only be “Riemann integrable inside the interval,” as long as it is bounded. It also tells us how to compute the integral.

Lemma 5.2.7. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function that is Riemann integrable on $[a', b']$ for all a', b' such that $a < a' < b' < b$. Then $f \in \mathcal{R}[a, b]$. Furthermore, if $a < a_n < b_n < b$ are such that $\lim a_n = a$ and $\lim b_n = b$, then*

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f.$$

Proof. Let $M > 0$ be a real number such that $|f(x)| \leq M$. Pick two sequences of numbers $a < a_n < b_n < b$ such that $\lim a_n = a$ and $\lim b_n = b$. Note that $M > 0$ and $(b-a) \geq (b_n - a_n)$. We thus have

$$-M(b-a) \leq -M(b_n - a_n) \leq \int_{a_n}^{b_n} f \leq M(b_n - a_n) \leq M(b-a).$$

Thus the sequence of numbers $\{\int_{a_n}^{b_n} f\}_{n=1}^{\infty}$ is bounded and by Bolzano-Weierstrass has a convergent subsequence indexed by n_k . Let us call L the limit of the subsequence $\{\int_{a_{n_k}}^{b_{n_k}} f\}_{k=1}^{\infty}$.

Lemma 5.2.1 says that the lower and upper integral are additive and the hypothesis says that f is integrable on $[a_{n_k}, b_{n_k}]$. Therefore

$$\underline{\int_a^b f} = \underline{\int_a^{a_{n_k}} f} + \underline{\int_{a_{n_k}}^{b_{n_k}} f} + \underline{\int_{b_{n_k}}^b f} \geq -M(a_{n_k} - a) + \int_{a_{n_k}}^{b_{n_k}} f - M(b - b_{n_k}).$$

We take the limit as k goes to ∞ on the right-hand side,

$$\int_a^b f \geq -M \cdot 0 + L - M \cdot 0 = L.$$

Next we use additivity of the upper integral,

$$\int_a^b f = \int_a^{a_{n_k}} f + \int_{a_{n_k}}^{b_{n_k}} f + \int_{b_{n_k}}^b f \leq M(a_{n_k} - a) + \int_{a_{n_k}}^{b_{n_k}} f + M(b - b_{n_k}).$$

We take the same subsequence $\{\int_{a_{n_k}}^{b_{n_k}} f\}_{k=1}^\infty$ and take the limit to obtain

$$\int_a^b f \leq M \cdot 0 + L + M \cdot 0 = L.$$

Thus $\int_a^b f = \int_a^b f = L$ and hence f is Riemann integrable and $\int_a^b f = L$. In particular, no matter what sequences $\{a_n\}$ and $\{b_n\}$ we started with and what subsequence we chose the L is the same number.

To prove the final statement of the lemma we use Theorem 2.3.7. We have shown that every convergent subsequence $\{\int_{a_{n_k}}^{b_{n_k}} f\}$ converges to $L = \int_a^b f$. Therefore, the sequence $\{\int_{a_n}^{b_n} f\}$ is convergent and converges to L . \square

Theorem 5.2.8. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function with finitely many discontinuities. Then $f \in \mathcal{R}[a, b]$.*

Proof. We divide the interval into finitely many intervals $[a_i, b_i]$ so that f is continuous on the interior (a_i, b_i) . If f is continuous on (a_i, b_i) , then it is continuous and hence integrable on $[c_i, d_i]$ for all $a_i < c_i < d_i < b_i$. By Lemma 5.2.7 the restriction of f to $[a_i, b_i]$ is integrable. By additivity of the integral (and induction) f is integrable on the union of the intervals. \square

Sometimes it is convenient (or necessary) to change certain values of a function and then integrate. The next result says that if we change the values only at finitely many points, the integral does not change.

Proposition 5.2.9. *Let $f: [a, b] \rightarrow \mathbb{R}$ be Riemann integrable. Let $g: [a, b] \rightarrow \mathbb{R}$ be a function such that $f(x) = g(x)$ for all $x \in [a, b] \setminus S$, where S is a finite set. Then g is a Riemann integrable function and*

$$\int_a^b g = \int_a^b f.$$

Sketch of proof. Using additivity of the integral, we split up the interval $[a, b]$ into smaller intervals such that $f(x) = g(x)$ holds for all x except at the endpoints (details are left to the reader).

Therefore, without loss of generality suppose that $f(x) = g(x)$ for all $x \in (a, b)$. The proof follows by Lemma 5.2.7, and is left as an exercise. \square

5.2.4 Exercises

Exercise 5.2.1: Let f be in $\mathcal{R}[a, b]$. Prove that $-f$ is in $\mathcal{R}[a, b]$ and

$$\int_a^b -f(x) \, dx = - \int_a^b f(x) \, dx.$$

Exercise 5.2.2: Let f and g be in $\mathcal{R}[a, b]$. Prove that $f + g$ is in $\mathcal{R}[a, b]$ and

$$\int_a^b f(x) + g(x) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx.$$

Hint: Use Proposition 5.1.7 to find a single partition P such that $U(P, f) - L(P, f) < \varepsilon/2$ and $U(P, g) - L(P, g) < \varepsilon/2$.

Exercise 5.2.3: Let $f: [a, b] \rightarrow \mathbb{R}$ be Riemann integrable. Let $g: [a, b] \rightarrow \mathbb{R}$ be a function such that $f(x) = g(x)$ for all $x \in (a, b)$. Prove that g is Riemann integrable and that

$$\int_a^b g = \int_a^b f.$$

Exercise 5.2.4: Prove the mean value theorem for integrals. That is, prove that if $f: [a, b] \rightarrow \mathbb{R}$ is continuous, then there exists a $c \in [a, b]$ such that $\int_a^b f = f(c)(b - a)$.

Exercise 5.2.5: If $f: [a, b] \rightarrow \mathbb{R}$ is a continuous function such that $f(x) \geq 0$ for all $x \in [a, b]$ and $\int_a^b f = 0$. Prove that $f(x) = 0$ for all x .

Exercise 5.2.6: If $f: [a, b] \rightarrow \mathbb{R}$ is a continuous function for all $x \in [a, b]$ and $\int_a^b f = 0$. Prove that there exists a $c \in [a, b]$ such that $f(c) = 0$ (Compare with the previous exercise).

Exercise 5.2.7: If $f: [a, b] \rightarrow \mathbb{R}$ and $g: [a, b] \rightarrow \mathbb{R}$ are continuous functions such that $\int_a^b f = \int_a^b g$. Then show that there exists a $c \in [a, b]$ such that $f(c) = g(c)$.

Exercise 5.2.8: Let $f \in \mathcal{R}[a, b]$. Let α, β, γ be arbitrary numbers in $[a, b]$ (not necessarily ordered in any way). Prove that

$$\int_\alpha^\gamma f = \int_\alpha^\beta f + \int_\beta^\gamma f.$$

Recall what $\int_a^b f$ means if $b \leq a$.

Exercise 5.2.9: Prove Corollary 5.2.3.

Exercise 5.2.10: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is bounded and has finitely many discontinuities. Show that as a function of x the expression $|f(x)|$ is bounded with finitely many discontinuities and is thus Riemann integrable. Then show that

$$\left| \int_a^b f(x) \, dx \right| \leq \int_a^b |f(x)| \, dx.$$

Exercise 5.2.11 (Hard): Show that the Thomae or popcorn function (see Example 3.2.12) is Riemann integrable. Therefore, there exists a function discontinuous at all rational numbers (a dense set) that is Riemann integrable.

In particular, define $f: [0, 1] \rightarrow \mathbb{R}$ by

$$f(x) := \begin{cases} 1/k & \text{if } x = m/k \text{ where } m, k \in \mathbb{N} \text{ and } m \text{ and } k \text{ have no common divisors,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Show that $\int_0^1 f = 0$.

If $I \subset \mathbb{R}$ is a bounded interval, then the function

$$\varphi_I(x) := \begin{cases} 1 & \text{if } x \in I, \\ 0 & \text{otherwise,} \end{cases}$$

is called an *elementary step function*.

Exercise 5.2.12: Let I be an arbitrary bounded interval (you should consider all types of intervals: closed, open, half-open) and $a < b$, then using only the definition of the integral show that the elementary step function φ_I is integrable on $[a, b]$, and find the integral in terms of a , b , and the endpoints of I .

When a function f can be written as

$$f(x) = \sum_{k=1}^n \alpha_k \varphi_{I_k}(x)$$

for some real numbers $\alpha_1, \alpha_2, \dots, \alpha_n$ and some bounded intervals I_1, I_2, \dots, I_n , then f is called a *step function*.

Exercise 5.2.13: Using the previous exercise, show that a step function is integrable on any interval $[a, b]$. Furthermore, find the integral in terms of a , b , the endpoints of I_k and the α_k .

Exercise 5.2.14: Let $f: [a, b] \rightarrow \mathbb{R}$ be increasing. a) Show that f is Riemann integrable. Hint: Use a uniform partition; each subinterval of same length. b) Use part a) to show that a decreasing function is Riemann integrable. c) Suppose $h = f - g$ where f and g are increasing functions on $[a, b]$. Show that h is Riemann integrable*.

Exercise 5.2.15 (Challenging): Suppose that $f \in \mathcal{R}[a, b]$, then the function that takes x to $|f(x)|$ is also Riemann integrable on $[a, b]$. Then show the same inequality as Exercise 5.2.10.

*Such an h is said to be of *bounded variation*.

5.3 Fundamental theorem of calculus

Note: 1.5 lectures

In this chapter we discuss and prove the *fundamental theorem of calculus*. The entirety of integral calculus is built upon this theorem, ergo the name. The theorem relates the seemingly unrelated concepts of integral and derivative. It tells us how to compute the antiderivative of a function using the integral and vice-versa.

5.3.1 First form of the theorem

Theorem 5.3.1. *Let $F : [a, b] \rightarrow \mathbb{R}$ be a continuous function, differentiable on (a, b) . Let $f \in \mathcal{R}[a, b]$ be such that $f(x) = F'(x)$ for $x \in (a, b)$. Then*

$$\int_a^b f = F(b) - F(a).$$

It is not hard to generalize the theorem to allow a finite number of points in $[a, b]$ where F is not differentiable, as long as it is continuous. This generalization is left as an exercise.

Proof. Let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$. For each interval $[x_{i-1}, x_i]$, use the mean value theorem to find a $c_i \in [x_{i-1}, x_i]$ such that

$$f(c_i)\Delta x_i = F'(c_i)(x_i - x_{i-1}) = F(x_i) - F(x_{i-1}).$$

Using the notation from the definition of the integral, we have $m_i \leq f(c_i) \leq M_i$ and so

$$m_i\Delta x_i \leq F(x_i) - F(x_{i-1}) \leq M_i\Delta x_i.$$

We sum over $i = 1, 2, \dots, n$ to get

$$\sum_{i=1}^n m_i\Delta x_i \leq \sum_{i=1}^n (F(x_i) - F(x_{i-1})) \leq \sum_{i=1}^n M_i\Delta x_i.$$

Notice that in the middle sum all the terms except the first and last cancel out and we end up with $F(x_n) - F(x_0) = F(b) - F(a)$. The sums on the left and on the right are the lower and the upper sum respectively. So

$$L(P, f) \leq F(b) - F(a) \leq U(P, f).$$

We take the supremum of $L(P, f)$ over all P and the left inequality yields

$$\underline{\int_a^b} f \leq F(b) - F(a).$$

Similarly, taking the infimum of $U(P, f)$ over all partitions P yields

$$F(b) - F(a) \leq \overline{\int_a^b f}.$$

As f is Riemann integrable, we have

$$\int_a^b f = \underline{\int_a^b f} \leq F(b) - F(a) \leq \overline{\int_a^b f} = \int_a^b f.$$

The inequalities must be equalities and we are done. \square

The theorem is used to compute integrals. Suppose we know that the function $f(x)$ is a derivative of some other function $F(x)$, then we can find an explicit expression for $\int_a^b f$.

Example 5.3.2: Suppose we are trying to compute

$$\int_0^1 x^2 dx.$$

We notice that x^2 is the derivative of $\frac{x^3}{3}$. We use the fundamental theorem to write

$$\int_0^1 x^2 dx = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}.$$

5.3.2 Second form of the theorem

The second form of the fundamental theorem gives us a way to solve the differential equation $F'(x) = f(x)$, where $f(x)$ is a known function and we are trying to find an F that satisfies the equation.

Theorem 5.3.3. *Let $f: [a, b] \rightarrow \mathbb{R}$ be a Riemann integrable function. Define*

$$F(x) := \int_a^x f.$$

First, F is continuous on $[a, b]$. Second, if f is continuous at $c \in [a, b]$, then F is differentiable at c and $F'(c) = f(c)$.

Proof. As f is bounded, there is an $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$. Suppose $x, y \in [a, b]$ with $x > y$. Then

$$|F(x) - F(y)| = \left| \int_a^x f - \int_a^y f \right| = \left| \int_y^x f \right| \leq M|x - y|.$$

By symmetry, the same also holds if $x < y$. So F is Lipschitz continuous and hence continuous.

Now suppose that f is continuous at c . Let $\varepsilon > 0$ be given. Let $\delta > 0$ be such that for $x \in [a, b]$ $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. In particular, for such x we have

$$f(c) - \varepsilon \leq f(x) \leq f(c) + \varepsilon.$$

Thus if $x > c$, then

$$(f(c) - \varepsilon)(x - c) \leq \int_c^x f \leq (f(c) + \varepsilon)(x - c).$$

When $c > x$, then the inequalities are reversed. Therefore, assuming that $c \neq x$ we get

$$f(c) - \varepsilon \leq \frac{\int_c^x f}{x - c} \leq f(c) + \varepsilon.$$

As

$$\frac{F(x) - F(c)}{x - c} = \frac{\int_a^x f - \int_a^c f}{x - c} = \frac{\int_c^x f}{x - c},$$

we have that

$$\left| \frac{F(x) - F(c)}{x - c} - f(c) \right| \leq \varepsilon.$$

The result follows. It is left to the reader to see why it is OK that we just have a non-strict inequality. \square

Of course, if f is continuous on $[a, b]$, then it is automatically Riemann integrable, F is differentiable on all of $[a, b]$ and $F'(x) = f(x)$ for all $x \in [a, b]$.

Remark 5.3.4. The second form of the fundamental theorem of calculus still holds if we let $d \in [a, b]$ and define

$$F(x) := \int_d^x f.$$

That is, we can use any point of $[a, b]$ as our base point. The proof is left as an exercise.

A common misunderstanding of the integral for calculus students is to think of integrals whose solution cannot be given in closed-form as somehow deficient. This is not the case. Most integrals we write down are not computable in closed-form. Even some integrals that we consider in closed-form are not really such. For example, how does a computer find the value of $\ln x$? One way to do it is to note that we define the natural log as the antiderivative of $1/x$ such that $\ln 1 = 0$. Therefore,

$$\ln x := \int_1^x 1/s \, ds.$$

Then we can numerically approximate the integral. So morally, we did not really “simplify” $\int_1^x 1/s \, ds$ by writing down $\ln x$. We simply gave the integral a name. If we require numerical answers, it is possible that we end up doing the calculation by approximating an integral anyway.

Another common function defined by an integral that cannot be evaluated symbolically is the erf function, defined as

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds.$$

This function comes up often in applied mathematics. It is simply the antiderivative of $(2/\sqrt{\pi})e^{-x^2}$ that is zero at zero. The second form of the fundamental theorem tells us that we can write the function as an integral. If we wish to compute any particular value, we numerically approximate the integral.

5.3.3 Change of variables

A theorem often used in calculus to solve integrals is the change of variables theorem. Let us prove it now. Recall that a function is continuously differentiable if it is differentiable and the derivative is continuous.

Theorem 5.3.5 (Change of variables). *Let $g: [a, b] \rightarrow \mathbb{R}$ be a continuously differentiable function. If $g([a, b]) \subset [c, d]$ and $f: [c, d] \rightarrow \mathbb{R}$ is continuous, then*

$$\int_a^b f(g(x)) g'(x) dx = \int_{g(a)}^{g(b)} f(s) ds.$$

Proof. As g , g' , and f are continuous, we know that $f(g(x))g'(x)$ is a continuous function on $[a, b]$, therefore it is Riemann integrable.

Define

$$F(y) := \int_{g(a)}^y f(s) ds.$$

By the second form of the fundamental theorem of calculus (using Exercise 5.3.4 below) F is a differentiable function and $F'(y) = f(y)$. Now we apply the chain rule and write

$$(F \circ g)'(x) = F'(g(x))g'(x) = f(g(x))g'(x).$$

We note that $F(g(a)) = 0$ and we use the first form of the fundamental theorem to obtain

$$\int_{g(a)}^{g(b)} f(s) ds = F(g(b)) = F(g(b)) - F(g(a)) = \int_a^b (F \circ g)'(x) dx = \int_a^b f(g(x))g'(x) dx.$$

□

The change of variables theorem is often used to solve integrals by changing them to integrals we know or which we can solve using the fundamental theorem of calculus.

Example 5.3.6: From an exercise, we know that the derivative of $\sin(x)$ is $\cos(x)$. Therefore we solve

$$\int_0^{\sqrt{\pi}} x \cos(x^2) dx = \int_0^{\pi} \frac{\cos(s)}{2} ds = \frac{1}{2} \int_0^{\pi} \cos(s) ds = \frac{\sin(\pi) - \sin(0)}{2} = 0.$$

However, beware that we must satisfy the hypotheses of the theorem. The following example demonstrates a common mistake for students of calculus. We must not simply move symbols around, we should always be careful that those symbols really make sense.

Example 5.3.7: Suppose we write down

$$\int_{-1}^1 \frac{\ln|x|}{x} dx.$$

It may be tempting to take $g(x) := \ln|x|$. Then take $g'(x) = \frac{1}{x}$ and try to write

$$\int_{g(-1)}^{g(1)} s ds = \int_0^0 s ds = 0.$$

This “solution” is incorrect, and it does not say that we can solve the given integral. First problem is that $\frac{\ln|x|}{x}$ is not Riemann integrable on $[-1, 1]$ (it is unbounded). The integral we wrote down simply does not make sense. Second, $\frac{\ln|x|}{x}$ is not even continuous on $[-1, 1]$. Finally, g is not continuous on $[-1, 1]$ either.

5.3.4 Exercises

Exercise 5.3.1: Compute $\frac{d}{dx} \left(\int_{-x}^x e^{s^2} ds \right)$.

Exercise 5.3.2: Compute $\frac{d}{dx} \left(\int_0^{x^2} \sin(s^2) ds \right)$.

Exercise 5.3.3: Suppose $F: [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on $[a, b] \setminus S$, where S is a finite set. Suppose there exists an $f \in \mathcal{R}[a, b]$ such that $f(x) = F'(x)$ for $x \in [a, b] \setminus S$. Show that $\int_a^b f = F(b) - F(a)$.

Exercise 5.3.4: Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Let $c \in [a, b]$ be arbitrary. Define

$$F(x) := \int_c^x f.$$

Prove that F is differentiable and that $F'(x) = f(x)$ for all $x \in [a, b]$.

Exercise 5.3.5: Prove integration by parts. That is, suppose that F and G are differentiable functions on $[a, b]$ and suppose that F' and G' are Riemann integrable. Then prove

$$\int_a^b F(x)G'(x) dx = F(b)G(b) - F(a)G(a) - \int_a^b F'(x)G(x) dx.$$

Exercise 5.3.6: Suppose that F , and G are differentiable functions defined on $[a, b]$ such that $F'(x) = G'(x)$ for all $x \in [a, b]$. Using the fundamental theorem of calculus, show that F and G differ by a constant. That is, show that there exists a $C \in \mathbb{R}$ such that $F(x) - G(x) = C$.

The next exercise shows how we can use the integral to “smooth out” a non-differentiable function.

Exercise 5.3.7: Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Let $\varepsilon > 0$ be a constant. For $x \in [a + \varepsilon, b - \varepsilon]$, define

$$g(x) := \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} f.$$

- Show that g is differentiable and find the derivative.
- Let f be differentiable and fix $x \in (a, b)$ (let ε be small enough). What happens to $g'(x)$ as ε gets smaller?
- Find g for $f(x) := |x|$, $\varepsilon = 1$ (you can assume that $[a, b]$ is large enough).

Exercise 5.3.8: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is continuous. Suppose that $\int_a^x f = \int_x^b f$ for all $x \in [a, b]$. Show that $f(x) = 0$ for all $x \in [a, b]$.

Exercise 5.3.9: Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $\int_a^x f = 0$ for all rational x in $[a, b]$. Show that $f(x) = 0$ for all $x \in [a, b]$.

Exercise 5.3.10: A function f is an odd function if $f(x) = -f(-x)$, and f is an even function if $f(x) = f(-x)$. Let $a > 0$. Assume f is continuous. Prove: a) If f is odd, then $\int_{-a}^a f = 0$. b) If f is even, then $\int_{-a}^a f = 2 \int_0^a f$.

Exercise 5.3.11: a) Show that $f(x) := \sin(1/x)$ is integrable on any interval (you can define $f(0)$ to be anything). b) Compute $\int_{-1}^1 \sin(1/x) dx$. (Mind the discontinuity)

Chapter 6

Sequences of Functions

6.1 Pointwise and uniform convergence

Note: 1–1.5 lecture

Up till now when we talked about sequences we always talked about sequences of numbers. However, a very useful concept in analysis is to use a sequence of functions. For example, a solution to some differential equation might be found by finding only approximate solutions. Then the real solution is some sort of limit of those approximate solutions.

When talking about sequences of functions, the tricky part is that there are multiple notions of a limit. Let us describe two common notions of a limit of a sequence of functions.

6.1.1 Pointwise convergence

Definition 6.1.1. For every $n \in \mathbb{N}$ let $f_n: S \rightarrow \mathbb{R}$ be a function. We say the sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise to $f: S \rightarrow \mathbb{R}$, if for every $x \in S$ we have

$$f(x) = \lim_{n \rightarrow \infty} f_n(x).$$

It is common to say that $f_n: S \rightarrow \mathbb{R}$ converges to f on $T \subset \mathbb{R}$ for some $f: T \rightarrow \mathbb{R}$. In that case we, of course, mean that $f(x) = \lim f_n(x)$ for every $x \in T$. We simply mean that the restrictions of f_n to T converge pointwise to f .

Example 6.1.2: The sequence of functions defined by $f_n(x) := x^{2n}$ converges to $f: [-1, 1] \rightarrow \mathbb{R}$ on $[-1, 1]$, where

$$f(x) = \begin{cases} 1 & \text{if } x = -1 \text{ or } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 6.1.

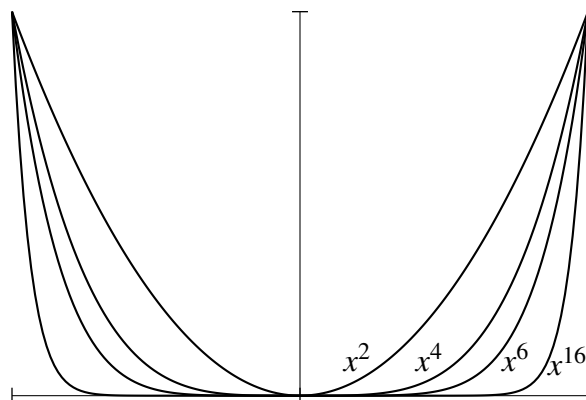


Figure 6.1: Graphs of $f_1, f_2, f_3,$ and f_8 for $f_n(x) := x^{2^n}$.

To see that this is so, first take $x \in (-1, 1)$. Then $0 \leq x^2 < 1$. We have seen before that

$$|x^{2^n} - 0| = (x^2)^n \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore $\lim f_n(x) = 0$.

When $x = 1$ or $x = -1$, then $x^{2^n} = 1$ for all n and hence $\lim f_n(x) = 1$. We also note that $\{f_n(x)\}$ does not converge for all other x .

Often, functions are given as a series. In this case, we use the notion of pointwise convergence to find the values of the function.

Example 6.1.3: We write

$$\sum_{k=0}^{\infty} x^k$$

to denote the limit of the functions

$$f_n(x) := \sum_{k=0}^n x^k.$$

When studying series, we have seen that on $x \in (-1, 1)$ the f_n converge pointwise to

$$\frac{1}{1-x}.$$

The subtle point here is that while $\frac{1}{1-x}$ is defined for all $x \neq 1$, and f_n are defined for all x (even at $x = 1$), convergence only happens on $(-1, 1)$.

Therefore, when we write

$$f(x) := \sum_{k=0}^{\infty} x^k$$

we mean that f is defined on $(-1, 1)$ and is the pointwise limit of the partial sums.

Example 6.1.4: Let $f_n(x) := \sin(xn)$. Then f_n does not converge pointwise to any function on any interval. It may converge at certain points, such as when $x = 0$ or $x = \pi$. It is left as an exercise that in any interval $[a, b]$, there exists an x such that $\sin(xn)$ does not have a limit as n goes to infinity.

Before we move to uniform convergence, let us reformulate pointwise convergence in a different way. We leave the proof to the reader, it is a simple application of the definition of convergence of a sequence of real numbers.

Proposition 6.1.5. *Let $f_n: S \rightarrow \mathbb{R}$ and $f: S \rightarrow \mathbb{R}$ be functions. Then $\{f_n\}$ converges pointwise to f if and only if for every $x \in S$, and every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that*

$$|f_n(x) - f(x)| < \varepsilon$$

for all $n \geq N$.

The key point here is that N can depend on x , not just on ε . That is, for each x we can pick a different N . If we can pick one N for all x , we have what is called uniform convergence.

6.1.2 Uniform convergence

Definition 6.1.6. Let $f_n: S \rightarrow \mathbb{R}$ be functions. We say the sequence $\{f_n\}$ converges uniformly to $f: S \rightarrow \mathbb{R}$, if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } x \in S.$$

Note that N now cannot depend on x . Given $\varepsilon > 0$ we must find an N that works for all $x \in S$. Because of Proposition 6.1.5, we see that uniform convergence implies pointwise convergence.

Proposition 6.1.7. *Let $\{f_n\}$ be a sequence of functions $f_n: S \rightarrow \mathbb{R}$. If $\{f_n\}$ converges uniformly to $f: S \rightarrow \mathbb{R}$, then $\{f_n\}$ converges pointwise to f .*

The converse does not hold.

Example 6.1.8: The functions $f_n(x) := x^{2n}$ do not converge uniformly on $[-1, 1]$, even though they converge pointwise. To see this, suppose for contradiction that they did. Take $\varepsilon := 1/2$, then there would have to exist an N such that $x^{2N} < 1/2$ for all $x \in [0, 1)$ (as $f_n(x)$ converges to 0 on $(-1, 1)$). But that means that for any sequence $\{x_k\}$ in $[0, 1)$ such that $\lim x_k = 1$ we have $x_k^{2N} < 1/2$. On the other hand x^{2N} is a continuous function of x (it is a polynomial), therefore we obtain a contradiction

$$1 = 1^{2N} = \lim_{k \rightarrow \infty} x_k^{2N} \leq 1/2.$$

However, if we restrict our domain to $[-a, a]$ where $0 < a < 1$, then f_n converges uniformly to 0 on $[-a, a]$. Again to see this note that $a^{2n} \rightarrow 0$ as $n \rightarrow \infty$. Thus given $\varepsilon > 0$, pick $N \in \mathbb{N}$ such that $a^{2n} < \varepsilon$ for all $n \geq N$. Then for any $x \in [-a, a]$ we have $|x| \leq a$. Therefore, for $n \geq N$

$$|x^{2n}| = |x|^{2n} \leq a^{2n} < \varepsilon.$$

6.1.3 Convergence in uniform norm

For bounded functions there is another more abstract way to think of uniform convergence. To every bounded function we assign a certain nonnegative number (called the uniform norm). This number measures the “distance” of the function from 0. Then we can “measure” how far two functions are from each other. We can then simply translate a statement about uniform convergence into a statement about a certain sequence of real numbers converging to zero.

Definition 6.1.9. Let $f: S \rightarrow \mathbb{R}$ be a bounded function. Define

$$\|f\|_u := \sup\{|f(x)| : x \in S\}.$$

$\|\cdot\|_u$ is called the *uniform norm*.

Proposition 6.1.10. A sequence of bounded functions $f_n: S \rightarrow \mathbb{R}$ converges uniformly to $f: S \rightarrow \mathbb{R}$, if and only if

$$\lim_{n \rightarrow \infty} \|f_n - f\|_u = 0.$$

Proof. First suppose that $\lim \|f_n - f\|_u = 0$. Let $\varepsilon > 0$ be given. Then there exists an N such that for $n \geq N$ we have $\|f_n - f\|_u < \varepsilon$. As $\|f_n - f\|_u$ is the supremum of $|f_n(x) - f(x)|$, we see that for all x we have $|f_n(x) - f(x)| < \varepsilon$.

On the other hand, suppose that f_n converges uniformly to f . Let $\varepsilon > 0$ be given. Then find N such that $|f_n(x) - f(x)| < \varepsilon$ for all $x \in S$. Taking the supremum we see that $\|f_n - f\|_u < \varepsilon$. Hence $\lim \|f_n - f\|_u = 0$. \square

Sometimes it is said that f_n converges to f in uniform norm instead of converges uniformly. The proposition says that the two notions are the same thing.

Example 6.1.11: Let $f_n: [0, 1] \rightarrow \mathbb{R}$ be defined by $f_n(x) := \frac{nx + \sin(nx^2)}{n}$. Then we claim that f_n converge uniformly to $f(x) := x$. Let us compute:

$$\begin{aligned} \|f_n - f\|_u &= \sup \left\{ \left| \frac{nx + \sin(nx^2)}{n} - x \right| : x \in [0, 1] \right\} \\ &= \sup \left\{ \frac{|\sin(nx^2)|}{n} : x \in [0, 1] \right\} \\ &\leq \sup \{1/n : x \in [0, 1]\} \\ &= 1/n. \end{aligned}$$

Using uniform norm, we define Cauchy sequences in a similar way as we define Cauchy sequences of real numbers.

Definition 6.1.12. Let $f_n: S \rightarrow \mathbb{R}$ be bounded functions. We say that the sequence is *Cauchy in the uniform norm* or *uniformly Cauchy* if for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for $m, k \geq N$ we have

$$\|f_m - f_k\|_u < \varepsilon.$$

Proposition 6.1.13. Let $f_n: S \rightarrow \mathbb{R}$ be bounded functions. Then $\{f_n\}$ is Cauchy in the uniform norm if and only if there exists an $f: S \rightarrow \mathbb{R}$ and $\{f_n\}$ converges uniformly to f .

Proof. Let us first suppose that $\{f_n\}$ is Cauchy in the uniform norm. Let us define f . Fix x , then the sequence $\{f_n(x)\}$ is Cauchy because

$$|f_m(x) - f_k(x)| \leq \|f_m - f_k\|_u.$$

Thus $\{f_n(x)\}$ converges to some real number so define

$$f(x) := \lim_{n \rightarrow \infty} f_n(x).$$

Therefore, f_n converges pointwise to f . To show that convergence is uniform, let $\varepsilon > 0$ be given. Pick a positive $\varepsilon' < \varepsilon$ and find an N such that for $m, k \geq N$ we have $\|f_m - f_k\|_u < \varepsilon'$. In other words for all x we have $|f_m(x) - f_k(x)| < \varepsilon'$. We take the limit as k goes to infinity. Then $|f_m(x) - f_k(x)|$ goes to $|f_m(x) - f(x)|$. Therefore for all x we get

$$|f_m(x) - f(x)| \leq \varepsilon' < \varepsilon.$$

And hence f_n converges uniformly.

For the other direction, suppose that $\{f_n\}$ converges uniformly to f . Given $\varepsilon > 0$, find N such that for all $n \geq N$ we have $|f_n(x) - f(x)| < \varepsilon/4$ for all $x \in S$. Therefore for all $m, k \geq N$ we have

$$|f_m(x) - f_k(x)| = |f_m(x) - f(x) + f(x) - f_k(x)| \leq |f_m(x) - f(x)| + |f(x) - f_k(x)| < \varepsilon/4 + \varepsilon/4.$$

Take supremum over all x to obtain

$$\|f_m - f_k\|_u \leq \varepsilon/2 < \varepsilon.$$

□

6.1.4 Exercises

Exercise 6.1.1: Let f and g be bounded functions on $[a, b]$. Show that

$$\|f + g\|_u \leq \|f\|_u + \|g\|_u.$$

Exercise 6.1.2: a) Find the pointwise limit $\frac{e^{x/n}}{n}$ for $x \in \mathbb{R}$.

b) Is the limit uniform on \mathbb{R} ?

c) Is the limit uniform on $[0, 1]$?

Exercise 6.1.3: Suppose $f_n: S \rightarrow \mathbb{R}$ are functions that converge uniformly to $f: S \rightarrow \mathbb{R}$. Suppose that $A \subset S$. Show that the sequence of restrictions $\{f_n|_A\}$ converges uniformly to $f|_A$.

Exercise 6.1.4: Suppose that $\{f_n\}$ and $\{g_n\}$ defined on some set A converge to f and g respectively pointwise. Show that $\{f_n + g_n\}$ converges pointwise to $f + g$.

Exercise 6.1.5: Suppose that $\{f_n\}$ and $\{g_n\}$ defined on some set A converge to f and g respectively uniformly on A . Show that $\{f_n + g_n\}$ converges uniformly to $f + g$ on A .

Exercise 6.1.6: Find an example of a sequence of functions $\{f_n\}$ and $\{g_n\}$ that converge uniformly to some f and g on some set A , but such that $\{f_n g_n\}$ (the multiple) does not converge uniformly to fg on A . Hint: Let $A := \mathbb{R}$, let $f(x) := g(x) := x$. You can even pick $f_n = g_n$.

Exercise 6.1.7: Suppose that there exists a sequence of functions $\{g_n\}$ uniformly converging to 0 on A . Now suppose that we have a sequence of functions $\{f_n\}$ and a function f on A such that

$$|f_n(x) - f(x)| \leq g_n(x)$$

for all $x \in A$. Show that $\{f_n\}$ converges uniformly to f on A .

Exercise 6.1.8: Let $\{f_n\}$, $\{g_n\}$ and $\{h_n\}$ be sequences of functions on $[a, b]$. Suppose that $\{f_n\}$ and $\{h_n\}$ converge uniformly to some function $f: [a, b] \rightarrow \mathbb{R}$ and suppose that $f_n(x) \leq g_n(x) \leq h_n(x)$ for all $x \in [a, b]$. Show that $\{g_n\}$ converges uniformly to f .

Exercise 6.1.9: Let $f_n: [0, 1] \rightarrow \mathbb{R}$ be a sequence of increasing functions (that is $f_n(x) \geq f_n(y)$ whenever $x \geq y$). Suppose that $f(0) = 0$ and that $\lim_{n \rightarrow \infty} f_n(1) = 0$. Show that $\{f_n\}$ converges uniformly to 0.

Exercise 6.1.10: Let $\{f_n\}$ be a sequence of functions defined on $[0, 1]$. Suppose that there exists a sequence of numbers $x_n \in [0, 1]$ such that

$$f_n(x_n) = 1.$$

Prove or disprove the following statements:

a) True or false: There exists $\{f_n\}$ as above that converges to 0 pointwise.

b) True or false: There exists $\{f_n\}$ as above that converges to 0 uniformly on $[0, 1]$.

Exercise 6.1.11: Fix a continuous $h: [a, b] \rightarrow \mathbb{R}$. Let $f(x) = h(x)$ for $x \in [a, b]$, $f(x) = h(a)$ for $x < a$ and $f(x) = h(b)$ for all $x > b$. First show that $f: \mathbb{R} \rightarrow \mathbb{R}$ is continuous. Now let f_n be the function g from Exercise 5.3.7 with $\varepsilon = 1/n$, defined on the interval $[a, b]$. Show that $\{f_n\}$ converges uniformly to h on $[a, b]$.

6.2 Interchange of limits

Note: 1–1.5 lectures

Large parts of modern analysis deal mainly with the question of the interchange of two limiting operations. It is easy to see that when we have a chain of two limits, we cannot always just swap the limits. For example,

$$0 = \lim_{n \rightarrow \infty} \left(\lim_{k \rightarrow \infty} \frac{n/k}{n/k + 1} \right) \neq \lim_{k \rightarrow \infty} \left(\lim_{n \rightarrow \infty} \frac{n/k}{n/k + 1} \right) = 1.$$

When talking about sequences of functions, interchange of limits comes up quite often. We treat two cases. First we look at continuity of the limit, and second we look at the integral of the limit.

6.2.1 Continuity of the limit

If we have a sequence of continuous functions, is the limit continuous? Suppose that f is the (pointwise) limit of f_n . If $\lim x_k = x$ we are interested in the following interchange of limits. The equality we have to prove (it is not always true) is marked with a question mark. In fact the limits to the left of the question mark might not even exist.

$$\lim_{k \rightarrow \infty} f(x_k) = \lim_{k \rightarrow \infty} \left(\lim_{n \rightarrow \infty} f_n(x_k) \right) \stackrel{?}{=} \lim_{n \rightarrow \infty} \left(\lim_{k \rightarrow \infty} f_n(x_k) \right) = \lim_{n \rightarrow \infty} f_n(x) = f(x).$$

In particular, we wish to find conditions on the sequence $\{f_n\}$ so that the above equation holds. It turns out that if we only require pointwise convergence, then the limit of a sequence of functions need not be continuous, and the above equation need not hold.

Example 6.2.1: Let $f_n: [0, 1] \rightarrow \mathbb{R}$ be defined as

$$f_n(x) := \begin{cases} 1 - nx & \text{if } x < 1/n, \\ 0 & \text{if } x \geq 1/n. \end{cases}$$

See Figure 6.2.

Each function f_n is continuous. Fix an $x \in (0, 1]$. If $n \geq 1/x$, then $x \geq 1/n$. Therefore for $n \geq 1/x$ we have $f_n(x) = 0$, and so

$$\lim_{n \rightarrow \infty} f_n(x) = 0.$$

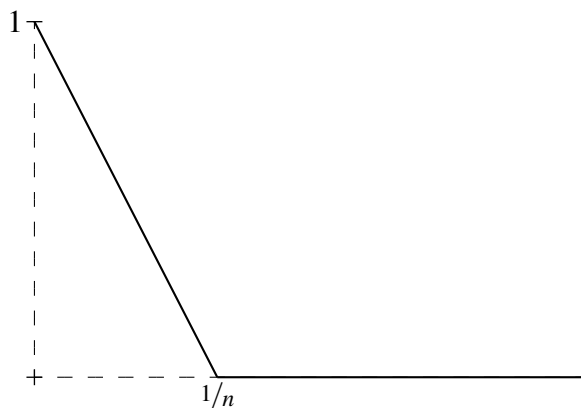
On the other hand if $x = 0$, then

$$\lim_{n \rightarrow \infty} f_n(0) = \lim_{n \rightarrow \infty} 1 = 1.$$

Thus the pointwise limit of f_n is the function $f: [0, 1] \rightarrow \mathbb{R}$ defined by

$$f(x) := \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases}$$

The function f is not continuous at 0.

Figure 6.2: Graph of $f_n(x)$.

If we, however, require the convergence to be uniform, the limits can be interchanged.

Theorem 6.2.2. *Let $\{f_n\}$ be a sequence of continuous functions $f_n: S \rightarrow \mathbb{R}$ converging uniformly to $f: S \rightarrow \mathbb{R}$. Then f is continuous.*

Proof. Let $x \in S$ be fixed. Let $\{x_n\}$ be a sequence in S converging to x .

Let $\varepsilon > 0$ be given. As f_k converges uniformly to f , we find a $k \in \mathbb{N}$ such that

$$|f_k(y) - f(y)| < \varepsilon/3$$

for all $y \in S$. As f_k is continuous at x , we find an $N \in \mathbb{N}$ such that for $m \geq N$ we have

$$|f_k(x_m) - f_k(x)| < \varepsilon/3.$$

Thus for $m \geq N$ we have

$$\begin{aligned} |f(x_m) - f(x)| &= |f(x_m) - f_k(x_m) + f_k(x_m) - f_k(x) + f_k(x) - f(x)| \\ &\leq |f(x_m) - f_k(x_m)| + |f_k(x_m) - f_k(x)| + |f_k(x) - f(x)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

Therefore $\{f(x_m)\}$ converges to $f(x)$ and hence f is continuous at x . As x was arbitrary, f is continuous everywhere. \square

6.2.2 Integral of the limit

Again, if we simply require pointwise convergence, then the integral of a limit of a sequence of functions need not be equal to the limit of the integrals.

Example 6.2.3: Let $f_n: [0, 1] \rightarrow \mathbb{R}$ be defined as

$$f_n(x) := \begin{cases} 0 & \text{if } x = 0, \\ n - n^2x & \text{if } 0 < x < 1/n, \\ 0 & \text{if } x \geq 1/n. \end{cases}$$

See Figure 6.3.

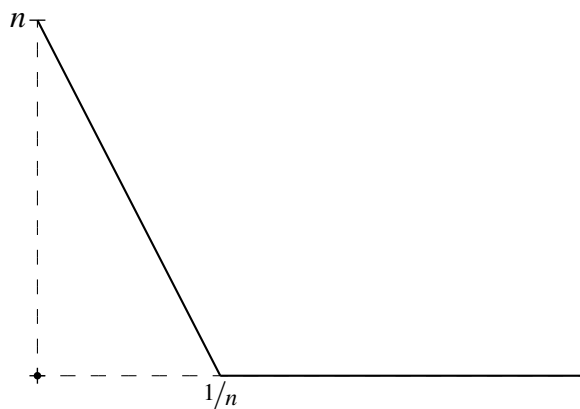


Figure 6.3: Graph of $f_n(x)$.

Each f_n is Riemann integrable (it is continuous on $(0, 1]$ and bounded), and it is easy to see

$$\int_0^1 f_n = \int_0^{1/n} (n - n^2x) dx = 1/2.$$

Let us compute the pointwise limit of $\{f_n\}$. Fix an $x \in (0, 1]$. For $n \geq 1/x$ we have $x \geq 1/n$ and so $f_n(x) = 0$. Therefore

$$\lim_{n \rightarrow \infty} f_n(x) = 0.$$

We also have $f_n(0) = 0$ for all n . Therefore the pointwise limit of $\{f_n\}$ is the zero function. Thus

$$1/2 = \lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx \neq \int_0^1 \left(\lim_{n \rightarrow \infty} f_n(x) \right) dx = \int_0^1 0 dx = 0.$$

But if we again require the convergence to be uniform, the limits can be interchanged.

Theorem 6.2.4. Let $\{f_n\}$ be a sequence of Riemann integrable functions $f_n: [a, b] \rightarrow \mathbb{R}$ converging uniformly to $f: [a, b] \rightarrow \mathbb{R}$. Then f is Riemann integrable and

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_a^b f_n.$$

Proof. Let $\varepsilon > 0$ be given. As f_n goes to f uniformly, we find an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $|f_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$ for all $x \in [a, b]$. In particular, by reverse triangle inequality $|f(x)| < \frac{\varepsilon}{2(b-a)} + |f_n(x)|$ for all x , hence f is bounded as f_n is bounded. Note that f_n is integrable and compute

$$\begin{aligned} \overline{\int_a^b f} - \underline{\int_a^b f} &= \overline{\int_a^b (f(x) - f_n(x) + f_n(x)) dx} - \underline{\int_a^b (f(x) - f_n(x) + f_n(x)) dx} \\ &= \overline{\int_a^b (f(x) - f_n(x)) dx} + \overline{\int_a^b f_n(x) dx} - \underline{\int_a^b (f(x) - f_n(x)) dx} - \underline{\int_a^b f_n(x) dx} \\ &= \overline{\int_a^b (f(x) - f_n(x)) dx} + \int_a^b f_n(x) dx - \underline{\int_a^b (f(x) - f_n(x)) dx} - \int_a^b f_n(x) dx \\ &= \overline{\int_a^b (f(x) - f_n(x)) dx} - \underline{\int_a^b (f(x) - f_n(x)) dx} \\ &\leq \frac{\varepsilon}{2(b-a)}(b-a) + \frac{\varepsilon}{2(b-a)}(b-a) = \varepsilon. \end{aligned}$$

The inequality follows from Proposition 5.1.8 and using the fact that for all $x \in [a, b]$ we have $\frac{-\varepsilon}{2(b-a)} < f(x) - f_n(x) < \frac{\varepsilon}{2(b-a)}$. As $\varepsilon > 0$ was arbitrary, f is Riemann integrable.

Finally we compute $\int_a^b f$. We apply Proposition 5.1.10 in the calculation. Again, for $n \geq M$ (where M is the same as above) we have

$$\begin{aligned} \left| \int_a^b f - \int_a^b f_n \right| &= \left| \int_a^b (f(x) - f_n(x)) dx \right| \\ &\leq \frac{\varepsilon}{2(b-a)}(b-a) = \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

Therefore $\{\int_a^b f_n\}$ converges to $\int_a^b f$. □

Example 6.2.5: Suppose we wish to compute

$$\lim_{n \rightarrow \infty} \int_0^1 \frac{nx + \sin(nx^2)}{n} dx.$$

It is impossible to compute the integrals for any particular n using calculus as $\sin(nx^2)$ has no closed-form antiderivative. However, we can compute the limit. We have shown before that $\frac{nx + \sin(nx^2)}{n}$ converges uniformly on $[0, 1]$ to x . By Theorem 6.2.4, the limit exists and

$$\lim_{n \rightarrow \infty} \int_0^1 \frac{nx + \sin(nx^2)}{n} dx = \int_0^1 x dx = 1/2.$$

Example 6.2.6: If convergence is only pointwise, the limit need not even be Riemann integrable. On $[0, 1]$ define

$$f_n(x) := \begin{cases} 1 & \text{if } x = p/q \text{ in lowest terms and } q \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

The function f_n differs from the zero function at finitely many points; there are only finitely many fractions in $[0, 1]$ with denominator less than or equal to n . So f_n is integrable and $\int_0^1 f_n = \int_0^1 0 = 0$. It is an easy exercise to show that f_n converges pointwise to the Dirichlet function

$$f(x) := \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{otherwise,} \end{cases}$$

which is not Riemann integrable.

6.2.3 Exercises

Exercise 6.2.1: While uniform convergence can preserve continuity, it does not preserve differentiability. Find an explicit example of a sequence of differentiable functions on $[-1, 1]$ that converge uniformly to a function f such that f is not differentiable. Hint: Consider $|x|^{1+1/n}$, show that these functions are differentiable, converge uniformly, and then show that the limit is not differentiable.

Exercise 6.2.2: Let $f_n(x) = \frac{x^n}{n}$. Show that f_n converges uniformly to a differentiable function f on $[0, 1]$ (find f). However, show that $f'(1) \neq \lim_{n \rightarrow \infty} f'_n(1)$.

Note: The previous two exercises show that we cannot simply swap limits with derivatives, even if the convergence is uniform. See also Exercise 6.2.7 below.

Exercise 6.2.3: Let $f: [0, 1] \rightarrow \mathbb{R}$ be a Riemann integrable (hence bounded) function. Find $\lim_{n \rightarrow \infty} \int_0^1 \frac{f(x)}{n} dx$.

Exercise 6.2.4: Show $\lim_{n \rightarrow \infty} \int_1^2 e^{-nx^2} dx = 0$. Feel free to use what you know about the exponential function from calculus.

Exercise 6.2.5: Find an example of a sequence of continuous functions on $(0, 1)$ that converges pointwise to a continuous function on $(0, 1)$, but the convergence is not uniform.

Note: In the previous exercise, $(0, 1)$ was picked for simplicity. For a more challenging exercise, replace $(0, 1)$ with $[0, 1]$.

Exercise 6.2.6: True/False; prove or find a counterexample to the following statement: If $\{f_n\}$ is a sequence of everywhere discontinuous functions on $[0, 1]$ that converge uniformly to a function f , then f is everywhere discontinuous.

Exercise 6.2.7: For a continuously differentiable function $f: [a, b] \rightarrow \mathbb{R}$, define

$$\|f\|_{C^1} := \|f\|_u + \|f'\|_u.$$

Suppose that $\{f_n\}$ is a sequence of continuously differentiable functions such that for every $\varepsilon > 0$, there exists an M such that for all $n, k \geq M$ we have

$$\|f_n - f_k\|_{C^1} < \varepsilon.$$

Show that $\{f_n\}$ converges uniformly to some continuously differentiable function $f: [a, b] \rightarrow \mathbb{R}$.

For the following two exercises let us define for a Riemann integrable function $f: [0, 1] \rightarrow \mathbb{R}$ the following number

$$\|f\|_{L^1} := \int_0^1 |f(x)| dx.$$

It is true that $|f|$ is integrable whenever f is, see Exercise 5.2.15. This norm defines another very common type of convergence called the L^1 -convergence, that is however a bit more subtle.

Exercise 6.2.8: Suppose that $\{f_n\}$ is a sequence of Riemann integrable functions on $[0, 1]$ that converges uniformly to 0. Show that

$$\lim_{n \rightarrow \infty} \|f_n\|_{L^1} = 0.$$

Exercise 6.2.9: Find a sequence of Riemann integrable functions $\{f_n\}$ on $[0, 1]$ that converges pointwise to 0, but

$$\lim_{n \rightarrow \infty} \|f_n\|_{L^1} \text{ does not exist (is } \infty).$$

Exercise 6.2.10 (Hard): Prove Dini's theorem: Let $f_n: [a, b] \rightarrow \mathbb{R}$ be a sequence of functions such that

$$0 \leq f_{n+1}(x) \leq f_n(x) \leq \dots \leq f_1(x) \quad \text{for all } n \in \mathbb{N}.$$

Suppose that $\{f_n\}$ converges pointwise to 0. Show that $\{f_n\}$ converges to zero uniformly.

Exercise 6.2.11: Suppose that $f_n: [a, b] \rightarrow \mathbb{R}$ is a sequence of functions that converges pointwise to a continuous $f: [a, b] \rightarrow \mathbb{R}$. Suppose that for any $x \in [a, b]$ the sequence $\{|f_n(x) - f(x)|\}$ is monotone. Show that the sequence $\{f_n\}$ converges uniformly.

Exercise 6.2.12: Find a sequence of Riemann integrable functions $f_n: [0, 1] \rightarrow \mathbb{R}$ such that f_n converges to zero pointwise, and such that a) $\{\int_0^1 f_n\}_{n=1}^\infty$ increases without bound, b) $\{\int_0^1 f_n\}_{n=1}^\infty$ is the sequence $-1, 1, -1, 1, -1, 1, \dots$

6.3 Picard's theorem

Note: 1–2 lectures (can be safely skipped)

A course such as this one should have a *pièce de résistance* caliber theorem. We pick a theorem whose proof combines everything we have learned. It is more sophisticated than the fundamental theorem of calculus, the first highlight theorem of this course. The theorem we are talking about is Picard's theorem* on existence and uniqueness of a solution to an ordinary differential equation. Both the statement and the proof are beautiful examples of what one can do with all that we have learned. It is also a good example of how analysis is applied as differential equations are indispensable in science.

6.3.1 First order ordinary differential equation

Modern science is described in the language of *differential equations*. That is equations that involve not only the unknown, but also its derivatives. The simplest nontrivial form of a differential equation is the so-called *first order ordinary differential equation*

$$y' = F(x, y).$$

Generally we also specify that $y(x_0) = y_0$. The solution of the equation is a function $y(x)$ such that $y(x_0) = y_0$ and $y'(x) = F(x, y(x))$.

When F involves only the x variable, the solution is given by the fundamental theorem of calculus. On the other hand, when F depends on both x and y we need far more firepower. It is not always true that a solution exists, and if it does, that it is the unique solution. Picard's theorem gives us certain sufficient conditions for existence and uniqueness.

6.3.2 The theorem

We need a definition of continuity in two variables. First, a point in the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is denoted by an ordered pair (x, y) . To make matters simple, let us give the following sequential definition of continuity.

Definition 6.3.1. Let $U \subset \mathbb{R}^2$ be a set and $F: U \rightarrow \mathbb{R}$ be a function. Let $(x, y) \in U$ be a point. The function F is *continuous* at (x, y) if for every sequence $\{(x_n, y_n)\}_{n=1}^{\infty}$ of points in U such that $\lim x_n = x$ and $\lim y_n = y$, we have that

$$\lim_{n \rightarrow \infty} F(x_n, y_n) = F(x, y).$$

We say F is continuous if it is continuous at all points in U .

*Named for the French mathematician Charles Émile Picard (1856–1941).

Theorem 6.3.2 (Picard's theorem on existence and uniqueness). *Let $I, J \subset \mathbb{R}$ be closed bounded intervals, let I_0 and J_0 be their interiors, and let $(x_0, y_0) \in I_0 \times J_0$. Suppose $F: I \times J \rightarrow \mathbb{R}$ is continuous and Lipschitz in the second variable, that is, there exists a number L such that*

$$|F(x, y) - F(x, z)| \leq L|y - z| \quad \text{for all } y, z \in J, x \in I.$$

Then there exists an $h > 0$ and a unique differentiable function $f: [x_0 - h, x_0 + h] \rightarrow J \subset \mathbb{R}$, such that

$$f'(x) = F(x, f(x)) \quad \text{and} \quad f(x_0) = y_0. \quad (6.1)$$

Proof. Suppose that we could find a solution f . By the fundamental theorem of calculus we can integrate the equation $f'(x) = F(x, f(x))$, $f(x_0) = y_0$, and write (6.1) as the integral equation

$$f(x) = y_0 + \int_{x_0}^x F(t, f(t)) dt. \quad (6.2)$$

The idea of our proof is that we will try to plug in approximations to a solution to the right-hand side of (6.2) to get better approximations on the left hand side of (6.2). We hope that in the end the sequence will converge and solve (6.2) and hence (6.1). The technique below is called *Picard iteration*, and the individual functions f_k are called the *Picard iterates*.

Without loss of generality, suppose that $x_0 = 0$ (exercise below). Another exercise tells us that F is bounded as it is continuous. Let $M := \sup\{|F(x, y)| : (x, y) \in I \times J\}$. Without loss of generality, we can assume $M > 0$ (why?). Pick $\alpha > 0$ such that $[-\alpha, \alpha] \subset I$ and $[y_0 - \alpha, y_0 + \alpha] \subset J$. Define

$$h := \min \left\{ \alpha, \frac{\alpha}{M + L\alpha} \right\}. \quad (6.3)$$

Now note that $[-h, h] \subset I$.

Set $f_0(x) := y_0$. We define f_k inductively. Assuming that $f_{k-1}([-h, h]) \subset [y_0 - \alpha, y_0 + \alpha]$, we see that $F(t, f_{k-1}(t))$ is a well defined function of t for $t \in [-h, h]$. Further if f_{k-1} is continuous on $[-h, h]$, then $F(t, f_{k-1}(t))$ is continuous as a function of t on $[-h, h]$ (left as an exercise). Define

$$f_k(x) := y_0 + \int_0^x F(t, f_{k-1}(t)) dt,$$

and f_k is continuous on $[-h, h]$ by the fundamental theorem of calculus. To see that f_k maps $[-h, h]$ to $[y_0 - \alpha, y_0 + \alpha]$, we compute for $x \in [-h, h]$

$$|f_k(x) - y_0| = \left| \int_0^x F(t, f_{k-1}(t)) dt \right| \leq M|x| \leq Mh \leq M \frac{\alpha}{M + L\alpha} \leq \alpha.$$

We now define f_{k+1} and so on, and we have defined a sequence $\{f_k\}$ of functions. We need to show that it converges to a function f that solves the equation (6.2) and therefore (6.1).

We wish to show that the sequence $\{f_k\}$ converges uniformly to some function on $[-h, h]$. First, for $t \in [-h, h]$ we have the following useful bound

$$|F(t, f_n(t)) - F(t, f_k(t))| \leq L|f_n(t) - f_k(t)| \leq L\|f_n - f_k\|_u,$$

where $\|f_n - f_k\|_u$ is the uniform norm, that is the supremum of $|f_n(t) - f_k(t)|$ for $t \in [-h, h]$. Now note that $|x| \leq h \leq \frac{\alpha}{M+L\alpha}$. Therefore

$$\begin{aligned} |f_n(x) - f_k(x)| &= \left| \int_0^x F(t, f_{n-1}(t)) dt - \int_0^x F(t, f_{k-1}(t)) dt \right| \\ &= \left| \int_0^x F(t, f_{n-1}(t)) - F(t, f_{k-1}(t)) dt \right| \\ &\leq L\|f_{n-1} - f_{k-1}\|_u |x| \\ &\leq \frac{L\alpha}{M+L\alpha} \|f_{n-1} - f_{k-1}\|_u. \end{aligned}$$

Let $C := \frac{L\alpha}{M+L\alpha}$ and note that $C < 1$. Taking supremum on the left-hand side we get

$$\|f_n - f_k\|_u \leq C \|f_{n-1} - f_{k-1}\|_u.$$

Without loss of generality, suppose that $n \geq k$. Then by induction we can show that

$$\|f_n - f_k\|_u \leq C^k \|f_{n-k} - f_0\|_u.$$

For $x \in [-h, h]$ we have

$$|f_{n-k}(x) - f_0(x)| = |f_{n-k}(x) - y_0| \leq \alpha.$$

Therefore

$$\|f_n - f_k\|_u \leq C^k \|f_{n-k} - f_0\|_u \leq C^k \alpha.$$

As $C < 1$, $\{f_n\}$ is uniformly Cauchy and by Proposition 6.1.13 we obtain that $\{f_n\}$ converges uniformly on $[-h, h]$ to some function $f: [-h, h] \rightarrow \mathbb{R}$. The function f is the uniform limit of continuous functions and therefore continuous.

We now need to show that f solves (6.2). First, as before we notice

$$|F(t, f_n(t)) - F(t, f(t))| \leq L|f_n(t) - f(t)| \leq L\|f_n - f\|_u.$$

As $\|f_n - f\|_u$ converges to 0, then $F(t, f_n(t))$ converges uniformly to $F(t, f(t))$ where $t \in [-h, h]$. Then the convergence is then uniform (why?) on $[0, x]$ (or $[x, 0]$ if $x < 0$) if $x \in [-h, h]$. Therefore,

$$\begin{aligned} y_0 + \int_0^x F(t, f(t)) dt &= y_0 + \int_0^x F(t, \lim_{n \rightarrow \infty} f_n(t)) dt \\ &= y_0 + \int_0^x \lim_{n \rightarrow \infty} F(t, f_n(t)) dt && \text{(by continuity of } F) \\ &= \lim_{n \rightarrow \infty} \left(y_0 + \int_0^x F(t, f_n(t)) dt \right) && \text{(by uniform convergence)} \\ &= \lim_{n \rightarrow \infty} f_{n+1}(x) = f(x). \end{aligned}$$

We apply the fundamental theorem of calculus to show that f is differentiable and its derivative is $F(x, f(x))$. It is obvious that $f(0) = y_0$.

Finally, what is left to do is to show uniqueness. Suppose $g: [-h, h] \rightarrow \mathbb{R}$ is another solution. As before we use the fact that $|F(t, f(t)) - F(t, g(t))| \leq L\|f - g\|_u$. Then

$$\begin{aligned} |f(x) - g(x)| &= \left| y_0 + \int_0^x F(t, f(t)) dt - \left(y_0 + \int_0^x F(t, g(t)) dt \right) \right| \\ &= \left| \int_0^x F(t, f(t)) - F(t, g(t)) dt \right| \\ &\leq L\|f - g\|_u |x| \leq Lh\|f - g\|_u \leq \frac{L\alpha}{M + L\alpha} \|f - g\|_u. \end{aligned}$$

As before, $C = \frac{L\alpha}{M + L\alpha} < 1$. By taking supremum over $x \in [-h, h]$ on the left hand side we obtain

$$\|f - g\|_u \leq C\|f - g\|_u.$$

This is only possible if $\|f - g\|_u = 0$. Therefore, $f = g$, and the solution is unique. \square

6.3.3 Examples

Let us look at some examples. The proof of the theorem gives us an explicit way to find an h that works. It does not, however, give us the best h . It is often possible to find a much larger h for which the theorem works.

The proof also gives us the Picard iterates as approximations to the solution. So the proof actually tells us how to obtain the solution, not just that the solution exists.

Example 6.3.3: Consider

$$f'(x) = f(x), \quad f(0) = 1.$$

That is, we let $F(x, y) = y$, and we are looking for a function f such that $f'(x) = f(x)$. We pick any I that contains 0 in the interior. We pick an arbitrary J that contains 1 in its interior. We can use $L = 1$. The theorem guarantees an $h > 0$ such that there exists a unique solution $f: [-h, h] \rightarrow \mathbb{R}$. This solution is usually denoted by

$$e^x := f(x).$$

We leave it to the reader to verify that by picking I and J large enough the proof of the theorem guarantees that we are able to pick α such that we get any h we want as long as $h < 1/2$. We omit the calculation.

Of course, we know (though we have not proved) that this function exists as a function for all x , so an arbitrary h ought to work. By same reasoning as above, we can see that no matter what x_0 and y_0 are, the proof guarantees an arbitrary h as long as $h < 1/2$. Fix such an h . We get a unique function defined on $[x_0 - h, x_0 + h]$. After defining the function on $[-h, h]$ we find a solution on the

interval $[0, 2h]$ and notice that the two functions must coincide on $[0, h]$ by uniqueness. We can thus iteratively construct the exponential for all $x \in \mathbb{R}$. Note that up until now we did not yet have proof of the existence of the exponential function.

Let us see the Picard iterates for this function. We start with $f_0(x) := 1$. Then

$$\begin{aligned} f_1(x) &= 1 + \int_0^x f_0(s) \, ds = x + 1, \\ f_2(x) &= 1 + \int_0^x f_1(s) \, ds = 1 + \int_0^x s + 1 \, ds = \frac{x^2}{2} + x + 1, \\ f_3(x) &= 1 + \int_0^x f_2(s) \, ds = 1 + \int_0^x \frac{s^2}{2} + s + 1 \, ds = \frac{x^3}{6} + \frac{x^2}{2} + x + 1. \end{aligned}$$

We recognize the beginning of the Taylor series for the exponential.

Example 6.3.4: Suppose we have the equation

$$f'(x) = (f(x))^2 \quad \text{and} \quad f(0) = 1.$$

From elementary differential equations we know that

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

is the solution. Do note that the solution is only defined on $(-\infty, 1)$. That is, we are able to use $h < 1$, but never a larger h . Note that the function that takes y to y^2 is not Lipschitz as a function on all of \mathbb{R} . As we approach $x = 1$ from the left the solution becomes larger and larger. The derivative of the solution grows as y^2 , and therefore the L required will have to be larger and larger as y_0 grows. Thus if we apply the theorem with x_0 close to 1 and $y_0 = \frac{1}{1-x_0}$ we find that the h that the proof guarantees will be smaller and smaller as x_0 approaches 1.

By picking α correctly, the proof of the theorem guarantees $h = 1 - \sqrt{3}/2 \approx 0.134$ (we omit the calculation) for $x_0 = 0$ and $y_0 = 1$, even though we see from above that any $h < 1$ should work.

Example 6.3.5: Suppose we start with the equation

$$f'(x) = 2\sqrt{|f(x)|}, \quad f(0) = 0.$$

Note that $F(x, y) = 2\sqrt{|y|}$ is not Lipschitz in y (why?). Therefore the equation does not satisfy the hypotheses of the theorem. The function

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

is a solution, but $g(x) = 0$ is also a solution. So a solution is not unique.

Example 6.3.6: Consider $y' = \varphi(x)$ where $\varphi(x) := 0$ if $x \in \mathbb{Q}$ and $\varphi(x) := 1$ if $x \notin \mathbb{Q}$. The equation has no solution regardless of the initial conditions we choose. A solution would have derivative φ , but φ does not have the intermediate value property at any point (why?). So no solution can exist by Darboux's theorem. Therefore some continuity hypothesis is necessary on F for existence of a solution.

6.3.4 Exercises

Exercise 6.3.1: Let $I, J \subset \mathbb{R}$ be intervals. Let $F : I \times J \rightarrow \mathbb{R}$ be a continuous function of two variables and suppose that $f : I \rightarrow J$ be a continuous function. Show that $F(x, f(x))$ is a continuous function on I .

Exercise 6.3.2: Let $I, J \subset \mathbb{R}$ be closed bounded intervals. Show that if $F : I \times J \rightarrow \mathbb{R}$ is continuous, then F is bounded.

Exercise 6.3.3: We have proved Picard's theorem under the assumption that $x_0 = 0$. Prove the full statement of Picard's theorem for an arbitrary x_0 .

Exercise 6.3.4: Let $f'(x) = xf(x)$ be our equation. Start with the initial condition $f(0) = 2$ and find the Picard iterates f_0, f_1, f_2, f_3, f_4 .

Exercise 6.3.5: Suppose that $F : I \times J \rightarrow \mathbb{R}$ is a function that is continuous in the first variable, that is, for any fixed y the function that takes x to $F(x, y)$ is continuous. Further, suppose that F is Lipschitz in the second variable, that is, there exists a number L such that

$$|F(x, y) - F(x, z)| \leq L|y - z| \quad \text{for all } y, z \in J, x \in I.$$

Show that F is continuous as a function of two variables. Therefore, the hypotheses in the theorem could be made even weaker.

Chapter 7

Metric Spaces

7.1 Metric spaces

Note: 1.5 lectures

As mentioned in the introduction, the main idea in analysis is to take limits. In chapter 2 we learned to take limits of sequences of real numbers. And in chapter 3 we learned to take limits of functions as a real number approached some other real number.

We want to take limits in more complicated contexts. For example, we want to have sequences of points in 3-dimensional space. We wish to define continuous functions of several variables. We even want to define functions on spaces that are a little harder to describe, such as the surface of the earth. We still want to talk about limits there.

Finally, we have seen the limit of a sequence of functions in chapter 6. We wish to unify all these notions so that we do not have to reprove theorems over and over again in each context. The concept of a metric space is an elementary yet powerful tool in analysis. And while it is not sufficient to describe every type of limit we find in modern analysis, it gets us very far indeed.

Definition 7.1.1. Let X be a set, and let $d: X \times X \rightarrow \mathbb{R}$ be a function such that

- (i) $d(x, y) \geq 0$ for all x, y in X ,
- (ii) $d(x, y) = 0$ if and only if $x = y$,
- (iii) $d(x, y) = d(y, x)$,
- (iv) $d(x, z) \leq d(x, y) + d(y, z)$ (*triangle inequality*).

Then the pair (X, d) is called a *metric space*. The function d is called the *metric* or sometimes the *distance function*. Sometimes we just say X is a metric space if the metric is clear from context.

The geometric idea is that d is the distance between two points. Items (i)–(iii) have obvious geometric interpretation: distance is always nonnegative, the only point that is distance 0 away from

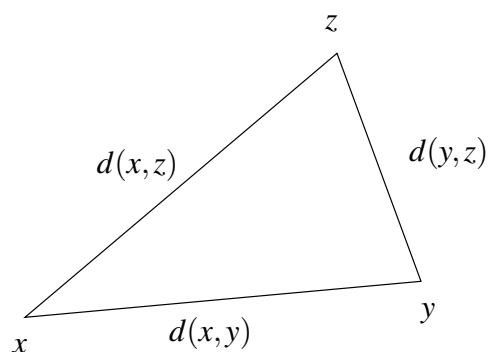


Figure 7.1: Diagram of the triangle inequality in metric spaces.

x is x itself, and finally that the distance from x to y is the same as the distance from y to x . The triangle inequality (iv) has the interpretation given in Figure 7.1.

For the purposes of drawing, it is convenient to draw figures and diagrams in the plane and have the metric be the standard distance. However, that is only one particular metric space. Just because a certain fact seems to be clear from drawing a picture does not mean it is true. You might be getting sidetracked by intuition from euclidean geometry, whereas the concept of a metric space is a lot more general.

Let us give some examples of metric spaces.

Example 7.1.2: The set of real numbers \mathbb{R} is a metric space with the metric

$$d(x, y) := |x - y|.$$

Items (i)–(iii) of the definition are easy to verify. The triangle inequality (iv) follows immediately from the standard triangle inequality for real numbers:

$$d(x, z) = |x - z| = |x - y + y - z| \leq |x - y| + |y - z| = d(x, y) + d(y, z).$$

This metric is the *standard metric on \mathbb{R}* . If we talk about \mathbb{R} as a metric space without mentioning a specific metric, we mean this particular metric.

Example 7.1.3: We can also put a different metric on the set of real numbers. For example, take the set of real numbers \mathbb{R} together with the metric

$$d(x, y) := \frac{|x - y|}{|x - y| + 1}.$$

Items (i)–(iii) are again easy to verify. The triangle inequality (iv) is a little bit more difficult. Note that $d(x, y) = \varphi(|x - y|)$ where $\varphi(t) = \frac{t}{t+1}$ and note that φ is an increasing function (positive

derivative). Hence

$$\begin{aligned} d(x, z) &= \varphi(|x - z|) = \varphi(|x - y + y - z|) \leq \varphi(|x - y| + |y - z|) \\ &= \frac{|x - y| + |y - z|}{|x - y| + |y - z| + 1} = \frac{|x - y|}{|x - y| + |y - z| + 1} + \frac{|y - z|}{|x - y| + |y - z| + 1} \\ &\leq \frac{|x - y|}{|x - y| + 1} + \frac{|y - z|}{|y - z| + 1} = d(x, y) + d(y, z). \end{aligned}$$

Here we have an example of a nonstandard metric on \mathbb{R} . With this metric we see for example that $d(x, y) < 1$ for all $x, y \in \mathbb{R}$. That is, any two points are less than 1 unit apart.

An important metric space is the n -dimensional *euclidean space* $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$. We use the following notation for points: $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. We also simply write $0 \in \mathbb{R}^n$ to mean the vector $(0, 0, \dots, 0)$. Before making \mathbb{R}^n a metric space, let us prove an important inequality, the so-called Cauchy-Schwarz inequality.

Lemma 7.1.4 (Cauchy-Schwarz inequality). *Take $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$. Then*

$$\left(\sum_{j=1}^n x_j y_j \right)^2 \leq \left(\sum_{j=1}^n x_j^2 \right) \left(\sum_{j=1}^n y_j^2 \right).$$

Proof. Any square of a real number is nonnegative. Hence any sum of squares is nonnegative:

$$\begin{aligned} 0 &\leq \sum_{j=1}^n \sum_{k=1}^n (x_j y_k - x_k y_j)^2 \\ &= \sum_{j=1}^n \sum_{k=1}^n (x_j^2 y_k^2 + x_k^2 y_j^2 - 2x_j x_k y_j y_k) \\ &= \left(\sum_{j=1}^n x_j^2 \right) \left(\sum_{k=1}^n y_k^2 \right) + \left(\sum_{j=1}^n y_j^2 \right) \left(\sum_{k=1}^n x_k^2 \right) - 2 \left(\sum_{j=1}^n x_j y_j \right) \left(\sum_{k=1}^n x_k y_k \right) \end{aligned}$$

We relabel and divide by 2 to obtain

$$0 \leq \left(\sum_{j=1}^n x_j^2 \right) \left(\sum_{j=1}^n y_j^2 \right) - \left(\sum_{j=1}^n x_j y_j \right)^2,$$

which is precisely what we wanted. □

Example 7.1.5: Let us construct standard metric for \mathbb{R}^n . Define

$$d(x, y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{j=1}^n (x_j - y_j)^2}.$$

For $n = 1$, the real line, this metric agrees with what we did above. Again, the only tricky part of the definition to check is the triangle inequality. It is less messy to work with the square of the metric. In the following, note the use of the Cauchy-Schwarz inequality.

$$\begin{aligned}
 d(x, z)^2 &= \sum_{j=1}^n (x_j - z_j)^2 \\
 &= \sum_{j=1}^n (x_j - y_j + y_j - z_j)^2 \\
 &= \sum_{j=1}^n \left((x_j - y_j)^2 + (y_j - z_j)^2 + 2(x_j - y_j)(y_j - z_j) \right) \\
 &= \sum_{j=1}^n (x_j - y_j)^2 + \sum_{j=1}^n (y_j - z_j)^2 + \sum_{j=1}^n 2(x_j - y_j)(y_j - z_j) \\
 &\leq \sum_{j=1}^n (x_j - y_j)^2 + \sum_{j=1}^n (y_j - z_j)^2 + 2\sqrt{\sum_{j=1}^n (x_j - y_j)^2 \sum_{j=1}^n (y_j - z_j)^2} \\
 &= \left(\sqrt{\sum_{j=1}^n (x_j - y_j)^2} + \sqrt{\sum_{j=1}^n (y_j - z_j)^2} \right)^2 = (d(x, y) + d(y, z))^2.
 \end{aligned}$$

Taking the square root of both sides we obtain the correct inequality.

Example 7.1.6: An example to keep in mind is the so-called *discrete metric*. Let X be any set and define

$$d(x, y) := \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

That is, all points are equally distant from each other. When X is a finite set, we can draw a diagram, see for example Figure 7.2. Things become subtle when X is an infinite set such as the real numbers.

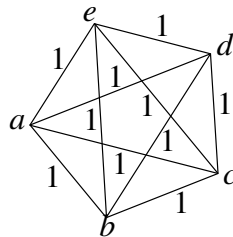


Figure 7.2: Sample discrete metric space $\{a, b, c, d, e\}$, the distance between any two points is 1.

While this particular example seldom comes up in practice, it gives a useful “smell test.” If you make a statement about metric spaces, try it with the discrete metric. To show that (X, d) is indeed a metric space is left as an exercise.

Example 7.1.7: Let $C([a, b], \mathbb{R})$ be the set of continuous real-valued functions on the interval $[a, b]$. Define the metric on $C([a, b], \mathbb{R})$ as

$$d(f, g) := \sup_{x \in [a, b]} |f(x) - g(x)|.$$

Let us check the properties. First, $d(f, g)$ is finite as $|f(x) - g(x)|$ is a continuous function on a closed bounded interval $[a, b]$, and so is bounded. It is clear that $d(f, g) \geq 0$, it is the supremum of nonnegative numbers. If $f = g$ then $|f(x) - g(x)| = 0$ for all x and hence $d(f, g) = 0$. Conversely if $d(f, g) = 0$, then for any x we have $|f(x) - g(x)| \leq d(f, g) = 0$ and hence $f(x) = g(x)$ for all x and $f = g$. That $d(f, g) = d(g, f)$ is equally trivial. To show the triangle inequality we use the standard triangle inequality.

$$\begin{aligned} d(f, h) &= \sup_{x \in [a, b]} |f(x) - g(x)| = \sup_{x \in [a, b]} |f(x) - h(x) + h(x) - g(x)| \\ &\leq \sup_{x \in [a, b]} (|f(x) - h(x)| + |h(x) - g(x)|) \\ &\leq \sup_{x \in [a, b]} |f(x) - h(x)| + \sup_{x \in [a, b]} |h(x) - g(x)| = d(f, h) + d(h, g). \end{aligned}$$

When treating $C([a, b], \mathbb{R})$ as a metric space without mentioning a metric, we mean this particular metric. Notice that $d(f, g) = \|f - g\|_\infty$, the uniform norm of Definition 6.1.9.

This example may seem esoteric at first, but it turns out that working with spaces such as $C([a, b], \mathbb{R})$ is really the meat of a large part of modern analysis. Treating sets of functions as metric spaces allows us to abstract away a lot of the gubby detail and prove powerful results such as Picard’s theorem with less work.

Oftentimes it is useful to consider a subset of a larger metric space as a metric space itself. We obtain the following proposition, which has a trivial proof.

Proposition 7.1.8. *Let (X, d) be a metric space and $Y \subset X$, then the restriction $d|_{Y \times Y}$ is a metric on Y .*

Definition 7.1.9. If (X, d) is a metric space, $Y \subset X$, and $d' := d|_{Y \times Y}$, then (Y, d') is said to be a *subspace* of (X, d) .

It is common to simply write d for the metric on Y , as it is the restriction of the metric on X . Sometimes we will say that d' is the *subspace metric* and that Y has the *subspace topology*.

A subset of the real numbers is bounded whenever all its elements are at most some fixed distance from 0. We also define bounded sets in a metric space. When dealing with an arbitrary metric space there may not be some natural fixed point 0. For the purposes of boundedness it does not matter.

Definition 7.1.10. Let (X, d) be a metric space. A subset $S \subset X$ is said to be *bounded* if there exists a $p \in X$ and a $B \in \mathbb{R}$ such that

$$d(p, x) \leq B \quad \text{for all } x \in S.$$

We say that (X, d) is bounded if X itself is a bounded subset.

For example, the set of real numbers with the standard metric is not a bounded metric space. It is not hard to see that a subset of the real numbers is bounded in the sense of chapter 1 if and only if it is bounded as a subset of the metric space of real numbers with the standard metric.

On the other hand, if we take the real numbers with the discrete metric, then we obtain a bounded metric space. In fact, any set with the discrete metric is bounded.

7.1.1 Exercises

Exercise 7.1.1: Show that for any set X , the discrete metric ($d(x, y) = 1$ if $x \neq y$ and $d(x, x) = 0$) does give a metric space (X, d) .

Exercise 7.1.2: Let $X := \{0\}$ be a set. Can you make it into a metric space?

Exercise 7.1.3: Let $X := \{a, b\}$ be a set. Can you make it into two distinct metric spaces? (define two distinct metrics on it)

Exercise 7.1.4: Let the set $X := \{A, B, C\}$ represent 3 buildings on campus. Suppose we wish to our distance to be the time it takes to walk from one building to the other. It takes 5 minutes either way between buildings A and B . However, building C is on a hill and it takes 10 minutes from A and 15 minutes from B to get to C . On the other hand it takes 5 minutes to go from C to A and 7 minutes to go from C to B , as we are going downhill. Do these distances define a metric? If so, prove it, if not, say why not.

Exercise 7.1.5: Suppose that (X, d) is a metric space and $\varphi: [0, \infty] \rightarrow \mathbb{R}$ is an increasing function such that $\varphi(t) \geq 0$ for all t and $\varphi(t) = 0$ if and only if $t = 0$. Also suppose that φ is subadditive, that is $\varphi(s+t) \leq \varphi(s) + \varphi(t)$. Show that with $d'(x, y) := \varphi(d(x, y))$, we obtain a new metric space (X, d') .

Exercise 7.1.6: Let (X, d_X) and (Y, d_Y) be metric spaces.

a) Show that $(X \times Y, d)$ with $d((x_1, y_1), (x_2, y_2)) := d_X(x_1, x_2) + d_Y(y_1, y_2)$ is a metric space.

b) Show that $(X \times Y, d)$ with $d((x_1, y_1), (x_2, y_2)) := \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$ is a metric space.

Exercise 7.1.7: Let X be the set of continuous functions on $[0, 1]$. Let $\varphi: [0, 1] \rightarrow (0, \infty)$ be continuous. Define

$$d(f, g) := \int_0^1 |f(x) - g(x)| \varphi(x) dx.$$

Show that (X, d) is a metric space.

Exercise 7.1.8: Let (X, d) be a metric space. For nonempty bounded subsets A and B let

$$d(x, B) := \inf\{d(x, b) : b \in B\} \quad \text{and} \quad d(A, B) := \sup\{d(a, B) : a \in A\}.$$

Now define the Hausdorff metric as

$$d_H(A, B) := \max\{d(A, B), d(B, A)\}.$$

Note: d_H can be defined for arbitrary nonempty subsets if we allow the extended reals.

a) Let $Y \subset \mathcal{P}(X)$ be the set of bounded nonempty subsets. Show that (Y, d_H) is a metric space. b) Show by example that d itself is not a metric. That is, d is not always symmetric.

7.2 Open and closed sets

Note: 2 lectures

7.2.1 Topology

It is useful to define a so-called *topology*. That is we define closed and open sets in a metric space. Before doing so, let us define two special sets.

Definition 7.2.1. Let (X, d) be a metric space, $x \in X$ and $\delta > 0$. Then define the *open ball* or simply *ball* of radius δ around x as

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

Similarly we define the *closed ball* as

$$C(x, \delta) := \{y \in X : d(x, y) \leq \delta\}.$$

When we are dealing with different metric spaces, it is sometimes convenient to emphasize which metric space the ball is in. We do this by writing $B_X(x, \delta) := B(x, \delta)$ or $C_X(x, \delta) := C(x, \delta)$.

Example 7.2.2: Take the metric space \mathbb{R} with the standard metric. For $x \in \mathbb{R}$, and $\delta > 0$ we get

$$B(x, \delta) = (x - \delta, x + \delta) \quad \text{and} \quad C(x, \delta) = [x - \delta, x + \delta].$$

Example 7.2.3: Be careful when working on a subspace. Suppose we take the metric space $[0, 1]$ as a subspace of \mathbb{R} . Then in $[0, 1]$ we get

$$B(0, 1/2) = B_{[0,1]}(0, 1/2) = [0, 1/2).$$

This is of course different from $B_{\mathbb{R}}(0, 1/2) = (-1/2, 1/2)$. The important thing to keep in mind is which metric space we are working in.

Definition 7.2.4. Let (X, d) be a metric space. A set $V \subset X$ is *open* if for every $x \in V$, there exists a $\delta > 0$ such that $B(x, \delta) \subset V$. See Figure 7.3. A set $E \subset X$ is *closed* if the complement $E^c = X \setminus E$ is open. When the ambient space X is not clear from context we say V is open in X and E is closed in X .

If $x \in V$ and V is open, then we say that V is an *open neighborhood* of x (or sometimes just *neighborhood*).

Intuitively, an open set is a set that does not include its “boundary.” Note that not every set is either open or closed, in fact generally most subsets are neither.

Example 7.2.5: The set $[0, 1) \subset \mathbb{R}$ is neither open nor closed. First, every ball in \mathbb{R} around 0, $(-\delta, \delta)$ contains negative numbers and hence is not contained in $[0, 1)$ and so $[0, 1)$ is not open. Second, every ball in \mathbb{R} around 1, $(1 - \delta, 1 + \delta)$ contains numbers strictly less than 1 and greater than 0 (e.g. $1 - \delta/2$ as long as $\delta < 2$). Thus $\mathbb{R} \setminus [0, 1)$ is not open, and so $[0, 1)$ is not closed.

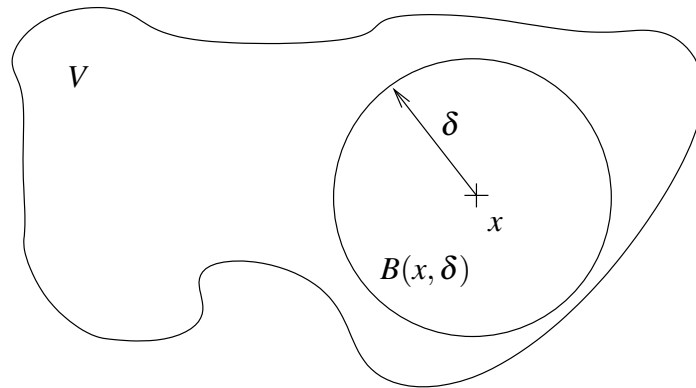


Figure 7.3: Open set in a metric space. Note that δ depends on x .

Proposition 7.2.6. Let (X, d) be a metric space.

(i) \emptyset and X are open in X .

(ii) If V_1, V_2, \dots, V_k are open then

$$\bigcap_{j=1}^k V_j$$

is also open. That is, finite intersection of open sets is open.

(iii) If $\{V_\lambda\}_{\lambda \in I}$ is an arbitrary collection of open sets, then

$$\bigcup_{\lambda \in I} V_\lambda$$

is also open. That is, union of open sets is open.

Note that the index set in (iii) is arbitrarily large. By $\bigcup_{\lambda \in I} V_\lambda$ we simply mean the set of all x such that $x \in V_\lambda$ for at least one $\lambda \in I$.

Proof. The set X and \emptyset are obviously open in X .

Let us prove (ii). If $x \in \bigcap_{j=1}^k V_j$, then $x \in V_j$ for all j . As V_j are all open, there exists a $\delta_j > 0$ for every j such that $B(x, \delta_j) \subset V_j$. Take $\delta := \min\{\delta_1, \dots, \delta_k\}$ and note that $\delta > 0$. We have $B(x, \delta) \subset B(x, \delta_j) \subset V_j$ for every j and thus $B(x, \delta) \subset \bigcap_{j=1}^k V_j$. Thus the intersection is open.

Let us prove (iii). If $x \in \bigcup_{\lambda \in I} V_\lambda$, then $x \in V_\lambda$ for some $\lambda \in I$. As V_λ is open then there exists a $\delta > 0$ such that $B(x, \delta) \subset V_\lambda$. But then $B(x, \delta) \subset \bigcup_{\lambda \in I} V_\lambda$ and so the union is open. \square

Example 7.2.7: The main thing to notice is the difference between items (ii) and (iii). Item (ii) is not true for an arbitrary intersection, for example $\bigcap_{n=1}^{\infty} (-1/n, 1/n) = \{0\}$, which is not open.

The proof of the following analogous proposition for closed sets is left as an exercise.

Proposition 7.2.8. *Let (X, d) be a metric space.*

(i) \emptyset and X are closed in X .

(ii) If $\{E_\lambda\}_{\lambda \in I}$ is an arbitrary collection of closed sets, then

$$\bigcap_{\lambda \in I} E_\lambda$$

is also closed. That is, intersection of closed sets is closed.

(iii) If E_1, E_2, \dots, E_k are closed then

$$\bigcup_{j=1}^k E_j$$

is also closed. That is, finite union of closed sets is closed.

We have not yet shown that the open ball is open and the closed ball is closed. Let us show this fact now to justify the terminology.

Proposition 7.2.9. *Let (X, d) be a metric space, $x \in X$, and $\delta > 0$. Then $B(x, \delta)$ is open and $C(x, \delta)$ is closed.*

Proof. Let $y \in B(x, \delta)$. Let $\alpha := \delta - d(x, y)$. Of course $\alpha > 0$. Now let $z \in B(y, \alpha)$. Then

$$d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \alpha = d(x, y) + \delta - d(x, y) = \delta.$$

Therefore $z \in B(x, \delta)$ for every $z \in B(y, \alpha)$. So $B(y, \alpha) \subset B(x, \delta)$ and $B(x, \delta)$ is open.

The proof that $C(x, \delta)$ is closed is left as an exercise. □

Again be careful about what is the ambient metric space. As $[0, 1/2)$ is an open ball in $[0, 1]$, this means that $[0, 1/2)$ is an open set in $[0, 1]$. On the other hand $[0, 1/2)$ is neither open nor closed in \mathbb{R} .

A useful way to think about an open set is a union of open balls. If U is open, then for each $x \in U$, there is a $\delta_x > 0$ (depending on x of course) such that $B(x, \delta_x) \subset U$. Then $U = \bigcup_{x \in U} B(x, \delta_x)$.

The proof of the following proposition is left as an exercise. Note that there are other open and closed sets in \mathbb{R} .

Proposition 7.2.10. *Let $a < b$ be two real numbers. Then (a, b) , (a, ∞) , and $(-\infty, b)$ are open in \mathbb{R} . Also $[a, b]$, $[a, \infty)$, and $(-\infty, b]$ are closed in \mathbb{R} .*

7.2.2 Connected sets

Definition 7.2.11. A nonempty metric space (X, d) is *connected* if the only subsets that are both open and closed are \emptyset and X itself.

When we apply the term *connected* to a nonempty subset $A \subset X$, we simply mean that A with the subspace topology is connected.

In other words, a nonempty X is connected if whenever we write $X = X_1 \cup X_2$ where $X_1 \cap X_2 = \emptyset$ and X_1 and X_2 are open, then either $X_1 = \emptyset$ or $X_2 = \emptyset$. So to test for disconnectedness, we need to find nonempty disjoint open sets X_1 and X_2 whose union is X . For subsets, we state this idea as a proposition.

Proposition 7.2.12. *Let (X, d) be a metric space. A nonempty set $S \subset X$ is not connected if and only if there exist open sets U_1 and U_2 in X , such that $U_1 \cap U_2 \cap S = \emptyset$, $U_1 \cap S \neq \emptyset$, $U_2 \cap S \neq \emptyset$, and*

$$S = (U_1 \cap S) \cup (U_2 \cap S).$$

Proof. If U_j is open in X , then $U_j \cap S$ is open in S in the subspace topology (with subspace metric). To see this, note that if $B_X(x, \delta) \subset U_j$, then as $B_S(x, \delta) = S \cap B_X(x, \delta)$, we have $B_S(x, \delta) \subset U_j \cap S$. The proof follows by the above discussion.

The proof of the other direction follows by using Exercise 7.2.12 to find U_1 and U_2 from two open disjoint subsets of S . \square

Example 7.2.13: Let $S \subset \mathbb{R}$ be such that $x < z < y$ with $x, y \in S$ and $z \notin S$. Claim: S is not connected. Proof: Notice

$$((-\infty, z) \cap S) \cup ((z, \infty) \cap S) = S. \quad (7.1)$$

Proposition 7.2.14. *A set $S \subset \mathbb{R}$ is connected if and only if it is an interval or a single point.*

Proof. Suppose that S is connected (so also nonempty). If S is a single point then we are done. So suppose that $x < y$ and $x, y \in S$. If z is such that $x < z < y$, then $(-\infty, z) \cap S$ is nonempty and $(z, \infty) \cap S$ is nonempty. The two sets are disjoint. As S is connected, we must have their union is not S , so $z \in S$.

Suppose that S is bounded, connected, but not a single point. Let $\alpha := \inf S$ and $\beta := \sup S$ and note that $\alpha < \beta$. Suppose $\alpha < z < \beta$. As α is the infimum, then there is an $x \in S$ such that $\alpha \leq x < z$. Similarly there is a $y \in S$ such that $\beta \geq y > z$. We have shown above that $z \in S$, so $(\alpha, \beta) \subset S$. If $w < \alpha$, then $w \notin S$ as α was the infimum, similarly if $w > \beta$ then $w \notin S$. Therefore the only possibilities for S are (α, β) , $[\alpha, \beta)$, $(\alpha, \beta]$, $[\alpha, \beta]$.

The proof that an unbounded connected S is an interval is left as an exercise.

On the other hand suppose that S is an interval. Suppose that U_1 and U_2 are open subsets of \mathbb{R} , $U_1 \cap S$ and $U_2 \cap S$ are nonempty, and $S = (U_1 \cap S) \cup (U_2 \cap S)$. We will show that $U_1 \cap S$ and $U_2 \cap S$ contain a common point, so they are not disjoint, and hence S must be connected. Suppose that there is $x \in U_1 \cap S$ and $y \in U_2 \cap S$. We can assume that $x < y$. As S is an interval $[x, y] \subset S$. Let $z := \inf(U_2 \cap [x, y])$. If $z = x$, then $z \in U_1$. If $z > x$, then for any $\delta > 0$ the ball $B(z, \delta) = (z - \delta, z + \delta)$ contains points that are not in U_2 , and so $z \notin U_2$ as U_2 is open. Therefore, $z \in U_1$. As U_1 is open, $B(z, \delta) \subset U_1$ for a small enough $\delta > 0$. As z is the infimum of $U_2 \cap [x, y]$, there must exist some $w \in U_2 \cap [x, y]$ such that $w \in [z, z + \delta) \subset B(z, \delta) \subset U_1$. Therefore $w \in U_1 \cap U_2 \cap [x, y]$. So $U_1 \cap S$ and $U_2 \cap S$ are not disjoint and hence S is connected. \square

Example 7.2.15: In many cases a ball $B(x, \delta)$ is connected. But this is not necessarily true in every metric space. For a simplest example, take a two point space $\{a, b\}$ with the discrete metric. Then $B(a, 2) = \{a, b\}$, which is not connected as $B(a, 1) = \{a\}$ and $B(b, 1) = \{b\}$ are open and disjoint.

7.2.3 Closure and boundary

Sometime we wish to take a set and throw in everything that we can approach from the set. This concept is called the closure.

Definition 7.2.16. Let (X, d) be a metric space and $A \subset X$. Then the *closure* of A is the set

$$\bar{A} := \bigcap \{E \subset X : E \text{ is closed and } A \subset E\}.$$

That is, \bar{A} is the intersection of all closed sets that contain A .

Proposition 7.2.17. Let (X, d) be a metric space and $A \subset X$. The closure \bar{A} is closed. Furthermore if A is closed then $\bar{A} = A$.

Proof. First, the closure is the intersection of closed sets, so it is closed. Second, if A is closed, then take $E = A$, hence the intersection of all closed sets E containing A must be equal to A . \square

Example 7.2.18: The closure of $(0, 1)$ in \mathbb{R} is $[0, 1]$. Proof: Simply notice that if E is closed and contains $(0, 1)$, then E must contain 0 and 1 (why?). Thus $[0, 1] \subset E$. But $[0, 1]$ is also closed. Therefore the closure $\overline{(0, 1)} = [0, 1]$.

Example 7.2.19: Be careful to notice what ambient metric space you are working with. If $X = (0, \infty)$, then the closure of $(0, 1)$ in $(0, \infty)$ is $(0, 1]$. Proof: Similarly as above $(0, 1]$ is closed in $(0, \infty)$ (why?). Any closed set E that contains $(0, 1)$ must contain 1 (why?). Therefore $(0, 1] \subset E$, and hence $\overline{(0, 1)} = (0, 1]$ when working in $(0, \infty)$.

Let us justify the statement that the closure is everything that we can “approach” from the set.

Proposition 7.2.20. Let (X, d) be a metric space and $A \subset X$. Then $x \in \bar{A}$ if and only if for every $\delta > 0$, $B(x, \delta) \cap A \neq \emptyset$.

Proof. Let us prove the two contrapositives. Let us show that $x \notin \bar{A}$ if and only if there exists a $\delta > 0$ such that $B(x, \delta) \cap A = \emptyset$.

First suppose that $x \notin \bar{A}$. We know \bar{A} is closed. Thus there is a $\delta > 0$ such that $B(x, \delta) \subset \bar{A}^c$. As $A \subset \bar{A}$ we see that $B(x, \delta) \subset A^c$ and hence $B(x, \delta) \cap A = \emptyset$.

On the other hand suppose that there is a $\delta > 0$ such that $B(x, \delta) \cap A = \emptyset$. Then $B(x, \delta)^c$ is a closed set and we have that $A \subset B(x, \delta)^c$, but $x \notin B(x, \delta)^c$. Thus as \bar{A} is the intersection of closed sets containing A , we have $x \notin \bar{A}$. \square

We can also talk about what is in the interior of a set and what is on the boundary.

Definition 7.2.21. Let (X, d) be a metric space and $A \subset X$, then the *interior* of A is the set

$$A^\circ := \{x \in A : \text{there exists a } \delta > 0 \text{ such that } B(x, \delta) \subset A\}.$$

The *boundary* of A is the set

$$\partial A := \bar{A} \setminus A^\circ.$$

Example 7.2.22: Suppose $A = (0, 1]$ and $X = \mathbb{R}$. Then it is not hard to see that $\bar{A} = [0, 1]$, $A^\circ = (0, 1)$, and $\partial A = \{0, 1\}$.

Example 7.2.23: Suppose $X = \{a, b\}$ with the discrete metric. Let $A = \{a\}$, then $\bar{A} = A^\circ$ and $\partial A = \emptyset$.

Proposition 7.2.24. *Let (X, d) be a metric space and $A \subset X$. Then A° is open and ∂A is closed.*

Proof. Given $x \in A^\circ$ we have $\delta > 0$ such that $B(x, \delta) \subset A$. If $z \in B(x, \delta)$, then as open balls are open, there is an $\varepsilon > 0$ such that $B(z, \varepsilon) \subset B(x, \delta) \subset A$, so z is in A° . Therefore $B(x, \delta) \subset A^\circ$ and so A° is open.

As A° is open, then $\partial A = \bar{A} \setminus A^\circ = \bar{A} \cap (A^\circ)^c$ is closed. □

The boundary is the set of points that are close to both the set and its complement.

Proposition 7.2.25. *Let (X, d) be a metric space and $A \subset X$. Then $x \in \partial A$ if and only if for every $\delta > 0$, $B(x, \delta) \cap A$ and $B(x, \delta) \cap A^c$ are both nonempty.*

Proof. If $x \notin \bar{A}$, then there is some $\delta > 0$ such that $B(x, \delta) \subset \bar{A}^c$ as \bar{A} is closed. So $B(x, \delta)$ contains no points of A .

Now suppose that $x \in A^\circ$, then there exists a $\delta > 0$ such that $B(x, \delta) \subset A$, but that means that $B(x, \delta)$ contains no points of A^c .

Finally suppose that $x \in \bar{A} \setminus A^\circ$. Let $\delta > 0$ be arbitrary. By Proposition 7.2.20 $B(x, \delta)$ contains a point from A . Also, if $B(x, \delta)$ contained no points of A^c , then x would be in A° . Hence $B(x, \delta)$ contains a points of A^c as well. □

We obtain the following immediate corollary about closures of A and A^c . We simply apply Proposition 7.2.20.

Corollary 7.2.26. *Let (X, d) be a metric space and $A \subset X$. Then $\partial A = \bar{A} \cap \overline{A^c}$.*

7.2.4 Exercises

Exercise 7.2.1: *Prove Proposition 7.2.8. Hint: consider the complements of the sets and apply Proposition 7.2.6.*

Exercise 7.2.2: *Finish the proof of Proposition 7.2.9 by proving that $C(x, \delta)$ is closed.*

Exercise 7.2.3: *Prove Proposition 7.2.10.*

Exercise 7.2.4: *Suppose that (X, d) is a nonempty metric space with the discrete topology. Show that X is connected if and only if it contains exactly one element.*

Exercise 7.2.5: *Show that if $S \subset \mathbb{R}$ is a connected unbounded set, then it is an (unbounded) interval.*

Exercise 7.2.6: *Show that every open set can be written as a union of closed sets.*

Exercise 7.2.7: a) Show that E is closed if and only if $\partial E \subset E$. b) Show that U is open if and only if $\partial U \cap U = \emptyset$.

Exercise 7.2.8: a) Show that A is open if and only if $A^\circ = A$. b) Suppose that U is an open set and $U \subset A$. Show that $U \subset A^\circ$.

Exercise 7.2.9: Let X be a set and d, d' be two metrics on X . Suppose that there exists an $\alpha > 0$ and $\beta > 0$ such that $\alpha d(x, y) \leq d'(x, y) \leq \beta d(x, y)$ for all $x, y \in X$. Show that U is open in (X, d) if and only if U is open in (X, d') . That is, the topologies of (X, d) and (X, d') are the same.

Exercise 7.2.10: Suppose that $\{S_i\}$, $i \in \mathbb{N}$ is a collection of connected subsets of a metric space (X, d) . Suppose that there exists an $x \in X$ such that $x \in S_i$ for all $i \in \mathbb{N}$. Show that $\bigcup_{i=1}^{\infty} S_i$ is connected.

Exercise 7.2.11: Let A be a connected set. a) Is \bar{A} connected? Prove or find a counterexample. b) Is A° connected? Prove or find a counterexample. Hint: Think of sets in \mathbb{R}^2 .

The definition of open sets in the following exercise is usually called the *subspace topology*. You are asked to show that we obtain the same topology by considering the subspace metric.

Exercise 7.2.12: Suppose (X, d) is a metric space and $Y \subset X$. Show that with the subspace metric on Y , a set $U \subset Y$ is open (in Y) whenever there exists an open set $V \subset X$ such that $U = V \cap Y$.

Exercise 7.2.13: Let (X, d) be a metric space. a) For any $x \in X$ and $\delta > 0$, show $\overline{B(x, \delta)} \subset C(x, \delta)$. b) Is it always true that $\overline{B(x, \delta)} = C(x, \delta)$? Prove or find a counterexample.

Exercise 7.2.14: Let (X, d) be a metric space and $A \subset X$. Show that $A^\circ = \bigcup\{V : V \subset A \text{ is open}\}$.

7.3 Sequences and convergence

Note: 1 lecture

7.3.1 Sequences

The notion of a sequence in a metric space is very similar to a sequence of real numbers.

Definition 7.3.1. A *sequence* in a metric space (X, d) is a function $x: \mathbb{N} \rightarrow X$. As before we write x_n for the n th element in the sequence and use the notation $\{x_n\}$, or more precisely

$$\{x_n\}_{n=1}^{\infty}.$$

A sequence $\{x_n\}$ is *bounded* if there exists a point $p \in X$ and $B \in \mathbb{R}$ such that

$$d(p, x_n) \leq B \quad \text{for all } n \in \mathbb{N}.$$

In other words, the sequence $\{x_n\}$ is bounded whenever the set $\{x_n : n \in \mathbb{N}\}$ is bounded.

If $\{n_j\}_{j=1}^{\infty}$ is a sequence of natural numbers such that $n_{j+1} > n_j$ for all j then the sequence $\{x_{n_j}\}_{j=1}^{\infty}$ is said to be a *subsequence* of $\{x_n\}$.

Similarly we also define convergence. Again, we will be cheating a little bit and we will use the definite article in front of the word *limit* before we prove that the limit is unique.

Definition 7.3.2. A sequence $\{x_n\}$ in a metric space (X, d) is said to *converge* to a point $p \in X$, if for every $\varepsilon > 0$, there exists an $M \in \mathbb{N}$ such that $d(x_n, p) < \varepsilon$ for all $n \geq M$. The point p is said to be the *limit* of $\{x_n\}$. We write

$$\lim_{n \rightarrow \infty} x_n := p.$$

A sequence that converges is said to be *convergent*. Otherwise, the sequence is said to be *divergent*.

Let us prove that the limit is unique. Note that the proof is almost identical to the proof of the same fact for sequences of real numbers. Many results we know for sequences of real numbers can be proved in the more general settings of metric spaces. We must replace $|x - y|$ with $d(x, y)$ in the proofs and apply the triangle inequality correctly.

Proposition 7.3.3. A convergent sequence in a metric space has a unique limit.

Proof. Suppose that the sequence $\{x_n\}$ has the limit x and the limit y . Take an arbitrary $\varepsilon > 0$. From the definition we find an M_1 such that for all $n \geq M_1$, $d(x_n, x) < \varepsilon/2$. Similarly we find an M_2 such that for all $n \geq M_2$ we have $d(x_n, y) < \varepsilon/2$. Now take an n such that $n \geq M_1$ and also $n \geq M_2$

$$\begin{aligned} d(y, x) &\leq d(y, x_n) + d(x_n, x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

As $d(y, x) < \varepsilon$ for all $\varepsilon > 0$, then $d(x, y) = 0$ and $y = x$. Hence the limit (if it exists) is unique. \square

The proofs of the following propositions are left as exercises.

Proposition 7.3.4. *A convergent sequence in a metric space is bounded.*

Proposition 7.3.5. *A sequence $\{x_n\}$ in a metric space (X, d) converges to $p \in X$ if and only if there exists a sequence $\{a_n\}$ of real numbers such that*

$$d(x_n, p) \leq a_n \quad \text{for all } n \in \mathbb{N},$$

and

$$\lim_{n \rightarrow \infty} a_n = 0.$$

Proposition 7.3.6. *Let $\{x_n\}$ be a sequence in a metric space (X, d) .*

(i) *If $\{x_n\}$ converges to $p \in X$, then every subsequence $\{x_{n_k}\}$ converges to p .*

(ii) *If for some $K \in \mathbb{N}$ the K -tail $\{x_n\}_{n=K+1}^\infty$ converges to $p \in X$, then $\{x_n\}$ converges to p .*

7.3.2 Convergence in euclidean space

It is useful to note what convergence means in the euclidean space \mathbb{R}^n .

Proposition 7.3.7. *Let $\{x^j\}_{j=1}^\infty$ be a sequence in \mathbb{R}^n , where we write $x^j = (x_1^j, x_2^j, \dots, x_n^j) \in \mathbb{R}^n$. Then $\{x^j\}_{j=1}^\infty$ converges if and only if $\{x_k^j\}_{j=1}^\infty$ converges for every k , in which case*

$$\lim_{j \rightarrow \infty} x^j = \left(\lim_{j \rightarrow \infty} x_1^j, \lim_{j \rightarrow \infty} x_2^j, \dots, \lim_{j \rightarrow \infty} x_n^j \right).$$

Proof. For $\mathbb{R} = \mathbb{R}^1$ the result is immediate. So let $n > 1$.

Let $\{x^j\}_{j=1}^\infty$ be a convergent sequence in \mathbb{R}^n , where we write $x^j = (x_1^j, x_2^j, \dots, x_n^j) \in \mathbb{R}^n$. Let $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ be the limit. Given $\varepsilon > 0$, there exists an M such that for all $j \geq M$ we have

$$d(x, x^j) < \varepsilon.$$

Fix some $k = 1, 2, \dots, n$. For $j \geq M$ we have

$$|x_k - x_k^j| = \sqrt{(x_k - x_k^j)^2} \leq \sqrt{\sum_{\ell=1}^n (x_\ell - x_\ell^j)^2} = d(x, x^j) < \varepsilon. \quad (7.2)$$

Hence the sequence $\{x_k^j\}_{j=1}^\infty$ converges to x_k .

For the other direction suppose that $\{x_k^j\}_{j=1}^\infty$ converges to x_k for every $k = 1, 2, \dots, n$. Hence, given $\varepsilon > 0$, pick an M , such that if $j \geq M$ then $|x_k - x_k^j| < \varepsilon/\sqrt{n}$ for all $k = 1, 2, \dots, n$. Then

$$d(x, x^j) = \sqrt{\sum_{k=1}^n (x_k - x_k^j)^2} < \sqrt{\sum_{k=1}^n \left(\frac{\varepsilon}{\sqrt{n}}\right)^2} = \sqrt{\sum_{k=1}^n \frac{\varepsilon^2}{n}} = \varepsilon.$$

The sequence $\{x^j\}$ converges to $x \in \mathbb{R}^n$ and we are done. \square

7.3.3 Convergence and topology

The topology, that is, the set of open sets of a space encodes which sequences converge.

Proposition 7.3.8. *Let (X, d) be a metric space and $\{x_n\}$ a sequence in X . Then $\{x_n\}$ converges to $x \in X$ if and only if for every open neighborhood U of x , there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $x_n \in U$.*

Proof. First suppose that $\{x_n\}$ converges. Let U be an open neighborhood of x , then there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U$. As the sequence converges, find an $M \in \mathbb{N}$ such that for all $n \geq M$ we have $d(x, x_n) < \varepsilon$ or in other words $x_n \in B(x, \varepsilon) \subset U$.

Let us prove the other direction. Given $\varepsilon > 0$ let $U := B(x, \varepsilon)$ be the neighborhood of x . Then there is an $M \in \mathbb{N}$ such that for $n \geq M$ we have $x_n \in U = B(x, \varepsilon)$ or in other words, $d(x, x_n) < \varepsilon$. \square

A set is closed when it contains the limits of its convergent sequences.

Proposition 7.3.9. *Let (X, d) be a metric space, $E \subset X$ a closed set and $\{x_n\}$ a sequence in E that converges to some $x \in X$. Then $x \in E$.*

Proof. Let us prove the contrapositive. Suppose $\{x_n\}$ is a sequence in X that converges to $x \in E^c$. As E^c is open, Proposition 7.3.8 says there is an M such that for all $n \geq M$, $x_n \in E^c$. So $\{x_n\}$ is not a sequence in E . \square

When we take a closure of a set A , we really throw in precisely those points that are limits of sequences in A .

Proposition 7.3.10. *Let (X, d) be a metric space and $A \subset X$. If $x \in \bar{A}$, then there exists a sequence $\{x_n\}$ of elements in A such that $\lim x_n = x$.*

Proof. Let $x \in \bar{A}$. We know by Proposition 7.2.20 that given $1/n$, there exists a point $x_n \in B(x, 1/n) \cap A$. As $d(x, x_n) < 1/n$, we have that $\lim x_n = x$. \square

7.3.4 Exercises

Exercise 7.3.1: *Let (X, d) be a metric space and let $A \subset X$. Let E be the set of all $x \in X$ such that there exists a sequence $\{x_n\}$ in A that converges to x . Show that $E = \bar{A}$.*

Exercise 7.3.2: *a) Show that $d(x, y) := \min\{1, |x - y|\}$ defines a metric on \mathbb{R} . b) Show that a sequence converges in (\mathbb{R}, d) if and only if it converges in the standard metric. c) Find a bounded sequence in (\mathbb{R}, d) that contains no convergent subsequence.*

Exercise 7.3.3: *Prove Proposition 7.3.4.*

Exercise 7.3.4: *Prove Proposition 7.3.5.*

Exercise 7.3.5: Suppose that $\{x_n\}_{n=1}^{\infty}$ converges to x . Suppose that $f: \mathbb{N} \rightarrow \mathbb{N}$ is a one-to-one and onto function. Show that $\{x_{f(n)}\}_{n=1}^{\infty}$ converges to x .

Exercise 7.3.6: If (X, d) is a metric space where d is the discrete metric. Suppose that $\{x_n\}$ is a convergent sequence in X . Show that there exists a $K \in \mathbb{N}$ such that for all $n \geq K$ we have $x_n = x_K$.

Exercise 7.3.7: A set $S \subset X$ is said to be dense in X if for every $x \in X$, there exists a sequence $\{x_n\}$ in S that converges to x . Prove that \mathbb{R}^n contains a countable dense subset.

Exercise 7.3.8 (Tricky): Suppose $\{U_n\}_{n=1}^{\infty}$ be a decreasing ($U_{n+1} \subset U_n$ for all n) sequence of open sets in a metric space (X, d) such that $\bigcap_{n=1}^{\infty} U_n = \{p\}$ for some $p \in X$. Suppose that $\{x_n\}$ is a sequence of points in X such that $x_n \in U_n$. Does $\{x_n\}$ necessarily converge to p ? Prove or construct a counterexample.

Exercise 7.3.9: Let $E \subset X$ be closed and let $\{x_n\}$ be a sequence in X converging to $p \in X$. Suppose $x_n \in E$ for infinitely many $n \in \mathbb{N}$. Show $p \in E$.

Exercise 7.3.10: Take $\mathbb{R}^* = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ be the extended reals. Define $d(x, y) := \frac{|x-y|}{1+|x-y|}$ if $x, y \in \mathbb{R}$, define $d(\infty, x) := d(-\infty, x) = 1$ for all $x \in \mathbb{R}$, and let $d(\infty, -\infty) := 2$. a) Show that (\mathbb{R}^*, d) is a metric space. b) Suppose that $\{x_n\}$ is a sequence of real numbers such that $x_n \geq n$ for all n . Show that $\lim x_n = \infty$ in (\mathbb{R}^*, d) .

Exercise 7.3.11: Suppose that $\{V_n\}_{n=1}^{\infty}$ is a collection of open sets in (X, d) such that $V_{n+1} \supset V_n$. Let $\{x_n\}$ be a sequence such that $x_n \in V_{n+1} \setminus V_n$ and suppose that $\{x_n\}$ converges to $p \in X$. Show that $p \in \partial V$ where $V = \bigcup_{n=1}^{\infty} V_n$.

Exercise 7.3.12: Prove Proposition 7.3.6.

7.4 Completeness and compactness

Note: 2 lectures

7.4.1 Cauchy sequences and completeness

Just like with sequences of real numbers we define Cauchy sequences.

Definition 7.4.1. Let (X, d) be a metric space. A sequence $\{x_n\}$ in X is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $k \geq M$ we have

$$d(x_n, x_k) < \varepsilon.$$

The definition is again simply a translation of the concept from the real numbers to metric spaces. So a sequence of real numbers is Cauchy in the sense of chapter 2 if and only if it is Cauchy in the sense above, provided we equip the real numbers with the standard metric $d(x, y) = |x - y|$.

Proposition 7.4.2. A convergent sequence in a metric space is Cauchy.

Proof. Suppose $\{x_n\}$ converges to x . Given $\varepsilon > 0$ there is an M such that for $n \geq M$ we have $d(x, x_n) < \varepsilon/2$. Hence for all $n, k \geq M$ we have $d(x_n, x_k) \leq d(x_n, x) + d(x, x_k) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. \square

Definition 7.4.3. Let (X, d) be a metric space. We say that X is *complete* or *Cauchy-complete* if every Cauchy sequence $\{x_n\}$ in X converges to an $x \in X$.

Proposition 7.4.4. The space \mathbb{R}^n with the standard metric is a complete metric space.

Proof. For $\mathbb{R} = \mathbb{R}^1$ this was proved in chapter 2.

Take $n > 1$. Let $\{x^j\}_{j=1}^\infty$ be a Cauchy sequence in \mathbb{R}^n , where we write $x^j = (x_1^j, x_2^j, \dots, x_n^j) \in \mathbb{R}^n$. As the sequence is Cauchy, given $\varepsilon > 0$, there exists an M such that for all $i, j \geq M$ we have

$$d(x^i, x^j) < \varepsilon.$$

Fix some $k = 1, 2, \dots, n$, for $i, j \geq M$ we have

$$|x_k^i - x_k^j| = \sqrt{(x_k^i - x_k^j)^2} \leq \sqrt{\sum_{\ell=1}^n (x_\ell^i - x_\ell^j)^2} = d(x^i, x^j) < \varepsilon. \quad (7.3)$$

Hence the sequence $\{x_k^j\}_{j=1}^\infty$ is Cauchy. As \mathbb{R} is complete the sequence converges; there exists an $x_k \in \mathbb{R}$ such that $x_k = \lim_{j \rightarrow \infty} x_k^j$.

Write $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. By Proposition 7.3.7 we have that $\{x^j\}$ converges to $x \in \mathbb{R}^n$ and hence \mathbb{R}^n is complete. \square

7.4.2 Compactness

Definition 7.4.5. Let (X, d) be a metric space and $K \subset X$. The set K is said to be *compact* if for any collection of open sets $\{U_\lambda\}_{\lambda \in I}$ such that

$$K \subset \bigcup_{\lambda \in I} U_\lambda,$$

there exists a finite subset $\{\lambda_1, \lambda_2, \dots, \lambda_k\} \subset I$ such that

$$K \subset \bigcup_{j=1}^k U_{\lambda_j}.$$

A collection of open sets $\{U_\lambda\}_{\lambda \in I}$ as above is said to be a *open cover* of K . So a way to say that K is compact is to say that *every open cover of K has a finite subcover*.

Proposition 7.4.6. Let (X, d) be a metric space. A compact set $K \subset X$ is closed and bounded.

Proof. First, we prove that a compact set is bounded. Fix $p \in X$. We have the open cover

$$K \subset \bigcup_{n=1}^{\infty} B(p, n) = X.$$

If K is compact, then there exists some set of indices $n_1 < n_2 < \dots < n_k$ such that

$$K \subset \bigcup_{j=1}^k B(p, n_j) = B(p, n_k).$$

As K is contained in a ball, K is bounded.

Next, we show a set that is not closed is not compact. Suppose that $\bar{K} \neq K$, that is, there is a point $x \in \bar{K} \setminus K$. If $y \neq x$, then for n with $1/n < d(x, y)$ we have $y \notin C(x, 1/n)$. Furthermore $x \notin K$, so

$$K \subset \bigcup_{n=1}^{\infty} C(x, 1/n)^c.$$

As a closed ball is closed, $C(x, 1/n)^c$ is open, and so we have an open cover. If we take any finite collection of indices $n_1 < n_2 < \dots < n_k$, then

$$\bigcup_{j=1}^k C(x, 1/n_j)^c = C(x, 1/n_k)^c$$

As x is in the closure, we have $C(x, 1/n_k) \cap K \neq \emptyset$, so there is no finite subcover and K is not compact. \square

We prove below that in finite dimensional euclidean space every closed bounded set is compact. So closed bounded sets of \mathbb{R}^n are examples of compact sets. It is not true that in every metric space, closed and bounded is equivalent to compact. There are many metric spaces where closed and bounded is not enough to give compactness, see Exercise 7.4.8. However, see Exercise 7.4.12.

A useful property of compact sets in a metric space is that every sequence has a convergent subsequence. Such sets are sometimes called *sequentially compact*. Let us prove that in the context of metric spaces, a set is compact if and only if it is sequentially compact. First we prove a lemma.

Lemma 7.4.7 (Lebesgue covering lemma*). *Let (X, d) be a metric space and $K \subset X$. Suppose that every sequence in K has a subsequence convergent in K . Given an open cover $\{U_\lambda\}_{\lambda \in I}$ of K , there exists a $\delta > 0$ such that for every $x \in K$, there exists a $\lambda \in I$ with $B(x, \delta) \subset U_\lambda$.*

Proof. Let us prove the lemma by contrapositive. If the conclusion is not true, then there is an open cover $\{U_\lambda\}_{\lambda \in I}$ of K with the following property. For every $n \in \mathbb{N}$ there exists an $x_n \in K$ such that $B(x_n, 1/n)$ is not a subset of any U_λ . Given any $x \in K$, there is a $\lambda \in I$ such that $x \in U_\lambda$. Hence there is an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U_\lambda$. Take M such that $1/M < \varepsilon/2$. If $y \in B(x, \varepsilon/2)$ and $n \geq M$, then by triangle inequality

$$B(y, 1/n) \subset B(y, 1/M) \subset B(x, \varepsilon) \subset U_\lambda. \quad (7.4)$$

In other words, for all $n \geq M$, $x_n \notin B(x, \varepsilon/2)$. Hence the sequence cannot have a subsequence converging to x . As $x \in K$ was arbitrary we are done. \square

Theorem 7.4.8. *Let (X, d) be a metric space. Then $K \subset X$ is a compact set if and only if every sequence in K has a subsequence converging to a point in K .*

Proof. Let $K \subset X$ be a set and $\{x_n\}$ a sequence in K . Suppose that for each $x \in K$, there is a ball $B(x, \alpha_x)$ for some $\alpha_x > 0$ such that $x_n \in B(x, \alpha_x)$ for only finitely many $n \in \mathbb{N}$. Then

$$K \subset \bigcup_{x \in K} B(x, \alpha_x).$$

Any finite collection of these balls is going to contain only finitely many x_n . Thus for any finite collection of such balls there is an $x_n \in K$ that is not in the union. Therefore, K is not compact.

So if K is compact, then there exists an $x \in K$ such that for any $\delta > 0$, $B(x, \delta)$ contains x_k for infinitely many $k \in \mathbb{N}$. The ball $B(x, 1)$ contains some x_k so let $n_1 := k$. If n_{j-1} is defined, then there must exist a $k > n_{j-1}$ such that $x_k \in B(x, 1/j)$, so define $n_j := k$. Notice that $d(x, x_{n_j}) < 1/j$. By Proposition 7.3.5, $\lim x_{n_j} = x$.

For the other direction, suppose that every sequence in K has a subsequence converging in K . Take an open cover $\{U_\lambda\}_{\lambda \in I}$ of K . Using the Lebesgue covering lemma above, we find a $\delta > 0$ such that for every x , there is a $\lambda \in I$ with $B(x, \delta) \subset U_\lambda$.

*Named after the French mathematician Henri Léon Lebesgue (1875 – 1941). The number δ is sometimes called the Lebesgue number of the cover.

Pick $x_1 \in K$ and find $\lambda_1 \in I$ such that $B(x_1, \delta) \subset U_{\lambda_1}$. If $K \subset U_{\lambda_1}$, we stop as we have found a finite subcover. Otherwise, there must be a point $x_2 \in K \setminus U_{\lambda_1}$. Note that $d(x_2, x_1) \geq \delta$. There must exist some $\lambda_2 \in I$ such that $B(x_2, \delta) \subset U_{\lambda_2}$. We work inductively. Suppose that λ_{n-1} is defined. Either $U_{\lambda_1} \cup U_{\lambda_2} \cup \cdots \cup U_{\lambda_{n-1}}$ is a finite cover of K , in which case we stop, or there must be a point $x_n \in K \setminus (U_{\lambda_1} \cup U_{\lambda_2} \cup \cdots \cup U_{\lambda_{n-1}})$. Note that $d(x_n, x_j) \geq \delta$ for all $j = 1, 2, \dots, n-1$. Next, there must be some $\lambda_n \in I$ such that $B(x_n, \delta) \subset U_{\lambda_n}$.

Either at some point we obtain a finite subcover of K or we obtain an infinite sequence $\{x_n\}$ as above. For contradiction suppose that there is no finite subcover and we have the sequence $\{x_n\}$. For all n and k , $n \neq k$, we have $d(x_n, x_k) \geq \delta$, so no subsequence of $\{x_n\}$ can be Cauchy. Hence no subsequence of $\{x_n\}$ can be convergent, which is a contradiction. \square

Example 7.4.9: The Bolzano-Weierstrass theorem for sequences of real numbers (Theorem 2.3.8) says that any bounded sequence in \mathbb{R} has a convergent subsequence. Therefore any sequence in a closed interval $[a, b] \subset \mathbb{R}$ has a convergent subsequence. The limit must also be in $[a, b]$ as limits preserve non-strict inequalities. Hence a closed bounded interval $[a, b] \subset \mathbb{R}$ is compact.

Proposition 7.4.10. *Let (X, d) be a metric space and let $K \subset X$ be compact. Suppose that $E \subset K$ is a closed set, then E is compact.*

Proof. Let $\{x_n\}$ be a sequence in E . It is also a sequence in K . Therefore it has a convergent subsequence $\{x_{n_j}\}$ that converges to some $x \in K$. As E is closed the limit of a sequence in E is also in E and so $x \in E$. Thus E must be compact. \square

Theorem 7.4.11 (Heine-Borel*). *A closed bounded subset $K \subset \mathbb{R}^n$ is compact.*

Proof. For $\mathbb{R} = \mathbb{R}^1$ if $K \subset \mathbb{R}$ is closed and bounded, then any sequence $\{x_k\}$ in K is bounded, so it has a convergent subsequence by Bolzano-Weierstrass theorem (Theorem 2.3.8). As K is closed, the limit of the subsequence must be an element of K . So K is compact.

Let us carry out the proof for $n = 2$ and leave arbitrary n as an exercise. As $K \subset \mathbb{R}^2$ is bounded, there exists a set $B = [a, b] \times [c, d] \subset \mathbb{R}^2$ such that $K \subset B$. If we show that B is compact, then K , being a closed subset of a compact B , is also compact.

Let $\{(x_k, y_k)\}_{k=1}^{\infty}$ be a sequence in B . That is, $a \leq x_k \leq b$ and $c \leq y_k \leq d$ for all k . A bounded sequence of real numbers has a convergent subsequence so there is a subsequence $\{x_{k_j}\}_{j=1}^{\infty}$ that is convergent. The subsequence $\{y_{k_j}\}_{j=1}^{\infty}$ is also a bounded sequence so there exists a subsequence $\{y_{k_{j_i}}\}_{i=1}^{\infty}$ that is convergent. A subsequence of a convergent sequence is still convergent, so $\{x_{k_{j_i}}\}_{i=1}^{\infty}$ is convergent. Let

$$x := \lim_{i \rightarrow \infty} x_{k_{j_i}} \quad \text{and} \quad y := \lim_{i \rightarrow \infty} y_{k_{j_i}}.$$

By Proposition 7.3.7, $\{(x_{k_{j_i}}, y_{k_{j_i}})\}_{i=1}^{\infty}$ converges to (x, y) . Furthermore, as $a \leq x_k \leq b$ and $c \leq y_k \leq d$ for all k , we know that $(x, y) \in B$. \square

*Named after the German mathematician Heinrich Eduard Heine (1821–1881), and the French mathematician Félix Édouard Justin Émile Borel (1871–1956).

7.4.3 Exercises

Exercise 7.4.1: Let (X, d) be a metric space and A a finite subset of X . Show that A is compact.

Exercise 7.4.2: Let $A = \{1/n : n \in \mathbb{N}\} \subset \mathbb{R}$. a) Show that A is not compact directly using the definition. b) Show that $A \cup \{0\}$ is compact directly using the definition.

Exercise 7.4.3: Let (X, d) be a metric space with the discrete metric. a) Prove that X is complete. b) Prove that X is compact if and only if X is a finite set.

Exercise 7.4.4: a) Show that the union of finitely many compact sets is a compact set. b) Find an example where the union of infinitely many compact sets is not compact.

Exercise 7.4.5: Prove Theorem 7.4.11 for arbitrary dimension. Hint: The trick is to use the correct notation.

Exercise 7.4.6: Show that a compact set K is a complete metric space.

Exercise 7.4.7: Let $C([a, b], \mathbb{R})$ be the metric space as in Example 7.1.7. Show that $C([a, b], \mathbb{R})$ is a complete metric space.

Exercise 7.4.8 (Challenging): Let $C([0, 1], \mathbb{R})$ be the metric space of Example 7.1.7. Let 0 denote the zero function. Then show that the closed ball $C(0, 1)$ is not compact (even though it is closed and bounded). Hints: Construct a sequence of distinct continuous functions $\{f_n\}$ such that $d(f_n, 0) = 1$ and $d(f_n, f_k) = 1$ for all $n \neq k$. Show that the set $\{f_n : n \in \mathbb{N}\} \subset C(0, 1)$ is closed but not compact. See chapter 6 for inspiration.

Exercise 7.4.9 (Challenging): Show that there exists a metric on \mathbb{R} that makes \mathbb{R} into a compact set.

Exercise 7.4.10: Suppose that (X, d) is complete and suppose we have a countably infinite collection of nonempty compact sets $E_1 \supset E_2 \supset E_3 \supset \dots$ then prove $\bigcap_{j=1}^{\infty} E_j \neq \emptyset$.

Exercise 7.4.11 (Challenging): Let $C([0, 1], \mathbb{R})$ be the metric space of Example 7.1.7. Let K be the set of $f \in C([0, 1], \mathbb{R})$ such that f is equal to a quadratic polynomial, i.e. $f(x) = a + bx + cx^2$, and such that $|f(x)| \leq 1$ for all $x \in [0, 1]$, that is $f \in C(0, 1)$. Show that K is compact.

Exercise 7.4.12 (Challenging): Let (X, d) be a complete metric space. Show that $K \subset X$ is compact if and only if K is closed and such that for every $\varepsilon > 0$ there exists a finite set of points x_1, \dots, x_n with $K \subset \bigcup_{j=1}^n B(x_j, \varepsilon)$. Note: Such a set K is said to be totally bounded, so we can say that in a complete metric space a set is compact if and only if it is closed and totally bounded.

Exercise 7.4.13: Take $\mathbb{N} \subset \mathbb{R}$ using the standard metric. Find an open cover of \mathbb{N} such that the conclusion of the Lebesgue covering lemma does not hold.

Exercise 7.4.14: Prove the general Bolzano-Weierstrass theorem: Any bounded sequence $\{x_k\}$ in \mathbb{R}^n has a convergent subsequence.

7.5 Continuous functions

Note: 1 lecture

7.5.1 Continuity

Definition 7.5.1. Let (X, d_X) and (Y, d_Y) be metric spaces and $c \in X$. Then $f: X \rightarrow Y$ is *continuous at c* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that whenever $x \in X$ and $d_X(x, c) < \delta$, then $d_Y(f(x), f(c)) < \varepsilon$.

When $f: X \rightarrow Y$ is continuous at all $c \in X$, then we simply say that f is a *continuous function*.

The definition agrees with the definition from chapter 3 when f is a real-valued function on the real line, if we take the standard metric on \mathbb{R} .

Proposition 7.5.2. Let (X, d_X) and (Y, d_Y) be metric spaces. Then $f: X \rightarrow Y$ is continuous at $c \in X$ if and only if for every sequence $\{x_n\}$ in X converging to c , the sequence $\{f(x_n)\}$ converges to $f(c)$.

Proof. Suppose that f is continuous at c . Let $\{x_n\}$ be a sequence in X converging to c . Given $\varepsilon > 0$, there is a $\delta > 0$ such that $d(x, c) < \delta$ implies $d(f(x), f(c)) < \varepsilon$. So take M such that for all $n \geq M$, we have $d(x_n, c) < \delta$, then $d(f(x_n), f(c)) < \varepsilon$. Hence $\{f(x_n)\}$ converges to $f(c)$.

On the other hand suppose that f is not continuous at c . Then there exists an $\varepsilon > 0$, such that for every $n \in \mathbb{N}$ there exists an $x_n \in X$, with $d(x_n, c) < 1/n$ such that $d(f(x_n), f(c)) \geq \varepsilon$. Then $\{x_n\}$ converges to c , but $\{f(x_n)\}$ does not converge to $f(c)$. \square

Example 7.5.3: Suppose $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is a polynomial. That is,

$$f(x, y) = \sum_{j=0}^d \sum_{k=0}^{d-j} a_{jk} x^j y^k = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \cdots + a_{0d}y^d,$$

for some $d \in \mathbb{N}$ (the degree) and $a_{jk} \in \mathbb{R}$. Then we claim that f is continuous. Let $\{(x_n, y_n)\}_{n=1}^{\infty}$ be a sequence in \mathbb{R}^2 that converges to $(x, y) \in \mathbb{R}^2$. We have proved that this means that $\lim x_n = x$ and $\lim y_n = y$. So by Proposition 2.2.5 we have

$$\lim_{n \rightarrow \infty} f(x_n, y_n) = \lim_{n \rightarrow \infty} \sum_{j=0}^d \sum_{k=0}^{d-j} a_{jk} x_n^j y_n^k = \sum_{j=0}^d \sum_{k=0}^{d-j} a_{jk} x^j y^k = f(x, y).$$

So f is continuous at (x, y) , and as (x, y) was arbitrary f is continuous everywhere. Similarly, a polynomial in n variables is continuous.

7.5.2 Compactness and continuity

Continuous maps do not map closed sets to closed sets. For example, $f: (0, 1) \rightarrow \mathbb{R}$ defined by $f(x) := x$ takes the set $(0, 1)$, which is closed in $(0, 1)$, to the set $(0, 1)$, which is not closed in \mathbb{R} . On the other hand continuous maps do preserve compact sets.

Lemma 7.5.4. *Let (X, d_X) and (Y, d_Y) be metric spaces and $f: X \rightarrow Y$ a continuous function. If $K \subset X$ is a compact set, then $f(K)$ is a compact set.*

Proof. A sequence in $f(K)$ can be written as $\{f(x_n)\}_{n=1}^\infty$, where $\{x_n\}_{n=1}^\infty$ is a sequence in K . The set K is compact and therefore there is a subsequence $\{x_{n_i}\}_{i=1}^\infty$ that converges to some $x \in K$. By continuity,

$$\lim_{i \rightarrow \infty} f(x_{n_i}) = f(x) \in f(K).$$

So every sequence in $f(K)$ has a subsequence convergent to a point in $f(K)$, and $f(K)$ is compact by Theorem 7.4.8. \square

As before, $f: X \rightarrow \mathbb{R}$ achieves an *absolute minimum* at $c \in X$ if

$$f(x) \geq f(c) \quad \text{for all } x \in X.$$

On the other hand, f achieves an *absolute maximum* at $c \in X$ if

$$f(x) \leq f(c) \quad \text{for all } x \in X.$$

Theorem 7.5.5. *Let (X, d) be a compact metric space and $f: X \rightarrow \mathbb{R}$ a continuous function. Then f is bounded and in fact f achieves an absolute minimum and an absolute maximum on X .*

Proof. As X is compact and f is continuous, we have that $f(X) \subset \mathbb{R}$ is compact. Hence $f(X)$ is closed and bounded. In particular, $\sup f(X) \in f(X)$ and $\inf f(X) \in f(X)$. That is because both the sup and inf can be achieved by sequences in $f(X)$ and $f(X)$ is closed. Therefore there is some $x \in X$ such that $f(x) = \sup f(X)$ and some $y \in X$ such that $f(y) = \inf f(X)$. \square

7.5.3 Continuity and topology

Let us see how to define continuity in terms of the topology, that is, the open sets. We have already seen that topology determines which sequences converge, and so it is no wonder that the topology also determines continuity of functions.

Lemma 7.5.6. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f: X \rightarrow Y$ is continuous at $c \in X$ if and only if for every open neighbourhood U of $f(c)$ in Y , the set $f^{-1}(U)$ contains an open neighbourhood of c in X .*

Proof. Suppose that f is continuous at c . Let U be an open neighbourhood of $f(c)$ in Y , then $B_Y(f(c), \varepsilon) \subset U$ for some $\varepsilon > 0$. As f is continuous, then there exists a $\delta > 0$ such that whenever x is such that $d_X(x, c) < \delta$, then $d_Y(f(x), f(c)) < \varepsilon$. In other words,

$$B_X(c, \delta) \subset f^{-1}(B_Y(f(c), \varepsilon)) \subset f^{-1}(U).$$

and $B_X(c, \delta)$ is an open neighbourhood of c .

For the other direction, let $\varepsilon > 0$ be given. If $f^{-1}(B_Y(f(c), \varepsilon))$ contains an open neighbourhood W of c , it contains a ball. That is, there is some $\delta > 0$ such that

$$B_X(c, \delta) \subset W \subset f^{-1}(B_Y(f(c), \varepsilon)).$$

That means precisely that if $d_X(x, c) < \delta$ then $d_Y(f(x), f(c)) < \varepsilon$ and so f is continuous at c . \square

Theorem 7.5.7. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f: X \rightarrow Y$ is continuous if and only if for every open $U \subset Y$, $f^{-1}(U)$ is open in X .*

The proof follows from Lemma 7.5.6 and is left as an exercise.

Example 7.5.8: Theorem 7.5.7 tells us that if E is closed, then $f^{-1}(E) = X \setminus f^{-1}(E^c)$ is also closed. Therefore if we have a continuous function $f: X \rightarrow \mathbb{R}$, then the *zero set* of f , that is, $f^{-1}(0) = \{x \in X : f(x) = 0\}$, is closed. An entire field of mathematics, *algebraic geometry*, is the study of zero sets of polynomials.

Similarly the set where f is nonnegative, that is, $f^{-1}([0, \infty)) = \{x \in X : f(x) \geq 0\}$ is closed. On the other hand the set where f is positive, $f^{-1}((0, \infty)) = \{x \in X : f(x) > 0\}$ is open.

7.5.4 Exercises

Exercise 7.5.1: Consider $\mathbb{N} \subset \mathbb{R}$ with the standard metric. Let (X, d) be a metric space and $f: X \rightarrow \mathbb{N}$ a continuous function. a) Prove that if X is connected, then f is constant (the range of f is a single value). b) Find an example where X is disconnected and f is not constant.

Exercise 7.5.2: Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by $f(0, 0) := 0$, and $f(x, y) := \frac{xy}{x^2 + y^2}$ if $(x, y) \neq (0, 0)$. a) Show that for any fixed x , the function that takes y to $f(x, y)$ is continuous. Similarly for any fixed y , the function that takes x to $f(x, y)$ is continuous. b) Show that f is not continuous.

Exercise 7.5.3: Suppose that $f: X \rightarrow Y$ is continuous for metric spaces (X, d_X) and (Y, d_Y) . Let $A \subset X$. a) Show that $f(\overline{A}) \subset \overline{f(A)}$. b) Show that the subset can be proper.

Exercise 7.5.4: Prove Theorem 7.5.7. Hint: Use Lemma 7.5.6.

Exercise 7.5.5: Suppose that $f: X \rightarrow Y$ is continuous for metric spaces (X, d_X) and (Y, d_Y) . Show that if X is connected, then $f(X)$ is connected.

Exercise 7.5.6: Prove the following version of the intermediate value theorem. Let (X, d) be a connected metric space and $f: X \rightarrow \mathbb{R}$ a continuous function. Suppose that there exist $x_0, x_1 \in X$ and $y \in \mathbb{R}$ such that $f(x_0) < y < f(x_1)$. Then prove that there exists a $z \in X$ such that $f(z) = y$. Hint: See Exercise 7.5.5.

Exercise 7.5.7: A continuous function $f: X \rightarrow Y$ for metric spaces (X, d_X) and (Y, d_Y) is said to be proper if for every compact set $K \subset Y$, the set $f^{-1}(K)$ is compact. Suppose that a continuous $f: (0, 1) \rightarrow (0, 1)$ is proper and $\{x_n\}$ is a sequence in $(0, 1)$ that converges to 0. Show that $\{f(x_n)\}$ has no subsequence that converges in $(0, 1)$.

Exercise 7.5.8: Let (X, d_X) and (Y, d_Y) be metric space and $f: X \rightarrow Y$ be a one-to-one and onto continuous function. Suppose that X is compact. Prove that the inverse $f^{-1}: Y \rightarrow X$ is continuous.

Exercise 7.5.9: Take the metric space of continuous functions $C([0, 1], \mathbb{R})$. Let $k: [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ be a continuous function. Given $f \in C([0, 1], \mathbb{R})$ define

$$\varphi_f(x) := \int_0^1 k(x, y)f(y) dy.$$

a) Show that $T(f) := \varphi_f$ defines a function $T: C([0, 1], \mathbb{R}) \rightarrow C([0, 1], \mathbb{R})$. b) Show that T is continuous.

Exercise 7.5.10: Let (X, d) be a metric space. a) If $p \in X$, show that $f: X \rightarrow \mathbb{R}$ defined by $f(x) := d(x, p)$ is continuous. b) Define a metric on $X \times X$ as in Exercise 7.1.6 part b, and show that $g: X \times X \rightarrow \mathbb{R}$ defined by $g(x, y) := d(x, y)$ is continuous. c) Show that if K_1 and K_2 are compact subsets of X , then there exists a $p \in K_1$ and $q \in K_2$ such that $d(p, q)$ is minimal, that is, $d(p, q) = \inf\{d(x, y) : x \in K_1, y \in K_2\}$.

7.6 Fixed point theorem and Picard's theorem again

Note: 1 lecture (optional)

In this section we prove the fixed point theorem for contraction mappings. As an application we prove Picard's theorem, which we proved without metric spaces in §6.3. The proof we present here is similar, but the proof goes a lot smoother with metric spaces and the fixed point theorem.

Definition 7.6.1. Let (X, d) and (X', d') be metric spaces. $f: X \rightarrow X'$ is said to be a *contraction* (or a *contractive map*) if it is a k -Lipschitz map for some $k < 1$, i.e. if there exists a $k < 1$ such that

$$d'(f(x), f(y)) \leq kd(x, y) \quad \text{for all } x, y \in X.$$

If $f: X \rightarrow X$ is a map, $x \in X$ is called a *fixed point* if $f(x) = x$.

Theorem 7.6.2 (Contraction mapping principle or Fixed point theorem). *Let (X, d) be a nonempty complete metric space and $f: X \rightarrow X$ a contraction. Then f has a fixed point.*

The words *complete* and *contraction* are necessary. See Exercise 7.6.6.

Proof. Pick any $x_0 \in X$. Define a sequence $\{x_n\}$ by $x_{n+1} := f(x_n)$.

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq kd(x_n, x_{n-1}) \leq \cdots \leq k^n d(x_1, x_0).$$

Suppose $m \geq n$, then

$$\begin{aligned} d(x_m, x_n) &\leq \sum_{i=n}^{m-1} d(x_{i+1}, x_i) \\ &\leq \sum_{i=n}^{m-1} k^i d(x_1, x_0) \\ &= k^n d(x_1, x_0) \sum_{i=0}^{m-n-1} k^i \\ &\leq k^n d(x_1, x_0) \sum_{i=0}^{\infty} k^i = k^n d(x_1, x_0) \frac{1}{1-k}. \end{aligned}$$

In particular the sequence is Cauchy (why?). Since X is complete we let $x := \lim x_n$ and we claim that x is our unique fixed point.

Fixed point? Note that f is continuous because it is a contraction (why?). Hence

$$f(x) = \lim f(x_n) = \lim x_{n+1} = x.$$

Unique? Let y be a fixed point.

$$d(x, y) = d(f(x), f(y)) = kd(x, y).$$

As $k < 1$ this means that $d(x, y) = 0$ and hence $x = y$. The theorem is proved. \square

Note that the proof is constructive. Not only do we know that a unique fixed point exists. We also know how to find it. Let us use the theorem to prove the classical Picard theorem on the existence and uniqueness of ordinary differential equations.

Consider the equation

$$\frac{dy}{dx} = F(x, y).$$

Given some x_0, y_0 we are looking for a function $f(x)$ such that $f'(x_0) = y_0$ and such that

$$f'(x) = F(x, f(x)).$$

There are some subtle issues, for example how long does the solution exist. Look at the equation $y' = y^2$, $y(0) = 1$. Then $y(x) = \frac{1}{1-x}$ is a solution. While F is a reasonably “nice” function and in particular exists for all x and y , the solution “blows up” at $x = 1$. For more examples related to Picard’s theorem see §6.3.

Theorem 7.6.3 (Picard’s theorem on existence and uniqueness). *Let $I, J \subset \mathbb{R}$ be compact intervals, let I_0 and J_0 be their interiors, and let $(x_0, y_0) \in I_0 \times J_0$. Suppose $F : I \times J \rightarrow \mathbb{R}$ is continuous and Lipschitz in the second variable, that is, there exists an $L \in \mathbb{R}$ such that*

$$|F(x, y) - F(x, z)| \leq L|y - z| \quad \text{for all } y, z \in J, x \in I.$$

Then there exists an $h > 0$ and a unique differentiable function $f : [x_0 - h, x_0 + h] \rightarrow J \subset \mathbb{R}$, such that

$$f'(x) = F(x, f(x)) \quad \text{and} \quad f(x_0) = y_0.$$

Proof. Without loss of generality assume $x_0 = 0$. As $I \times J$ is compact and $F(x, y)$ is continuous, it is bounded. Let $M := \sup\{|F(t, x)| : (t, x) \in I \times J\}$. Pick $\alpha > 0$ such that $[-\alpha, \alpha] \subset I$ and $[y_0 - \alpha, y_0 + \alpha] \subset J$. Let

$$h := \min \left\{ \alpha, \frac{\alpha}{M + L\alpha} \right\}.$$

Note $[-h, h] \subset I$. Define the set

$$Y := \{f \in C([-h, h], \mathbb{R}) : f([-h, h]) \subset [y_0 - \alpha, y_0 + \alpha]\}.$$

Here $C([-h, h], \mathbb{R})$ is equipped with the standard metric $d(f, g) := \sup\{|f(x) - g(x)| : x \in [-h, h]\}$. With this metric we have shown in an exercise that $C([-h, h], \mathbb{R})$ is a complete metric space.

Exercise 7.6.1: *Show that $Y \subset C([-h, h], \mathbb{R})$ is closed.*

Define a mapping $T : Y \rightarrow C([-h, h], \mathbb{R})$ by

$$T(f)(x) := y_0 + \int_0^x F(t, f(t)) dt.$$

Exercise 7.6.2: Show that if $f: [-h, h] \rightarrow J$ is continuous then $F(t, f(t))$ is continuous on $[-h, h]$ as a function of t . Use this to show that T is well defined and that $T(f) \in C([-h, h], \mathbb{R})$.

Let $f \in Y$ and $|x| \leq h$. As F is bounded by M we have

$$\begin{aligned} |T(f)(x) - y_0| &= \left| \int_0^x F(t, f(t)) dt \right| \\ &\leq |x|M \leq hM \leq \frac{\alpha M}{M + L\alpha} \leq \alpha. \end{aligned}$$

Therefore, $T(Y) \subset Y$. We can thus consider T as a mapping of Y to Y .

We claim T is a contraction. First, for $x \in [-h, h]$ and $f, g \in Y$ we have

$$|F(x, f(x)) - F(x, g(x))| \leq L|f(x) - g(x)| \leq Ld(f, g).$$

Therefore,

$$\begin{aligned} |T(f)(x) - T(g)(x)| &= \left| \int_0^x F(t, f(t)) - F(t, g(t)) dt \right| \\ &\leq |t|Ld(f, g) \leq hLd(f, g) \leq \frac{L\alpha}{M + L\alpha} d(f, g). \end{aligned}$$

We can assume $M > 0$ (why?). Then $\frac{L\alpha}{M + L\alpha} < 1$, and the claim is proved by taking supremum of the left hand side above to obtain $d(T(f), T(g)) \leq \frac{L\alpha}{M + L\alpha} d(f, g)$.

We apply the fixed point theorem (Theorem 7.6.2) to find a unique $f \in Y$ such that $T(f) = f$, that is,

$$f(x) = y_0 + \int_0^x F(t, f(t)) dt. \quad (7.5)$$

By the fundamental theorem of calculus, f is differentiable and $f'(x) = F(x, f(x))$. Finally, $f(0) = y_0$.

Exercise 7.6.3: We have shown that f is the unique function in Y . Why is it the unique continuous function $f: [-h, h] \rightarrow J$ that solves (7.5) above?

□

7.6.1 Exercises

Exercise 7.6.4: Suppose $X = X' = \mathbb{R}$ with the standard metric. Let $0 < k < 1$, $b \in \mathbb{R}$. a) Show that the map $F(x) = kx + b$ is a contraction. b) Find the fixed point and show directly that it is unique.

Exercise 7.6.5: Suppose $X = X' = [0, 1/4]$ with the standard metric. a) Show that $f(x) = x^2$ is a contraction, and find the best (smallest) k that works. b) Find the fixed point and show directly that it is unique.

Exercise 7.6.6: a) Find an example of a contraction of non-complete metric space with no fixed point. b) Find a 1-Lipschitz map of a complete metric space with no fixed point.

Exercise 7.6.7: Consider $y' = y^2$, $y(0) = 1$. Start with $f_0(x) = 1$. Find a few iterates (at least up to f_2). Prove that the limit of f_n is $\frac{1}{1-x}$.

Further Reading

- [BS] Robert G. Bartle and Donald R. Sherbert, *Introduction to real analysis*, 3rd ed., John Wiley & Sons Inc., New York, 2000.
- [DW] John P. D'Angelo and Douglas B. West, *Mathematical Thinking: Problem-Solving and Proofs*, 2nd ed., Prentice Hall, 1999.
- [R1] Maxwell Rosenlicht, *Introduction to analysis*, Dover Publications Inc., New York, 1986. Reprint of the 1968 edition.
- [R2] Walter Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1976. International Series in Pure and Applied Mathematics.
- [T] William F. Trench, *Introduction to real analysis*, Pearson Education, 2003. http://ramanujan.math.trinity.edu/wtrench/texts/TRENCH_REAL_ANALYSIS.PDF.

Index

- absolute convergence, 72
- absolute maximum, 92, 181
- absolute minimum, 92, 181
- absolute value, 31
- additive property of the integral, 125
- algebraic geometry, 182
- Archimedean property, 27
- arithmetic-geometric mean inequality, 30

- ball, 164
- bijection, 15
- bijective, 15
- Bolzano's intermediate value theorem, 94
- Bolzano's theorem, 94
- Bolzano-Weierstrass theorem, 61, 179
- boundary, 168
- bounded above, 21
- bounded below, 21
- bounded function, 33, 92
- bounded interval, 35
- bounded sequence, 39, 171
- bounded set, 22, 162
- bounded variation, 132

- Cantor's theorem, 17, 35
- cardinality, 15
- Cartesian product, 13
- Cauchy in the uniform norm, 143
- Cauchy sequence, 65, 175
- Cauchy series, 70
- Cauchy-complete, 67, 175
- Cauchy-Schwarz inequality, 159
- chain rule, 106
- change of variables theorem, 136

- closed ball, 164
- closed interval, 35
- closed set, 164
- closure, 168
- cluster point, 63, 79
- compact, 176
- comparison test for series, 73
- complement, 10
- complement relative to, 9
- complete, 67, 175
- completeness property, 22
- composition of functions, 15
- conditionally convergent, 72
- connected, 166
- constant sequence, 39
- continuous at c , 86, 180
- continuous function, 86, 180
- continuous function of two variables, 151
- continuously differentiable, 113
- Contraction mapping principle, 184
- converge, 40, 171
- convergent sequence, 40, 171
- convergent series, 68
- converges, 80
- converges absolutely, 72
- converges in uniform norm, 142
- converges pointwise, 139
- converges uniformly, 141
- countable, 16
- countably infinite, 16

- Darboux sum, 118
- Darboux's theorem, 112

- decreasing, 111
- Dedekind completeness property, 22
- DeMorgan's theorem, 10
- density of rational numbers, 27
- derivative, 103
- difference quotient, 103
- differentiable, 103
- differential equation, 151
- Dini's theorem, 150
- direct image, 14
- Dirichlet function, 89
- discontinuity, 89
- discontinuous, 89
- discrete metric, 160
- disjoint, 10
- distance function, 157
- divergent sequence, 40, 171
- divergent series, 68
- diverges, 80
- domain, 14

- element, 8
- elementary step function, 132
- empty set, 8
- equal, 9
- euclidean space, 159
- even function, 138
- existence and uniqueness theorem, 152, 185
- extended real numbers, 29

- field, 22
- finite, 15
- finitely many discontinuities, 128
- first derivative, 115
- first order ordinary differential equation, 151
- Fixed point theorem, 184
- function, 13
- fundamental theorem of calculus, 133

- graph, 14
- greatest lower bound, 22

- half-open interval, 35
- harmonic series, 71
- Hausdorff metric, 163
- Heine-Borel theorem, 178

- image, 14
- increasing, 111
- induction, 12
- induction hypothesis, 12
- infimum, 22
- infinite, 15
- injection, 15
- injective, 15
- integers, 9
- integration by parts, 137
- interior, 168
- intermediate value theorem, 94
- intersection, 9
- interval, 35
- inverse function, 15
- inverse image, 14
- irrational, 26

- L'Hopital's rule, 114
- Lagrange form, 115
- least upper bound, 21
- least-upper-bound property, 22
- Lebesgue covering lemma, 177
- Lebesgue number, 177
- Leibniz rule, 105
- limit, 80
- limit inferior, 57
- limit of a function, 80
- limit of a sequence, 40, 171
- limit superior, 57
- linearity of series, 71
- linearity of the derivative, 105
- linearity of the integral, 127
- Lipschitz continuous, 101
- lower bound, 21
- lower Darboux integral, 119

- lower Darboux sum, 118
- mapping, 13
- maximum, 29
- Maximum-minimum theorem, 92
- mean value theorem, 110
- mean value theorem for integrals, 131
- member, 8
- metric, 157
- metric space, 157
- minimum, 29
- Minimum-maximum theorem, 92
- monotone decreasing sequence, 42
- monotone increasing sequence, 42
- monotone sequence, 42
- monotonic sequence, 42
- monotonicity of the integral, 127

- n times differentiable, 115
- naïve set theory, 8
- natural numbers, 9
- negative, 23
- neighborhood, 164
- nonnegative, 23
- nonpositive, 23
- nth derivative, 115
- nth Taylor polynomial for f , 115

- odd function, 138
- one-to-one, 15
- onto, 15
- open ball, 164
- open cover, 176
- open interval, 35
- open neighborhood, 164
- open set, 164
- ordered field, 23
- ordered set, 21

- p-series, 74
- p-test, 74
- partial sums, 68

- partition, 117
- Picard iterate, 152
- Picard iteration, 152
- Picard's theorem, 152, 185
- pointwise convergence, 139
- polynomial, 87
- popcorn function, 90, 132
- positive, 23
- power set, 16
- principle of induction, 12
- principle of strong induction, 13
- product rule, 106
- proper, 183
- proper subset, 9

- quotient rule, 106

- range, 14
- range of a sequence, 39
- ratio test for sequences, 55
- ratio test for series, 76
- rational numbers, 9
- real numbers, 21
- refinement of a partition, 119
- relative maximum, 109
- relative minimum, 109
- remainder term in Taylor's formula, 115
- restriction, 84
- reverse triangle inequality, 32
- Riemann integrable, 121
- Riemann integral, 121
- Rolle's theorem, 109

- second derivative, 115
- sequence, 39, 171
- sequentially compact, 177
- series, 68
- set, 8
- set building notation, 9
- set theory, 8
- set-theoretic difference, 9

- set-theoretic function, 13
- squeeze lemma, 47
- standard metric on \mathbb{R}^n , 159
- standard metric on \mathbb{R} , 158
- step function, 132
- strictly decreasing, 111
- strictly increasing, 111
- subadditive, 162
- subcover, 176
- subsequence, 44, 171
- subset, 9
- subspace, 161
- subspace metric, 161
- subspace topology, 161, 170
- supremum, 21
- surjection, 15
- surjective, 15
- symmetric difference, 18

- tail of a sequence, 44
- Taylor polynomial, 115
- Taylor's theorem, 115
- Thomae function, 90, 132
- topology, 164
- totally bounded, 179
- triangle inequality, 31, 157

- unbounded interval, 35
- uncountable, 16
- uniform convergence, 141
- uniform norm, 142
- uniform norm convergence, 142
- uniformly Cauchy, 143
- uniformly continuous, 98
- union, 9
- universe, 8
- upper bound, 21
- upper Darboux integral, 119
- upper Darboux sum, 118

- Venn diagram, 10

- well ordering property, 12

- zero set, 182