Math 330 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

Last update: December 1, 2016

Version: **2016-12-01**

Contents

1 Before you start

History of Updates: *2016-08-22*

Date	*Topic*
2016-08-22	*None yet for Fall 2016*

1.1 About this document

Remark 1.1 (The purpose of this document)**.** This write-up was originally written in 2005 under the title "Introduction to Abstract Math – A Journey to Approximation Theory" and parts of it now serve as lecture notes for the course "Math 330: Number systems" which is held at the Department of Mathematical Sciences at Binghamton University.

These notes serve at least two purposes:

a. They provide background material on topics that cannot found in sufficient detail in [1] B/G (Beck/Geoghegan): The Art of Proof. This document will often be simply referred to as "B/G". It serves as the primary reference for the first two thirds of the Math 330 course.

b. They cover material which is beyond the scope of [1] B/G such as

- material on lim inf and lim sup
- convergence, continuity and compactness in metric spaces
- applications of Zorn's lemma

These topics are usually covered in the last third of my Math 330 class.

Remark 1.2 (Acknowledgements)**.** The early chapters of this document draw on the following chapters of [4] Bryant, Kirby Course Notes for MAD 2104:

Ch.1, section 1: Introduction to Sets
Ch.1, section 2: Introduction to Functions
Ch.2: Logic
Ch.3: Methods of Proofs
Ch.4, section 1: Set Operations
Ch.4, section 2: Properties of Functions

Moreover such a document cannot be written with the intent to supplement the [1] B/G book without strongly borrowing from it.

Remark 1.3 (How to navigate this document)**.** Scrutinize the table of contents, including the headings for the subchapters. You will find many entries there tagged with a directive.

For example, the reference to ch.8 (Real functions) has been tagged with **(Understand this!)**,
the first subchapter ch.8.1 (Operations on real functions) has no tag,
the second subchapter ch.8.2 (Maxima, suprema, limsup . . .) has been tagged with **(Study this!)**,
the third subchapter ch.8.3 (Sequences of sets and indicator functions and their liminf and limsup) has no tag.

All directives apply to the entire subtree and a lower level directive overrides the "parent directives". Example: the "Understand this!" directive of subsection 10.2.4: Continuity of Polynomials overrides the "Study this!" directive of subsection 10.2 on Continuity.
Accordingly, when you do not see any comment, back up in the table of contents: first to the parent entry, then to its parent entry . . . until you find one.

a. "Study this" directive: When you read "Study this", you must understand the material in depth. You will need to do this with paper and pencil in hand and make an effort not only to understand what the definitions and theorems are all about – not a minor undertaking because some of the subject matter is quite abstract – but aso make an effort to follow the proofs at least from a birds eye perspective.

b. "Understand this" directive: When you read "Understand this", you should know the definitions, propositions and theorems without worrying about proofs. Chances are that the material will be referred to from truly important sections of this write-up and is primarily needed for their understanding.

b. "Skim this" directive: When you read "Skim this", you should understand how the material is structured. You may find it easier to do some of your homework. A good example is chapter 3 on logic. It has been marked as "Skim this" but some of the later subchapters override this as "Understand this" and you will have problems doing so unless you can find your way around in the material that precedes them.

b. "Skip this" directive: When you read "Skip this", you need not worry about the content.

You will find almost every week reading assignments as part of your homework. The reading is due prior to when it is needed in class, both for this document and the Beck/Geoghegan text. I assume that you did your reading and I will assume in particular that you have learned the definitions so that I can move along at a fast pace except for some topics that I will focus on in detail. For this document it means that you should do the "Study this" and "Understand this" material as I indicated above.

Remark 1.4 (How to navigate this document)**.** My theory is that, particularly in Math, more words take a lot less time to understand than a skeletal write-up like the one given in the course text. Accordingly, almost all of the "Study this" and "Understand this"material provided in this document comes with quite detailed proofs. Those proofs are there for you to study.

Some of those proofs, notably those in prop. 6.2, make use of "$\Leftrightarrow$" to show that two sets are equal. You should study this technique but, as you will hear me say many times in class, I recommend that you abstain from using "$\Leftrightarrow$" between statements in your proofs. Chances are that you very likely lack the experience to do so without error.

Some of the material was written from scratch, other material was pulled in from a document that was written as early as 10 years ago. I have make an attempt to make the entire document more homogeneous but there will be some inconsistencies. Your help in pointing out to me the most notable trouble spots would be deeply appreciated.

There are differences in style: the original document was written in a much more colloquial style as it was addressed to talented high school students who had expressed a special interest in studying college level math.

This is a living document: material will be added as I find the time to do so. Be sure to check the latest PDF frequently. You certainly should do so when an announcement was made that this document contains new additions and/or corrections.

1.2 How to properly write a proof (Study this!)

Study this brief chapter to understand some of the dos and don'ts when submitting your homework.

To prove an equation such as $A = Z$ you are asked to do one of the following:

a.

$$\begin{aligned} A &= B \quad \text{(use ....)} \\ &= C \quad \text{(use ....)} \\ &= D \quad \text{(use ....)} \\ &\ldots\ldots\ldots\ldots \\ &= Z \quad \text{(use ....)} \end{aligned}$$

You then conclude from the transitivity of equality that $A = Z$ is indeed true.

b. *You transform the left side (L.S.) and the right side (R.S.) separately and show that in each case you obtain the same item, say M:*

Left side:

$$\begin{aligned} A &= B \quad \text{(use ....)} \\ &= C \quad \text{(use ....)} \\ &= D \quad \text{(use ....)} \\ &\ldots\ldots\ldots\ldots \\ &= M \quad \text{(use ....)} \end{aligned}$$

Right side:

$$\begin{aligned} Z &= Y \quad \text{(use ....)} \\ &= X \quad \text{(use ....)} \\ &= W \quad \text{(use ....)} \\ &\ldots\ldots\ldots\ldots \\ &= M \quad \text{(use ....)} \end{aligned}$$

and you rightfully conclude that the proof is done because it follows from $A = M$ and $Z = M$ that $A = Z$.

c. *Instead you may choose to proceed as follows*

$$\begin{aligned} A &= Z \quad \textit{(that's what you want to prove)} \\ B &= Y \quad \textit{(you do with both } A \textit{ and } Z \textit{ the same operation ......)} \\ C &= X \quad \textit{(you do with both } B \textit{ and } Y \textit{ the same operation ......)} \\ D &= W \quad \textit{(you do with both } C \textit{ and } X \textit{ the same operation ......)} \\ &\ldots\ldots\ldots\ldots \\ M &= M \quad \textit{(you do with both } L \textit{ and } N \textit{ the same operation ......)} \end{aligned}$$

What is potentially wrong with that last approach?

*In the abstract the issue is that when using method **a** or **b** you take in each step an equation that is known to be true or that you assume to be true and you rightfully conclude by the use of transitivity that you have proved what you wanted to be true.*

When you use method c you take an equation that you want to be true ($A = Z$) but have not yet proved to be so. If you are wrong then doing the same thing to both sides may potentially lead to two things that are equal.

*Here is a simple example that demonstrates why **method c is not allowed** for a mathematical proof. This method will be used in two different ways to prove that $-2 = 2$.*

First proof:

$$\begin{aligned} -2 &= 2 \quad \textit{(want to prove)} \\ -2 \cdot 0 &= 2 \cdot 0 \quad \textit{(multiply both sides from the right w. 0)} \\ 0 &= 0 \quad \textit{(B/G ax.1.2 (additive neutral element)} \end{aligned}$$

We are done. ■

Second proof:

$$\begin{aligned} -2 &= 2 \quad \textit{(want to prove)} \\ (-2)^2 &= 2^2 \quad \textit{(square both sides)} \\ 4 &= 4 \quad \textit{(obvious)} \end{aligned}$$

We are done. ■

*Now you know why you must never use method **c**.* [1]

[1] You will learn later in this chapter about injective functions which guarantee that if you do an operation (apply a function) to two different items then the results will also be different. If method c was restricted to only such operations then there would not be a problem. In the two "proofs" that show $-2 = 2$ we use operations that are not injective: In the first proof the assignment $x \mapsto 0 \cdot x$ throws everything into the same result zero. The second proof employs the assignment $x \mapsto x^2$ which maps two numbers x, y that differ by sign only to the same squared value $x^2 = y^2$.

2 Preliminaries on sets and numbers (Understand this!)

2.1 Sets and basic set operations

Ask a mathematician how her or his Math is different from the kind of Math you learn in high school, in fact, from any kind of Math you find outside textbooks for mathematicians and theoretical physicists. One of the answers you are likely to get is that Math is not so much about numbers but also about other objects, among them sets and functions. Once you know about those, you can tackle sets of functions, set functions, sets of set functions, ...

An entire book can be filled with a mathematically precise theory of sets. [2] *For our purposes the following "naive" definition suffices:*

Definition 2.1 (Sets)**.** A **set** is a collection of stuff called **members** or **elements** which satisfies the following rules: The order in which you write the elements does not matter and if you list an element two or more times then **it only counts once**.

We write a set by enclosing within curly braces the elements of the set. This can be done by listing all those elements or giving instructions that describe those elements. For example, to denote by X the set of all integer numbers between 18 and 24 we can write either of the following:

$$X := \{18, 19, 20, 21, 22, 23, 24\} \quad \text{or} \quad X := \{n : n \text{ is an integer and } 18 \leqq n \leqq 24\}$$

Both formulas clearly define the same collection of all integers between 18 and 24. On the left the elements of X are given by a complete list, on the right **setbuilder notation**, i.e., instructions that specify what belongs to the set, is used instead.

It is customary to denote sets by capital letters and their elements by small letters but this is not a hard and fast rule. You will see many exceptions to this rule in this document.

We write $x_1 \in X$ to denote that an item x_1 is an element of the set X and $x_2 \notin X$ to denote that an item x_2 is not an element of the set X

For the above example we have $20 \in X$, $27 - 6 \in X$, $38 \notin X$, 'Jimmy' $\notin X$.

Example 2.1 (No duplicates in sets)**.** The following collection of alphabetic letters is a set:

$$S_1 = \{a, e, i, o, u\}$$

and so is this one:

$$S_2 = \{a, e, e, i, i, i, o, o, o, o, u, u, u, u, u\}$$

Did you notice that those two sets are equal?

Remark 2.1. The symbol n in the definition of $X = \{n : n \text{ is an integer and } 18 \leqq n \leqq 24\}$ is a **dummy variable** in the sense that it does not matter what symbol you use. The following sets all are equal to X:

$$\{x : x \text{ is an integer and } 18 \leqq x \leqq 24\},$$
$$\{\alpha : \alpha \text{ is an integer and } 18 \leqq \alpha \leqq 24\},$$
$$\{\mathfrak{z} : \mathfrak{z} \text{ is an integer and } 18 \leqq \mathfrak{z} \leqq 24\}$$

[2] See remark 2.2 ("Russell's Antinomy") below.

Remark 2.2 (Russell's Antinomy)**.** Care must be taken so that, if you define a set with the use of setbuilder notation, no inconsistencies occur. Here is an example of a definition of a set that leads to contradictions.

$$A := \{B : B \text{ is a set and } B \notin B\} \tag{2.1}$$

What is wrong with this definition? To answer this question let us find out whether or not this set A is a member of A. Assume that A belongs to A. The condition to the right of the colon states that $A \notin A$ is required for membership in A, so our assumption $A \in A$ must be wrong. In other words, we have established "by contradiction" that $A \notin A$ is true. But this is not the end of it: Now that we know that $A \notin A$ it follows that $A \in A$ because A contains **all** sets that do not contain themselves.

In other words, we have proved the impossible: both $A \in A$ and $A \notin A$ are true! There is no way out of this logical impossibility other than excluding definitions for sets such as the one given above. It is very important for mathematicians that their theories do not lead to such inconsistencies and examples as the one above have lead to very complicated theories about "good sets". It is possible for a mathematician to specialize in the field of axiomatic set theory (actually, there are several set theories) which endeavors to show that the sets are of any relevance in mathematical theories do not lead to any logical contradictions.

The great majority of mathematicians take the "naive" approach to sets which is not to worry about accidentally defining sets that lead to contradictions and we will take that point of view in this document.

Definition 2.2 (empty set)**.** $\emptyset$ or {} denotes the **empty set**. It is the one set that does not contain any elements.

Remark 2.3 (Elements of the empty set and their properties)**.** You can state anything you like about the elements of the empty sets as there are none. The following statements all are true:

a: If $x \in \emptyset$ then x is a positive number.
b: If $x \in \emptyset$ then x is a negative number.
c: Define $a \sim b$ if and only if both are integers and $a - b$ is an even number.
For any $x, y, z \in \emptyset$ it is true that
c1: $x \sim x$,
c2: if $x \sim y$ then $y \sim x$,
c3: if $x \sim y$ and $y \sim z$ then $x \sim z$.
d: Let A be any set. If $x \in \emptyset$ then $x \in A$.

As you will learn later, **c:** means that "$\sim$" is an equivalence relation (see def.4.3 on p.70) and **d:** means that the empty set is a subset (see the next definition) of any other set.

Definition 2.3 (subsets and supersets)**.** We say that a set A is a **subset** of the set B and we write $A \subseteq B$ if any element of A also belongs to B. Equivalently we say that B is a **superset** of the set A and we write $B \supseteq A$. We also say that B includes A or A is included by B. Note that $A \subseteq A$ and $\emptyset \subseteq A$ is true for any set A.

If $A \subseteq B$ but $A \neq B$, i.e., there is at least one $x \in B$ such that $x \notin A$, we can emphasize that by saying that A is a **strict subset** of B. We write "$A \subsetneq B$" or "$A \subset B$". Alternatively we say that B is a **strict superset** of A and we write "$B \supsetneq A$") or "$B \supset A$".

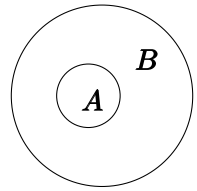


Figure 2.1: Set inclusion: $A \subseteq B, \ B \supseteq A$

Two sets A and B are equal means that they both contain the same elements. In other words, $A = B$ *iff* $A \subseteq B$ *and* $B \subseteq A$.

> *"**iff**" is a short for "if and only if": P iff Q for two statements P and Q means that if P is valid then Q is valid and vice versa.* [3]

Definition 2.4 (unions, intersections and disjoint unions)**.** Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or that either one contains any elements!

The **union** $A \cup B$ (pronounced "A union B") is defined as the set of all elements which belong to A or B or both.

The **intersection** $A \cap B$ (pronounced "A intersection B") is defined as the set of all elements which belong to both A and B.

We call A and B **disjoint** if $A \cap B = \emptyset$. We then usually write $A \uplus B$ (pronounced "A disjoint union B") rather than $A \cup B$.

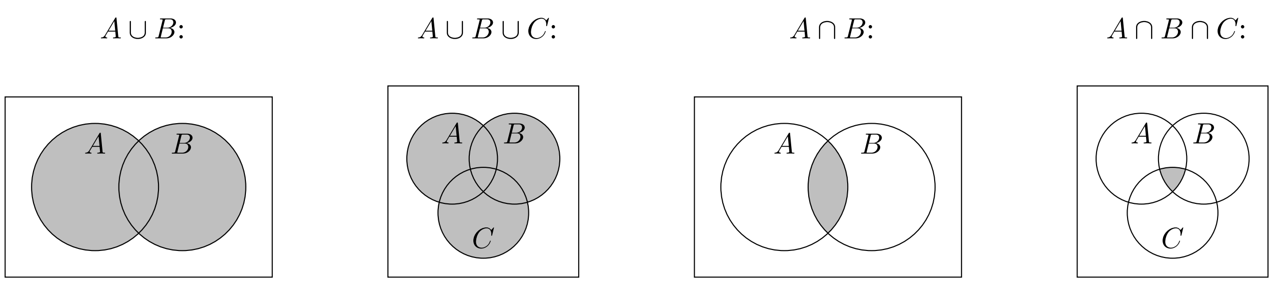


Figure 2.2: Union and intersection of sets

Definition 2.5 (set differences and symmetric differences)**.** Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or that either one contains any elements!

[3] A formal definition of "if and only if" will be given in def.def-x:logic-stmt-equiv on p.33 where we will also introduce the symbolic notation $P \Leftrightarrow Q$.

The **difference set** or **set difference** $A \setminus B$ (pronounced "A minus B") is defined as the set of all elements which belong to A but not to B:

$$A \setminus B := \{x \in A : x \notin B\} \tag{2.2}$$

The **symmetric difference** $A \triangle B$ (pronounced "A delta B") is defined as the set of all elements which belong to either A or B but not to both A and B:

$$A \triangle B := (A \cup B) \setminus (A \cap B) \tag{2.3}$$

Definition 2.6 (Universal set)**.** Usually there always is a big set Ω that contains everything we are interested in and we then deal with all kinds of subsets $A \subseteq \Omega$. Such a set is called a **"universal" set**.

For example, in this document, we often deal with real numbers and our universal set will then be $\mathbb{R}$*.*

If there is a universal set, it makes perfect sense to talk about the complement of a set:

Definition 2.7 (Complement of a set)**.** The **complement** of a set A consists of all elements of Ω which do not belong to A. We write $A^{\complement}$. or $\complement A$ In other words:

$$A^{\complement} := \complement A := \Omega \setminus A = \{\omega \in \Omega : x \notin A\} \tag{2.4}$$

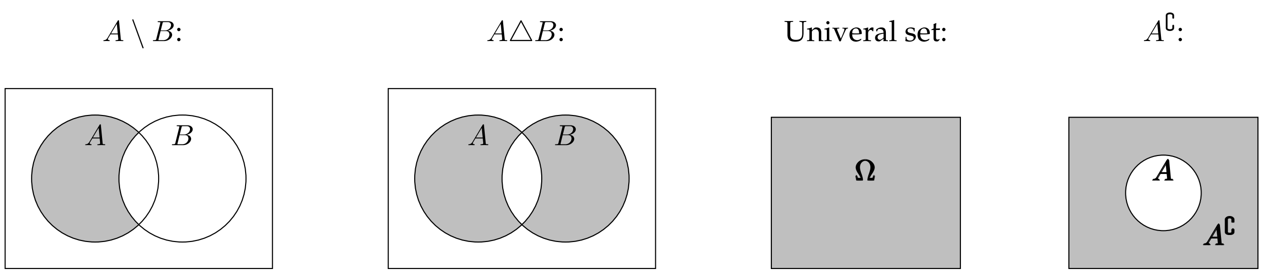


Figure 2.3: Difference, symmetric difference, universal set, complement

Remark 2.4 (Complement of empty, all)**.** Note that for any kind of universal set Ω it is true that

$$\Omega^{\complement} = \emptyset, \qquad \emptyset^{\complement} = \Omega \tag{2.5}$$

Example 2.2 (Complement of a set relative to the unit interval)**.** Assume we are exclusively dealing with the unit interval, i.e., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. [4] Let $a \in [0,1]$ and $\delta > 0$ and

$$A = \{x \in [0,1] : a - \delta < x < a + \delta\} \tag{2.6}$$

the δ–neighborhood [5] of a (with respect to $[0,1]$ because numbers outside the unit interval are not considered part of our universe). Then the complement of A is

$$A^{\complement} = \{x \in [0,1] : x \leqq a - \delta \text{ or } x \geqq a + \delta\}.$$

[4] $\mathbb{R}$ is the set of all real numbers, i.e., the kind of numbers that make up the x-axis and y-axis in a beginner's calculus course (see remark 2.6 ("Classification of numbers") on p.14).

[5] Neighborhoods of a point will be discussed in the chapter on the topology of $\mathbb{R}^n$ (see (10.4) on p.156) In short, the δ–neighborhood of a is the set of all points with distance less than δ from a.

Draw some Venn diagrams to visualize the following formulas.

Proposition 2.1. *Let A, B, X be sets and assume $A \subseteq X$. Then*

$$A \cup \emptyset = A; \qquad A \cap \emptyset = \emptyset \tag{2.7a}$$
$$A \cup \Omega = \Omega; \qquad A \cap \Omega = A \tag{2.7b}$$
$$A \cup A^{\complement} = \Omega; \qquad A \cap A^{\complement} = \emptyset \tag{2.7c}$$
$$A \triangle B = (A \setminus B) \uplus (B \setminus A) \tag{2.7d}$$
$$A \setminus A = \emptyset \tag{2.7e}$$
$$A \triangle \emptyset = A; \qquad A \triangle A = \emptyset \tag{2.7f}$$
$$X \triangle A = X \setminus A \tag{2.7g}$$
$$A \cup B = (A \triangle B) \uplus (A \cap B) \tag{2.7h}$$
$$A \cap B = (A \cup B) \setminus (A \triangle B) \tag{2.7i}$$
$$A \triangle B = \emptyset \ \textit{ if and only if } \ B = A \tag{2.7j}$$

Proof: Left as an exercise.

■

Definition 2.8 (Power set)**.** The **power set**

$$2^{\Omega} \ := \ \mathfrak{P}(\Omega) \ := \ \{A : A \subseteq \Omega\}$$

of a set Ω is the set of all its subsets.

Remark 2.5. Note that $\emptyset \in 2^{\Omega}$ for any set Ω, even if $\Omega = \emptyset$: $2^{\emptyset} = \{\emptyset\}$. It follows that the power set of the empty set is not empty.

A lot more will be said about sets once families are defined.

2.2 Numbers

Remark 2.6 (Classification of numbers)**.** [6]

We call numbers without decimal points such as $3, -29, 0, 3000000, 3 \cdot 10^6, -1, \dots$ **integers** and we write $\mathbb{Z}$ for the set of all integers.

Numbers in the set $\mathbb{N} = \{1, 2, 3, \dots\}$ of all strictly positive integers are called **natural numbers**.

A number that is an integer or can be written as a fraction is called a **rational number** and we write $\mathbb{Q}$ for the set of all rational numbers. Examples of rational numbers are

$$\tfrac{3}{4}, \ -0.75, \ -\tfrac{1}{3}, \ .\bar{3}, \ \tfrac{13}{4}, \ -5, \ 2.99\bar{9}, \ -37\tfrac{2}{7}.$$

[6] The classification of numbers in this section is not meant to be mathematically exact. For this consult, e.g., [1] B/G (Beck/Geoghegan).

The bar on top of the rightmost part of a decimal such as "$.\bar{3}$" means that this part should be repeated over and over again, e.g., $.\bar{3} = 0.33333333333\ldots$ and $1.234\overline{567} = 1.234567567567\ldots$.

Note that a mathematician does not care whether a rational number is written as a fraction "$\frac{numerator}{denominator}$" or as a decimal. The following all are representations of one third

$$0.\bar{3} \;=\; .\bar{3} \;=\; 0.33333333333\ldots \;=\; \frac{1}{3} \;=\; \frac{2}{6} \tag{2.8}$$

and here are several equivalent ways of expressing the number minus four:

$$-\,4 \;=\; -4.000 \;=\; -3.\bar{9} \;=\; -\frac{12}{3} \;=\; -\frac{400}{100} \tag{2.9}$$

We call the barred portion of the decimal digits the **period** of the number and we also talk about **repeating decimals**. The number of digits in the barred portion is called the **period length**. This period length can be bigger than one. For example, the number $1.234\overline{567}$ from above has period length 3 and the number $0.1\overline{45}$ has period length 2.

You may have heard that there are numbers which cannot be expressed as integers or fractions or numbers with a finite amount of decimals to the right of the decimal point. Examples for that are $\sqrt{2}$ *and* $\boldsymbol{\pi}$. Those "**irrational numbers**" (really, that what we call them) fill the gaps between the rational numbers. In fact, there is a simple way (but not easy to prove) of characterizing irrational numbers: Rational numbers are those that can be expressed with at most finitely many digits to the right of the decimal point, including repeating decimals. You can find the underlying theory and exact proofs in B/G ch.12. Irrational numbers must then be those with infinitely many decimal digits without any continually repeating patterns.

Example 2.3. To illustrate that repeating decimals are in fact rational numbers we convert $x = 0.1\overline{45}$ into a fraction:

$$99x \;= 100x - x \;=\; 14.5\overline{45} - 0.1\overline{45} =\; 14.4$$

It follows that $x = 144/990$ and that's definitely a fraction which you can simplify if you like.

Now we can finally give an informal definition of the most important kind of numbers: We call any kind of number, either rational or irrational, a **real number** and we write $\mathbb{R}$ for the set of all real numbers. It can be shown that there are a lot more irrational numbers than rational numbers, even though $\mathbb{Q}$ is a **dense subset** in $\mathbb{R}$ in the following sense: No matter how small an interval $(a,b) = \{x \in \mathbb{R} : a < x < b\}$ of real numbers you choose, it will contain infinitely many rational numbers.

Definition 2.9 (Types of numbers)**.** We summarize what was said sofar about the classification of numbers:

$\mathbb{N} := \{1, 2, 3, \ldots\}$ denotes the set of **natural numbers**.

$\mathbb{Z} := \{0, \pm1, \pm2, \pm3, \ldots\}$ denotes the set of all **integers**.

$\mathbb{Q} := \{n/d : n \in \mathbb{Z}, d \in \mathbb{N}\}$ denotes the set of all **rational numbers**.

$\mathbb{R} := \{$all integers or decimal numbers with finitely or infinitely many decimal digits$\}$ denotes the set of all **real numbers**.

$\mathbb{R} \setminus \mathbb{Q} = \{$all real numbers which cannot be written as fractions of integers$\}$ denotes the set of all **irrational numbers**. There is no special symbol for irrational numbers. Example: $\sqrt{2}$ and π are irrational.

Here are some customary abbreviations about often referenced sets of numbers:

$\mathbb{N}_0 := \mathbb{Z}_+ := \mathbb{Z}_{\geq 0} := \{0, 1, 2, 3, \dots\}$ denotes the set of non–negative integers,

$\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geqq 0\}$ denotes the set of all non–negative real numbers,

$\mathbb{R}^+ := \mathbb{R}_{>0} := \{x \in \mathbb{R} : x > 0\}$ denotes the set of all positive real numbers,

$\mathbb{R}^\star := \mathbb{R}_{\neq 0} := \{x \in \mathbb{R} : x \neq 0\}$.

Definition 2.10 (Intervals of real numbers)**.** We use the following notation for intervals of real numbers a, b:

$[a, b] := \{x \in \mathbb{R} : a \leqq x \leqq b\}$ is called the **closed interval** with endpoints a and b.

$]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ is called the **open interval** with endpoints a and b.

$[a, b[:= \{x \in \mathbb{R} : a \leqq x < b\}$ and $]a, b] := \{x \in \mathbb{R} : a < x \leqq b\}$ are called **half-open intervals** with endpoints a and b.

We further define the following intervals of "infinite length":

$$\begin{aligned}]-\infty, a] :=& \{x \in \mathbb{R} : x \leqq a\}, \quad]-\infty, a[:= \{x \in \mathbb{R} : x < a\}, \\ ]a, \infty[:=& \{x \in \mathbb{R} : x > a\}, \quad [a, \infty[:= \{x \in \mathbb{R} : x \geqq a\}, \quad [-\infty, \infty[:= \mathbb{R} \end{aligned} \tag{2.10}$$

Finally we define $[a, b[:=]a, b[:=]a, b] := \emptyset$ for $a \geqq b$ and $[a, b] := \emptyset$ for $a > b$.

Assumption 2.1 (Square roots are always assumed non–negative)**.** Remember that for any number a it is true that

$$a \cdot a = (-a)(-a) = a^2 \qquad \text{e.g.,} \quad 2^2 = (-2)^2 = 4$$

or that, expressed in form of square roots, for any number $b \geqq 0$

$$(+\sqrt{b})(+\sqrt{b}) = (-\sqrt{b})(-\sqrt{b}) = b.$$

We will always assume that "$\sqrt{b}$" is the **positive** value unless the opposite is explicitly stated. Example: $\sqrt{9} = +3$, not -3.

Proposition 2.2 (The Triangle Inequality for real numbers)**.** *The following inequality is used all the time in mathematical analysis to show that the size of a certain expression is limited from above:*

$$Triangle\ Inequality : \boxed{|\boldsymbol{a} + \boldsymbol{b}| \leqq |\boldsymbol{a}| + |\boldsymbol{b}|} \tag{2.11}$$

This inequality is true for any two real numbers a and b.

It is easy to prove this: just look separately at the three cases where both numbers are non-negative, both are negative or where one of each is positive and negative. ■

Proposition 2.3 (The Triangle Inequality for n real numbers)**.** *The above inequality also holds true for more than two real numbers: Let $n \in \mathbb{N}$ such that $n \geqq 2$. Let $a_1, a_2, \dots, a_n \in \mathbb{N}$. Then*

$$|a_1 + a_2 + \ldots + a_n| \leqq |a_1| + |a_2| + \ldots + |a_n| \tag{2.12}$$

The proof will be done by induction, a principle which is defined first:

Definition 2.11 (Principle of proof by mathematical induction)**.** Actually, "definition" is a misnomer. This principle is a mathematical statement that follows from the structure of the natural numbers which have a starting point to the "left" (a smallest element 1) and then progress in the well understood sequence [7]

$$2,\ 3,\ 4,\ \ldots,\ k-1,\ k,\ k+1,\ \ldots$$

This is the principle: Let us assume that we know that some statement can be proved to be true in the following two situations:

A. Base case. The statement is true for some (small) k_0; usually that means $k_0 = 0$ or $k_0 = 1$

B. Induction Step. We prove the following for all $k \in \mathbb{N}_0$ such that $k \geqq k_0$: if the property is true for k ("**Induction Assumption**") then it will also be true for $k+1$

C. Conclusion: Then the property is true for any $k \in \mathbb{N}_0$ such that $k \geqq k_0$.

Either you have been explained this principle before and say "Oh, that – what's the big deal?" or you will be mighty confused. So let me explain how it works by walking you through the proof of the triangle inequality for n real numbers (2.12).

Proof of the triangle inequality for n real numbers:

A. For $k_0 = 2$, inequality 2.12 was already shown (see 2.11), so we found a k_0 for which the property is true.

B. Let us assume that 2.12 is true for some $k \geqq 2$. We now must prove the inequality for $k+1$ numbers $a_1, a_2, \dots, a_k, a_{k+1} \in \mathbb{N}$: We abbreviate

$$A \ := \ a_1 + a_2 + \ldots + a_k; \qquad B \ := \ |a_1| + |a_2| + \ldots + |a_k|$$

then our induction assumption for k numbers is that $|A| \leqq B$. We know the triangle inequality is valid for the two variables A and a_{k+1} and it follows that $|A + a_{k+1}| \leqq |A| + |a_{k+1}|$. Look at both of those inequalities together and you have

$$|A + a_{k+1}| \leqq |A| + |a_{k+1}| \leqq B + |a_{k+1}| \tag{2.13}$$

In other words,

$$|(a_1 + a_2 + \ldots + a_k) + a_{k+1}| \ \leqq \ B + |a_{k+1}| \ = \ (|a_1| + |a_2| + \ldots + |a_k|) + |a_{k+1}| \tag{2.14}$$

[7] The first two chapters of [1] B/G (Beck/Geoghegan) use the "axiomatic" method to develop the mathematical structure of integers and natural numbers and give an exact proof of the induction principle.

and this is (2.12) for $k+1$ rather than k numbers: We have shown the validity of the triangle inequality for $k+1$ items under the assumption that it is valid for k items. It follows from the induction principle that the inequality is valid for any $k \geqq k_0 = 2$. ■

To summarize what we did in all of part B: We were able to show the validity of the triangle inequality for $k+1$ numbers under the assumption that it was valid for k numbers.

Remark 2.7 (Why induction works)**.** But how can we from all of the above conclude that the triangle inequality works for all $n \in \mathbb{N}$ such that $n \geqq k_0 = 2$? That's much simpler to demonstrate than what we just did.

Step 1:	We know that it is true for $k_0 = 2$ because that was actually proved in A.
Step 2:	But according to B, if it is true for k_0, it is also true for the successor $k_0 + 1 = 3$.
Step 3:	But according to B, if it is true for $k_0 + 1$, it is also true for the successor $(k_0 + 1) + 1 = 4$.
Step 4:	But according to B, if it is true for $k_0 + 2$, it is also true for the successor $(k_0 + 2) + 1 = 5$.
......	....
Step 53,920:	But according to B, if it is true for $k_0 + 53{,}918$, it is also true for the successor $(k_0 + 53{,}918) + 1 = 53{,}921$.
......	....

And now you understand why it is true for any natural number $n \geqq k_0$. ■

2.3 Addenda to chapter 2

2.3.1 Bounded sets in $\mathbb{Z}$

The material given here complements ch.2.4 (The Well-Ordering Principle) of [1] Beck/Geoghegan.

Definition 2.12 (Upper and lower bounds, maxima and minima in $\mathbb{Z}$)**.** Let $A \subseteq \mathbb{Z}$. [8] Let $l, u \in \mathbb{Z}$. We call l a **lower bound** of A if $l \leqq a$ for all $a \in A$.

We call u an **upper bound** of A if $u \geqq a$ for all $a \in A$.

We call A **bounded above** if this set has an upper bound and we call A **bounded below** if A has a lower bound. We call A **bounded** if A is both bounded above and bounded below.

A **minimum** (min) of A is a lower bound l of A such that $l \in A$. A **maximum** (max) of A is an upper bound u of A such that $u \in A$.

The next proposition states that min and max are unique if they exist. This makes it possible to write $\min(A)$ or $\min A$ for the minimum of A and $\max(A)$ or $\max A$ for the maximum of A.

Proposition 2.4. *Let $A \subseteq \mathbb{Z}$. If A has a maximum then it is unique. If A has a minimum then it is unique.*

Proof: Homework!

[8] The definitions given here also apply to subsets of $\mathbb{Q}$ and $\mathbb{R}$. See def.8.5 on p.115.

■

Definition 2.13 (Translation and dilation of sets)**.** We use here the following notation for a set $A \subseteq \mathbb{Z}$.

Let $\alpha, b \in \mathbb{Z}$. We define

$$\lambda A + b \ := \{\lambda a + b : a \in A\}. \tag{2.15}$$

In particular, for $\lambda = \pm 1$ we obtain

$$A + b \ = \{a + b : a \in A\}, \tag{2.16}$$
$$-A \ = \{-a : a \in A\}. \tag{2.17}$$

Remark 2.8. Note that the above makes sense for any algebraic structure [9] with binary operations "+" and/or "·", e.g., groups or vector spaces (in the latter case "$(\lambda, a) \mapsto \lambda a$" would be the scalar product between a real number (scalar) λ and a vector a.

Theorem 2.1 (Generalization of the Well-Ordering Principle)**.**

a. *Let A be a nonempty subset of $\mathbb{Z}$ which is bounded below. Then A possesses a minimum in $\mathbb{Z}$.*
b. *Let B be a nonempty subset of $\mathbb{Z}$ which is bounded above. Then B possesses a maximum in $\mathbb{Z}$.*
c. *Let C be a nonempty bounded subset of $\mathbb{Z}$. Then C possesses both minimum and maximum in $\mathbb{Z}$.*

Proof (outline):

a. *If A has 1 as a lower bound then $A \subseteq \mathbb{N}$ and the theorem simply is the Well-Ordering Principle (B/G theorem 2.32). Next we just assume that A is bounded below. Let $a_\star$ be a lower bound of A. Let $A' := A - a_\star + 1$. Then $a' \geq 1$ for all $a' \in A'$ and it follows from the Well-Ordering Principle that the minimum $\min(A')$ of A' exists. It is easy to see from $\min(A') \in A'$ that then $m := \min(A') + a_\star - 1 \in A$ and that m is a lower bound of A because $a_\star$ is a lower bound of A. It follows that $m = \min(A)$.*

b. *We assume that B is bounded above. Let $b^\star$ be an upper bound of B. Let $B' := -B$. Then B' has $-b^\star$ as a lower bound and it follows from the already proven part **a** that the minimum $\min(B')$ of B' exists. Let $m := -\min(B')$. It follows from $\min(B') \in B'$ that $m \in -B' = -(-B) = B$ and it follows from $\min(B') \leqq b'$ for all $b' \in B'$ that $m \geqq b$ for all $b \in B$. But then m must be the maximum of A.*

c. *is a trivial consequence of **a** and **b*** ■

Example 2.4 (The Well-Ordering Principle is not true in $\mathbb{Q}$ and $\mathbb{R}$)**.**

a. $\mathbb{R}$: The set $A := \{x \in \mathbb{R} : x^2 < 2\}$ is bounded in $\mathbb{R}$ (by ± 2) but has neither min (would have to be $-\sqrt{2} \notin A$) nor max (would have to be $+\sqrt{2} \notin A$).
but: $-\sqrt{2}$ is a special lower bound: it is the **greatest lower bound** $\inf(A)$ of A and $\sqrt{2}$ is a special upper bound: it is the **least** (smallest) **upper bound** $\sup(A)$ of A.

b. $\mathbb{Q}$: The set $B := \{x \in \mathbb{Q} : x^2 < 2\} = A \cap \mathbb{Q}$ is bounded in $\mathbb{Q}$ (by ± 2) and also has neither min nor max for the same reasons as A.
Further: $-\sqrt{2}$ is **not** a lower bound of B and $\sqrt{2}$ is **not** an upper bound of B because those numbers are not in our "universe" $\mathbb{Q}$. The set B has neither min, max, inf, sup!

[9] See the optional chapter 13 (Algebraic structures) on p.226

2.3.2 Examples and exercises for sets

Exercise 2.1. Let $X = \{x, y, \{x\}, \{x, y\}\ \}$. True or false?

a. $\{x\} \in X$ **c.** $\{\ \{x\}\ \} \in X$ **e.** $y \in X$ **g.** $\{y\} \in X$
b. $\{x\} \subseteq X$ **d.** $\{\ \{x\}\ \} \subseteq X$ **f.** $y \subseteq X$ **h.** $\{y\} \subseteq X$

For the subsequent exercises refer to example 4.4 for the preliminary definition of cardinality of a set and to def.4.1 (Cartesian Product of two sets) for the definition of Cartesian product. You find both in ch.4.1 (Cartesian products and relations) on p.69

Exercise 2.2. Find the cardinality of each of the following sets:

a. $A = \ \{x, y,\ \{x\},\ \{x, y\}\ \}$ **c.** $C = \ \{u, v, v, v, u\}$ **e.** $E = \ \{\sin(k\pi/2) : k \in \mathbb{Z}\}$
b. $B = \ \{1,\ \{0\},\ \{1\}\ \}$ **d.** $D = \ \{3z - 10 : z \in \mathbb{Z}\}$ **f.** $F = \ \{\pi x : x \in \mathbb{R}\}$

Exercise 2.3. Let $X = \{x, y, \{x\}, \{x, y\}\ \}$ and $Y = \{x, \{y\}\ \}$. True or false?

a. $x \in X \cap Y$ **c.** $x \in X \cup Y$ **e.** $x \in X \setminus Y$ **g.** $x \in X \Delta Y$
b. $\{y\} \in X \cap Y$ **d.** $\{y\} \in X \cup Y$ **f.** $\{y\} \in X \setminus Y$ **h.** $\{y\} \in X \Delta Y$

Exercise 2.4. Let $X = \ \{1, 2, 3, 4\}$ and let $Y = \ \{x, y\}$.

a. What is $X \times Y$? **c.** What is card$(X \times Y)$? **e.** Is $(x, 3) \in X \times Y$? **g.** Is $3 \cdot x \in X \times Y$?
b. What is $Y \times X$? **d.** What is card$(Y \times X)$? **f.** Is $(x, 3) \in Y \times X$? **h.** Is $2 \cdot y \in Y \times X$?

3 Logic (Skim this!)

This chapter uses material presented in ch.2 (Logic) and ch.3 (Methods of Proofs) of [4] Bryant, Kirby Course Notes for MAD 2104.

3.1 Prologue: Notation for functions

The material presented in this section will be discussed in greater detail in chapter 4 (unctions and relations) on p.69. It is presented here because the definition of a function is needed to properly discuss statement functions.

Note 3.1 (Motivation for a good function definition)**.** When discussing logic we deal with statement functions (predicates) (see def.3.4 on p.24) and we are in the same predicament as when discussing some run of the mill functions known from calculus such as $f_1(x) = \sqrt{x}$ and $f_2(x,y) = \ln(x-y)$: Sometimes $f_1(x)$ means the entire graph, i.e., the entire collection of pairs $\big(x, \sqrt{x}\big)$ and sometimes it just refers to the function value $\sqrt{x}$ for a "fixed but arbitrary" number x. In case of the function $f_2(x)$: Sometimes $f_2(x,y)$ means the entire graph, i.e., the entire collection of pairs $\big((x,y), \ln(x-y)\big)$ and sometimes it just refers to the function value $\ln(x-y)$ for a pair of "fixed but arbitrary" numbers (x,y).

This issue is addressed in the material of ch.4.2 on p.71 which precedes the mathematically precise definition of a function (def.4.6 on p.74). You are encouraged to look at it once you have read the remainder of this short section as ch.4.2 contains everything you see here.

To get to a usable definition of a function there are several things to consider. In the following $f_1(x)$ and $f_2(x,y)$ again denote the functions $f_1(x) = \sqrt{x}$ and $f_2(x,y) = \ln(x-y)$.

- **a.** The source of all allowable arguments (x–values in case of $f_1(x)$ and (x,y)–values in case of $f_2(x,y)$) will be called the domain of the function. The domain is explicitly specified as part of a function definition and it may be chosen for whatever reason to be only a subset of all arguments for which the function value is a valid expression. In case of the function $f_1(x)$ this means that the domain must be restricted to a subset of the interval $[0,\infty[$ because the square root of a negative number cannot be taken. In case of the function $f_2(x,y)$ this means that the domain must be restricted to a subset of $\{\ (x,y) : x,y \in \mathbb{R}\ and\ x-y > 0\}$ because logarithms are only defined for strictly positive numbers.
- **b.** The set to which all possible function values belong will be called the codomain of the function. As is the case for the domain, the codomain also is explicitly specified as part of a function definition. It may be chosen as any <u>superset</u> of the set of all function values for which the argument belongs to the domain of the function. In case of the function $f_1(x)$ this means that we are OK if the codomain is a superset of the interval $[0,\infty[$. Such a set is big enough because square roots are never negative. It is OK to specify the interval $]-3.5,\infty$ or even the set $\mathbb{R}$ of all real numbers as the codomain. In case of the function $f_2(x,y)$ this means that we are OK if the codomain contains $\mathbb{R}$. Not that it would make a lot of sense but the set $\mathbb{R} \cup \{$ all inhabitants of Chicago $\}$ also is an acceptable choice for the codomain.

c. A function $y = f(x)$ is not necessarily something that maps (assigns) numbers or pairs of numbers to numbers but domain and codomain can be a very different kind of animal. In this chapter on logic you will learn about statement functions $A(x)$ which assign arguments x from an arbitrary set $\mathscr{U}$, called the universe of discourse, to statements $A(x)$, i.e., sentences that are either true or false.

d. Considering all that was said so far you can now think of the graph of a function $f(x)$ with domain D and codomain C (see the beginning of this chapter) as the set $\Gamma_f := \{\big(x, f(x)\big) : x \in D\}$. Alternatively you can characterize this function by the assignment rule which specifies how the function value $f(x)$ depends on any given argument $x \in D$. We write "$x \mapsto f(x)$ to indicate this. You can also write instead $f(x) =$ whatever the actual function value will be. This is possible if you do not write about functions in general but about specific functions such as $f_1(x) = \sqrt{x}$ and $f_2(x, y) = \ln(x - y)$. We further write "$f : C \to D$" as a short way of saying that the function $f(x)$ has domain C and codomain D.

In case of the function $f_1(x) = \sqrt{x}$ for which we choose the interval $X := [2.5, 7]$ as the domain (small enough because $X \subseteq [0, \infty[$) and $Y :=]1, 3[$ as the codomain (big enough because $1 < \sqrt{x} < 3$ for any $x \in X$) we specify this function as

$$\text{either } f_1 : X \to Y, \quad x \mapsto \sqrt{x} \qquad\qquad \text{or } f_1 : X \to Y, \quad f(x) = \sqrt{x}.$$

Let us choose $U := \{(x, y) : x, y \in \mathbb{R} \;\; 1 \leqq x \leqq 10 \text{ and } y < -2\}$ as the domain and $V := [0, \infty[$ as the codomain for $f_2(x, y) = \ln(x - y)$. These choices are OK because $x - y \geqq 1$ for any $(x, y) \in U$ and hence $ln(x - y) \geqq 0$, i.e., $f_2(x, y) \in V$ for all $(x, y \in U$. We specify this function as

$$\text{either } f_2 : U \to V, \quad (x, y) \mapsto \ln(x - y) \qquad\qquad \text{or } f_2 : U \to V, \quad f(x, y) = \ln(x - y).$$

We incorporate the above into the following preliminary definition.

Definition 3.1 (Preliminary definition of a function)**.** A **function** f consists of two nonempty sets X and Y and an assignment rule $x \mapsto f(x)$ which assigns any $x \in X$ uniquely to some $y \in Y$. We write $f(x)$ for this assigned value and call it the function value of the argument x. X is called the **domain** and Y is called the **codomain** of f. We write

$$f : X \to Y, \qquad x \mapsto f(x). \tag{3.1}$$

We read "$a \mapsto b$" as "a is assigned to b" or "a maps to b" and refer to $\mapsto$ as the **maps to operator** or **assignment operator**.

Remark 3.1. The name given to the argument variable is irrelevant. Let f_1, f_2, X, Y, U, V be as defined in **d** of note 3.1. The function

$$g_1 : X \to Y, \quad p \mapsto \sqrt{p}$$

is identical to the function f_1. The function

$$g_2 : U \to V, \quad (t, s) \mapsto \ln(t - s)$$

is identical to the function f_2 and so is the function

$$g_3 : U \to V, \quad (s,t) \mapsto \ln(s-t).$$

The last example tells you that you can swap function names as long as you do it consistently in all places.

Note 3.2 (Textual variables)**.** It was mentioned in **c** above that the input variables and function values need not necessarily numbers but they can also be textual. For example, the domain of a function may consist of the first names of certain persons.

A note on textual variables: If the variable is the last name of the person James Joice and valid input for the function $F : p \mapsto$ "Each morning p writes two pages.") then we write interchangeably Joyce or 'Joyce'. Quotes are generally avoided unless they add clarity.

In the above example "Each morning 'Joyce' writes two pages." emphasizes that Joyce is the replacement of a parameter where as F('Joyce') does not seem to improve the simpler notation F(Joyce) and you will most likely see the expression F(Joyce) = "Each morning 'Joyce' writes two pages."

We also need the definition of a cartesian product. [10]

Definition 3.2 (Preliminary definition: cartesian product)**.** Let X and Y be two sets The set

$$X \times Y \ := \{(x,y) : x \in X, y \in Y\} \tag{3.2}$$

is called the **cartesian product** of X and Y.

Note that the order is important: (x,y) and (y,x) are different unless $x = y$.

We write X^2 as an abbreviation for$X \times X$.

This definition generalizes to more than two sets as follows: Let $X_1, X_2, \dots, X_n$ be sets. The set

$$X_1 \times X_2 \cdots \times X_n \ := \{(x_1, x_2, \dots, x_n) : x_j \in X_j \text{ for each } j = 1, 2, \dots n\} \tag{3.3}$$

is called the cartesian product of $X_1, X_2, \dots, X_n$.

We write X^n as an abbreviation for$X \times X \times \cdots \times X$.

Remark 3.2. The domains given in **a** and **d** of note 3.1 are subsets of the cartesian product $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x,y) : x, y \in \mathbb{R}\}$.

[10] See ch.4.1 (Cartesian products and relations) on p.69 for the real thing and examples.

3.2 Statements and statement functions

Definition 3.3 (Statements)**.** A **statement** [11] is a sentence or collection of sentences that is either true or false. We write T or **true** for "true" and F or **false** for "false" and we refer to those constants as **truth values**

Example 3.1. The following are examples of statements:

- **a.** "Dogs are mammals" (a true statement)
- **b.** "Roses are mammals. 7 is a number." This is a false statement which also could have been written as a single sentence: "Roses are mammals and 7 is a number"
- **c.** "I own 5 houses" (a statement because this sentence is either true or false depending on whether I told the truth or I lied)
- **d.** "The sum of any two even integers is even" (a true statement)
- **e.** "The sum of any two even integers is even **and** Roses are mammals" (a false statement)
- **f.** "**Either** the sum of any two even integers is even **or** Roses are mammals" (a true statement)

Example 3.2. The following are **not** statements:

- **a.** "Who is invited for dinner?"
- **b.** "$2x = 27$" (the variable x must be bound (specified) to determine whether this sentence is true or false: It is true for $x = 13.5$ and it is false for $x = 33$)
- **c.** "$x^2 + y^2 = 34$" (both variables x and y must be bound to determine whether this sentence is true or false It is true for $x = 5$ and $y = 3$ and it is false for $x = 7.8$ and $y = 2$)
- **d.** "Stop bothering me!"

For the remainder of the entire chapter on logic we define

$$\mathscr{S} := \textit{the set of all statements} \tag{3.4}$$

$\mathscr{S}$ will appear as the codomain of statement functions.

Be sure to understand the material of ch.3.1 (Prologue: Notation for functions) on p.21) before continuing.

Definition 3.4 (Statement functions (predicates))**.** We need to discuss some preliminaries before arriving at the definition of a statement function. Let A be a sentence or collection of sentences which contains one or more variables (placeholders) such that, if each of those variables is assigned a specific value, it is either true or false, i.e., it is an element of the set $\mathscr{S}$ of all statements. If A contains n variables $x_1, x_2, \dots, x_n$ and if they are **bound**, i.e., assigned to the specific values $x_1 = x_{10},\ x_2 = x_{20}, \dots, x_n = x_{n0}$, we write $A(x_{10}, x_{20}, \dots, x_{n0})$ for the resulting statement.

To illustrate this let A := "x is green and y and z like each other". If we know the specific values for the variables x, y, z then this sentence will be true or false. For example $A(\text{this lime, Tim, Fred})$ is true or false depending on whether Tim and Fred do or do not like each other.

There are restrictions for the choice of $x_1 = x_{10}, x_2 = x_{20}, \dots, x_n = x_{n0}$: Associated with each variable x_j in A is a set $\mathscr{U}_j$ which we call the **universe of discourse**, in short, **UoD**, for the jth

[11] usually called a **proposition** in a course on logic but we do not use this term as in mathematics "proposition" means a theorem of lesser importance.

variable in A. Each value x_{j0} $(j = 1, 2, \dots n)$ must be chosen in such a way that $x_{j0} \in \mathscr{U}_j$. If this is not the case then the expression $A(x_{10}, x_{20}, \dots, x_{n0})$ is called **inadmissible** and we refuse to deal with it.

What was said can be rephrased as follows: We have an assignment $(x_1, x_2, \dots, x_n) \mapsto A(x_1, x_2, \dots, x_n)$ which results in a statement, i.e., an element of $\mathscr{S}$ (see (3.4)) just as long as $x_{j0} \in \mathscr{U}_j$. In other words we have a function

$$A : \mathscr{U}_1 \times \mathscr{U}_2 \times \cdots \times \mathscr{U}_n \to \mathscr{S}, \quad (x_1, x_2, \dots, x_n) \mapsto A(x_1, x_2, \dots, x_n) \tag{3.5}$$

in the sense of def. 3.1. with the cartesian product of the UoDs for $x_1, \dots, x_n$ as domain and $\mathscr{S}$ as codomain. We call such a function a **statement function** [12] or **predicate**.

Note 3.3 (Relaxed notation for statement functions)**.** You should remember that a statement function is a function in the sense of def.3.1 but we we will often use the simpler notation

A := "some text that contains the placeholders $x_1, x_2, \dots, x_n$ and evaluates to **true** or **false** once all x_j are bound"

together with the specification of each UoD $\mathscr{U}_j$ rather than the formal notation

$$A : \mathscr{U}_1 \times \mathscr{U}_2 \times \cdots \times \mathscr{U}_n \to \mathscr{S}, \quad (x_1, x_2, \dots, x_n) \mapsto A(x_1, x_2, \dots, x_n).$$

If A contains two or more variables then the formal notation has an advantage. There is no doubt when looking at an evaluation such as $A(5.5, 7, -3, 8)$ which placeholder in the string corresponds to 5.5, which one corresponds to 7 etc. When employing the relaxed notation then we decide this according to the following

Left to right rule for statement functions: If the string A contains n different place holders then the expression $A(x_{10}, x_{20}, \dots x_{n0})$ implies the following: If the name of the first (leftmost) place holder in A is x then each occurrence of x is bound to the value x_{10}. If the name of the first of the remaining place holders in A is y then each occurrence of y is bound to the value x_{20}, ... After $n-1$ steps the remaining placeholders all have the same name, say z and each occurrence of z is bound to the value x_{n0}. If there is any confusion about what is first, what is second, ... then this will be indicated when A is specified or when its variables are bound for the first time.

Example 3.3. In def.3.4 A = "x is green and y and z like each other" was used to illustrate the concept of a statement function. We never showed how to write the actual statement function. We must decide the UoDs for x, y, z and we define them as follows.
UoD for x: $\mathscr{U}_x :=$ all plants and animals in the U.S.,
UoDs for y and z: $\mathscr{U}_y := \mathscr{U}_z :=$ all BU majors in actuarial science.

a. Here is the formal definition: Let A be the statement function

$$A : \mathscr{U}_x \times \mathscr{U}_y \times \mathscr{U}_z \to \mathscr{S}, \quad (x, y, z) \mapsto A(x, y, z) := \text{"}x \text{ is green and } y \text{ and } z \text{ like each other"}$$

[12] A statement function is usually called a **proposition function** in a course on logic. As previously mentioned, we do not use the term "proposition" in this document because in most brances of mathematics it refers to a theorem of lesser importance.

b. Here is the relaxed definition: Let A be the statement function

A := "x is green and y and z like each other" with UoDs $\mathscr{U}_x$ for x, $\mathscr{U}_y$ for y and $\mathscr{U}_z$ for z.

The example above and all those below for statement functions of more than a single variable employ the left to right rule.

Adhering to the left to right rule is not a big deal because

We will restrict ourselves in this document from now on to statement functions of one or two variables.

Example 3.4. Let $A(t)$ = "$t - 4.7$ is an integer". Then $A : \mathbb{R} \to \mathscr{S},\ x \mapsto A(x)$ is a one parameter statement function with UoD $\mathbb{R}$ and x as the variable. Note that it is immaterial that we wrote t in the equation and x in the "$\mapsto$" expression because with deal with a dummy variable and we have employed its name consistently in both cases. We have

- **a.** A(Honda) = "'Honda' $- 4.7$ is an integer" is inadmissible because a car brand is not part of our universe of discourse.
- **b.** If $u_0 \in \mathscr{U}$ then $A(u_0)$ = "$u_0 - 4.7$ is an integer" is a statement which evaluates to true or false depending on that fixed but unknown value of u_0.
- **c.** If $n \in \mathscr{U}$ then $A(n)$ is the statement(!) "$n - 4.7$ is an integer". It does not matter that this expression looks exactly like the original A: The expression $A(n)$ implies that the parameter inside the sentence collection A which happens to be named "n" has been bound to a fixed (but unspecified) value also denoted by n.

Example 3.5. Let $B(x,y)$:= "$x^2 - y + 2 = 11$". Then $B : \mathbb{R}\times]1,100[\to \mathscr{S},\ (x,y) \mapsto B(x,y)$ is a two parameter statement function with UoD $\mathbb{R}$ for x and UoD $]1,100[$ for the variable y. Then

- **a.** $B(4,-2)$ = "$4^2 - (-2) + 2 = 11$" (a false statement) because x is the leftmost item in B.
- **b.** $B(z,10)$ = "$z^2 - 10 + 2 = 11$" (true or false depending on z).
- **c.** **BE CAREFUL:** If $x, y \in \mathbb{R}$ then $B(y,x)$ = "$y^2 - x + 2 = 11$" and **NOT** "$x^2 - y + 2 = 11$" because the "evaluate left to right" rule matters, not any similarity or even coincidence between the symbols inside the sentence collection and in the evaluation $B(\cdot,\cdot)$

Example 3.6. The following are predicates:

- **a.** P := "$2x = 27$" (see example 3.2.**b**), UoD $\mathscr{U} := \{x \in \mathbb{R} : x > 10\}$
- **b.** Q := "$x^2 + y^2 = 34$" (example 3.2.**c**), UoD $\mathscr{V} := \{(x,y) : x, y \in \mathbb{R}$ and $x < y\}$
- **c.** R := "$x^2 + y^2 = 34$ and $xy > 100$", UoDs are $\mathscr{W}_x := \mathscr{W}_y := [-50, 25]$.

Note the following for **c**: $R(-30, 20)$ evaluates to a false statement because $(-30) \cdot 20 > 100$ is false. $R(30, 20)$ does not evaluate to any kind of statement: It is an inadmissible expression because $30 \notin \mathscr{W}_x$.

d. The sentence "Stop bothering x!" is **not** a statement function because this imperative will not be true or false even if x is bound to a specific value.

Example 3.7. Let B := "$x + 7 = 16$ and d is a dog". Let $\mathscr{U}_x := \mathbb{N}$ and $\mathscr{U}_d := \{d : d \text{ is a vegetable or animal}\}$.

B becomes a statement function of two variables x and d if we specify that the UoD for x is $\mathscr{U}_x$ and the UoD for d is $\mathscr{U}_d$

Assume for the following that Robby is an animal.

a. $B(9, \text{Robby})$ is the statement "$9 + 7 = 16$ and Robby is a dog". It is true in case Robby is a dog and false in case Robby is not a dog.
b. $B(20, \text{Robby})$ is the statement "$20 + 7 = 16$ and Robby is a dog" which is false regardless of what Robby might be because $20 + 7 = 16$ by itself is false.
c. $B(d, F)$ is the statement "$d + 7 = 16$ **and** $F$is a dog": which is true or false depending on the fixed but unspecified values of d and F. Note that d corresponds to the leftmost variable x inside B and not to the second variable d!
d. $B(x)$ is not a valid expression as we do not allow "partial evaluation" of a predicate. [13]

3.3 Logic operations and their truth tables

We now resume our discussion of statements.

3.3.1 Overview of logical operators

*Statements can be connected with **logical operators**, also called **connectives**, to form another statement, i.e., something that is either **true** or **false**.*

Here is an overview of the important connectives. [14] *Their meaning will be explained subsequently, once we define compound statements and compound statement functions.*

negation:	$\neg A$	***not*** A
conjunction:	$A \wedge B$	A ***and*** B
double arrow (biconditional):	$A \leftrightarrow B$	A ***double arrow*** B
logical equivalence:	$A \Leftrightarrow B$	A ***if and only if*** B
disjunction (inclusive or):	$A \vee B$	A ***or*** B
exclusive or:	A **xor** B	***either*** A ***or*** B, ***exactly one of*** A ***or*** B
arrow:	$A \to B$	A ***arrow*** B, ***if*** A ***then*** B
implication:	$A \Rightarrow B$	A ***implies*** B, ***if*** A ***then*** B

[13] To indicate that we consider d as fixed but arbitrary and want to interpret "$x + 7 = 16$ and d is a dog" as a statement function of only x as a variable we could have introduced the notation $B(\cdot, d) : x \mapsto B(x, d)$. Similarly, to indicate that we consider x as fixed but arbitrary and want to interpret "$x + 7 = 16$ and d is a dog" as a statement function of only d as a variable we could have introduced the notation $B(x, \cdot) : d \mapsto B(x, d)$. We choose not to overburden the reader with this additional notation. Rather, this situation can be handled by defining two new predicates $C : x \mapsto C(x) :=$ "$x + 7 = 16$ and z is a dog" and $D : d \mapsto D(d) :=$ "$z + 7 = 16$ and d is a dog" and then state that z is not a variable but a fixed (but unspecified) value.

[14] This order is rather unusual in that usually you would discuss biconditional and logical equivalence operators last, but logical equivalence between two statements A and B is what we think of when saying "A if and only if B" and it helps to understand what this phrase means in the context of logic as early as possible.

Notations 3.1 (use of symbols vs descriptive English)**.**

a. In the entire chapter on logic we generally use for logical operators their symbols like "$\neg$" or "$\Rightarrow$" in formulas but we use their corresponding English expressions (**not** and **implies** in this case) in connection with constructs which contain English language.

For example we would write $\neg(A \vee \neg B)$ rather than **not**$(A$ **or not** $B)$ but we would write "$d + 7 = 16$ **and** F is a dog" rather than "$(d + 7 = 16) \wedge (F$is a dog)"

b. Outside chapter 3 symbols are not used at all for logical operators. We use boldface such as "**and**" rather than just plain type face only to make it visually easier to understand the structure of a mathematical construct which employs connectives.

Definition 3.5 (Compound statements)**.** A statement which does not contain any logical operators is called a **simple statement** and one that employs logical operators is called a **compound statement**.

Similarly statement functions which contain logical operators are called **compound statement functions**.

Example 3.8. Statements **e** and **f** of example 3.1 are examples of compound statements.

In **e** the two simple statements "The sum of any two even integers is even" and "Roses are mammals" are connected by **and**.

In **f** the two simple statements "The sum of any two even integers is even" and "Roses are mammals" are connected by **either** ... **or**.

3.3.2 Negation and conjunction, truth tables and tautologies (Understand this!)

We now give the definition of the first two logical operators which were introduced in the table of section 3.3.1.

Definition 3.6 (Negation)**.** The **negation operator** is represented by the symbol "$\neg$" and it reverses the truth value of a statement A, i.e., if A is **true** then $\neg(A)$ is **false** and if A is **false** then $\neg(A)$ is **true**.

(3.6) This is expressed in this "truth table" for $\neg A$:[15]

A	¬ A
F	T
T	F

Example 3.9. Let A := "Rover is a horse". Then $\neg A$ = "Rover is **not** a horse" and $\neg\neg A = \neg(\neg A)$ = "Rover is a horse" = A.

Let us not quibble here about whether $\neg\neg A$ is not in reality the statement "Rover is not not a horse" which admittedly means the same as "Rover is a horse" but looks different.

There is no question about the fact that the T/F values for A and $\neg\neg A$ are the same. Just compare column 1 with column 3.

A	$\neg A$	$\neg(\neg A)$
F	T	F
T	F	T

[15] The definition of a truth table will be given shortly. See def.3.8 on p.29.

Note that we did not use any specifics about A. We derived the T/F values for $\neg\neg A$ from those in the second column by applying the definition of the $\neg$ operator to the statement $B := \neg A$.

In other words we have proved that the statements A and $\neg\neg A$ are **logically equivalent** in the sense that one of them is true whenever the other one is true and vice versa.

*All operators discussed subsequently are **binary operators**, i.e., they connect two input parameters (statements) A, B and four rather than two rows are needed to show what will happen for each of the four combinations A: **false** and B: **false**, A: **false** and B: **true**, A: **true** and B: **false**, A: **true** and B: **true**.*

*In contrast, the already discussed negation operator "¬" is a **unary operators**, i.e., it has a single input parameter. We will keep referring to "¬" as a connective even though there are no two or more items that can be connected.*

Definition 3.7 (Conjunction)**.** The **conjunction operator** is represented by the symbols "$\wedge$" or "**and**". The expression A **and** B is **true** if and only if both A and B are **true**.

(3.7) Truth table for A **and** B:

A	B	$A \wedge B$
F	F	F
F	T	F
T	F	F
T	T	T

The **and** connective generalizes to more than two statements $A_1, A_2, \ldots, A_n$ in the obvious manner:

$A_1 \wedge A_2 \wedge \cdots \wedge A_n$ is **true** if and only if each one of $A_1, A_2, \ldots, A_n$ is **true** and **false** otherwise.

Definition 3.8 (Truth table)**.** A **truth table** contains the symbols for statements in the header, i.e., the top row and shows in subsequent rows how their truth values relate.

It contains in the leftmost columns statements which you may think of as varying inputs and it contains in the columns to the right compound statements which were built from those inputs by the use of logical operators. We have a row for each possible combination of truth values for the input statements. Such a combination then determines the truth value for each of the other statements.

When we count rows we start with zero for the header which contains the statement names. Row 1 is the first row which contains T/F values.

An example for a truth table is the following table which you encountered in the definition above 3.7 of the conjunction operator:

A	B	$A \wedge B$
F	F	F
F	T	F
T	F	F
T	T	T

Here the input statements are A and B. The compound statement $A \wedge B$ is built from those inputs with the use of the $\wedge$ operator. We have 4 possible T/F combinations for A and B and each one of those determines the truth value of $A \wedge B$. For example, row 2 contains A:F and B:T and from this we obtain F as the corresponding truth value of $A \wedge B$.

Some truth tables have more than two inputs. If there are three statements A, B, C from wich the compound statements that interest us are built then there will be $2^3 = 8$ rows to hold all possible combinations of truth values and for n inputs there will be 2^n rows.

Definition 3.9 (Logically impossible)**.** The statements A and B in the truth table of def.3.8 were of a generally nature and all four T/F combinations had to be considered. If we deal with statements which are more specific but have some variability because they contain place holders [16] then there may be dependencies that rule out certain combinations as nonsensical. For example let x be some fixed but unspecified number and look at a truth table which has the statements $A := A(x) :=$ "$x > 5$" and $B := B(x) :=$ "$x > 7$" as input. It is clearly impossible that A is false and B is true, no matter what value x may have.

We call such combinations **logically impossible** or **contradictory**. We abbreviate "logically impossible" with **L/I**.

Both truth tables indicate that the combination A:F and B:T is logically impossible for $A =$ "$x > 5$" and $B =$ "$x > 7$".

A	B	$A \wedge B$
F	F	F
F	T	L/I
T	F	F
T	T	T

A	B	$A \wedge B$
F	F	F
T	F	F
T	T	T

Remark 3.3. It was mentioned in the definition of logically impossible T/F combinations that there had to be some relationship between the inputs, i.e., some placeholders or some fixed but unspecified constants to make this an interesting definitions.

Consider what happens if you have two statements A and B for which this is not the case. For example, let A := "All tomatoes are blue" (obviously false) and B := "Arkansas is a state of the U.S.A." (obviously true).

For those two specific statements we know upfront that we have A:F and B:T, so why bother with the other three cases? In other words, the appropriate truth table is either of those two:

A	B	$A \wedge B$
F	F	L/I
F	T	F
T	F	L/I
T	T	L/I

A	B	$A \wedge B$
F	T	F

Remark 3.4. We chose for a more compact notation to place "L/I" into one of the statement columns but be aware that the L/I attribute really belongs to certain combinations of the T/F values of the inputs. In other words,

the L/I attribute belongs to certain rows of the truth table. A more accurate way would be to place L/I into a separate status column and place "N/A" or "-" or nothing into all columns other than those for the inputs:

Status	A	B	$A \wedge B$
	F	F	F
L/I	F	T	-
	T	F	F
	T	T	T

Of course more than two input statements can be involved when discussing logical impossibility. The following example will show this.

Example 3.10. Let U, V, W, Z be the statement functions

$U := x \mapsto U(x) :=$ "$x \in [0, 4]$",
$V := x \mapsto V(x) :=$ "$x \notin \emptyset$",
$W := x \mapsto W(x) :=$ "$x < -1$",
$Z := x \mapsto Z(x) :=$ "$x > 2$"

[16] e.g., if we have a statement function $P : x \mapsto P(x)$ and we look at the statements $P(x_0)$ for which x_0 belongs to the UoD of P or a certain subset thereof

with UoD $\mathbb{R}$ in each case. Let Q be a statement function that is built from U, V, W, Z with the help of logical operators.

We observe the following:

- **a.** $V(x)$ is always true because the empty set does not contain any elements.
- **a′.** In other words, there is no x in the UoD for which $V(x)$ is false.
- **b.** There is no x in the UoD for which $W(x)$ and $Z(x)$ can both be true.

The following rows in the resulting truth table yield an L/I regardless whether we enter a truth value of T or F into anyone of the "•" entries.

$U(x)$	$V(x)$	$W(y)$	$Z(x)$	$Q(x)$
•	F	•	•	L/I
•	•	T	T	L/I

Remark 3.5. As in example 3.10 above let

$U := U(x) :=$ "$x \in [0, 4]$", $V := V(x) :=$ "$x \notin \emptyset$", $W := W(x) :=$ "$x < -1$", $Z := Z(x) :=$ "$x > 2$".

a. The statement [17] $Q(x) := \neg(U(x) \wedge V(x)) \wedge W(x) \wedge Z(x)$ can never be true, regardless of x.

To see this directly note again that $V(x)$ is trivially true for any x because the emptyset by definition does not contain any elements. It follows that $U(x) \wedge V(x)$ means "$x \in [0, 4]$" and $Q(x)$ means "$x < 0$" $\wedge$ "$x > 4$" $\wedge$ "$x < -1$" $\wedge$ "$x > 2$" which is equivalent to "$x < -1$" $\wedge$ "$x > 4$" and certainly false for any x in the UoD, i.e., $x \in \mathbb{R}$.

Alternatively we can use the results from example 3.10 where we found out that $W(x)$ and $Z(x)$ cannot both be true at the same time.

The remaining rows in the resulting truth table yield an F for $Q(x)$ regardless of the truth values of $U(x)$ and $V(x)$ because $W(x) \wedge Z(x)$ is false, hence $Q(x) =$ whatever $\wedge\big(W(x) \wedge Z(x)\big)$ is false for those remaining rows.

$U(x)$	$V(x)$	$W(y)$	$Z(x)$	$Q(x)$
•	•	F	F	F
•	•	F	T	F
•	•	T	F	F

b. Let $R : x \mapsto R(x) := \neg Q(x)$ be the statement function with UoD $\mathbb{R}$ which represents for each x in the UoD the opposite of Q. Because $Q(x)$ is false for all x, $R(x)$ is true for all x in the universe of discourse for x.

Statements which are true or false under all circumstances like the statements $R(x)$ and $Q(x)$ from the remark above deserve special names.

Definition 3.10 (Tautologies and contradictions)**.** A **tautology** is a statement which is true under all circumstances, i.e., under all combinations of truth values which are not logically impossible.

A **contradiction** is a statement which is false under all circumstances.

We write T_0 for the tautology "$1 = 1$" and F_0 for the contradiction "$1 = 0$". This gives us a convenient way to incorporate statements which are true or false under all circumstances into formulas that build compound statements.

[17] It is tough to come up with some decent examples of compound statements if the only operators at your disposal so far are negation and conjunction.

Example 3.11. Here are some examples of tautologies.

a. The statements $R(x)$ of remark 3.5 are tautologies.

b. T_0 is a boring example of a tautology. So is any true statement without any variables such as "$9 + 12 = 21$" and "a cat is not a cow".

c. There are formulas involving arbitrary statements which are tautologies. We will show that for any two statements A and B the statement $P := \neg\big(A \wedge \neg A\big)$ is a tautology.

Here are some examples of contradictions.

d. The statements $Q(x)$ of remark 3.5 are contradictions.

e. F_0 is a boring example of a contradiction. So is any false statement without any variables such as "$9 + 12 = 50$" and "a dog is a whale".

f. There are formulas involving arbitrary statements which are contradictions. We will show that for any two statements A and B the statement $Q := \big(A \wedge \neg A\big) \wedge B$ is a contradiction.

*Proof of **c** and **f**:*

$P = \neg\big(A \wedge \neg A\big)$ (last column) has entries all T, hence P is a tautology.
$Q = \big(A \wedge \neg A\big) \wedge B$ (next to last column) has entries all F, hence Q is a contradiction.

A	B	$\neg A$	$A \wedge \neg A$	$(A \wedge \neg A) \wedge B$	$\neg(A \wedge \neg A)$
F	*F*	*T*	*F*	*F*	*T*
F	*T*	*T*	*F*	*F*	*T*
T	*F*	*F*	*F*	*F*	*T*
T	*T*	*F*	*F*	*F*	*T*

■

We now continue with the conjunction operator.

Example 3.12. In the following let x, y be two (fixed but arbitrary) integers and let $A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$". Be sure to understand that $A(x)$ and $B(y)$ are in fact statements and not predicates, because the symbols x, y are bound from the start and hence cannot be considered variables of the predicates $A :=$ "$x \in \mathbb{N}$" and $B :=$ "$y \in \mathbb{Z}$ **and** $y > 0$".

We will reuse the statements $A(x)$ and $B(y)$ in examples for the subsequently defined logical operators.

a. If no assumptions are made about a relationship between x and y then all four T/F combinations are possible and, to explore conjunction, we must deal with the full truth table

$A(x)$	$B(y)$	$A(x) \wedge B(y)$
F	F	F
F	T	F
T	F	F
T	T	T

b. On the other hand, if $x < y$ then the truth of $A(x)$ implies that of $B(y)$ because if y is an integer which dominates some natural number x then we have $y > x \geqq 1 > 0$, i.e., y is an integer bigger than zero, i.e., truth of $A(x)$ and falseness of $B(y)$ are incompatible.

It follows that the combination T/F is L/I. We discard the corresponding row and restrict ourselves to the truth table

$A(x)$	$B(y)$	$A(x) \wedge B(y)$
F	F	F
F	T	F
T	T	T

c. Even better, if $x = y$, i.e., we compare truth/falsehood of $A(x)$ with that of $B(x)$, we only need to worry about the two combinations F/F and T/T for the following reason: The set of positive integers is the set $\{1, 2, \dots\}$ and this is, by definition, the set $\mathbb{N}$ of all natural numbers. This means that the statements "$x \in \mathbb{N}$" and "$y \in \mathbb{Z}$ **and** $y > 0$" are just two different ways of expressing the same thing.

It follows that either both $A(x)$ and $B(x)$ are true or both are false. We discard the logically impossible combinations F/T and T/F and restrict ourselves to the truth table

$A(x)$	$B(x)$	$A(x) \wedge B(x)$
F	F	F
T	T	T

3.3.3 Biconditional and logical equivalence operators – Part 1

Definition 3.11 (Double arrow operator (biconditional))**.** The **double arrow operator** [18] is represented by the symbol "$\leftrightarrow$" and read "A double arrow B". $A \leftrightarrow B$ is **true** if and only if either both A and B are **true or** both A and B are **false**.

(3.8) Truth table for $A \leftrightarrow B$:

A	B	$A \leftrightarrow B$
F	F	T
F	T	F
T	F	F
T	T	T

Definition 3.12 (Logical equivalence operator)**.** Two statements A and B are **logically equivalent** if the statement $A \leftrightarrow B$ is a tautology, i.e., if the combinations A:**true**, B:**false** and A:**false**, B:**true** both are logically impossible.

We write $A \Leftrightarrow B$ and we say "A if and only if B" to indicate that A and B are logically equivalent.

(3.9) Truth table for $A \Leftrightarrow B$:

A	B	$A \Leftrightarrow B$
F	F	T
F	T	L/I
T	F	L/I
T	T	T

The discussion of the $\leftrightarrow$ and $\Leftrightarrow$ operators will be continued in ch.3.3.6 (Biconditional and logical equivalence operators – Part 2) on p.41

[18] [4] Bryant, Kirby Course Notes for MAD 2104 calls this operator the **equivalence operator** but we abstain from that terminology because "A is equivalent to B" has a different meaning and is written $A \Leftrightarrow B$.

3.3.4 Inclusive and exclusive or

Definition 3.13 (Disjunction)**.** The **disjunction operator** is represented by the symbols "$\vee$" or "**or**". The expression A **or** B is **true** if and only if either A or B is **true**.

(3.10) Truth table for $A \;\vee\; B$:

A	B	$A \vee B$
F	F	F
F	T	T
T	F	T
T	T	T

The **or** connective generalizes to more than two statements $A_1, A_2, \dots, A_n$ in the obvious manner:

$A_1 \vee A_2 \vee \dots \vee A_n$ is **true** if and only if at least one of $A_1, A_2, \dots, A_n$ is **true** and **false** otherwise, i.e., if each of the A_k is **false**.

Example 3.13. As in example 3.12 let $x, y \in \mathbb{Z}$ and let
$A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$"

a. If no assumptions are made about a relationship between x and y then all four T/F combinations are possible and, to explore conjunction, we must deal with the full truth table

$A(x)$	$B(y)$	$A(x) \vee B(y)$
F	F	F
F	T	T
T	F	T
T	T	T

b. Let $x < y$. We have seen in example 3.12.**b** that the combination T/F is impossible and we can restrict ourselves to the simplified truth table

$A(x)$	$B(y)$	$A(x) \vee B(y)$
F	F	F
F	T	T
T	T	T

c. Now let $x = y$. We have seen in example 3.12.**c** that either both $A(x)$ and $B(y) = B(x)$ are true or both are false. Because the combinations F/T and T/F are impossible we can restrict ourselves to the simplified truth table

$A(x)$	$B(x)$	$A(x) \wedge B(x)$
F	F	F
T	T	T

Definition 3.14 (Exclusive or)**.** The **exclusive or operator** is represented by the symbol "**xor**". [19] A **xor** B is **true** if and only if either A or B is **true** (but not both as is the case for the inclusive or).

(3.11) Truth table for A **xor** B:

A	B	A **xor** B
F	F	F
F	T	T
T	F	T
T	T	F

Example 3.14. As in example 3.12 let $x, y \in \mathbb{Z}$ and let
$A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$"

[19] Some documents such as [4] Bryant, Kirby Course Notes for MAD 2104. also use the symbol $\oplus$.

a. If no assumptions are made about a relationship between x and y then all four T/F combinations are possible and, to explore conjunction, we must deal with the full truth table

$A(x)$	$B(y)$	$A(x)$ **xor** $B(y)$
F	F	F
F	T	T
T	F	T
T	T	F

b. Let $x < y$. We have seen in example 3.12.**b** that the combination T/F is impossible and we can restrict ourselves to the simplified truth table

$A(x)$	$B(y)$	$A(x)$ **xor** $B(y)$
F	F	F
F	T	T
T	T	F

c. Now let $x = y$. We have seen in example 3.12.**c** that either both $A(x)$ and $B(y) = B(x)$ are true or both are false. Because the combinations F/T and T/F are impossible we can restrict ourselves to the simplified truth table

$A(x)$	$B(x)$	$A(x)$ **xor** $B(x)$
F	F	F
T	T	F

This last truth table is remarkable. The truth values for $A(x)$ **xor** $B(x)$ are **false** in each row, hence it is a contradiction as defined in def.3.10 on p.31.

Remark 3.6. Note that the truth values for $A \leftrightarrow B$ are the exact opposites of those for A **xor** B:

$A \leftrightarrow B$ is true exactly when both A and B have the same truth value whereas A **xor** B is true exactly when A and B have opposite truth values. In other words,

$A \leftrightarrow B$ is true whenever $\neg[A \text{ xor } B]$ is true and false whenever $\neg[A \text{ xor } B]$ is false.

Exercise 3.1. use that last remark to prove that for any two statements A and B the compound statement

$$[A \leftrightarrow B] \leftrightarrow \neg[A \textbf{ xor } B]$$

is a tautology.

3.3.5 Arrow and implication operators

Definition 3.15 (Arrow operator)**.** The **arrow operator** [20] is represented by the symbol "$\to$". We read $A \to B$ as "A arrow B" but see remark 3.8 below for the interpretation "if A then B".

(3.12) Truth table for $A \to B$:

A	B	$A \to B$
F	F	T
F	T	T
T	F	F
T	T	T

In other words, $A \to B$ is **false** if and only if A is **true** and B is **false**.

Definition 3.16 (Implication operator)**.** We say that A **implies** B and we write

$$A \Rightarrow B \tag{3.13}$$

[20] [4] Bryant, Kirby Course Notes for MAD 2104 calls this operator the **implication operator** but we abstain from that terminology because "A implies B" has a different meaning and is written $A \Rightarrow B$.

for two statements A and B if the statement $A \to B$ is a tautology, i.e., if the combination A: **true**, B: **false** is logically impossible.

(3.14) Truth table for $A \Rightarrow B$:

A	B	$A \Rightarrow B$
F	F	T
F	T	T
T	F	L/I
T	T	T

Remark 3.7. There are several ways to express $A \Rightarrow B$ in plain english:

Short form:	Interpret this as:
A implies B	The truth of A implies the truth of B
if A then B	if A is true then B is true
A only if B	A is true only if B is true
B if A	B is true if A is true
B whenever A	B is true whenever A is true
A is sufficient for B	The truth of A is sufficient for the truth of B
B is necessary for A	The truth of B is necessary for the truth of A

Theorem 3.1 (Transitivity of "$\Rightarrow$"). *Let A, B, C be three statements such that $A \Rightarrow B$ and $B \Rightarrow C$. Then $A \Rightarrow C$.*

Proof:

*$A \Rightarrow B$ means that the combination A:T, B:F is logically impossible because otherwise $A \to B$ would have a truth value of F and we would not have a tautology. Hence we can drop row 5 from the truth table on the right. Similarly we can drop row 7 because it contains the combination B:T, C:F which contradicts our assumption that $B \Rightarrow C$. But those are the only rows for which $A \to C$ yields **false** because only they contain the combination A:T, C:F. It follows that $A \to C$ is a tautology, i.e., $A \Rightarrow C$.*

	A	B	C
1	F	F	F
2	F	F	T
3	F	T	F
4	F	T	T
5	T	F	F
6	T	F	T
7	T	T	F
8	T	T	T

■

Theorem 3.2 (Transitivity of "$\to$"). *Let A, B, C be three statements.*

Then $\big[(A \to B) \wedge (B \to C)\big] \;\Rightarrow\; (A \to C)$.

Proof: We must show that $\big[(A \to B) \wedge (B \to C)\big] \;\to\; (A \to C)$ *is a tautology. We do this by brute force and compute the truth table.*

A	B	C	$A \to B$	$B \to C$	$P :=$ $(A \to B) \wedge (B \to C)$	$A \to C$	$P \to (A \to C)$
F	F	F	T	T	T	T	T
F	F	T	T	T	T	T	T
F	T	F	T	F	F	T	T
F	T	T	T	T	T	T	T
T	F	F	F	T	F	F	T
T	F	T	F	T	F	T	T
T	T	F	T	F	F	F	T
T	T	T	T	T	T	T	T

We see that the last column with the truth values for $\big[(A \to B) \wedge (B \to C)\big] \;\to\; (A \to C)$ contains ***true*** *everywhere and we have proved that this statement is a tautology.* ■

Definition 3.17. In the context of $A \to B$ and $A \Rightarrow B$ we call A the **premise** or the **hypothesis** [21] and we call B the **conclusion.** [22]

We call $B \to A$ the **converse** of $A \to B$ and we call $\neg B \to \neg A$ the **contrapositive** of $A \to B$.

We call $B \Rightarrow A$ the **converse** of $A \Rightarrow B$ and we call $\neg B \Rightarrow \neg A$ the **contrapositive** of $A \Rightarrow B$.

Remark 3.8.

- **a.** The difference between $A \to B$ and $A \Rightarrow B$ is that $A \Rightarrow B$ implies a relation between the premise A and the conclusion B which renders the T/F combination A:T, B:F logically impossible, i.e., the pared down truth table has only **true** entries in the $A \Rightarrow B$ column. In other words, $A \Rightarrow B$ is the statement $A \to B$ in case the latter is a tautology as defined in def.3.10 on p.31.
- **b.** Both $A \to B$ and $A \Rightarrow B$ are interpreted as "if A then B" but we prefer in general to say "A arrow B" for $A \to B$ because outside the realm of logic $A \Rightarrow B$ is what mathematicians use when they refer to "If ... then " constructs to state and prove theorems.

Example 3.15. The converse of "if x is a dog then x is a mammal" is "if x is a mammal then x is a dog". You see that, regardless whether you look at it in the context of $\to$ or $\Rightarrow$, a "**if** . . . **then**" statement can be true whereas its converse will be false and vice versa.

The contrapositive of "if x is a dog then x is a mammal" is "if x is not a mammal then x is a not a dog". Switching to the contrapositive did not switch the truth value of the "**if** . . . **then**" statement. This is not an accident: see the Contrapositive Law (3.40) on p.45.

Remark 3.9. What is the connection between the truth tables for $A \to B$, $A \Rightarrow B$ and modeling "if A then B"?

We answer this question as follows:

a. If the premise A is guaranteed to be false, you should be allowed to conclude from it anything you like:

[21] also called the **antecedent**
[22] Another word for conclusion is **consequent** .

Consider the following statements which are obviously false:

F_1 : "The average weight of a 30 year old person is 7 ounces",
F_2 : "The number 12.7 is an integer",
F_3 : "There are two odd integers m and n such that $m + n$ is odd",
F_4 : "All continuous functions are differentiable" [23]

and some that are known to be true:

T_1 : "The moon orbits the earth",
T_2 : "The number 12.7 is not an integer",
T_3 : "If m and n are even integers then $m + n$ is even",
T_4 : "All differentiable functions are continuous"

a1. What about the statement "**if** F_3 **then** T_1": "If There are two odd integers m and n such that $m+n$ is odd then the moon orbits the earth"? This may not make a lot of sense to you, but consider this:

The truth of "**if** F_3 **then** T_1" **is not the same as** the truth of just F_1. No absolute claim is made that the moon orbits the earth. You are only asked to concede such is the case under the assumption that two odd integers can be found whose sum is odd. But we know that no such integers exist, i.e., we are dealing with a vacuous premise and there is no obligation on our part to show that the moon indeed orbits the earth! Because of this we should have no problem to accept the validity of "**if** F_3 **then** T_1". Keep in mind though that knowing that **if** F_3 **then** T_1 will not help to establish the truth or falseness of T_1!

a2. Now what about the statement "**if** F_3 **then** F_2": "If There are two odd integers m and n such that $m+n$ is odd then the number 12.7 is an integer"? The truth of this implication should be much easier to understand than allowing to conclude something false from something false:

When was the last time that someone bragged "Yesterday I did xyz" and you responded with something like "If you did xyz then I am the queen of Sheba" in the serene knowledge that there is no way that this person could have possibly done xyz? You know that you have no burden of proof to show that you are the queen of Sheba because you did not make this an absolute claim: You hedged that such is only the case if it is true that the other person in fact did xyz yesterday.

So, yes, the argument "**if** F_3 **then** F_2". sounds OK and we should accept it as true but, as in the case of "**if** F_3 **then** T_1". this has no bearing on the truth or falseness of F_2.

To summarize, "**if** F **then** B". should be true, no matter what you plug in for B. We thus have obtained the first two rows of a sensible truth table for A → B:

A	B	A → B
F	F	T
F	T	T

b. Is it OK to say that if the premise A is true then we may infer that the conclusion B is also true? Definitely! There is nothing wrong with "**if** T_2 **then** T_4", i.e., the statement "If The number 12.7 is not an integer then all differentiable functions are continuous"

[23] A counterexample is the function $f(x) = |x|$ because it is continuous everywhere but not differentiable at $x = 0$.

We can add the fourth row but we do not have #3 yet:

A	B	A → B
F	F	T
F	T	T
T	F	??
T	T	T

c. Is it OK to say that, if the premise A is true, we may say in parallel that A implies B even if the conclusion B is false? No way! Let's assume that Jane is a goldfish. Then A: "Jane is a fish" is true and B: "Jane is a rocket scientist" is false. It is definitely NOT OK to say, under those circumstances, "If Jane is a fish" then Jane is a rocket scientist". Contrast that with this modification that fits case b: "If Jane is a fish' then Jane is **not** a rocket scientist". No one should have a problem with that! We now can complete row #3: T → F is false.

We now have the complete truth table for $A \to B$ and it matches the one in def.3.15:

A	B	$A \to B$
F	F	T
F	T	T
T	F	F
T	T	T

The truth table (3.14) for $A \Rightarrow B$ is then derived from that for $A \to B$ by demanding that A and B be such that $A \to B$ cannot be false, i.e, the combination A:F, B:T must be logically impossible:

A	B	$A \Rightarrow B$
F	F	T
F	T	T
T	F	L/I
T	T	T

We arrived in this remark at the truth tables for $A \to B$ and $A \Rightarrow B$ based on what seems to be reasonable. But the discipline of logic is as exacting a subject as abstract math and the process had to be done in reverse: We first had to **define** $A \to B$ and $A \Rightarrow B$ by means of the truth tables given in def.3.15 and def.3.16 and from there we justified why these operators appropriately model "if A then B".

Example 3.16. As in example 3.12 let $x, y \in \mathbb{Z}$ and let
$A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$"

a. If no assumptions are made about a relationship between x and y then all four T/F combinations are possible and, to explore conjunction, we must deal with the full truth table

$A(x)$	$B(y)$	$A(x) \to B(y)$
F	F	T
F	T	T
T	F	F
T	T	T

b. Let $x < y$. We have seen in example 3.12.**b** that the combination T/F is impossible and we can restrict ourselves to the simplified truth table

$A(x)$	$B(y)$	$A(x) \to B(y)$
F	F	T
F	T	T
T	T	T

c. Now let $x = y$. We have seen in example 3.12.**c** that either both $A(x)$ and $B(y) = B(x)$ are true or both are false. Because the combinations F/T and T/F are impossible we can restrict ourselves to the simplified truth table

$A(x)$	$B(x)$	$A(x) \to B(x)$
F	F	T
T	T	T

We see that $A(x) \to B(y)$ is a tautology in case that $x < y$ or $x = y$.

We have seen that some work was involved to show that the "$A(x) \to B(y)$" statement of the last example is a tautology. How do we interpret this?

> *If you show that a "**if** P **then** Q" statement is a tautology then you have demonstrated that a true premise necessarily results in a true conclusion. You have "**proved**" the validity of the conclusion Q from the validity of the hypothesis P.*

*The next example is a modification of the previous one. We replace the statements $A(x)$ and $B(y)$ with statement functions $x \mapsto A(x), y \mapsto B(y), (x,y) \mapsto C(x,y)$. and replace $A(x) \to B(y)$ with an equivalent $\to$ statement which involves those three statement functions. Our goal is now to show that this new **if**... **then** statement is a tautology for all x and y which belong to their universes of discourse.*

Example 3.17. Let $\mathscr{U}_x := \mathscr{U}_y := \mathbb{Z}$ be the UoDs for the variables x and y.

Let $A : \mathscr{U}_x \to \mathscr{S}$ with $x \mapsto$ "$x \in \mathbb{N}$",
$B : \mathscr{U}_y \to \mathscr{S}$ with $y \mapsto$ "$y \in \mathbb{Z}$ **and** $y > 0$",
$C : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}$ with $(x,y) \mapsto$ "$x < y$".

Let us try to show that for any x in the UoD of x and y in the UoD of y, i.e., for any two integers x and y, the function value $T(x,y)$ of the statement function

(3.15) $\quad T : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}$ with $(x,y) \mapsto T(x,y) := \left[\big(A(x) \wedge C(x,y)\big) \ \to \ B(y)\right]$ is a tautology.

Note that

a. The last arrow in (3.15) is the arrow operator $\to$, not the function assignment operator $\mapsto$.
b. if we can demonstrate that (3.15) is correct then we can replace $\big(A(x) \wedge C(x,y)\big) \ \to \ B(y)$ with $\big(A(x) \wedge C(x,y)\big) \ \Rightarrow \ B(y)$. We interpret this as having proved the (trivial) Theorem: It is true for all integers x and y that if $x \in \mathbb{N}$ and $x < y$ then $y \in \mathbb{Z}$ **and** $y > 0$.

The trick is of course to think of x and y not as placeholders but as fixed but unspecified integers. Then $A(x), B(y)$ and $C(x,y)$ are ordinary statements and we can build truth tables just as always. Observe that we now have three "inputs" $A(x), B(y)$ and $C(x,y)$ and the full truth table contains nine entries.

We need not worry about numbers x and y whose combination (x,y) results in the falseness of the premise $A(x) \wedge C(x,y)$ because **false** $\to$ $B(y)$ always results in **true**. In other words we do not worry about any combination of x and y for which at least one of $A(x)$, $C(x,y)$ is false. To phrase it differently we focus on such x and y for which we have that both $A(x)$, $C(x,y)$ are true and eliminate all other rows from the truth table. There are only two cases to consider: either $B(y)$ is **false** or $B(y)$ is **true**:

$A(x)$	$C(x,y)$	$B(y)$	$A(x) \wedge C(x,y)$	$\big(A(x) \wedge C(x,y)\big) \to B(y)$
T	T	F	T	F
T	T	T	T	T

The proof is done if it can be shown that the first row is a logically impossible. We now look at the components $A(x)$, $C(x,y)$, $B(y)$ in context. We have seen in example 3.12**b.** that the assumed truth of $C(x,y)$ together with that of $A(x)$ is incompatible with $B(y)$ being false. This eliminates the first row from that last truth table and what remains is

$A(x)$	$C(x,y)$	$B(y)$	$A(x) \wedge C(x,y)$	$\big(A(x) \wedge C(x,y)\big) \to B(y)$
T	T	T	T	T

In other words we obtain the value **true** for all non-contradictory combinations in the last column of the truth table and this proves (3.15).

Remark 3.10. Let us compare example 3.16.**b** with example 3.17. Besides using statements in the former and predicates in the latter a more subtle difference is that, because x and y were assumed to be known from the outset,

example 3.16.**b** allowed us to formulate a truth table in which none of the statements had to explicitly refer to the condition $x < y$.

In contrast to this we had to introduce in example 3.17 the predicate $C =$ "$x < y$" to bring this condition into the truth tables

Was there any advantage of switching from statements to predicates and adding a significant amount of complexity in doing so? The answer is yes but it will only become clear when we introduce quantifiers for statement functions.

We will come back to the subject of proofs in chapter 3.7.1 (Building blocks of mathematical theories) on p.56.

3.3.6 Biconditional and logical equivalence operators – Part 2 (Understand this!)

This chapter continues the discussion of the $\leftrightarrow$ and $\Leftrightarrow$ operators from ch.3.3.3 (Biconditional and logical equivalence operators – Part 1) on p.33.

Remark 3.11.

a. Equivalence $A \Leftrightarrow B$ provides a "**replacement principle for statements**": Logically equivalent statements are not "semantically identical" but they cannot be distinguished as far as their "logic content", i.e., the circumstances under which they are true or false are concerned.

b. Note that $A \Leftrightarrow B$ means the same as the following: A is true whenever B is true and A is false whenever B is false because this is the same as saying that, in a truth table that contains entries for A and B, each row either has the value T in both columns or the value F in both columns. This in turn is the same as saying that the column for $A \leftrightarrow B$ has T in each row, i.e., $A \leftrightarrow B$ is a tautology.

b'. There is not much value to **b** if A and B are simple statements but things become a lot more interesting if compound statements like $A := \neg(P \wedge Q)$ and $B := \neg P \vee \neg Q$ are looked at.

We illustrate the above remark with the following theorem.

Theorem 3.3 (De Morgan's laws for statements)**.** *Let A and B be statements. Then we have the following* *logical equivalences:*

$$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B, \tag{3.16}$$

$$\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B. \tag{3.17}$$

Those formulas generalize to n statements $A_1, A_2, \dots, A_n$ as follows:

$$\neg(A_1 \wedge A_2 \wedge \cdots \wedge A_n) \Leftrightarrow \neg A_1 \vee \neg A_2 \vee \cdots \vee \neg A_n, \tag{3.18}$$

$$\neg(A_1 \vee A_2 \vee \cdots \vee A_n) \Leftrightarrow \neg A_1 \wedge \neg A_2 \wedge \cdots \wedge \neg A_n. \tag{3.19}$$

Proof of 3.16: Here is the truth table for both $\neg(A \wedge B)$ and $\neg A \vee \neg B$ depending on the truth values of A and B.

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$\neg A \vee \neg B$	$[\neg(A \wedge B)] \leftrightarrow [\neg A \vee \neg B]$
F	F	F	T	T	T	T	T
F	T	F	T	T	F	T	T
T	F	F	T	F	T	T	T
T	T	T	F	F	F	F	T

This proves the validity of 3.16. Note that the last column of the truth table is superfluous because getting T in each row follows from the fact that the rows of the statement to the left and the one to the right of "$\leftrightarrow$" both contain the same entries T-T-T-F. The column has been included because it illustrates what was said in remark 3.11.

Proof of 3.17: Homework. ■

Example 3.18. As in example 3.12 let $x, y \in \mathbb{Z}$ and let
$A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$"

a. If no assumptions are made about a relationship between x and y then the full truth table needs all four entries and we obtain

$A(x)$	$B(y)$	$A(x) \leftrightarrow B(y)$
F	F	T
F	T	F
T	F	F
T	T	T

b. Let $x < y$. We have seen in example 3.12 that the combination T/F is impossible and we can restrict ourselves to the simplified truth table

$A(x)$	$B(y)$	$A(x) \rightarrow B(y)$
F	F	T
F	T	F
T	T	T

c. Now let $x = y$. We have seen in example 3.12.**c** that then either $A(x)$ and $B(y) = B(x)$ must both be true or they must both be false. Because the combinations F/T and T/F are impossible we can restrict ourselves to the simplified truth table

$A(x)$	$B(x)$	$A(x) \leftrightarrow B(x)$
F	F	T
T	T	T

It follows that for any given number x the statement $A(x) \leftrightarrow B(x)$ is always true, irrespective of the truth values of $A(x)$ and $B(x)$. Hence $A(x) \leftrightarrow B(x)$ is a tautology and we can write $A(x) \Leftrightarrow B(x)$ for all x.

3.3.7 More examples of tautologies and contradictions (Understand this!)

Now that we have all logical operators at our disposal we can give more examples of tautologies and contradictions.

Example 3.19. In the following let P, Q, R be three arbitrary statements, let x, y be two (fixed but arbitrary) integers and let $A(x) :=$ "$x \in \mathbb{N}$" and $B(y) :=$ "$y \in \mathbb{Z}$ **and** $y > 0$". (see example 3.12 on p. 32).

a. Tautologies:

T_0,
$A_1 :=$ "$5 + 7 = 12$",
$A_2 :=$ "Any integer is even **or** odd",
$A_3 := P \lor \neg P$ (Tertium non datur or law of the excluded middle),
$A_4 := P \lor T_0$,
$A_5 := (P \land Q) \lor (P \land \neg Q)$,
$A_6 := (P \to Q) \leftrightarrow (\neg P \lor Q)$ (Implication is logically equivalent to an **or** statement),
$A_7 := [$"$x < y$"$\land A(x)] \to B(y)$ (see 3.16.**b** on p.39),
$A_8 := A(x) \leftrightarrow B(x)$ (see 3.16.**c**).

Note that we can express the fact that A_6, A_7, A_8 are tautologies as follows:

$$(P \to Q) \Leftrightarrow (\neg P \lor Q), \quad [\text{“}x < y\text{”} \land A(x)] \Rightarrow B(y), \quad A(x) \Leftrightarrow B(x).$$

b. Contradictions:

F_0,
$B_1 :=$ "$5 + 7 = 15$",
$B_2 :=$ "There are some non-zero numbers x such that $x = 2x$",
$B_3 := P \land \neg P$,
$B_3 := P \land F_0$,
$B_4 := F_0 \land (P \lor \neg P)$,
$B_5 := [\neg P \lor \neg Q] \land [P \land Q]$,
$B_6 := A(x)$ **xor** $B(x)$ (see 3.14.**c** on p. 34).

Proof that A_3 is a tautology:

P	$\neg P$	$P \lor \neg P$
F	*T*	*T*
T	*F*	*T*

Proof that A_4 is a tautology:

P	T_0	$P \vee T_0$
F	T	T
T	T	T

*Note that even though there are two inputs, P and T_0, there are only two valid combinations of truth values because the only choice for T_0 is **true**.*

Proof that A_6 is a tautology:

P	Q	$P \to Q$	$\neg P$	$\neg P \vee Q$	$(P \to Q) \leftrightarrow (\neg P \vee Q)$
F	F	T	T	T	T
F	T	T	T	T	T
T	F	F	F	F	T
T	T	T	F	T	T

■

Remark 3.12. The interesting tautologies and contradictions are not those involving only specific statements such as $T_0, F_0, A_1, A_2, B_1, B_2$, from above but those statements like A_5, A_6, B_4 and B_5 which specify formulas relating the general statements P, Q and R.

3.4 Statement equivalences (Understand this!)

Symbolic logic has a collection of very useful statement equivalences which are given here. They were taken from ch.2 on logic, subchapter 2.4 (Important Logical Equivalences) of [4] Bryant, Kirby Course Notes for MAD 2104.

Theorem 3.4. *Let P, Q, R be statements.*

***a.** Identity Laws:*

$$P \wedge T_0 \Leftrightarrow P \tag{3.20}$$

$$P \vee F_0 \Leftrightarrow P \tag{3.21}$$

***b.** Domination Laws:*

$$P \vee T_0 \Leftrightarrow T_0 \tag{3.22}$$

$$P \wedge F_0 \Leftrightarrow F_0 \tag{3.23}$$

***c.** Idempotent Laws:*

$$P \vee P \Leftrightarrow P \tag{3.24}$$

$$P \wedge P \Leftrightarrow P \tag{3.25}$$

***d.** Double Negation Law:*

$$\neg(\neg P) \Leftrightarrow P \tag{3.26}$$

e. *Commutative Laws:*

$$P \lor Q \Leftrightarrow Q \lor P \tag{3.27}$$
$$P \land Q \Leftrightarrow Q \land P \tag{3.28}$$

f. *Associative Laws:*

$$(P \lor Q) \lor R \Leftrightarrow P \lor (Q \lor R) \tag{3.29}$$
$$\textit{hence} \ \ (P \lor Q) \lor R \Leftrightarrow P \lor Q \lor R \tag{3.30}$$
$$(P \land Q) \land R \Leftrightarrow P \land (Q \land R) \tag{3.31}$$
$$\textit{hence} \ \ (P \land Q) \land R \Leftrightarrow P \land Q \land R \tag{3.32}$$

g. *Distributive Laws:*

$$P \lor (Q \land R) \Leftrightarrow (P \lor Q) \land (P \lor R) \tag{3.33}$$
$$P \land (Q \lor R) \Leftrightarrow (P \land Q) \lor (P \land R) \tag{3.34}$$

h. *De Morgan's Laws:* [24]

$$\neg(P \land Q) \Leftrightarrow \neg P \lor \neg Q \tag{3.35}$$
$$\neg(P \lor Q) \Leftrightarrow \neg P \land \neg Q \tag{3.36}$$

i. *Absorption Laws:*

$$P \land (P \lor Q) \Leftrightarrow P \tag{3.37}$$
$$P \lor (P \land Q) \Leftrightarrow P \tag{3.38}$$

j. *Implication Law:*

$$(P \to Q) \Leftrightarrow (\neg P \lor Q) \tag{3.39}$$

You should remember this formula because the fact that implication can be expressed as an OR statement is often extremely useful when showing that two statements are logically equivalent.

k. *Contrapositive Laws:*

$$(P \to Q) \Leftrightarrow (\neg Q \to \neg P) \tag{3.40}$$
$$(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P) \tag{3.41}$$

l. *Tautology:*

$$(P \lor \neg P) \Leftrightarrow T_0 \tag{3.42}$$

m. *Contradiction:*

$$(P \land \neg P) \Leftrightarrow F_0 \tag{3.43}$$

[24] This is theorem 3.3 (De Morgan's laws for statements).

n. *Equivalence:* (3.44) $$(P \to Q) \land (Q \to P) \Leftrightarrow (P \leftrightarrow Q)$$

The proof for only some of the laws stated above are given here. You can prove all others by writing out the truth tables to show that left and right sides of the $\ldots \Leftrightarrow \ldots$ *statements are indeed logically equivalent.*

Proof of ***h*** *(De Morgan's laws):*
See theorem 3.3 on p.41.

Proof of ***j*** *(implication law):*
We prove (3.39) using a truth table:

We see that the entries T-T-F-T in the $\neg P \lor Q$ *column match those given for* $P \to Q$ *in def.3.15 on p.35 of the arrow operator. This proves the logical equivalence of those statements.*

P	Q	$\neg P$	$\neg P \lor Q$
F	F	T	T
F	T	T	T
T	F	F	F
T	T	F	T

Proof of ***k*** *(contrapositive law for* $\to$*):*
We prove (3.40) with the help of the previously given laws ***a*** *through* ***j****:*

$$(P \to Q) \overset{(j)}{\Leftrightarrow} (\neg P \lor Q) \overset{(e)}{\Leftrightarrow} (Q \lor \neg P) \overset{(d)}{\Leftrightarrow} (\neg(\neg Q) \lor \neg P) \overset{(j)}{\Leftrightarrow} (\neg Q \to \neg P)$$

■

Example 3.20. Use the logical equivalences of thm.3.4 to prove that $\neg\big(\neg A \land (A \land B)\big)$ is a tautology.

Solution:

	$\neg\big(\neg A \land (A \land B)\big)$	
$\Leftrightarrow$	$\neg(\neg A) \lor \neg(A \land B)$	*De Morgan's Law* (3.35)
$\Leftrightarrow$	$A \lor (\neg A \lor \neg B)$	*De Morgan* (3.35) + *Double negation* (3.26)
$\Leftrightarrow$	$(A \lor \neg A) \lor \neg B$	*Associative law* (3.29)
$\Leftrightarrow$	$T_0 \lor \neg B$	*Tautology* (3.42)
$\Leftrightarrow$	T_0	*Commutative Law* (3.27) + *Domination Law* (3.22) ■

Example 3.21. Find a simple expression for the negation of the statement "**if** you come before 6:00 **then** I'll take you to the movies".

Solution: Let A := *"You come before 6:00" and* B := *"I'll take you to the movies". Our task is to find a simple logical equivalent to* $\neg(A \to B)$*. We proceed as follows:*

$$\neg(A \to B) \overset{(j)}{\Leftrightarrow} \neg(\neg A \lor B) \overset{(h)}{\Leftrightarrow} (\neg(\neg A) \land \neg B) \overset{(d)}{\Leftrightarrow} (A \land \neg B)$$

This translates into the statement "you come before 6:00 ***and*** *I won't take you to the movies".*

■

Remark 3.13. Now that we accept that such logical expressions are DEFINED by their truth tables, we must accept the following: if two logical expressions with two statements A and B as input have the same truth table, then they are logically equivalent and we may interchangeably use one or the other in a proof.

3.5 The connection between formulas for statements and for sets (Understand this!)

Given statements a, b and sets A, B you may have the impression that there are connections between $a \wedge b$ and $A \cap B$, between $a \vee b$ and $A \cup B$, between $\neg a$ and $A^{\complement}$, etc. We will briefly explore this.

In this chapter we switch to small letters for statements and statement functions and use capital letters to denote sets. You have already seen an example in the introduction.

We assume the existence of a universal set $\mathscr{U}$ of which all sets are subsets.

All statements will be of the form $a(x) =$ "$x \in A$" for some set $A \subset \mathscr{U}$. In other words we associate with such a set A the following statement function:

$$a : \mathscr{U} \to \mathscr{S}, \qquad x \mapsto a(x) =: \text{“}x \in A\text{''} \tag{3.45}$$

This relationship establishes a correspondence between the subset A of $\mathscr{U}$ and the predicate $a =$ "$x \in A$" with UoD $\mathscr{U}$. We write $a \cong A$ for this correspondence.

Example 3.22. Let $a \cong A$ and $b \cong B$.

We have

a. $T_0 \cong \mathscr{U}$, $F_0 \cong \emptyset$

b. $\neg a : x \mapsto \neg a(x) = \neg$"$x \in A$" evaluates to a true statement if and only if $x \notin A$, i.e. $x \in A^{\complement}$. Hence $\neg a \cong A^{\complement}$.

c. $a \wedge b : x \mapsto a(x) \wedge b(x) =$ "$x \in A$ **and** $x \in B$" evaluates to a true statement if and only if $x \in A \cap B$. Hence $a \wedge b \cong A \cap B$.

d. $a \vee b : x \mapsto a(x) \vee b(x) =$ "$x \in A$ **or** $x \in B$" evaluates to a true statement if and only if $x \in A \cup B$. Hence $a \vee b \cong A \cup B$.

We expand the table of formulas for statements given in thm 3.4 on p.44 of ch.3.4 (Statement equivalences) with a third column which shows the corresponding relation for sets. Having a translation of statement relations to set relations allows you to use Venn diagrams as a visualization aid.

Theorem 3.5. *For a set $\mathscr{U}$ Let p, q, r be statement functions and let $P, Q, R \subseteq \mathscr{U}$ such that $p \cong P$, $q \cong Q$, $r \cong R$. Then we have the following:*

***a.** Identity:*	(3.46)	$p \wedge T_0 \Leftrightarrow p$	$P \cap \mathscr{U} = P$
	(3.47)	$p \vee F_0 \Leftrightarrow p$	$P \cup \emptyset = P$
***b.** Domination:*	(3.48)	$p \vee T_0 \Leftrightarrow T_0$	$P \cup \mathscr{U} = \mathscr{U}$
	(3.49)	$p \wedge F_0 \Leftrightarrow F_0$	$P \cap \emptyset = \emptyset$
***c.** Idempotency:*	(3.50)	$p \vee p \Leftrightarrow p$	$P \cup P = P$
	(3.51)	$p \wedge p \Leftrightarrow p$	$P \cap P = P$
***d.** Double Negation:*	(3.52)	$\neg(\neg p) \Leftrightarrow p$	$(P^{\complement})^{\complement} = P$
***e.** Commutative:*	(3.53)	$p \vee q \Leftrightarrow q \vee p$	$P \cup Q = Q \cup P$
	(3.54)	$p \wedge q \Leftrightarrow q \wedge p$	$P \cap Q = Q \cap P$
***f.** Associative:*	(3.55)	$(p \vee q) \vee r \Leftrightarrow p \vee (q \vee r)$	$(P \cup Q) \cup R = P \cup (Q \cup R)$
	(3.56)	$(p \wedge q) \wedge r \Leftrightarrow p \wedge (q \wedge r)$	$(P \cap Q) \cap R = P \cap (Q \cap R)$
***g.** Distributive:*	(3.57)	$p \vee (q \wedge r) \Leftrightarrow (p \vee q) \wedge (p \vee r)$	$P \cup (Q \cap R) = (P \cup Q) \cap (P \cup R)$
	(3.58)	$p \wedge (q \vee r) \Leftrightarrow (p \wedge q) \vee (p \wedge r)$	$P \cap (Q \cup R) = (P \cap Q) \cup (P \cap R)$
***h.** De Morgan:*	(3.59)	$\neg(p \wedge q) \Leftrightarrow \neg p \vee \neg q$	$(P \cap Q)^{\complement} = P^{\complement} \cup Q^{\complement}$
	(3.60)	$\neg(p \vee q) \Leftrightarrow \neg p \wedge \neg q$	$(P \cup Q)^{\complement} = P^{\complement} \cap Q^{\complement}$
***i.** Absorption:*	(3.61)	$p \wedge (p \vee q) \Leftrightarrow p$	$P \cap (P \cup Q) = P$
	(3.62)	$p \vee (p \wedge q) \Leftrightarrow p$	$P \cup (P \cap Q) = P$

j1. *Implication 1:*

(3.63) $(p \to q) \Leftrightarrow (\neg p \vee q)$ $\qquad (P \setminus Q)^{\complement} = P^{\complement} \cup Q$

Interpretation: $p(x) \to q(x)$, *i.e.,* "$x \in P$" $\to$ "$x \in Q$" *is* ***true*** *if and only if* $p(x)$*:T,* $q(x)$*:F is L/I., i.e., if and only if* $x \notin P \cap Q^{\complement} = P \setminus Q$*, i.e.,* $x \in (P \setminus Q)^{\complement}$.

(3.64) $p \Rightarrow q$

j2. *Implication 2:*

$P \setminus Q = \emptyset$, *i.e.,* $P \subseteq Q$

Note that we are not dealing with $p \to q$ *but with* $p \Rightarrow q$ *where we assume for all* x *a relation between* p *and* q *which renders* $p(x)$*:T,* $q(x)$*:F logically impossible.*

k. *Contrapositive:*

(3.65) $(P \to Q) \Leftrightarrow (\neg Q \to \neg P)$ $\qquad P^{\complement} \cup Q = Q \cup P^{\complement}$

(3.66) $(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P)$ $\qquad P \subseteq Q \Leftrightarrow Q^{\complement} \subseteq P^{\complement}$

l. *Tautology:*

(3.67) $(P \vee \neg P) \Leftrightarrow T_0$ $\qquad P \cup P^{\complement} = \mathscr{U}$

m. *Contradiction:*

(3.68) $(P \wedge \neg P) \Leftrightarrow F_0$ $\qquad P \cap P^{\complement} = \emptyset$

n1. *Equivalence 1:*

(3.69) $(p \to q) \wedge (q \to p) \Leftrightarrow (p \leftrightarrow q)$

$(P^{\complement} \cup Q) \cap (Q^{\complement} \cup P) = \{x : x$ *both in* P, Q ***or*** x *neither in* P *nor in* $Q\}$

n2. *Equivalence 2:*

(3.70) $(p \Rightarrow q) \wedge (q \Rightarrow p) \Leftrightarrow (p \Leftrightarrow q)$

$(P \subseteq Q)$ ***and*** $(Q \subseteq P) \Leftrightarrow (P = Q)$

Proof: The set equalities are evident except for the following:

Proof of Equivalence 1:

$$
\begin{aligned}
(P^{\complement} \cup Q) \cap (Q^{\complement} \cup P) &= \left[(P^{\complement} \cup Q) \cap Q^{\complement}\right] \cup \left[(P^{\complement} \cup Q) \cap P\right] \\
&= (P^{\complement} \cap Q^{\complement}) \cup (Q \cap Q^{\complement}) \cup (P^{\complement} \cap P) \cup (Q \cap P) \\
&= (P^{\complement} \cap Q^{\complement}) \cup (Q \cap P) \\
&= \{x : x \textit{ neither in } P \textit{ nor in } Q \textbf{ or } x \textit{ both in } P, Q\}.
\end{aligned}
$$

■

3.6 Quantifiers for statement functions

This chapter has been kept rather brief. You can find more about quantifiers in ch.2 on logic, subchapter ch.2.3 (Predicates and Quantifiers) of [4] Bryant, Kirby Course Notes for MAD 2104.

3.6.1 Quantifiers for one-variable statement functions

Definition 3.18 (Quantifiers)**.** Let $A : \mathscr{U} \to \mathscr{S},\ x \mapsto A(x)$ be a statement function of a single variable x with UoD $\mathscr{U}$ for x.

a. The **universal quantification** of the predicate A is the statement

(3.71) $$\text{“For all } x\ A(x)\text{''}, \quad \text{written } \forall x A(x).$$

The above is a short for "$A(x)$ is true for each $x \in \mathscr{U}$". We call the symbol $\forall$ the **universal quantifier** symbol.

b. The **existential quantification** of the predicate A is the statement

(3.72) $$\text{“For some } x\ A(x)\text{''}, \quad \text{written } \exists x A(x).$$

The above is a short for "There exists $x \in \mathscr{U}$ such that $A(x)$ is true". [25] We call the symbol $\exists$ the **existential quantifier** symbol.

c. The **unique existential quantification** of the predicate A is the statement

(3.73) $$\text{“There exists unique } x \text{ such that } A(x)\text{''}, \quad \text{written } \exists! x A(x).$$

The above is a short for "There exists a unique $x \in \mathscr{U}$ such that $A(x)$ is true". [26] We call the symbol $\exists!$ the **unique existential quantifier**

symbol.

Example 3.23. Let $A : [-3, 3] \to \mathscr{S}$ be the statement function $x \mapsto$ "$x^2 - 4 = 0$".

Let $C := \forall x A(x)$ $D := \exists x A(x)$ and $E := \exists! x A(x)$. Then

[25] Equivalently, "$A(x)$ is true for some $x \in \mathscr{U}$" or "$A(x)$ is true for at least one $x \in \mathscr{U}$".
[26] Equivalently, "$A(x)$ is true for exactly one $x \in \mathscr{U}$".

$C =$ "for all $x \in [-3, 3]$ it is true that $x^2 - 4 = 0$"

$D =$ "there is at least one $x \in [-3, 3]$ such that $x^2 - 4 = 0$"

$E =$ "there is exactly one $x \in [-3, 3]$ such that $x^2 - 4 = 0$"

Note that each of C, D, E is in fact a statement because each one is either true or false: Clearly the zeroes of the function $f(x) = x^2 - 4$ in the interval $-3 \leqq x \leqq 3$ are $x = \pm 2$. It follows that D is a true statement and A and C are false statements.

Example 3.24. Let $\mathscr{U} := \{$ all human beings $\}$ be the UoD for the following three predicates:

$S(x) :=$ "x is a student at NYU",
$C(x) :=$ "x cheats when taking tests",
$H(x) :=$ "x is honest",

Let us translate the following three english verbiage statements into formulas:

$A_1 :=$ "All humans are NYU students",
$A_2 :=$ "All NYU students cheat on tests",
$A_3 :=$ "Any NYU student who cheats on tests is not honest".

Solution:

$A_1 = \forall x \; S(x)$,
$A_2 = \forall x \; [S(x) \to C(x)]$,
$A_3 = \forall x \; [(S(x) \wedge C(x)) \to \neg H(x)]$.

Example 3.25. We continue example 3.24.

Let us simplify $A_3 = \forall x \; [(S(x) \wedge C(x)) \to \neg H(x)]$.

It is clear that "$A(x)$ is true for all x" is equivalent to "There is no x such that $A(x)$ is false". In other words, we have for any statement function A the following:

$$\forall x \; A(x) \; \Leftrightarrow \; \neg \big[\; \exists x \; (\neg A(x)) \; \big].$$

But A_3 is the form $\forall x \; A(x)$: replace $A(x)$ with $(S(x) \wedge C(x)) \to \neg H(x)$.

It follows that

$$A_3 \; \Leftrightarrow \; \neg \big[\; \exists x \; (\neg \big(S(x) \wedge C(x)) \to \neg H(x) \big)) \; \big].$$

What a mess! let us drop the "(x)" everywhere and the above becomes

$$A_3 \; \Leftrightarrow \; \neg \big[\; \exists x \; (\neg \big(S \wedge C) \to \neg H \big)) \; \big].$$

We have seen in example 3.21 on p.46 that for any two statements P and Q the equivalence $\neg(P \to Q) \; \Leftrightarrow \; (P \wedge \neg Q)$ is true.

Let us apply this with $P := S \wedge C$ and $Q := \neg H$. We obtain

$$A_3 \; \Leftrightarrow \neg \big[\; \exists x \; \big((S \wedge C) \wedge \neg(\neg H) \big) \; \big]. \; \Leftrightarrow \; \neg \big[\; \exists x \; \big(S \wedge C \wedge H \big) \; \big].$$

where we obtained the last equivalence by applying the double negation law to $\neg(\neg H)$ and the associative law for $\wedge$ to remove the parentheses from $(S \wedge C) \wedge H$.

As a last step we bring back the "(x)" terms and obtain

$$A_3 \Leftrightarrow \neg\exists x\, [S(x) \wedge C(x) \wedge H(x)].$$

In other words, A_3 means "There is no one who is an NYU student and who cheats on tests and is honest". This should make sense if you remember the original meaning of A_3: "Any NYU student who cheats on tests is not honest"

3.6.2 Quantifiers for two-variable statement functions

We now discuss quantifiers for statement functions of two variables. Things become a lot more interesting because we can mix up $\forall$, $\exists$ and $\exists!$.

Unless mentioned otherwise B denotes the statement function of two variables

(3.74)
$$B : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}, \quad x \mapsto B(x,y)$$

It follows that the unverses of discourse are $\mathscr{U}_x$ for x and $\mathscr{U}_y$ for y.

We need a quantifier for each variable to bind the expression $B(x,y)$ with placeholders x and y into a statement, i.e., into something that will be true or false. This done by example as follows:

Definition 3.19 (Doubly quantified expressions)**.** Here is a table of statements involving two quantifiers and their meanings.

a.	$\forall x \forall y B(x,y)$	"for all $x \in \mathscr{U}_x$ and for all $y \in \mathscr{U}_y$ (we have the truth of) $B(x,y)$",
b.	$\forall x \exists y B(x,y)$	"for all $x \in \mathscr{U}_x$ there exists (at least one) $y \in \mathscr{U}_y$ such that $B(x,y)$",
c.	$\exists x \forall y B(x,y)$	"there exists (at least one) $x \in \mathscr{U}_x$ such that for all $y \in \mathscr{U}_y$ $B(x,y)$",
d.	$\exists! x \forall y B(x,y)$	"there exists exactly one $x \in \mathscr{U}_x$ such that for all $y \in \mathscr{U}_y$ $B(x,y)$",
e.	$\exists x \exists y B(x,y)$	"there exists (at least one) $x \in \mathscr{U}_x$ and (at least one) $y \in \mathscr{U}_y$ such that $B(x,y)$",

Example 3.26. Let $\mathscr{U}_x := \mathbb{N}, \mathscr{U}_y := \mathbb{Z}$ and $B : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}, \quad (x,y) \mapsto B(x,y) :=$ "$x + y = 1$". Then

a.	$\forall x \forall y B(x,y)$	**false**
b.	$\forall x \exists y B(x,y)$	**true**: for the given x choose $y := 1 - x$.
c.	$\exists y \forall x B(x,y)$	**false**
d.	$\forall y \exists x B(x,y)$	**false**: If you choose $y > 0$ then the only x that satisfies the equation $x + y = 1$ is $x = 1 - y \leq 0$, i.e., $x \notin \mathbb{N}$, the UoD for x.
e.	$\exists! x \forall y B(x,y)$	**false**
f.	$\exists x \exists y B(x,y)$	**true**: choose $x := 10$ and $y := -9$.

Understand the different outcomes of **b**, **c** and **d** and remember this:

a.	The order in which the qualifiers are applied is important. $\forall x \exists y$ generally does not mean the same as $\exists y \forall x$.
b.	Interchanging variable names in the qualifiers is not OK. $\forall x \exists y$ generally does not mean the same as $\forall y \exists x$.

Proposition 3.1. *Note the following:*

$$\forall x \forall y B(x,y) \Leftrightarrow \forall y \forall x B(x,y) \tag{3.75}$$
$$\exists x \exists y B(x,y) \Leftrightarrow \exists y \exists x B(x,y) \tag{3.76}$$
$$\forall x \exists y B(x,y) \not\Leftrightarrow \exists y \forall x B(x,y) \tag{3.77}$$
$$\exists y \forall x B(x,y) \Rightarrow \forall x \exists y B(x,y) \tag{3.78}$$

Proof: *(3.75) and (3.76) follow from* ***a*** *and* ***e*** *in def. 3.19 and we saw an example for (3.75) in the previous example.*

The last item is not so obvious. We argue as follows: Assume that $\exists y \forall x B(x,y)$ is true. Then there is some $y_0 \in \mathscr{U}_y$ such that $B(x, y_0$ is true for all $x \in \mathscr{U}_x$.

Why does that imply the truth of $\forall x \exists y B(x,y)$, i.e., for all $x \in \mathscr{U}_x$ you can pick some $y \in \mathscr{U}_y$ such that $B(x,y)$ is true? Here is the answer: Pick y_0. This works because, by assumption, $B(x,y_0)$ is true for all $x \in \mathscr{U}_x$. ■

Remark 3.14. The last part of the proof of (3.78) is worth a closer look:

"$\forall x \exists y \dots$" only tells you that for all x there will be some y which generally depends on x, something we sometimes emphasize using "functional notation" $y = y(x)$.

"$\exists y \forall x \dots$" does more: it postulates the existence of some y_0 which is suitable for each x in its UoD. The assignment $y(x) = y_0$ is constant in x!

Remark 3.15 (Partially quantified statement functions)**.** Given a statement function

$$B : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}, \quad x \mapsto B(x,y)$$

with two place holders x and y, we can elect to use only one quantifier for either x or y. If we only quantify x then we only bind x and y still remains a placeholder and if we only quantify y then we only bind y and x still remains a placeholder.

Example 3.27. Let $\mathscr{U}_x :=$ { all students at this party } and $\mathscr{U}_y :=$ { "Linear Algebra", Discrete Mathematics", "Multivariable Calculus", "Ordinary Differential Equations", "Complex Variables", "Graph Theory", "Real Analysis" }.

Let $A :=$ "x studies y" be the two-variable statement function with UoD $\mathscr{U}_x$ for x and UoD $\mathscr{U}_y$ for y, i.e.,

$A : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}, \quad (x,y) \mapsto A(x,y) =$ "x studies y".

Then $B := \forall x\ A(x,y)$ is the one-variable predicate

$B : \mathscr{U}_y \to \mathscr{S}, \quad y \mapsto B(y) =$ "all students at this party study y"

and $C := \exists! y\ A(x,y)$ is the one-variable predicate

$C : \mathscr{U}_x \to \mathscr{S}, \quad x \mapsto C(x) =$ "x studies exactly one of the courses listed in $\mathscr{U}_y$".

3.6.3 Quantifiers for statement functions of more than two variables

Remark 3.16. Although this document limits its scope to statement functions of one or two variables (see the note before remark 3.4 in ch.3.2 (Statements and statement functions)) we discuss briefly the use of quantifiers for predicates

$$A : \mathscr{U}_1 \times \mathscr{U}_2 \times \cdots \times \mathscr{U}_n \to \mathscr{S}, \quad (x_1, x_2, \ldots, x_n) \mapsto A(x_1, x_2, \ldots, x_n).$$

with n place holders.

Each one of those variables needs to be bound by one of the quantifiers $\forall, \exists, \exists!$ in order to obtain a statement, i.e., something that is either true or false.

Example 3.28 (Continuity vs uniform continuity)**.** This example demonstrates the effect of switching a $\forall$ quantifier with an $\exists$ quantifier for a predicate with four variables. One quantification corresponds to ordinary continuity and the other corresponds to uniform continuity of a function. Do not worry if you do not understand how this example relates to continuity. You will learn about this in chapter 10.

Let $a < b$ be two real numbers and let $f :]a, b[\to \mathbb{R}$ be a function which maps each x in its domain $]a, b[$ to a real number $y = f(x)$.

Let $\mathscr{U}_\varepsilon := \mathscr{U}_\delta :=]0, \infty[$ and $\mathscr{U}_x := \mathscr{U}_{x'} :=]a, b[$. Let $P : \mathscr{U}_x \times \mathscr{U}_{x'} \times \mathscr{U}_\delta \times \mathscr{U}_\varepsilon \to \mathscr{S}$ be the predicate

$$(x, x', \delta, \varepsilon) \mapsto P(x, x', \delta, \varepsilon) \ := \ \text{“if } |x - x'| < \delta \text{ then } |f(x) - f(x')| < \varepsilon\text{“}.$$

Let $A := \forall\varepsilon\ \forall x\ \exists\delta\ \forall x' P(x, x', \delta, \varepsilon)$ Then A being true is equivalent to saying that the function f is continuous at each point $x \in]a, b[$. [27]

Let $B := \forall\varepsilon\ \exists\delta\ \forall x\ \forall x' P(x, x', \delta, \varepsilon)$. Then B being true is equivalent to saying that the function f is uniformly continuous in $]a, b[$. [28]

The difference between A and B is that in statement A the variable δ whose existence is required may depend on both ε and x, i.e., $\delta = \delta(\varepsilon, x$

On the other hand, to satisfy B, a δ must be found which still may depend on ε but it must be suitable for <u>all</u> $x \in]a, b[$, i.e., $\delta = \delta(\varepsilon$.

Remark 3.17 (Partially quantified statement functions)**.** What was said in remark 3.15 about partial qualification of two-variable predicates generalizes to more than two variables: If A is a statement function with n variables and we use quantifiers for only $m < n$ of those variables then $n - m$ variables in the resulting expression remain unbound and this expression becomes a statement function of those unbound variables.

For example, if $A(w, x, y, z)$ is a four-variable predicate then $B : (x, z) \mapsto [\forall y\ \neg\exists w\ A(w, x, y, z)]$ defines a two-variable predicate B which inherits the UoDs for x and z from the original statement function A.

---

[27] See def.10.24 (ε-δ continuity) on p.175.
[28] See def.10.28 (Uniform continuity of functions) on p.182.

3.6.4 Quantifiers and negation (Understand this!)

Negation of statements involving quantifiers is governed by

Theorem 3.6 (De Morgan's laws for quantifiers)**.** *Let A be a statement function with UoD $\mathscr{U}$. Then*

a.	$\neg\big(\forall x A(x)\big) \Leftrightarrow \exists x \neg A(x)$	*"It is **not** true that $A(x)$ is true for all x" $\Leftrightarrow$ "There is some x for which $A(x)$ is **not** true"*
b.	$\neg\big(\exists x A(x)\big) \Leftrightarrow \forall x \neg A(x)$	*"There is **no** x for which $A(x)$ is true" $\Leftrightarrow$ "$A(x)$ is **not** true for all x "*

*Proof of **a**: Not given here but you can find it in ch.2 on logic, subchapter 3.11 (De Morgan's Laws for Quantifiers) of [4] Bryant, Kirby Course Notes for MAD 2104.*

*Proof of **b**: Let $\mathscr{U}_x$ be the UoD for x.*

The truth of $\neg\big(\exists x A(x)\big)$ means that $\exists x A(x)$ is false, i.e., $A(x)$ is false for all $x \in \mathscr{U}_x$. This is equivalent to stating that $\neg A(x)$ is true for all $x \in \mathscr{U}_x$ and this is by definition, the truth of $\forall x \neg A(x)$. ■

You can use the formulas above for negation of statements of more than one variable with more than one quantifier using the following method, demonstrated here by example.

Example 3.29. Negate the statement $\exists x \forall y P(x,y)$, i.e., move the $\neg$ operator of $\neg\exists x \forall y P(x,y)$ to the right past all quantifiers.

The key is to introduce an intermittent predicate $A : x \mapsto A(x) := \big[\forall y P(x,y)\big]$. We obtain

$$\big[\neg\exists x \forall y P(x,y)\big] \Leftrightarrow \big[\neg\exists x A(x)\big] \overset{\textbf{(b)}}{\Leftrightarrow} \big[\forall x \neg A(x)\big] \Leftrightarrow \big[\forall x (\neg\forall y P(x,y))\big]$$
$$\overset{\textbf{(a)}}{\Leftrightarrow} \big[\forall x (\exists y \neg P(x,y))\big]$$

Example 3.30. As in example 3.29, negate the statement $\exists x \forall y P(x,y)$ but do so using parentheses instead of explicitly defining an intermittent predicate.

Here is the solution:

$$\big[\neg\exists x \forall y P(x,y)\big] \Leftrightarrow \big[\neg\exists x \big(\forall y P(x,y)\big)\big] \overset{\textbf{(b)}}{\Leftrightarrow} \big[\forall x \neg\big(\forall y P(x,y)\big)\big] \Leftrightarrow \big[\forall x (\neg\forall y P(x,y))\big]$$
$$\overset{\textbf{(a)}}{\Leftrightarrow} \big[\forall x (\exists y \neg P(x,y))\big]$$

3.7 Proofs (Understand this!)

*We have informally discussed proofs in examples 3.16 and 3.17 of chapter 3.3.5 (Arrow and implication operators) on p.35 and seen in two simple cases how a proof can be done by building a single truth table for an **if** ... **then** statement and showing that it is a tautology. In this chapter we take a deeper look at the concept of "proof".*

Many subjects discussed here follow closely ch.3 (Methods of Proofs) of [4] Bryant, Kirby Course Notes for MAD 2104.

3.7.1 Building blocks of mathematical theories

Some of the terminology definitions in notations 3.2 and 3.4 were taken almost literally from ch.3 (Methods of Proofs), subchapter 1 (Logical Arguments and Formal Proofs) of [4] Bryant, Kirby Course Notes for MAD 2104.

Notations 3.2 (Axioms, rules of inferences and assertions)**.**

- **a.** An **axiom** is a statement that is true by definition. No justification such as a proof needs to be given.
- **b.** A **rule of inference** is a logical rule that is used to deduce the truth of a statement from the truth of others.
- **c.** For some statements it is not clear whether they are true for false. Even if a statement is known to be true there might be someone like a student taking a test who is given the task to demonstrate, i.e., prove its truth. In this context we call a statement an **assertion** and we call it a **valid assertion** if it can be shown to be true. An assertion which is not known to be true by anyone is often called a **conjecture**.

Example 3.31. Let $A :=$ "all continuous functions are differentiable" (known to be false [29]) and $B :=$ "all differentiable functions are continuous" (known to be true). A homework problem in calculus may ask the students to figure out which of the four statements $A, \neg A, B, \neg B$ are valid assertions and give proofs to that effect.

Remark 3.18.

a. Goldbach's conjecture states that every even integer greater than 2 can be expressed as the sum of two primes, i.e., integers p greater than 1 which can be divided evenly by no natural number other than p ($p/p = 1$) or 1 ($p/1 = p$). Goldbach came up with this in 1742, more than 250 years ago. No one has been able until now to either prove the validity of this assertion or provide a counterexample to prove its falsehood.

b. Fermat's conjecture was that there are no four numbers $a, b, c, n \in \mathbb{N}$ such that $n > 2$ and $a^n + b^n = c^n$. [30] This was stated by Pierre de Fermat in 1637 who then claimed that he had a proof. Unfortunately he never got around to write it down. A successful proof was finally published in 1994 by Andrew Wiles. Accordingly, Fermat's conjecture was rechristened Fermat's Last Theorem.

[29] see remark 3.9 on p.37 in ch.3.3.5 (Arrow and implication operators).

[30] We have an elementary counterexample for $n = 2$: $3^2 + 4^2 = 25 = 5^2$.

Notations 3.3 (Proofs)**.** A **proof** is the demonstration that an assertion is valid. This demonstration must be detailed enough so that a person with sufficient expert knowledge can understand that we do indeed have a statement which is true for all logically possible combinations of T/F values. To show that the arguments given in this demonstration are valid, available tools are

- **a**. the rules of inference which wil be discussed in section 3.7.2 (Rules of Inference) on p.59
- **b**. logical equivalences for statements (see ch.3.3.6 (Biconditional and logical equivalence operators – Part 2) on p.ch.41).

> In almost all cases the assertion in question is of the form "**if** P **then** C". Proving it means showing that the statement $P \to C$ is a tautology, i.e., it can be replaced by the stronger $P \Rightarrow C$ statement. The proof then consists of the demonstration that the combination P: **true**, C: **false** can be ruled out as logically impossible. In other words, assuming P: **true**, i.e., the truth of the premise, it must be shown that C: **true**, i.e., the conclusion then also is necessarily true.

Usually a proof is broken down into several "sub-proofs" which can be proved separately and where some or all of those steps again will be broken down into several steps ... You can picture this as a hierarchical upside down tree with a single node at the top. At the most detailed level at the bottom we have the leaf nodes. The proof of the entire statement is represented by that top node.

Notations 3.4 (Theorems, lemmata and corollaries)**.**

- **a.** A **theorem** is an assertion that can be proved to be true using definitions, axioms, previously proven theorems, and rules of inference.
- **b.** A **lemma** (plural: lemmata) is a theorem whose main importance is that it can used to prove other theorems.
- **c.** A **corollary** is a theorem whose truth is a fairly easy consequence of another theorem.

Remark 3.19 (Terminology is different outside logic)**.** The terminology given in the above definitions is specific to the subject of mathematical logic. In other branches of mathematics and hence outside this chapter 3 different meanings are attached to those terms:

> Each one of **lemma**, **proposition**, **theorem**, **corollary** is a theorem as defined above in notations 3.2, i.e., a statement that can be proved to be true. We distinguish those terms by comparing them to propositions:

- **a.** Theorems are considered more important than propositions.
- **b.** The main purpose of a lemma is to serve as a tool to prove other propositions or theorems.
- **c.** A corollary is a fairly easy consequence of some lemma, proposition, theorem or other corollary.

> *It was mentioned as a footnote to the definition of a statement (def. 3.3 on p.24) that what we call a statement, [4] Bryant, Kirby calls a proposition and that we deviate from that approach because mathematics outside logic uses "proposition" to denote a theorem of lesser importance.*

Any mathematical theory must start out with a collection of undefined terms and axioms that specify certain properties of those undefined terms.

There is no way to build a theory without undefined terms because the following will happen if you try to define every term: You define T_2 in terms of T_1, then you define T_3 in terms of T_2, etc. Two possibilities:

1. *Each of $T_1, T_2, T_3, \dots$ are different and you end up with an infinite sequence of definitions.*
2. *At least one of those terms is repeated and there will be a circular chain of definitions.*

Neither case is acceptable if you want to specify the foundations of a mathematical system.

Example 3.32. Here are a few important examples of mathematical systems and their ingredients.

a. In Euclid's geometry of the plane some of the undefined terms are "point", "line segment" and "line". The five Euclidean axioms specify certain properties which relate those undefined terms. You may have heard of the fifth axiom, Euclid's parallel postulate. It has been reproduced here with small alterations from Wikipedia's "Euclidean geometry" entry: [31]
(It is postulated that) "if a line segment falling on two line segments makes the interior angles on the same side less than two right angles, the two line segments, if produced indefinitely, meet on that side on which are the angles less than the two right angles".

b. In the so called Zermelo-Fraenkel set theory which serves as the foundation for most of the math that has been done in the last 100 years, the concept of a "set" and the relation "is an element of" ($\in$) are undefined terms.

c. Chapters 1 and 2 of [1] Beck/Geoghegan list several axioms which stipulate the existence of a nonempty set called $\mathbb{Z}$ whose elements are called "integers" which you can "add" and "multiply". Certain algebraic properties such as "$a+b = b+a$" and "$c \cdot (a+b) = (c \cdot a) + (c \cdot a)$" are given as true and so is the existence of an additive neutral unit "0" and a multiplicative neutral unit "1". Besides those algebraic properties the existence of a strict subset $\mathbb{N}$ called "positive integers" is assumed which has, among others, the property that any $z \in \mathbb{Z}$ either satisfies $z \in \mathbb{Z}$ or $-z \in \mathbb{Z}$ or $z = 0$. Finally there is the induction axiom which states that if you create the sequence 1, $1+1$, $(1+1)+1$, $\dots$ then you capture all of $\mathbb{N}$. This axiom is the basis for the principle of proof by mathematical induction (see def.2.11 on p. 17).

Once we have the undefined terms and axioms for a mathematical system, we can begin defining new terms and proving theorems (or lemmas, or corollaries) within the system.

Remark 3.20 (Axioms vs. definitions)**.** You can define anything you want but if you are not careful you may have a logical contradiction and the set of all items that satisfy that definition is empty. In contrast, axioms will postulate the existence of an item or an entire collection of items which satisfy all axioms. If the axioms contradict each other we have a theory which is inconsistent and the only way to deal with it is to discard it and rework its foundations An example for this was set theory in its early stages. Anything that you could phrase as "Let A be the set which contains ..." was fair game to define a set. We saw in remark 2.2 (Russell's Antinomy) on p.11 that this lead to problems so serious that they caused some of the leading mathematicians of the time to revisit the foundations of mathematics.

[31] https://en.wikipedia.org/wiki/Euclidean_geometry#Axioms

Example 3.33. For example you can define an oddandeven integer to be any $z \in \mathbb{Z}$ which satisfies that $z - 212$ is an even number and $z + 48$ is an odd number and you can prove great things for such z. The problem is of course that the set of all oddandeven integers is empty! We have a definition which is useless for all practical purposes, but no mathematical harm is done.

On the other hand, if you add as an additional axiom for $\mathbb{Z}$ in example 3.32.**c** that $\mathbb{Z}$ must contain one or more oddandeven integers then you are in a conundrum because you postulated the existence of a set $\mathbb{Z}$ which satisfies all axioms and the existence of such a set is logically impossible!

xxxxxxx

p.23 When we analyze arguments or logical expression ...

From Kirby, p.48: An assertion involving predicates is valid if it is true for every element in the universe of discourse.

An assertion involving predicates is satisfiable if there is a universe and an interpretation for which the assertion is true. Otherwise it is unsatisfiable.

The scope of a quantifier is the part of an assertion in which the variable is bound by the quantifier.

xxxxxxxxxxxxxxxxxxxxxxxx

p.59/60:

An argument is valid if it is uses only the given hypotheses together with the axioms, definitions, previously proven assertions, and the rules of inference, which are listed above. In those rules in which there is more than one hypothesis, the order of the hypotheses is not important. For example, modus tollens could be just as well stated:

xxxxxxxxxxxxxxxxxxxxx

p.60 Remark 1.4.1. An argument of the form

$h_1 h_2 \dots h_n \quad \therefore c$ *is valid if and only if the proposition* $[h_1 \wedge h2 \wedge \ \wedge hn] \to c$ *is a tautology.*

xxxxxxxxxxxxxxxxxxxxxxxx

"argument" is implicitly defined here (p.58): What is a proof? A proof is a demonstration, or == argument ==, that shows beyond a shadow of a doubt that a given assertion is a logical consequence of our axioms and definition

3.7.2 Rules of Inference

Remark 3.21 (Most important rules of inference)**.** In Notations 3.2 on p.56 we described the term "rule of inference" as "a logical rule that is used to deduce the truth of a statement from the truth of others". The most important rules of inference are those that allow you to draw a conclusion of the form "if A is true then I am allowed to deduce the the truth of C." This basically amounts to having is a list of premises $A_1, A_2, \dots, A_n$ and a conclusion C such that

(3.79) the compound statement $\big[A_1 \wedge A_2 \wedge \dots \wedge A_n\big] \to C$ is a tautology.

In other words, the column for the conclusion C in the truth table for this statement must have the value **true** for each combination of truth values which is not logically impossible.

Observe that the order of the premises does not matter because the **and** connective is commutative.

Theorem 3.7. *Let $P_1, P_2, \dots, P_n$ and C be statements. Then the statement $(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \to C$ is a tautology if and only if the following combination of truth values is logically impossible:*

(3.80) $$P_j \text{ is } \textbf{true} \text{ for each } j = 1, 2, \dots, n \text{ and } C \text{ is } \textbf{false}.$$

Proof:

*Let $P := (P_1 \wedge P_2 \wedge \cdots \wedge P_n)$. Then "$P_j$ is **true** for each $j = 1, 2, \dots, n$" means according to the definition of the $\wedge$ operator the same as the truth of P. Hence proving the theorem is equivalent to proving that the statement $P \to C$ is a tautology if and only if the combination of truth values*

(3.81) $$P \text{ is } \textbf{true} \text{ and } C \text{ is } \textbf{false} \text{ is logically impossible}.$$

*In other words, we must prove that $P \to C$ is a tautology if and only if the row with the combination P:T, C:F, i.e., row 3, is logically impossible and can be ignored. This is is obvious as row 3 is the only one for which $P \to C$ evaluates to **false**.*

	P	C	$P \to C$
1.	F	F	T
2.	F	T	T
3.	T	F	***false***
4.	T	T	T

■

Notations 3.5. Rules of inference are commonly written in the following form:

Your explanations go into this area	A_1 A_2 $\cdots$ A_n ——— $\therefore C$

Read "$\therefore$" as "therefore". The following, more compact notation can also be found:

$$\frac{A_1, A_2, \dots, A_n}{\therefore C}$$

Theorem 3.8 (The three most important inference rules)**.** *The following lists three inference rules, i.e., those arrow statements are indeed tautoloties:*

(3.82)	*Modus Ponens* *(Law of detachment - the mode that affirms the antecedent (the premise))*	$$\begin{array}{l} A \\ A \to C \\ \hline \therefore C \end{array}$$
(3.83)	*Modus Tollens* *(The mode that Denies the consequent (the conclusion))*	$$\begin{array}{l} \neg C \\ A \to C \\ \hline \therefore \neg A \end{array}$$
(3.84)	*Hypothetical syllogism*	$$\begin{array}{l} A \to B \\ B \to C \\ \hline \therefore A \to C \end{array}$$

Here is the compact notation:

Modus Ponens	*Modus Tollens*	*Hypothetical syllogism*
$\dfrac{A,\ A \to C}{\therefore C}$	$\dfrac{\neg C,\ A \to C}{\therefore \neg A}$	$\dfrac{A \to B,\ B \to C}{\therefore A \to C}$

Note that the proof that the hypothetical syllogism is a tautology was given in thm.3.1 on p.36

Proof:

■

Example 3.34. Here are five more inference rules.

(3.85)	Disjunction Introduction	$$\begin{array}{l} A \\ \hline \therefore A \vee B \end{array}$$
(3.86)	Conjunction elimination	$$\begin{array}{l} A \wedge B \\ \hline \therefore A \end{array}$$
(3.87)	Disjunctive syllogism	$$\begin{array}{l} A \vee B \\ \neg A \\ \hline \therefore B \end{array}$$

(3.88) Conjunction introduction

$$\begin{array}{l} A \\ B \\ \hline \therefore A \wedge B \end{array}$$

(3.89) Constructive dilemma

$$\begin{array}{l} (A \to B) \wedge (C \to D) \\ A \vee C \\ \hline \therefore B \vee D \end{array}$$

Compact notation:

Disjunction Introduction	Conjunction elimination	Disjunctive syllogism
$\begin{array}{l} A \\ \hline \therefore A \vee B \end{array}$	$\begin{array}{l} A \wedge B \\ \hline \therefore A \end{array}$	$\begin{array}{l} A \vee B, \neg A \\ \hline \therefore B \end{array}$
Conjunction introduction	Constructive dilemma	
$\begin{array}{l} A, B \\ \hline \therefore A \wedge B \end{array}$	$\begin{array}{l} (A \to B) \wedge (C \to D), A \vee C \\ \hline - \\ \therefore B \vee D \end{array}$	

None of the rules of inference that were given in this chapter involve quantifiers. You can find information about that topic in ch.2, section 1.6 (Rules of Inference for Quantifiers) of [4] Bryant, Kirby Course Notes for MAD 2104.

3.7.3 An example of a direct proof

We illustrate in detail a mathematical proof by applying some the tools you have learned so far in this chapter on logic. For an example we will prove the theorem that each polynomial is differentiable. We define a polynomial as a function $f(x) = \sum_{j=0}^{n} c_j x^j$ *for some* $n = 0, 1, 2, \dots$*, i.e., for some* $n \in \mathbb{Z}_{\geqq 0}$ *and we write* $\mathscr{D}$ *for the set of all differentiable functions. We now can formulate our theorem.*

Theorem 3.9. *Given the statements*

a: $A :=$ "$(n \in \mathbb{Z}_{\geqq 0}) \wedge (c_0 \in \mathbb{R}) \wedge (c_1 \in \mathbb{R}) \wedge \cdots \wedge (c_n \in \mathbb{R}) \wedge \big(f(x) = \sum_{j=0}^{n} c_j x^j\big)$",

b: $B :=$ "$f(x) \in \mathscr{D}$", *the following is valid:* $A \Rightarrow B$. [32]

Proof:

We first collect the necessary ingredients.

We define the following statements which serve as abbreviations so that the formulas we will build are reasonably compact.

a: $Z_j := \text{“}j \in \mathbb{Z}_{\geqq 0}\text{”}$,
b: $C_j := Z_j \wedge \text{“}c_j \in \mathbb{R}\text{”}$,
c: [33] $X_j := Z_j \wedge \text{“}x^j \in \mathscr{D}\text{”}$,
d: $D_j := Z_j \wedge \text{“}c_j x^j \in \mathscr{D}\text{”}$,
e: $E := Z_n \wedge \text{“}f(x) = \sum_{j=0}^{n} c_j x^j\text{”}$,
f: $B := \text{“}f(x) \in \mathscr{D}\text{”}$ *(repeated for convenient reference)*

We now can write our theorem as

$$(Z_n \wedge C_0 \wedge C_1 \wedge \cdots \wedge C_n \wedge E) \to B. \tag{3.90}$$

We assume that the following three theorems were proved previously, hence we may use them without giving a proof.

Theorem Thm-1: If $p(x)$ is a power of x, i.e., $p(x) = x^n$ for some $n = 0, 1, 2, \dots$, then is $p(x)$ differentiable.

We rewrite Thm-1 as an implication which uses the statements above. Let

$$A_1 := Z_n \ \wedge \ \text{“}p(x) = x^n\text{”}, \quad B_1 := X_n.$$

Then Thm-1 states that $A_1 \Rightarrow B_1$. [34]

Theorem Thm-2: The product of a constant (real number) and a differentiable function is differentiable.

We rewrite Thm-2 as an implication. Let
$A_2 := \text{“}c \in \mathbb{R}\text{”} \wedge \text{“}h(x) \in \mathscr{D}\text{”} \wedge \text{“}g(x) = c \cdot h(x)\text{”}$,
$B_2 := \text{“}h(x) \in \mathscr{D}\text{”}$,
Then Thm-2 states that $A_2 \Rightarrow B_2$.

Theorem Thm-3: The sum of differentiable functions is differentiable

[32] Note here and for the other theorems the use of $A_2 \Rightarrow B_2$ instead of $A_2 \to B_2$: We assume that Thm-2 has been proved, i.e., $A_2 \to B_2$ is a tautology.
[33] The expression x^j in **c** and **d** denotes the function $x \mapsto x^j$.
[34] As is the case for the theorem we want to prove, note here and for Thm-2 and Thm-3 below the use of $A_1 \Rightarrow B_1$ instead of $A_1 \to B_1$: Thm-1 has been proved already, i.e., we know that $A_1 \to B_1$ is a tautology.

We rewrite Thm-3 as an implication. Let

$$A_3 := \text{“}Z_n \wedge \text{“}h_1(x) \in \mathscr{D}\text{''} \wedge \text{“}h_2(x) \in \mathscr{D}\text{''} \wedge \cdots \wedge \text{“}h_n(x) \in \mathscr{D}\text{''} \wedge \text{ “}g(x) = \sum_{j=0}^{n} h_j(x)\text{ ''},$$

$$B_3 := \text{“}g(x) \in \mathscr{D}\text{''},$$

Then Thm-3 states that $A_3 \Rightarrow B_3$.

	Assertion	Reason
a:	$Z_0, Z_1, \dots Z_n$	*evident from* $\mathbb{Z}_{\geqq 0} = \{0, 1, 2, \dots\}$
b:	$C_0, C_1, \dots C_n$	*part of the premise of* $A \to B$ *(see (3.90))*
c:	$Z_j \to X_j \;\; (j = 0, 1, \dots n)$	*Thm-1 with* $n := j$
d:	$X_j \;\; (j = 0, 1, \dots n)$	***c*** *and modus ponens*
e:	$(Z_j \wedge C_j \wedge X_j) \to D_j \;\; (j = 0, 1, \dots n)$	*Thm-2 with* $c := c_j$ *and* $h(x) := x^j$
f:	$D_j \;\; (j = 0, 1, \dots n)$	***e*** *and modus ponens*
g:	E	*part of the premise of* $A \to B$
h:	$(Z_n \wedge D_0 \wedge D_1 \wedge \cdots \wedge D_n \wedge E) \to B$	***g*** *and Thm-3 with* $h_j(x) := c_j x^j$ *and* $g(x) := f(x)$
i:	B	***h*** *and modus ponens*

We have demonstrated that the truth of the premise A of our theorem implies that of its conclusion B and this proves the theorem. ■

Remark 3.22. Let us reflect on the steps involved in the proof above.

- **a:** Break down all statements involved – not only those in the theorem you want to prove but also in all theorems, axioms and definitions you reference – into reusable components and name those components with a symbol so that it is easier to understand what assertions you employ and how they lead to the truth of other assertions. Example: D_j references the component $Z_j \;\wedge\; \text{“}c_j x^j \in \mathscr{D}\text{''}$ (which itself references the component $Z_j = \text{“}j \in \mathbb{Z}''_{\geqq 0}$).
- **b:** Rewrite the theorem to be proved as an implication $A \Rightarrow B$.
- **c:** Do the same for the three other theorems that we assumed as already having been proved.
  The following is specific to our example but can be modified to other problems.
- **d:** Start by using the premise A and the definition $\mathbb{Z}_{\geqq 0} := \{0, 1, 2, \dots\}$ to get the first two rows. Show that what you have implies the truth of the premise of Thm-1 and then use the modus ponens inference rule to deduce the truth of its conclusion X_j. This allows X_j to become an additional assertion.
- **e:** Use that new assertion to obtain the truth of the premise of Thm-2 and then use again modus ponens to deduce the truth of its conclusion D_j. Now D_j becomes an additional assertion.
- **f:** Use that new assertion to obtain the truth of the premise of Thm-3 and then use again modus ponens to deduce the truth of D_j. Now D_j becomes an additional assertion.

3.7.4 Invalid proofs due to faulty arguments

Remark 3.23 (Fallacies in logical arguments)**.** People who are not very analytical often commit the following errors in their argumentation:

(3.91)	Affirming the Consequent (proving the wrong direction)	$P \to Q$ Q ——— $\therefore P$
(3.92)	Denying the Antecedent (indirect proof in the wrong direction)	$P \to Q$ $\neg P$ ——— $\therefore \neg Q$
(3.93)	Circular Reasoning	The argument incorporates use of the (not yet proven) conclusion

The reason that the above are fallacies stems from the fact that the above "rules of inferences" are not tautologies.

Example 3.35 (Fallacies in reasoning)**. a.** Affirming the Consequent:
"If you are a great mathematician then you can add $2+2$". It is true that you can add $2+2$. You conclude that you are a great mathematician.

b. Denying the Antecedent:
"If this animal is a cat then it can run quickly". This is not a cat. You conclude that this animal cannot run quickly.

c. Circular Reasoning: [35]
"If xy is divisible by 5 then x is divisible by 5 or y is divisible by 5".

The following incorrect proof uses the yet to be proven fact that the factors can be divided evenly by 5.

Proof:
If xy is divisible by 5 then $xy = 5k$ for some $k \in \mathbb{Z}$. But then $x = 5m$ or $y = 5n$ for some $m, n \in \mathbb{Z}$ (this is the spot where the conclusion was used). Hence x is divisible by 5 or y is divisible by 5. ■

3.8 Categorization of proofs (Understand this!)

*There are different methods by which you can attempt to prove an "**if** ... **then**" statement $P \Rightarrow Q$. They are:*

[35] This is example 1.8.3 in ch.3 (Methods of Proofs) of [4] Bryant, Kirby Course Notes for MAD 2104.

a. *Trivial proof*
b. *Vacuous proof*
c. *Direct proof*
d. *Proof by contrapositive*
e. *Indirect proof (proof by contradiction)*
f. *Proof by cases*

3.8.1 Trivial proofs

The underlying principle of a trivial proof is the following: If we know that the conclusion Q is true then any implication $P \Rightarrow Q$ is valid, regardless of the hypothesis P.

Example 3.36 (Trivial proof)**.** Prove that if it rains at least 60 days per year in Miami then $25 + 35 = 60$.

Proof: There is nothing to prove as it is known that $25 + 35 = 60$. It is irrelevant whether or not in rains (or snows, if you prefer) 60 days per year in Miami. ■

3.8.2 Vacuous proofs

The underlying principle of a vacuous proof is that a wrong premise allows you to conclude anything you want: Both P:F, Q:F and P:F, Q:T yield ***true*** *for $P \to Q$.*

For example, it was mentioned in remark 2.3 (Elements of the empty set and their properties) on p.11 that you can state anything you like about the elements of the empty set as there are none. The underlying principle of proving this kind of assertion is that of a vacuous proof. We prove here assertion ***d*** *of that remark.*

Theorem 3.10. *Let A be any set. Then $\emptyset \subseteq A$.*

Proof:

According to the definition of $\subseteq$ we must prove that if $x \in \emptyset$ then $x \in A$.

So let $x \in \emptyset$. We stop right here: "$x \in \emptyset$" is a false statement regardless of the nature of x because the empty set, by definition, does not contain any elements. It follows that $x \in A$. ■

Remark 3.24. You may ask: But is it not equally true that if $x \in \emptyset$ then $x \notin A$? The answer to that is YES, it is equally true that $x \in A$? and $x \notin A$?, but so what? First you'll find me an x that belongs to the empty set and **only then** am I required to show you that it both does and does not belong to A!

3.8.3 Direct proofs

In a direct proof of $P \Rightarrow Q$ we assume the truth of the hypothesis P and then employ logical equivalences, including the rules of inference, to show the truth of Q.

We proved in chapter 3.7.3 (An example of a direct proof) on p.62 that each polynomial is differentiable (theorem 3.9). That was an example of a direct proof.

3.8.4 Proof by contrapositive

A proof by contrapositive makes use of the logical equivalence $(P \Rightarrow Q) \;\Leftrightarrow\; (\neg Q \Rightarrow \neg P)$ *(see the contrapositive law (3.41) on p45). We give a direct proof of* $\neg Q \Rightarrow \neg P$*, i.e., we assume the falseness of* Q *and prove that then* P *must also be false. Here is an example.*

Theorem 3.11. *Let* A, B *be two subsets of some universal set* Ω *such that* $A \cap B^{\complement} = \emptyset$*. Then* $A \subseteq B$*.*

Proof: We prove the contrapositive instead: If $A \nsubseteq B$ *then* $A \cap B^{\complement} \neq \emptyset$*.*

So let us assume $A \nsubseteq B$*. This means that not every element of* A *also belongs to* B*. In other words, there exists some* $x \in A$ *such that* $x \notin B$*. But then* $x \in A \setminus B = A \cap B^{\complement}$*, i.e.,* $A \cap B^{\complement} \neq \emptyset$*.*

We have proved from the negated conclusion $A \nsubseteq B$ *the negated premise* $A \cap B^{\complement} \neq \emptyset$*.* ■

3.8.5 Proof by contradiction (indirect proof)

A proofs by contradiction are a generalization of proofs by contrapositive. We assume that it is possible for the implication $P \Rightarrow Q$ *that the premise* P *can be true and* Q *can be false at the same time and construct the assumption of the truth of* $P \cap \neg Q$ *a statement* R *such that both* R *and* $\neg R$ *must be true. Here is an example.*

Theorem 3.12. *Let* $A \subseteq \mathbb{Z}$ *with the following properties:*

(3.94) $$m, n \in A \;\Rightarrow\; m + n \in A,$$

(3.95) $$m, n \in A \;\Rightarrow\; mn \in A,$$

(3.96) $$0 \notin A,$$

(3.97) $$\text{if } n \in \mathbb{Z} \text{ then either } n \in A \text{ or } -n \in A \text{ or } n = 0.$$

Then $1 \in A$*.*

Proof by contradiction: Assume that A *is a set of integers with properties (3.94) – (3.97) but that* $1 \notin A$*. We will show that then* $1 \in A$ *must be true. This finishes the proof because it is impossible that both* $1 \notin A$ *and* $1 \in A$ *are true.*

a. *It follows from* $1 \notin A$ *and (3.97) and* $1 \neq 0$ *that* $-1 \in A$*.*
b. *It now follows from (3.95) that* $(-1) \cdot (-1) \in A$*, i.e.,* $1 \in A$*.*

We have reached our contradiction. ■

Remark 3.25. In this simple proof the statement R for which both R and $\neg R$ were shown to be true happens to be the conclusion $1 \in A$. This generally does not need to be the case.

3.8.6 Proof by cases

Sometimes an assumption P *is too messy to take on in its entirety and it is easier to break it down into two or more cases* $P_1, P_2, \dots, P_n$ *each of which only covers part of* P *but such that* $P_1 \vee P_2 \vee \dots \vee P_n$ *covers all* *of it, i.e., we assume*

$$P_1 \vee P_2 \vee \cdots \vee P_n \;\Leftrightarrow\; P. \tag{3.98}$$

Proof by cases then rests on the following theorem:

Theorem 3.13. *Let $P, Q, P_1 \vee P_2 \vee \cdots \vee P_n$ be statements such that (3.98) is true. Then*

$$\big(P \Rightarrow Q\big) \;\Leftrightarrow\; \big[\,(P_1 \Rightarrow Q) \;\vee\; (P_2 \Rightarrow Q) \;\vee\; \ldots\; (P_n \Rightarrow Q)\,\big]. \tag{3.99}$$

Proof (outline): You would do the proof by induction. Prove (3.99) first for $n = 2$ by expressing $A \to B$ as $\neg A \vee B$ and then building a truth table that compares $\big(\neg(P_1 \vee P_2)\big) \vee Q$ with $\neg P_1 \vee Q \vee \neg P_2 \vee Q$.

Then do the induction step in which (3.98) becomes $P_1 \vee P_2 \vee \cdots \vee P_{n+1} \;\Leftrightarrow\; P$ by setting $A := P_1 \vee P_2 \vee \cdots \vee P_n$ and this way reducing the proof of (3.99) for $n+1$ to that of 2 components. You make the validity of $\big(A \Rightarrow Q\big) \;\Leftrightarrow\; \big[\,(P_1 \Rightarrow Q) \;\vee\; (P_2 \Rightarrow Q) \;\vee\; \ldots\; (P_n \Rightarrow Q)\,\big]$ the induction assumption. ■

Theorem 3.14. *Prove that for any $x \in \mathbb{R}$ such that $x \neq 5$ we have*

$$\frac{x}{x-5} > 0 \;\Rightarrow\; \big[\,(x < 0) \text{ or } (x > 5)\,\big]. \tag{3.100}$$

Proof: There are two cases for which $x/(x-5) > 0$:
either both $x > 0$ and $x - 5 > 0$ or both $x < 0$ and $x - 5 < 0$. We write

$P :=$ "$x/(x-5) > 0$",[36] $P_1 := x > 0$ ***and*** $x - 5 > 0$, $P_2 := x < 0$ ***and*** $x - 5 < 0$. *Then $P = P_1 \vee P_2$.*

case 1. *P_1:*
Obviously $x > 0$ and $x - 5 > 0$ if and only if $x > 5$, so we have proved $P_1 \Rightarrow (x > 5)$.

case 2. *P_2:*
Obviously $x < 0$ and $x - 5 < 0$ if and only if $x < 0$, so we have proved $P_2 \Rightarrow (x < 0)$.

We now conclude from $P = P_1 \vee P_2$ and theorem 3.13 the validity of (3.100). ■

[36] $P :=$ "$x/(x-5) > 0$ **and** $x \neq 5$" if you want to be a stickler for precision

4 Functions and relations (Study this!)

4.1 Cartesian products and relations

Definition 4.1 (Cartesian Product of two sets)**.** The **cartesian product** of two sets A and B is

$$A \times B \ := \ \{(a,b) : a \in A, b \in B\},$$

i.e., it consists of all pairs (a,b) with $a \in A$ and $b \in B$.

Two elements (a_1, b_1) and (a_2, b_2) are called **equal** if and only if $a_1 = a_2$ and $b_1 = b_2$. In this case we write $(a_1, b_1) = (a_2, b_2)$.

It follows from this definition of equality that the pairs (a,b) and (b,a) are different unless $a = b$. In other words, the order of a and b is important. We express this by saying that the cartesian product consists of **ordered pairs**.

As a shorthand, we abbreviate $A^2 := A \times A$.

Example 4.1 (Coordinates in the plane)**.** Here is the most important example of a cartesian product of two sets. Let $A = B = \mathbb{R}$. Then $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \{(x,y) : x, y \in \mathbb{R}\}$ is the set of pairs of real numbers. I am sure you are familiar with what those are: They are just points in the plane, expressed by their x– and y–coordinates.
Examples of such points are are: $(1,0) \in \mathbb{R}^2$ (a point on the x–axis), $(0,1) \in \mathbb{R}^2$ (a point on the y–axis), $(1.234, -\sqrt{2}) \in \mathbb{R}^2$.

You should understand why we do not allow two pairs to be equal if we flip the coordinates: Of course $(1,0)$ and $(0,1)$ are different points in the xy–plane!

Remark 4.1 (Functions as subsets of cartesian products)**.** A function [37] $y = f(x)$ which assigns real numbers x to function values $f(x)$, e.g., $f(x) = x^2$, is characterized by its graph

$$\Gamma_f \ := \ \{\big(x, f(x)\big) : x \in \mathbb{R}\}$$

which is a subset of the cartesian product $\mathbb{R} \times \mathbb{R}$.

Remark 4.2 (Empty cartesian product)**.** Note that $A \times B = \emptyset$ if and only if $A = \emptyset$ or $B = \emptyset$ or both are empty.

Definition 4.2 (Relation)**.** Let X and Y be two sets and $R \subseteq X \times Y$ a subset of their cartesian product $X \times Y$. We call R a **relation** on (X, Y). A relation on (X, X) is simply called a relation on X. If $(x,y) \in R$ we say that x **and** y **are related** and we usually write xRy instead of $(x,y) \in R$.

A relation on X is **reflexive** if xRx for all $x \in X$. It is **symmetric** if x_1Rx_2 implies x_1Rx_2 for all $x_1, x_2 \in X$. It is **transitive** if x_1Rx_2 and x_2Rx_3 implies x_1Rx_3 for all $x_1, x_2, x_3 \in X$. It is **antisymmetric** if x_1Rx_2 and x_2Rx_1 implies $x_1 = x_2$ for all $x_1, x_2 \in X$.

Here are some examples of relations.

[37] The precise definition of a function will be given in section 4.2 on p.71.

Example 4.2 (Equality as a relation)**.** Given a set X let $R := \{(x,x) : x \in X\}$, i.e., xRy if and only if $x = y$. This defines a relation on X which is reflexive, symmetric and transitive.

Example 4.3 (Set inclusion as a relation)**.** Given a set X let $R := \{(A,B) : A, B \subseteq X \text{ and } A \subseteq B\}$, i.e., ARB if and only if $A \subseteq B$. This defines a relation which is reflexive, antisymmetric and transitive.

Example 4.4 (Cardinality as a relation)**.** Let X be a finite set, i.e., a set which only contains finitely many elements. For $A \subseteq X$ let card(A) be the number of its elements. [38] Let

$$R := \{(A,B) : A, B \subseteq X \text{ and } \operatorname{card}(A) = \operatorname{card}(B)\},$$

i.e., ARB if and only if A and B possess the same number of elements. This defines a relation on the power set 2^X of X which is reflexive, symmetric and transitive.

Example 4.5 (Empty relation)**.** Given two sets X and Y let $R := \emptyset$. This **empty relation** is the only relation on (X,Y) if X or Y is empty.

Example 4.6. Let $X := \mathbb{R}^2$ be the xy plane. For any point $\vec{x} = (x_1, x_2)$ in the plane let $\|\vec{x}\| := \sqrt{x_1^2 + x_2^2}$ be its length [39] and let $R := \{(\vec{x}, \vec{y}) \in \mathbb{R}^2 \times \mathbb{R}^2 : \|\vec{x}\| = \|\vec{y}\|\}$. In other words, two points in the plane are related when they have the same length: they are located on a circle with radius $r = \|\vec{x}\| = \|\vec{y}\|$. The relation R is reflexive, symmetric and transitive but not antisymmetric.

CORRECTION (Sept 11, 2016): As late as in the 2016-09-06 version it was stated incorrectly that "The relation R is reflexive and transitive but neither symmetric nor antisymmetric."

The relations given in examples 4.2, 4.4, 4.5 and 4.6 are reflexive, symmetric and transitive. Such relations are so important that they deserve a special name:

Definition 4.3 (Equivalence relation and equivalence classes)**.** Let R be a relation on a set X which is reflexive, symmetric and transitive. We call such a relation an **equivalence relation** on X. It is customary to write $x \sim y$ rather than xRy (or $(x,y) \in R$) and we say that x and y are **equivalent**

Given an equivalence relation "$\sim$" on a set X and $x \in X$ let

$$[x]_\sim := \{y \in X : y \sim x\} = \{\text{ all items equivalent to } x\}. \tag{4.1}$$

We call $[x]_\sim$ the **equivalence class** of x. If it is clear from the context what equivalence relation is referred to then we simply write $[x]$ instead of $[x]_\sim$.

Relations which are reflexive, antisymmetric and transitive like the relation of example 4.3 (set inclusion) allow to compare items for "bigger" and "smaller" or "before" and "after". They also deserve a special name:

Definition 4.4 (Partial Order Relation)**.** Let R be a relation on a set X which is reflexive, antisymmetric and transitive. We call such a relation a **partial ordering** of X. or a **partial order relation** on X. [40] It is customary to write "$x \preceq y$" or "$y \succeq x$" rather than "xRy" for a partial ordering R. We say that "x before y" or "y after x".

If "$x \preceq y$" defines a partial ordering on X then $(X, \preceq)$ is called a **partially ordered set** set or a **POset**.

[38] You will see later that card(X) is the cardinality of A (see def.5.7 on p.98).

[39] See def.9.3 on p.135. of the length or Euclidean norm of a vector in n-dimensional space.

[40] Some authors, Dudley among them, do not include reflexivity into the definition of a partial ordering and then distinguish instead between "strict partial orders" and "reflexive partial orders".

Remark 4.3. The properties of a partial ordering can now be phrased as follows:

(4.2) $$x \preceq x \text{ for all } x \in X \qquad \textbf{reflexivity},$$

(4.3) $$x \preceq y \text{ and } y \preceq x \quad \Rightarrow \quad y = x \qquad \textbf{antisymmetry},$$

(4.4) $$x \preceq y \text{ and } y \preceq z \quad \Rightarrow \quad x \preceq z \qquad \textbf{transitivity}.$$

Remark 4.4 (Partial orderings and reflexivity)**.** Note the following:

A. According to the above definition, the following are partial orderings of X:

1. $X = \mathbb{R}$ and $x \preceq y$ if and only if $x \leqq y$.
2. $X = 2^{\Omega}$ for some set Ω and $A \preceq B$ if and only if $A \subseteq B$ (example 4.3).
3. $X = \mathbb{R}$ and $x \succeq y$ if and only if $x \geqq y$.

B. The following relations are **not** partial orderings of X because none of them is reflexive.

4. $X = \mathbb{R}$ and $x \preceq y$ if and only if $x < y$.
5. $X = 2^{\Omega}$ for some set Ω and $A \preceq B$ if and only if $A \subset B$ (i.e., $A \subseteq B$ but $A \neq B$).
6. $X = \mathbb{R}$ and $x \succeq y$ if and only if $x > y$.

Note that each one of those three relations is antisymmetric. For example, let us look at $x < y$. It is indeed true that the premise $\left[x < y \textbf{ and } y < x\right]$ allows us to conclude that $y = x$ as there are no such numbers x and y and a premise that is known never to be true allows us to conclude anything we want!

C. An equivalence relation $\sim$ is a never a partial ordering of X except in the very uninteresting case where you have $x \sim y$ if and only if $x = y$.

D. A partial ordering of X, as any relation on X in general, is inherited by any subset $A \subseteq X$ as follows: Let $\preceq$ be a partial ordering on a set X and let $A \subseteq X$. We define a relation $\preceq_A$ on A as follows: Let $x, y \in A$. Then $x \preceq_A y$ if and only if $x \preceq y$.

Definition 4.5 (Inverse Relation)**.** Let X and Y be two sets and $R \subseteq X \times Y$ a relation on (X, Y). Let

$$R^{-1} := \{ (y, x) : (x, y) \in R \}.$$

Clearly R^{-1} is a subset of $Y \times X$ and hence a relation on (Y, X). We call R^{-1} the **inverse relation** to the relation R.

Example 4.7. Let $R := \{(x, x^3) : x \in \mathbb{R}\}$. Then R is the relation on $\mathbb{R}$ which represents the function $y = f(x) = x^3$. We obtain

$$R^{-1} = \{(x^3, x) : x \in \mathbb{R}\} = \{(y, y^{1/3}) : y \in \mathbb{R}\}$$

In other words, R^{-1} represents the inverse function $x = f^{-1}(y) = y^{1/3}$.

4.2 Functions (mappings) and families

4.2.1 Some preliminary observations about functions

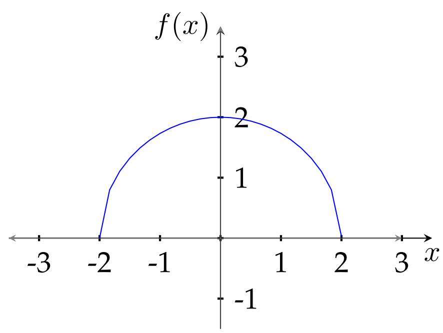
Remark 4.5 (A layman's definition of a function)**.** We look at the set $\mathbb{R}$ of all real numbers and the function $y = f(x) = \sqrt{4 - x^2}$ which associates with certain real numbers x (the "argument" or "independent variable") another real number $y = \sqrt{4-x^2}$ (the "function value" or "dependent variable"):

$$f(0) = \sqrt{(4-0} = 2,\ f(2) = f(-2) = \sqrt{(4-4} = 0,\ f(2/3) = f(-2/3) = \sqrt{(36-4)/9} = \sqrt{30}/3,\ \dots$$

You can think of this function as a rule or law which specifies what item y is obtained as the output or result if the item x is provided as input. [41]

Let us look a little bit closer at the function $y = f(x) = \sqrt{4-x^2}$ and its properties:

- **a**. For some real numbers x there is no function value: For example, if $x = 10$ then $4 - x^2 = -96$ is negative and the square root cannot be taken.
- **b**. For some other x, e.g., $x = 0$ or $x = 2/3$, there is a function value $f(x)$. A moment's reflection shows that the biggest possible set of potential arguments [42] is the interval $[-2, 2]$.
- **c**. For a given x there is never more than one function value $f(x)$. This property allows us to think of a function as an assignment rule: It assigns to certain arguments x a unique function value $f(x)$. We observed in **b** that $f(x)$ exists if and only if $x \in [-2, 2]$.
- **d**. Not every $y \in \mathbb{R}$ is suitable as a function value: A square root cannot be negative, hence no x exists such that $f(x) = -1$ or $f(x) = -\pi$.
- **e**. On the other hand there are numbers y such as $y = 2$ which are "hit" more than once by the function: $f(2) = f(-2) = 0$. [43]
- **f**. Graphs as drawings: We are used to look at the graphs of functions, Here is a picture of the graph of $f(x) = \sqrt{4-x^2}$.



[41] Real numbers were defined in section "Numbers" on p.14.

[42] This set is called by some authors the **natural domain** of the function (e.g., [2] Brewster/Geoghegan).

[43] Matter of fact, only for $y = 2$ there exists a single argument x such that $y = f(x)$ $(x = 0)$. All other y-values in the interval $[0, 2]$ are "mapped" to by two different arguments $x = \pm\sqrt{4-y^2}$.

g. Graphs as sets: Drawings as the one above have limited precision (the software should have drawn a perfect half circle with radius 2 about the origin but there seem to be wedges at $x \approx \pm 1.8$). Also, how would you draw a picture of a function which assigns a 3–dimensional vector [44] (x, y, z) to its distance $w = F(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ from the zero vector $(0, 0, 0)$? You would need four dimensions, one each for x, y, z, w, to draw the graph!
To express the graph of a function without a picture, let us look at a verbal description: The graph of a function $f(x)$ is the collection of the pairs $(x, f(x))$ for all points x which belong to the set $[-2, 2]$ of potential arguments (see **a**). In mathematical parlance: The graph of the function $f(x)$ is the set

$$\Gamma_f \ := \ \{\big(x, f(x)\big) : x \in D_f\}$$

(see remark 4.1 on p. 69).

We now make adjustments to some of those properties which will get us closer to the definition of a function as it is used in abstract mathematics.

Remark 4.6 (A better definition of a function)**.** We make the following alterations to remark 4.5.

- We require an upfront specification of the set A of items that will be allowed as input (arguments) for the function and we require that $y = f(x)$ makes sense for each $x \in A$. Given the function $y = f(x) = \sqrt{4 - x^2}$ from above this means that A must be a subset of $[-2, 2]$.
- We require an upfront specification of the set B of items that will be allowed as output (function values) for the function. This set must be so big that each $x \in A$ has a function value $y \in B$. We do not mind if B contains redundant y values. For $y = f(x) = \sqrt{4 - x^2}$ any superset of the closed interval $[0, 2]$ will do. We may choose $B := [0, 2]$ or $B := [-2, 2\pi]$ or $B := [0, 4]$ or $B := \mathbb{R} \cup \{$ all inhabitants of Chicago $\}$.

Doing so gives us the following: A function consists of three items: a set A of inputs, a set B of outputs and an assignment rule $x \mapsto f(x)$ with the following properties:

1. For **all** inputs $x \in A$ there is a function value $f(x) \in B$.
2. For any input $x \in A$ there is never more than one function value $f(x) \in B$. It follows from property 1 that each $x \in A$ uniquely determines its function value $y = f(x)$. This property allows us to think of a function as an assignment rule: It assigns to each $x \in A$ a unique function value $f(x) \in B$.
3. Not every $y \in B$ needs to be a function value $f(x)$ for some $x \in A$.
4. On the other hand there may be numbers y which are "hit" more than once by the function.

[44]Skip this example on first reading if you do not know about functions of several variables. You will find information about this in chapter 9 ("Vectors and vector spaces") on p.132.

5. The graph Γ_f of a function $f(x)$ is the collection of the pairs $(x, f(x))$ for all points x which belong to the set A, i.e.,

$$\Gamma_f := \{(x, f(x)) : x \in A\}. \tag{4.5}$$

Γ_f has the following properties:

5a. $\Gamma_f \subseteq A \times B$, i.e., Γ_f is a relation on (A, B) (see def.4.2 on p.69).

5b. For each $x \in A$ there exists a unique $y \in B$ such that $(x, y) \in \Gamma_f$

5c. If $x \mapsto g(x)$ is another function with inputs A and outputs B which is different from $x \mapsto f(x)$ (i.e., there is at least one $a \in A$ such that $f(a) \neq g(a)$) then the graphs Γ_f and Γ_g do not coincide

6. Conversely if A and B are two nonempty sets then any relation Γ on (A, B) which satifies **5a** and **5b** uniquely determines a function $x \mapsto f(x)$ with inputs A and outputs B as follows: For $a \in A$ we define $f(a)$ to be the element $b \in B$ for which $(a, b) \in \Gamma$. We know from **5b** that such b exists and is uniquely determined.

Here is a complicated way of looking at the example above: Let $X = [-2, 2]$ and $Y = \mathbb{R}$. Then $y = f(x) = \sqrt{4 - x^2}$ is a rule which "maps" each element $x \in X$ to a uniquely determined number $y \in Y$ which depends on x as follows: it is the square root of 4 minus the square of x.

Mathematicians are very lazy as far as writing is concerned and they figured out long ago that writing "depends on xyz" all the time not only takes too long, but also is aesthetically very unpleasing and makes statements and their proofs hard to understand. So they decided to write "(xyz)" instead of "depends on xyz" and the modern notion of a function or mapping $y = f(x)$ was born.

Here is another example: if you say $f(x) = x^2 - \sqrt{2}$, it's just a short for "I have a rule which maps a number x to a value $f(x)$ which depends on x in the following way: compute $x^2 - \sqrt{2}$." It is crucial to understand from which set X you are allowed to pick the "arguments" x and it is often helpful to state what kinds of objects $f(x)$ the x–arguments are associated with, i.e., what set Y they will belong to.

We now are ready to give the precise definition of a function.

4.2.2 Definition of a function and some basic properties

We have seen in remark 4.6 on p. 73 that a function can be thought of equivalently as an assignment rule $x \mapsto f(x)$ or as a graph. Mathematicians prefer the latter because "assignment rule" is a rather vague term (an undefined term in the sense of ch. 3.7.1 (Building blocks of mathematical theories) on p.56) whereas "graph" is entirely defined in the language of sets. This chapter starts with the official definition of a function. It then deals with the following concepts: composition of functions, injective, surjective, bijective and inverse functions, restriction and extension of functions.

Definition 4.6 (Mappings (functions))**.** Given are two arbitrary nonempty sets X and Y and a relation Γ on (X, Y) (see 4.2 on p.69) which satisfies the following:

(4.6) for each $x \in X$ there exists exactly one $y \in Y$ such that $(x, y) \in \Gamma$.

We call the triplet $f(\cdot) := (X, Y, \Gamma)$ a **function** or **mapping** from X into Y. The set X is called the **domain** or **preimage** and Y is called the **codomain** or **image set** or **target** of the mapping $f(\cdot)$. We will mostly use the words "domain" and "codomain" in this document.

Usually mathematicians simply write f instead of $f(\cdot)$ We mostly follow that convention but include the "$(\cdot)$" part if it helps you to see more easily in a formula that a function rather than an ordinary element of a set is involved.

Let $x \in X$. We write $f(x)$ for the uniquely determined $y \in Y$ such that $(x, y) \in \Gamma$. We write Γ_f or $\Gamma(f)$ if we want to stress that Γ is the relation associated with the function $f = (X, Y, \Gamma)$ and we call Γ the **graph** of the function f. Clearly

$$\Gamma = \Gamma_f = \Gamma(f) = \{(x, f(x)) : x \in X\}. \tag{4.7}$$

It is customary to write

$$f : X \to Y, \qquad x \mapsto f(x) \tag{4.8}$$

instead of $f = (X, Y, \Gamma)$ and we henceforth follow that convention. We abbreviate that to $f : X \to Y$ if it is clear or irrelevant how to compute $f(x)$ from x. We read "$a \mapsto b$" as "a is assigned to b" or "a maps to b" and refer to $\mapsto$ as the **maps to operator** or **assignment operator**.

Domain elements $x \in X$ are called **independent variables** or **arguments** and $f(x) \in Y$ is called the **function value** of x. The subset

$$f(X) := \{y \in Y : y = f(x) \text{ for some } x \in X\} = \{f(x) : x \in X\}$$

of Y is called the **range** or **image** of the function $f(\cdot)$. [45]

We say "f maps X into Y" and "f maps the domain value x to the function value $f(x)$".

Figure 4.1 on p.76 illustrates the graph of a function as a subset of $X \times Y$.

Remark 4.7 (Mappings vs. functions)**.** Mathematicians do not always agree 100% on their definitions. The issue of what is called a function and what is called a mapping is subject to debate. Some mathematicians call a mapping a function only if its codomain is a subset of the real numbers [46] but the majority does what I'll try to adhere to in this document: I use "mapping" and "function" interchangeably and I'll talk about **real functions** rather than just functions if the codomain is part of $\mathbb{R}$ (see (8.1) on p.114).

Remark 4.8. The symbol for the argument x in the definition of a function is a **dummy variable** in the sense that it does not matter what symbol you use.

The following each define the same function with domain $[0, \infty[$ and codomain $\mathbb{R}$ which assigns to any non-negative real number its (positive) square root:

$$\begin{aligned} f : [0, \infty[\to \mathbb{R}, &\qquad x \mapsto \sqrt{x}, \\ f : [0, \infty[\to \mathbb{R}, &\qquad y \mapsto \sqrt{y}, \\ f : [0, \infty[\to \mathbb{R}, &\qquad f(\gamma) = \sqrt{\gamma}. \end{aligned}$$

[45] We distinguish the image set (codomain) Y of $f(\cdot)$ from its image (range) $f(X)$ which is a subset of Y.

[46] or if the codomain is a subset of the complex numbers, but we won't discuss complex numbers in this document.

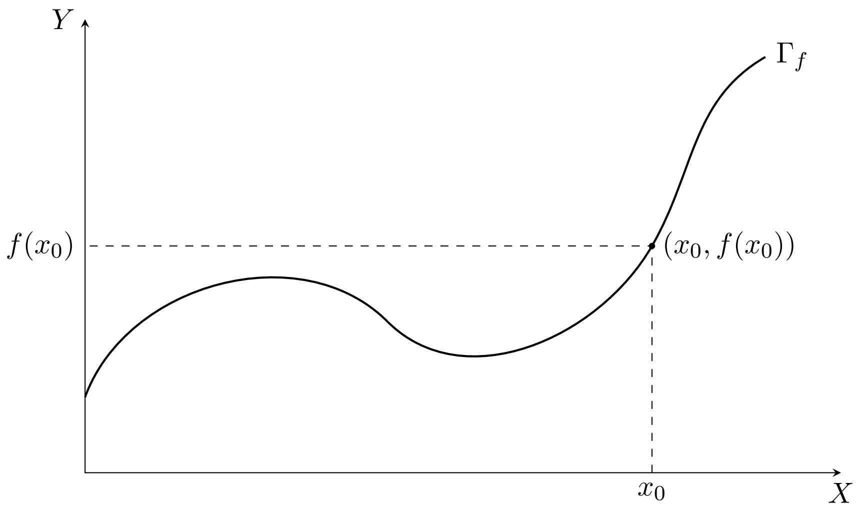


Figure 4.1: Graph of a function.

Matter of fact, not even the symbol you choose for the function matters as long as the operation (here: assign a number to its square root) is unchanged. In other words, the following still describe the same function as above:

$$\begin{aligned} &\varphi : [0,\infty[\to \mathbb{R}, && t \mapsto \sqrt{t},\\ &A : [0,\infty[\to \mathbb{R}, && x \mapsto \sqrt{x},\\ &g : [0,\infty[\to \mathbb{R}, && g(A) \;=\; \sqrt{A}. \end{aligned}$$

In contrast, the following three functions all are different from each other and none of them equals f because domain and/or codomain do not match:

$$\begin{aligned} &\psi :]0,\infty[\to \mathbb{R}, && x \mapsto \sqrt{x} \quad \text{(different domain).}\\ &B : [0,\infty[\to]0,\infty[, && x \mapsto \sqrt{x} \quad \text{(different codomain),}\\ &h : [0,1[\to [0,1[, && x \mapsto \sqrt{x} \quad \text{(different domain and codomain).} \end{aligned}$$

Definition 4.7 (Function composition)**.** Given are three nonempty sets X, Y and Z and two functions $f : X \to Y$ and $g : Y \to Z$. Given $x \in X$ we know the meaning of the expression $g\big(f(x)\big)$:

$y := f(x)$ is the function value of x for the function f, i.e., the unique $y \in Y$ such that $(x, y) \in \Gamma_f$ and

$z := g(y) = g\big(f(x)\big)$ is the function value of $f(x)$ for the function g, i.e., the unique $z \in Z$ such that $\big(f(x), z\big) = \big(f(x), g(f(x))\big) \in \Gamma_g$.

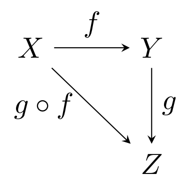
The set $\Gamma := \{\big(x, g(f(x)) : x \in X\big)\}$ is a relation on (X, Z) such that

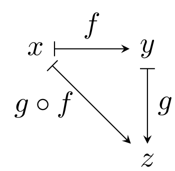
(4.9) for each $x \in X$ there exists exactly one $z \in Z$, namely, $z = g\big(f(x)\big)$, such that $(x, z) \in \Gamma$.

It follows that Γ is the graph of a function $h = (X, Z, \Gamma)$ with function values $h(x) = g\big(f(x)\big)$ for each $x \in X$. We call h the **composition** of f and g and we write $h = g \circ f$ ("g after f").

As far as notation is concerned it is OK to write either of $g \circ f(x)$ or $(g \circ f)(x)$. The additional parentheses may give a clearer presentation if f and/or g are fairly complex.

The following shows how you diagram the composition of two functions. The left picture shows the domains and codomains for each mapping and the left one the element assignments.





We have a special name for the "do nothing function" which assigns each argument to itself:

Definition 4.8 (identity mapping)**.** Given any non–empty set X, we use the symbol id_X for the **identity** mapping defined as

$$id_X : X \to X, \qquad x \mapsto x.$$

We drop the subscript if it is clear what set is referred to.

We now give some examples of mappings.

Example 4.8. Let $\Gamma := \{(x, x^3) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$. Then $f = (\mathbb{R}, \mathbb{R}, \Gamma)$ is the function

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^3.$$

Example 4.9. Let $\Gamma := \{(x, x^2 + 1) : x \in \mathbb{R}\}$. Then $g = (\mathbb{R}, \mathbb{R}, \Gamma)$ is the function

$$g : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^2 + 1.$$

Example 4.10. Let $\Gamma := \{(x, \ln(x)) : x \in]0, \infty[\}$. Here $\ln(x)$ denotes the natural logarithm of x. Then $h = (]0, \infty[, \mathbb{R}, \Gamma)$ is the function

$$h :]0, \infty[\to \mathbb{R}, \qquad x \mapsto \ln(x).$$

Example 4.11. Let $\Gamma := \{(x, \sqrt{x}) : x \in [0, \infty[\}$. Then $\varphi = ([0, \infty[, \mathbb{R}, \Gamma)$ is the function

$$\varphi : [0, \infty[\to \mathbb{R}, \qquad x \mapsto \sqrt{x}.$$

Example 4.12. Let $\Gamma := \{(x, \sqrt{x}) : x \in [0, \infty[\}$. We can consider Γ as a subset of $[0, \infty[\times \mathbb{R}$ but also as a subset of $[0, \infty[\times [0, \infty[$. In the first case we obtain a function $\varphi = ([0, \infty[, \mathbb{R}, \Gamma)$, i.e., the function

$$\varphi : [0, \infty[\to \mathbb{R}, \qquad x \mapsto \sqrt{x}.$$

In the second case we obtain a different(!) function $\psi = ([0, \infty[, [0, \infty[, \Gamma)$, i.e., the function

$$\psi : [0, \infty[\to [0, \infty[, \qquad x \mapsto \sqrt{x}.$$

If you have taken multivariable calculus or linear algebra then you know that functions need not necessarily map numbers to numbers but they can also map vectors to numbers, numbers to vectors (curves) or vectors to vectors.

Example 4.13. We define a function which maps two-dimensional vectors to numbers. Let $A := \{\big((x,y) \in \mathbb{R}^2 : x^2+y^2 \leqq 1\ \}$. Let $\Gamma := \{\big((x,y), \sqrt{1-x^2-y^2}\big) : (x,y) \in A\}$. Then $F = (A, \mathbb{R}, \Gamma)$ is the function

$$F : A \to \mathbb{R}, \qquad (x,y) \mapsto \sqrt{1-x^2-y^2}.$$

Note that the domain is not a set of real numbers but of points in the plane and that the graph of F is a set of points (x,y,z) in 3–dimensional space.

Example 4.14. We define a function which maps numbers to two-dimensional vectors (a curve in the plane). Let $\Gamma := \{\big(t, (\sin t, \cos t)\big) : t \in \mathbb{R}\ \}$. Then $G = (\mathbb{R}, \mathbb{R}^2, \Gamma)$ is the function

$$G : \mathbb{R} \to \mathbb{R}^2, \qquad t \mapsto (\sin t, \cos t).$$

Note that the codomain is not a set of real numbers but the Euclidean plane.

Example 4.15. Let $\Gamma := \{\big((x,y), (2x - y/3,\ x/6 + 4y)\big) : x,y \in \mathbb{R}\ \}$. Then $H = (\mathbb{R}^2, \mathbb{R}^2, \Gamma)$ is the function

$$H : \mathbb{R}^2 \to \mathbb{R}^2, \qquad (x,y) \mapsto (2x - y/3,\ x/6 + 4y).$$

Note that both domain and codomain are the Euclidean plane.

Skip the remainder of this example if you do not know about matrix multiplication.

Let A be the 2×2 matrix

$$A := \begin{pmatrix} 2 & -1/3 \\ 1/6 & 4 \end{pmatrix}.$$

We then obtain for any pair of numbers $\vec{x} = (x,y)^T$ [47] that

$$A\vec{x} = \begin{pmatrix} 2 & -1/3 \\ 1/6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - y/3 \\ x/6 + 4y \end{pmatrix}$$

As is customary in linear algebra we now think of $\mathbb{R}^2$ as the collection of column vectors $\{\begin{pmatrix} x \\ y \end{pmatrix} : x,y \in \mathbb{R}\}$ rather than the cartesian product $\mathbb{R}^2 \times \mathbb{R}^2$ which is the collection of row vectors $\{(x,y) : x,y \in \mathbb{R}\}$.

We now reformulate the last example in the framework of linear algebra. Skip the following example if you do not know about matrix multiplication.

[47] Here $(x,y)^T = \begin{pmatrix} x \\ y \end{pmatrix}$ is the transpose of $(x.y)$, i.e., the operation that switches rows and columns of any matrix. In particular it transforms a row vector into a column vector and vice versa.

Example 4.16. As is customary in linear algebra we now think of $\mathbb{R}^2$ as the collection of column vectors $\{\begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R}\}$ rather than the cartesian product $\mathbb{R}^2 \times \mathbb{R}^2$ which is the collection of row vectors $\{(x, y) : x, y \in \mathbb{R}\}$.

Let A be the 2×2 matrix

$$A := \begin{pmatrix} 2 & -1/3 \\ 1/6 & 4 \end{pmatrix}.$$

We then obtain for any pair of numbers $\vec{x} = (x, y)^T$ [48] that

$$A\vec{x} = \begin{pmatrix} 2 & -1/3 \\ 1/6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - y/3 \\ x/6 + 4y \end{pmatrix}$$

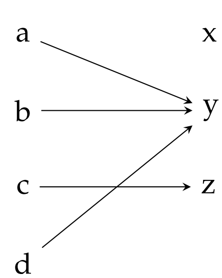
Let $\Gamma := \{ (\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 2x - y/3 \\ x/6 + 4y \end{pmatrix}) : x, y \in \mathbb{R} \}$. Then $H = (\mathbb{R}^2, \mathbb{R}^2, \Gamma)$ is the function

$$H : \mathbb{R}^2 \to \mathbb{R}^2, \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \end{pmatrix}.$$

Note that both domain and codomain are the Euclidean plane.

If you want to construct a counterexample to a mathematical statement concerning functions it often is best to construct functions with small domain and codomain so that you can draw a picture that completely describes the assignments. The next example will illustrate this.

Example 4.17. Let $X := \{a, b, c, d\}$, $Y := \{x, y, z\}$ and let $\Gamma := \{ (a, y), (b, y), (c, z), (d, y)\}$. Then $I = (X, Y, \Gamma)$ is the function which maps the elements of X to Y according to the following diagram:



Note that nothing was said about the nature of the elements of X and Y. Knowledge of it is not needed to make observations like the following: Look at items **3** and **4** of remark 4.6 (A better definition of a function) on p.73. Convince yourself that $x \in Y$ is an example for **3**: Not every element of Y needs to be a function value and that $y \in Y$ is an example for **4**: There may be elements of Y which are "hit" more than once by the function.

[48] Here $(x, y)^T = \begin{pmatrix} x \\ y \end{pmatrix}$ is the transpose of $(x.y)$, i.e., the operation that switches rows and columns of any matrix. In particular it transforms a row vector into a column vector and vice versa.

Example 4.18. This example represents a mathematical model for computing probabilities of the outcomes of rolling a fair die and demonstrates that probability can be thought of as a function that maps sets to numbers.

If we roll a die then the outcome will be an integer between 1 and 6, i.e., the state space for this random action will be $X := \{1, 2, 3, 4, 5, 6\}$. For $A \subseteq X$ let Prob(A) denote the probability that rolling the die results in an outcome $x \in A$.

For example Prob(an even number occurs) $=$ Prob$\big(\{2, 4, 6\}\big) = 50\% = 1/2$. Clearly we have for singletons consisting of a single outcome that

$$\text{Prob}\big(\{1\}\big) \;=\; \text{Prob}\big(\{2\}\big) \;=\; \cdots \;=\; \text{Prob}\big(\{6\}\big) \;=\; 1/6 \;=\; 16.\bar{6}\%.$$

Your everyday experience tells you that if $A = \{x_1, x_2, \dots, x_k\}$ where $x_j \in X$ for each index j (and hence $k \leqq 6$ because a set does not contain duplicates) then

$$\text{Prob}(A) \;=\; \text{Prob}\big(\{x_1\}\big) + \text{Prob}\big(\{x_2\}\big) + \cdots + \text{Prob}\big(\{x_k\}\big) \;=\; \sum_{j=1}^{k} \text{Prob}\big(\{x_j\}\big).$$

What if A is the event that the roll of the die does not result in any outcome, i.e., $A = \emptyset$? We do not worry about the die getting stuck in mid-air or the dog snatching it before we get a chance to see the outcome and consider this event impossible, i.e., Prob$(\emptyset) = 0$.

We now have a probability associated with every $A \subseteq X$, i.e., with every $A \in 2^X$ and can finally write this probability as a function. Let $\Gamma := \{(A, \text{Prob}(A)) : A \subseteq X\}$. Then $P = (2^X, [0, 1], \Gamma)$ is the function

$$P : 2^X \to [0, 1], \qquad A \mapsto \text{Prob}(A).$$

Why do we use $[0, 1]$ and not $\mathbb{R}$ as the codomain? The answer is that we could have done so but no event has a probablity that exceeds 100% or is negative, so $[0, 1]$ is big enough and by choosing this set as the codomain we do not deviate from standard presentation of mathematical probability theory.

To understand the next example you need to be familiar with the concepts of continuity, differentiability and antiderivatives (integrals) of functions of a single variable. Just skip the parts where you lack the background.

Example 4.19. Let $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{\infty\}$ and let $X :=]a, b[$ be the open (end points a, b are excluded) interval of all real numbers between a and b. Let $x_0 \in]a, b[$ be "fixed but arbitrary".

The following is known from calculus (see [11] Stewart, J: Single Variable Calculus): Let $f :]a, b[\to \mathbb{R}$ be a function which is continuous on $]a, b[$. Then

a. f is integrable, i.e., for any $\alpha, \beta \in \mathbb{R}$ such that $a < \alpha < \beta < b$ the **definite integral** $\displaystyle\int_{\alpha}^{\beta} f(u)du$ exists. For a definition of integrability see section 4.1 of [11] Stewart, J: Single Variable Calculus.

b. f has an **antiderivative** : There exists a function $F :]a, b[\to \mathbb{R}$ whose derivative $F'(\cdot)$ exists on all of $]a, b[$ and coincides with f, i.e., $F'(x) = f(x)$ for all $x \in]a, b[$.

c. This antiderivative satisfies $\int_{\alpha}^{\beta} f(u)du = F(\beta) - F(\alpha)$ for all $a < \alpha < \beta < b$ and it is **not** uniquely defined: If $C \in \mathbb{R}$ then $F(\cdot) + C$ is also an antiderivative of f.
On the other hand, if both F_1 and F_2 are antiderivatives for f then their difference $G(\cdot) := F_2(\cdot) - F_1(\cdot)$ has the derivative $G'(\cdot) = f(\cdot) - f(\cdot)$ which is constant zero on $]a, b[$. It follows that any two antiderivatives only differ by a constant.
To summarize the above: If we have one antiderivative F of f then any other antiderivative $\tilde{F}$ is of the form $\tilde{F}(\cdot) = F(\cdot) + C$ for some real number C.
This fact is commonly expressed by writing $\int f(u)du = F(x) + C$ for the **indefinite integral** (an integral without integration bounds).

d. It follows from **c** that if some $c_0 \in \mathbb{R}$ is given then there is only one antiderivative F such that $F(x_0) = c_0$.
Here is a quick proof: Let G be another antiderivative of f such that $G(x_0) = c_0$. Because $G - F$ is constant we have for all $x \in]a, b[$ that

$$G(x) - F(x) \; = \; \text{const} \; = \; G(x_0) - F(x_0) \; = \; 0,$$

i.e., $G = F$.

After those reminders about integration we are ready to define a function $I(\cdot)$ for which both domain and codomain are sets of functions.

Given are a, b and x_0 as above and $c_0 \in \mathbb{R}$. Let

$$\begin{aligned} \mathscr{F} &:= \{ f :]a, b[\to \mathbb{R} \text{ such that } f \text{ is continuous on }]a, b[\; \}, \\ \mathscr{G} &:= \{ g :]a, b[\to \mathbb{R} \text{ such that } g \text{ is differentiable on }]a, b[\text{ and } g(x_0) = c_0 \; \}. \end{aligned}$$

It follows from the preparatory remarks that for each $f \in \mathscr{F}$ there exists a unique $F \in \mathscr{G}$ which is an antiderivative for f. We now define a function $I : \mathscr{F} \to \mathscr{G}$ by specifying its graph as the set

$$\Gamma \; := \; \{ \big(f(\cdot), g(\cdot) \big) : f \in \mathscr{F}, g \in \mathscr{G} \text{ and } g \text{ is an antiderivative of } f \; \}.$$

In other words, $I = (\mathscr{F}, \mathscr{G}, \Gamma)$ is the function

$$I : \mathscr{F} \to \mathscr{G}, \qquad f(\cdot) \mapsto I(f)(\cdot) = \int f(u)du + C$$

where C is determined by the requirement that $I(f)(x_0) = c_0$.

Definition 4.9 (Surjective, injective, bijective)**.** Let $f : X \to Y$. As usual the graph of f is denoted Γ_f.

a. Surjectivity: In general it is not true that $f(X) = \{ f(x) : x \in X \}$ equals the entire codomain Y, i.e., that

(4.10) for each $y \in Y$ there exists at least one $x \in X$ such that $(x, y) \in \Gamma_f$.

But if $f(X) = Y$, i.e., if (4.10) holds, we call f **surjective** and we say that f maps X **onto** Y.

b. Injectivity: In general it is not true that if $y \in f(X)$ then $y = f(x)$ for a unique x, i.e., that if there is another $x_1 \in X$ such that also $y = f(x_1)$ then it follows that $x_1 = x$. But if this is the case, i.e., if

$$\text{for each } y \in Y \text{ there exists at most one } x \in X \text{such that } (x,y) \in \Gamma_f. \tag{4.11}$$

then we call f **injective** .

We can express (4.11) also as follows: If $x, x_1 \in X$ and $y \in Y$ are such that $(x,y) \in \Gamma_f$ and $(x_1, y) \in \Gamma_f$ then it follows that $x_1 = x$.

c. Bijectivity: Let $f : X \to Y$ be both injective and surjective. Such a function is called **bijective**.

It follows from (4.10) and (4.11) that f is bijective if and only if

$$\text{for each } y \in Y \text{ there exists exactly one } x \in X \text{such that } (x,y) \in \Gamma_f. \tag{4.12}$$

We rewrite (4.12) by employing Γ_f's inverse relation $\Gamma_f^{-1} = \{(y,x) : (x.y) \in \Gamma\}$ (see def. 4.5 on p.71) and obtain

$$\text{for each } y \in Y \text{ there exists exactly one } x \in X \text{such that } (y,x) \in \Gamma_f^{-1}. \tag{4.13}$$

But this implies, according to (4.6) in the definition of a function, that Γ_f^{-1} is the graph of a function $g := (Y, X, \Gamma_f^{-1})$ with domain Y and codomain X where, for a given $y \in Y$, $g(y)$ stands for the uniquely determined $x \in X$ such that $(y,x) \in \Gamma_f^{-1}$. Note that

$$\Gamma_f^{-1} = \Gamma_g. \tag{4.14}$$

We call $g(\cdot)$ the **inverse mapping** or **inverse function** of $f(\cdot)$ and write $f^{-1}(\cdot)$ instead of $g(\cdot)$.

Remark 4.9.

a. It follows from (4.14) that

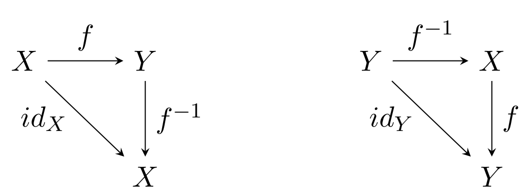
$$\Gamma_f^{-1} = \Gamma_{f^{-1}}. \tag{4.15}$$

b. Each $x \in X$ is mapped to $y = f(x)$ which is the only element of Y such that $f^{-1}(y) = x$,

c. Each $y \in Y$ is mapped to $x = f^{-1}(y)$ which is the only element of X such that $f(x) = y$.

d. It follows that $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.

In other words, $f^{-1} \circ f = id_X$ and $f \circ f^{-1} = id_Y$. Here is the picture:



Theorem 4.1 (Characterization of inverse functions)**.** *Let X and Y be nonempty sets and $f : X \to Y$. Then the following are equivalent:*

***a**. f is bijective.*
***b**. There exists $g : Y \to X$ such that both $g \circ f = id_X$ and $f \circ g = id_Y$.*

*Proof of **a** $\Rightarrow$ **b**: We have seen in part **d** of remark 4.9 that $g := f^{-1}$ satisfies **b**.*

*Proof of **b** $\Rightarrow$ **a**: We must show that f is both surjective and injective. First we show that f is surjective. Let $y \in Y$. we must find some $x \in X$ such that $f(x) = y$. Let $x := g(y)$. Then*

$$f(x) \;=\; f\big(g(y)\big) \;=\; f \circ g(y) \;=\; id_Y(y) \;=\; y.$$

We have $f(x) = y$ and this proves surjectivity. Now we show that f is injective. Let $x_1, x_2 \in X$ and $y \in Y$ such that $f(x_1) = f(x_2) = y$. We are done if we can prove that $x_1 = x_2$. We have

$$x_1 = id_X(x_1) = g \circ f(x_1) = g\big(f(x_1)\big) = g(y) = g\big(f(x_2)\big) = g \circ f(x_2) = id_X(x_2) = x_2,$$

i.e., $x_1 = x_2$ and this proves injectivity of f. ■

Remark 4.10. [Horizontal and vertical line tests] Let X and Y be nonempty sets and $f : X \to Y$. The following needs to be taken with a grain of salt because X and Y may not be sets of real numbers.

Let $R \subseteq X \times Y$. (4.6) on p.74 states that R is the graph of a function with domain X and codomain Y if and only if any "vertical line", i.e., any set $H_0 \subseteq X \times Y$ of the form $H(x_0) := \{(x_0, y) : y \in Y\}$ intersects R in exactly one "point".

a. (4.6) on p.74 states that R is the graph of a function with domain X and codomain Y if and only if it passes the "vertical line test": Any "vertical line", i.e., any set $V_0 \subseteq X \times Y$ of the form $V(x_0) := \{(x_0, y) : y \in Y\}$ for a fixed $x_0 \in X$ intersects R in **exactly one** "point".
b. (4.10) on p.81 states that R is the graph of a surjective function with domain X and codomain Y if and only if it passes in addition to the "vertical line test" the following "horizontal line test": any "horizontal line", i.e., any set $H_0 \subseteq X \times Y$ of the form $H(x_0) := \{(x, y_0) : x \in X\}$ for a fixed $y_0 \in Y$ intersects R in **at least one** "point".
c. (4.11) on p.82 states that R is the graph of an injective function with domain X and codomain Y if and only if it passes in addition to the "vertical line test" the following "horizontal line test": any "horizontal line", i.e., any set $H_0 \subseteq X \times Y$ of the form $H(x_0) := \{(x, y_0) : x \in X\}$ for a fixed $y_0 \in Y$ intersects R in **at most one** "point".
d. It follows from (4.12) on p.82 but also from the above that that R is the graph of an injective function with domain X and codomain Y if and only if it passes in addition to the "vertical line test" the following "horizontal line test": any "horizontal line", i.e., any set $H_0 \subseteq X \times Y$ of the form $H(x_0) := \{(x, y_0) : x \in X\}$ for a fixed $y_0 \in Y$ intersects R in **exactly one** "point". Note the symmetry between this test and the one for vertical lines. This is another indication that the inverse graph R^{-1} of a bijective function is a graph of a function (the inverse function f^{-1}).

Remark 4.11. Abstract math is about proving theorems and propositions and functions are very important tools for that. It may be very important to know or to show that a certain function is injective or surjective or both. But these properties depend on the choice of domain and codomain as this simple example shows.

Let $f : A \to B$ be the function $f(x) := x^2$.

$A = \mathbb{R}, B = \mathbb{R}$:	f is neither injective nor surjective
$A =]-2, 3[, B = [0, 9[$:	f is surjective but not injective
$A =]0, 3[, B = [0, 9]$:	f is injective but not surjective
$A =]0, 3[, B =]0, 9[$:	f is bijective

The above explains why specification of domain and codomain are required for a function.

Definition 4.10 (Restriction/Extension of a function)**.** Given are three non-empty sets A, X and Y such that $A \subseteq X$. Let $f : X \to Y$ a function with domain X. We define the **restriction of** f **to** A as the function

$$f\big|_A : A \to Y \quad \text{defined as} \quad f\big|_A(x) := f(x) \text{ for all } x \in A. \tag{4.16}$$

Conversely let $f : A \to Y$ and $\varphi : X \to Y$ be functions such that $f = \varphi\big|_A$. We then call φ an **extension** of f to X.

Example 4.20 (No duplicates in sets)**.** For an example let $X := \mathbb{R}$, $A := [0, 1]$ and $f(x) := 3x^2(x \in [0, 1])$. For any $\alpha \in \mathbb{R}$ the function $\varphi_\alpha : \mathbb{R} \to \mathbb{R}$ defined as $\varphi_\alpha(x) := 3x^2$ if $0 \leqq x \leqq 1$ and αx otherwise defines a different extension of f to $\mathbb{R}$.

Notations 4.1. As the only difference between f and $f\big|_A$ is the domain, it is customary to write f instead of $f\big|_A$ to make formulas look simpler if doing so does not give rise to confusions.

Remark 4.12. The restriction $f\big|_A$ is always uniquely determined by f. Such is usually not the case for extensions if A is a strict subset of X.

Many more properties of mappings will be discussed later. Now we look at families, sequences and some additional properties of sets.

4.2.3 Sequences, families and functions as families

Definition 4.11 (Indices)**.** Given is an expression of the form

$$X_\imath.$$

We say that X is **indexed by** or **subscripted by** or **tagged by** $\imath$. We call $\imath$ the **index** or **subsript** of X and we call $X_\imath$ an **indexed item** .

Remark 4.13. Both X and $\imath$ can occur in many different ways. Here is a collection of indexed items:

$$x_7,\ A_\alpha,\ k_T,\ \mathfrak{H}_{2/9},\ f_x,\ x_t, h_{\mathscr{A}},\ , i_{\mathbb{R}},\ H_{2\pi}$$

Some of the indices in this collection are highly unusual: Not only are some of them negative but they are fractions (e.g., $2/9$) or irrational (e.g., 2π) Others are not even numbers (e.g., α, T, x, t, $\mathscr{A}$ and $\mathbb{R}$) where it is not clear from the information available to us whether those indices are names of variables which represent numbers or whether they represent functions, sets or other mathematical objects. There is one exception: The index of $i_{\mathbb{R}}$ is the set of all real numbers.

We can turn any set into a "family" by tagging each of its members with an index. As an example, look at the following two indexed versions of the set S_2 from example 2.1 on p. 10:

$$F = (a_1, e_1, e_2, i_1, i_2, i_3, o_1, o_2, o_3, o_4, u_3, u_5, u_9, u_{11}, u_{99})$$
$$G = (a_k, e_{-\sqrt{2}}, e_1, i_{-6}, i_{\mathscr{B}}, i_{\mathbb{R}}, o_7, o_{2/3}, o_{-8}, o_3, u_A, u_B, u_C, u_D, u_E)$$

We note several things:

- ***a**. F has the kind of indices that we are familiar with: all of them are positive integers.*
- ***b**. Some of the indices in F occur multiple times. For example, 3 occurs as an index for i_3, o_3, u_3.*
- ***c**. All of the indices in G are unique.*
- ***d**. As in remark 4.13 some of the indices are very unusual.*

*The last point is not much of a problem as mathematicians are used to very unusual notation but point (**b**), the non-uniqueness of indices, is something that we want to avoid. We ask for the following: The indices of an indexed collection must belong to some set J and each index $\imath \in J$ must be used exactly once. Remember that this automatically takes care of the duplicate indices problem as a set never contains duplicate values (see def.2.1 on p. 10). We further demand that there is a set X such that each indexed item $x_\imath$ belongs to X.*

We now are ready to give the definition of a family:

Definition 4.12 (Indexed families)**.** Let J and X be non-empty set and assume that

for each $\imath \in J$ there exists exactly one indexed item $x_\imath \in X$.

Let $R := \{(\imath, x_\imath) : \imath \in J\}$. Then R is a relation on (J, X) which satisfies (4.6) of the definition of a function

$$F : J \to X, \qquad \imath \mapsto F(\imath)$$

(see def.4.6 on p.74) whose graph Γ_F equals R.

We write $(x_\imath)_{\imath \in J}$ for this function if we want to deal with the collection of indexed elements $x_\imath$ rather than the function $F(\cdot)$ or the relation R. Reasons for this will be given in rem.4.15 on p.86.

$(x_\imath)_{\imath \in J}$ is called an **indexed family** or simply a **family** in X and J is called the **index set** of the family. For each $\jmath \in J$, $x_\jmath$ is called a **member of the family** $(x_\imath)_{\imath \in J}$.

$\imath$ is a dummy variable: $(x_\imath)_{\imath \in J}$ and $(x_k)_{k \in J}$ describe the same family as long as $\imath \mapsto x_\imath$ and $k \mapsto x_k$ describe the same function $F : J \to X$. This should not come as a surprise to you if you recall remark 4.8 on p.75.

Remark 4.14. The codomain X does not occur in the notation $(x_\imath)_{\imath \in J}$. This is OK because we do now worry about surjectivity or injectivity of families and the only thing that matters about X is that it is big enough to contain each indexed item. You can think of the codomain as

$$X = \bigcup \left[x_\imath : \imath \in J \right] := \{x : x = x_{i_0} \text{ for some } i_0 \in I\}$$

if you like. [49]

Note 4.1 (Simplified notation for families)**.** If there is no confusion about the index set then it can be dropped from the specification of a family and we simply write $(x_\imath)_\imath$ instead of $(x_\imath)_{\imath\in J}$. Even the index outside the right parentheses may be omitted and we write $(x_\imath)$ if it is clear that we are talking about families.

For example, a proposition may start as follows: Let (A_α) and (B_α) be two families of subsets of Ω indexed by the same set. Then .....

It is clear from the formulation that we deal in fact with two families $(A_\alpha)_{\alpha\in J}$ and $(B_\alpha)_{\alpha\in J}$. Nothing is said about J, probably because the proposition is valid for any index set or because this set was fixed once and for all earlier on for the entire section.

Example 4.21. Here is an example of a family of subsets of $\mathbb{R}$ which are indexed by real numbers: Let $J = [0,1]$ and $X := 2^{\mathbb{R}}$. For $0 \leqq x \leqq 1$ let $A_x := [x, 2x]$ be the set of all real numbers between x and $2x$. Then $(A_x)_{x\in[0,1]}$ is such a family.

Remark 4.15. If a family is just some kind of function, why bother with yet another definition? The answer to this is that there are many occasions in which writing something as a collection of indexed items rather than as a function makes things easier to understand. For example, look at theorem 5.1 (De Morgan's Law) on p.93. One of the formulas there states that for any indexed family $(A_\alpha)_{\alpha\in I}$ of subsets of a universal set Ω it is true that

$$\Big(\bigcup_\alpha A_\alpha\Big)^{\complement} = \bigcap_\alpha A_\alpha^{\complement}.$$

Without the notion of a family you might have to say something like this: Let $A : I \to 2^\Omega$ be a function which assigns its arguments to subsets of Ω . Then

$$\Big(\bigcup_\alpha A(\alpha)\Big)^{\complement} = \bigcap_\alpha A(\alpha)^{\complement}.$$

The additional parentheses around the index α just add complexity to the formula.

Example 4.22 (Sequences as families)**.** You have worked with special families before: those where $J = \mathbb{N}$ or $J = \mathbb{Z}_{\geq 0}$ and X is a subset of the real numbers. Example: $x_n := 1/n$. Here

$$(x_n)_{n\in\mathbb{N}} \text{ corresponds to the indexed collection } 1,\ \frac{1}{2},\ \frac{1}{3},\ \dots,\ \frac{1}{n},\ \dots$$

Such families are called sequences:

Definition 4.13 (Sequences)**.** Let n_0 be an integer and let $J := \{n_0, n_{0+1} n_{0+2}, \dots\} = \{k \in \mathbb{Z} : k \geq n_0\}$. Let X be an arbitrary nonempty set. An indexed family in X with index set J is called a **sequence** in X with **start index** n_0.

As is true for families in general, the name of the index variable of a sequence does not matter as long as it is applied consistently. It does not matter whether you write $(x_j)_{j\in J}$ or $(x_n)_{n\in J}$ or $(x_\beta)_{\beta\in J}$.

[49] General unions and intersections will be defined in ch.5.1 (More on set operations). See def.5.1 on p.91.

Note 4.2 (Simplified notation for sequences)**.** It is customary to choose either of i, j, k, l, m, n as the symbol of the index variable of a sequence and to stay away from other symbols whenever possible.

If J is defined as above then is not unusual to see "$(x_n)_{n \geq n_0}$" instead of "$(x_n)_{n \in J}$". By default the index set for a sequence is $J = \mathbb{N} = \{1, 2, 3, 4, \dots\}$ and we can then write $(x_n)_n$ or just (x_n) (compare this to note 4.1 about simplified notation for families).

Here are two more examples of sequences:

Example 4.23 (Oscillating sequence)**.** $x_j := (-1)^j$ $(j \in \mathbb{N}_0)$
Try to understand why this is the sequence

$$x_0 = 1, \quad x_2 = -1, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 1, \quad x_5 = -1, \ \dots$$

Example 4.24 (Series (summation sequence))**.** Let $s_k := 1 + 2^{-1} + 2^{-2} + \dots + 2^{-k}$ $(k = 1, 2, 3, \dots)$:

$$\begin{aligned} s_1 &= 1, \quad s_2 = 1 + 1/2 = 2 - 1/2, \quad s_3 = 1 + 1/2 + 1/4 = 2 - 1/4, \quad \dots, \\ s_k &= 1 + 1/2 + \dots + 2^{k-1} = 2 - 2^{k-1}; \qquad s = 1 + 1/2 + 1/4 + 1/8 + \dots \quad \text{"infinite sum"}. \end{aligned}$$

You obtain s_{k+1} from $s_k = 2 - 2^{k-1}$ by cutting the difference 2^{k-1} to the number 2 in half (that would be 2^k) and adding that to s_k. It is intuitively obvious from $s_k = 2 - 2^{k-1}$ that the infinite sum s adds up to 2. Such an infinite sum is called a **series**. The precise definition of a series will be given later.

Remark 4.16. Having defined the family $(x_\imath)_{\imath \in J}$ as the function which maps $\imath \in J$ to $x_\imath$ means that a family distinguishes any two of its members $x_\imath$ and $x_\jmath$ by remembering what their indices are, even if they represent one and the same element of X: Think of "$(x_\imath)_{\imath \in J}$" as an abbreviation for

$$\text{“}\big((\imath, x_\imath)\big)_{\imath \in J}\text{”}. \tag{4.17}$$

Doing so should also make it much easier to see the equivalence of functions and families: (4.17) looks at its core very much like the graph $\{(\imath, x_\imath) : \imath \in J\}$ of the function $\imath \mapsto x_\imath$.

Remark 4.17 (Families and sequences can contain duplicates)**.** One of the important properties of sets is that they do not contain any duplicates (see def.2.1 (sets) on p.10). On the other hand remark 4.16 casually mentions that families and hence sequences as special kinds of families can contain duplicates. Let us look at this now more closely.

The two sets $A := \{31, 20, 20, 20, 31\}$ and $B := \{20, 31\}$ are equal. On the other hand let $J := \{\alpha, \beta, \pi, \star, Q\}$ and define the family $(w_\imath)_{\imath \in J}$ in B by its associated graph as follows:

$$\Gamma := \{(\alpha, 31), (\beta, 20), (\pi, 20), (\star, 20), (Q, 31)\}, \ \text{ i.e.,}, \ \ w_\alpha = 31, w_\beta = 20, , w_\pi = 20, w_\star = 20, w_Q = 31.$$

The three occurrences of 20 cannot be distinguished as elements of the set A. In contrast to this the items $(\beta, 20), (\pi, 20), (\star, 20)$ as elements of $\Gamma \subseteq J \times A = J \times B$ [50] are different from each other because two pairs (a, b) and (x, y) are equal only if $x = a$ and $y = b$.

[50] Be sure to understand that $J \times A = J \times B$!

In contrast to sets, families and sequences allow you to deal with duplicates.

We remember that, by definition 4.12 on p.85 of a family,

a family $(x_\imath)_{\imath \in J}$ in X is just the function $F : J \to X$ which maps $\imath \in J$ to $F(z) = x_z$. Conversely, let X, Y be nonempty sets and let $f : X \to Y$ be a function with domain X and codomain Y. For $x \in X$ let $f_x := f(x)$. Then f can be written as $(f_x)_{x \in X}$, i.e., as a family in Y with index set X. In other words, we have the following:

Proposition 4.1. *The following ways of specifying a function $f : X \to Y, \ \ x \mapsto f(x)$ are equivalent:*

a. *$f = (X, Y, \Gamma)$ is defined by its graph $\Gamma := \{(x, f(x)) : x \in X\}$.*
b. *f is defined by the following family in Y: $(f(x))_{x \in X}$*

Note that this is one case where we had to explicitly mention the codomain Y in the specification of the family.

There will be a lot more on sequences and series (sequences of sums) in later chapters, but we need to develop more concepts, such as convergence, to continue with this subject. Now let's get back to sets.

4.3 Addenda to chapter 4.1

4.3.1 Examples and exercises for functions and relations

Exercise 4.1. Injectivity and Surjectivity

- Let $f : \mathbb{R} \to [0, \infty[; \quad x \mapsto x^2$.
- Let $g : [0, \infty[\to [0, \infty[; \quad x \mapsto x^2$.
  In other words, g is same function as f as far as assigning function values is concerned, but that its domain was downsized to $[0, \infty[$.

Answer the following with **true** or **false**.

a.	f is surjective	**c.**	g is surjective
b.	f is injective	**d.**	g is injective

If your answer is **false** then give a specific counterexample.

Exercise 4.2 (Excercise 4.1 continued)**.** Let $A \subseteq \mathbb{R}$.

- Let $F_1 : A \to [-2, 20[; \quad x \mapsto x^2$.
- Let $F_2 : A \to [2, 20[; \quad x \mapsto x^2$.
- Let $G_1 : A \to [-2, 20[; \quad x \mapsto \sqrt{x}$.
- Let $G_2 : A \to [2, 20[; \quad x \mapsto \sqrt{x}$.
- Let $G_3 : A \to [-20, 2[; \quad x \mapsto \sqrt{x}$.
- Let $G_4 : A \to [-20, -2[; \quad x \mapsto \sqrt{x}$.

What choice of A makes

a.	F_1 surjective?	**c.**	F_2 surjective?	**e.**	G_1 surjective?	**g.**	G_2 surjective?
b.	F_1 injective?	**d.**	F_2 injective?	**f.**	G_1 injective?	**h.**	G_2 injective?

i.	G_3 surjective?	**k.**	G_4 surjective?
j.	G_3 injective?	**l.**	G_4 injective?

For the questions above

- Write **impossible** if no choice of $A \subseteq \mathbb{R}$ exists.
- Write **NAF** for any of $F_1, F_2, G_1, G_2, G_3, G_4$ which does **not define a function**.

Exercise 4.3. Find $f : X \to Y$ and $A \subseteq X$ such that $f(A^{\complement}) \neq f(A)^{\complement}$. Hint: use $f(x) = x^2$ and choose Y as a **one element only** set (which does not leave you a whole lot of choices for X). See example 4.17 on p.79.

Exercise 4.4. You will learn later in this course that
injective $\circ$ injective $=$ injective,
surjective $\circ$ surjective $=$ surjective.

The following illustrates that the reverse is not necessarily true.

Find functions $f : \{a\} \to \{b_1, b_2\}$ and $g : \{b_1, b_2\} \to \{a\}$ such that $h := g \circ f : \{a\}$ is bijective but such that it is **not true** that both f, g are injective and it is also **not true** that both f, g are surjective.

Hint: There are not a whole lot of possibilities. Draw possible candidates for f and g in arrow notation as on p.118. You should easily be able to figure out some examples. Again, think simple and look at example 4.17 on p.79.

Exercise 4.5. B/G Project 6.9.:

On $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ we define the relation $\sim$ as follows.

$$(m_1, n_1) \sim (m_2, n_2) \Leftrightarrow m_1 \cdot n_2 = n_1 \cdot m_2. \tag{4.18}$$

a. Prove that $\sim$ defines an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

Let

$$\mathcal{Q} := \{[(m, n)] : m, n \in \mathbb{Z} \text{ and } n \neq 0\} \tag{4.19}$$

be the set of all equivalence classes of $\sim$. We define two binary operations $\oplus$ and $\otimes$ on $\mathcal{Q}$ as follows;

$$[(m_1, n_1)] \oplus [(m_2, n_2)] := [(m_1 n_2 + m_2 n_1, n_1 n_2)], \tag{4.20}$$
$$[(m_1, n_1)] \otimes [(m_2, n_2)] := [(m_1 m_2, n_1 n_2)] \tag{4.21}$$

b. Prove that these binary operations are defined consistently: the right hand sides of (4.20) and (4.21) do not depend on the particular choice of elements picked from the sets $[(m_1, n_1)]$ and $[(m_2, n_2)]$. In other words, prove the following:

Let $(p_1, q_1) \sim (m_1, n_1)$ and $(p_2, q_2) \sim (m_2, n_2)$. Then

$$[(m_1 n_2 + m_2 n_1, n_1 n_2)] = [(p_1 q_2 + p_2 q_1, q_1 q_2)], \tag{4.22}$$
$$[(m_1 m_2, n_1 n_2)] = [(p_1 p_2, q_1 q_2)]. \tag{4.23}$$

or, equivalently, then

$$(m_1 n_2 + m_2 n_1, n_1 n_2) \sim (p_1 q_2 + p_2 q_1, q_1 q_2), \tag{4.24}$$
$$(m_1 m_2, n_1 n_2) \sim (p_1 p_2, q_1 q_2). \tag{4.25}$$

5 More on sets (Understand this!)

5.1 More on set operations (Study this!)

The material in this chapter thematically belongs to ch.2.1 on p.10 but it had to be deferred to this chapter as much of it deals with families of sets, i.e., families $(A_i)_i$

Definition 5.1 (Arbitrary unions and intersections)**.** Let J be a nonempty set and let $(A_i)_{i\in J}$ be a family of sets. We define

(5.1) $$\bigcup_{i\in I} A_i \;:=\; \bigcup\big[A_i : i\in I\big] \;:=\; \{x : x\in A_{i_0} \text{ for some } i_0\in I\},$$

(5.2) $$\bigcap_{i\in I} A_i \;:=\; \bigcap\big[A_i : i\in I\big] \;:=\; \{x : x\in A_{i_0} \text{ for each } i_0\in I\}.$$

We call $\bigcup\limits_{i\in I} A_i$ the **union** and $\bigcap\limits_{i\in I} A_i$ the **intersection** of the family $(A_i)_{i\in J}$

It is convenient to allow unions and intersections for the empty index set $J=\emptyset$. For intersections this requires the existence of a universal set Ω. We define

(5.3) $$\bigcup_{i\in\emptyset} A_i \;:=\emptyset, \qquad \bigcap_{i\in\emptyset} A_i \;:=\; \Omega.$$

Note that any statement concerning arbitrary families of sets such as the definition above covers finite lists $A_1, A_2, \dots, A_n$ of sets ($J=\{1,2,\dots,n\}$) and also sequences $A_1, A_2, \dots,$ of sets ($J=\mathbb{N}$).

Here are some examples of non-finite unions and intersections.

Example 5.1. For the following note that $[u,v]=\emptyset$ for $u>v$ and $]u,v[=\emptyset$ for $u\geqq v$ (see (2.10) on p.16). Let $a,b\in\mathbb{R}$. Then

(5.4) $$]a,b[= \bigcup_{n\in\mathbb{N}} [a+1/n, b-1/n],$$

(5.5) $$[a,b] \;=\; \bigcap_{n\in\mathbb{N}}]a-1/n, b+1/n[.$$

Proof: Homework!

■

Example 5.2. For any set A we have $A \;=\; \bigcup\limits_{a\in A}\{a\}$. According to (5.3) this also is true if $A=\emptyset$.

The following trivial lemma (a lemma is a "proof subroutine" which is not remarkable on its own but very useful as a reference for other proofs) is useful if you need to prove statements of the form $A\subseteq B$ or $A=B$ for sets A and B. It is a means to simplity the proofs of [1] B/G (Beck/Geoghegan), project 5.12. You must reference this lemma as the "inclusion lemma" when you use it in your homework or exams. Be sure to understand what it means if you choose $J=\{1,2\}$ (draw one or two Venn diagrams).

Lemma 5.1 (Inclusion lemma)**.** *Let J be an arbitrary, non-empty index set. Let $U \subseteq X_j, Y, Z_j, W$ $(j \in J)$ be sets such that $U \ \subseteq Y \ X_j \subseteq Y \subseteq Z_j \subseteq W$ for all $j \in J$. Then*

$$(5.6) \qquad U \ \subseteq \ \bigcap_{j \in J} X_j \ \subseteq \ Y \ \subseteq \bigcup_{j \in J} Z_j \ \subseteq W.$$

Proof:
Let $x \in U$. Then $x \in X_j$ for all $j \in J$, hence $xx \in \bigcap_{j \in J} X_j$.

Now let $x \in \bigcap_{j \in J} X_j$. Then $x \in X_j$ for all $j \in J$. But then $x \in Y$ for all $j \in J$ because $X_j \subseteq Y$ for all $j \in J$.

$x \in Y$ for all $j \in J$ implies that $x \in Y$ and the left side inclusion of the lemma is shown.

Now assume $x \in Y$. We note that $Y \subseteq Z_j$ for all $j \in J$ implies $x \in Z_j$ for all $j \in J$.

But then $x \in Z_j$ for at least one $j \in J$ (did you notice that we needed to assume $J \neq \emptyset$?)

It follows that $x \in \bigcup_{j \in J} Z_j$ and the middle inclusion of the lemma is shown.

Finally, assume $x \in \bigcup_{j \in J} Z_j$ It follows from the definitions of unions that there exists at least one $j_0 \in J$ such that $x \in Z_{j_0}$. But then $x \in W$ as W contains Z_{j_0}. x is an arbitrary element of $\bigcup_{j \in J} Z_j$ and if follows that $\bigcup_{j \in J} Z_j \ \subseteq W$. This finishes the proof of the rightmost inclusion. ■

Definition 5.2 (Disjoint families)**.** Let J be a nonempty set. We call a family of sets $(A_i)_{i \in J}$ a **mutually disjoint family** if any two different sets A_i, A_j have intersection $A_i \cap A_j = \emptyset$, i.e., if any two sets in that family are mutually disjoint.

Definition 5.3 (Partition)**.** Let $\mathfrak{A} \subseteq 2^{\Omega}$. We call $\mathfrak{A}$ a **partition** of Ω if $A \cap B = \emptyset$ for any two $A, B \in \mathfrak{A}$ and $\Omega \ = \ \biguplus \Big[A : A \in \mathfrak{A} \Big]$.

We extend this definition to arbitrary families and hence finite collections and sequences of subsets of Ω: Let J be an arbitrary non-empty set, let $(A_j)_{j \in J}$ be a family of subsets of Ω and let $\mathfrak{A} \ := \ \{A_j : j \in J\}$. We call $(A_j)_{j \in J}$ a partition of Ω if $\mathfrak{A}$ is a partition of Ω.
Note that duplicate non-empty sets cannot occur in a disjoint family of sets because otherwise the condition of mutual disjointness does not hold.

Example 5.3. Here are some examples of partitions.

a. For any set Ω the collection $\{ \ \{\omega\} : \omega \in \Omega\}$ is a partition of Ω.

b. The empty set is a partition of the empty set and it is its only partition. Do you see that this is a special case of **a**?

c. The set of half open intervals $\{ \ ]k, k+1] : k \in \mathbb{Z}\}$ is a partitioning of $\mathbb{R}$.

d. Given is a strictly increasing sequence $n_0 = 0 < n_1 < n_2 < \dots$ of non-negative integers. For $k \in \mathbb{N}$ let $A_k := \{j \in \mathbb{N} : n_{k-1} < j \leqq n_k\}$. Then the set $\{A_k : k \in \mathbb{N}\}$ is a partition of $\mathbb{N}$ (**not** of $\mathbb{Z}_{\geq 0}$!)

Theorem 5.1 (De Morgan's Law)**.** *Let there be a universal set Ω (see (2.6) on p.13). Then the following "duality principle" holds for any indexed family $(A_\alpha)_{\alpha \in I}$ of sets:*

(5.7) $$\textbf{\textit{a.}}\ \Big(\bigcup_\alpha A_\alpha\Big)^{\complement} = \bigcap_\alpha A_\alpha^{\complement} \qquad \textbf{\textit{b.}}\ \Big(\bigcap_\alpha A_\alpha\Big)^{\complement} = \bigcup_\alpha A_\alpha^{\complement}$$

To put this in words, the complement of an arbitrary union is the intersection of the complements and the complement of an arbitrary intersection is the union of the complements.

Generally speaking the formulas are a consequence of the duality principle for set operations which states that any true statement involving a family of subsets of a universal sets can be converted into its "dual" true statement by replacing all subsets with their complements, all unions with intersections and all intersections with unions.

Proof of De Morgan's law, formula a:

First we prove that $\complement(\bigcup_\alpha A_\alpha) \subseteq \bigcap_\alpha(\complement A_\alpha)$: Assume that $x \in \complement(\bigcup_\alpha A_\alpha)$. Then $x \notin (\bigcup_\alpha A_\alpha)$ which is the same as saying that x does not belong to any of the A_α. That means that x belongs to each $\complement A_\alpha$ and hence also to the intersection $\bigcap_\alpha(\complement A_\alpha)$.

Now we prove that the right hand side set of formula a contains the left hand side set. So let $x \in \bigcap(\complement A_\alpha)$. Then x belongs to each of the $\complement A_\alpha$ and hence to none of the A_α. Then it also does not belong to the union of all the A_α and must therefore belong to the complement $\complement(\bigcup_\alpha A_\alpha)$. This completes the proof of formula a. The proof of formula b is not given here because the mechanics are the same. ■

You should draw the Venn diagrams involving just two sets A_1 and A_2 for both formulas a and b so that you understand the visual representation of De Morgan's law.

Proposition 5.1 (Distributivity of unions and intersections)**.** *Let $(A_i)_{i \in I}$ be an arbitrary family of sets and let B be a set. Then*

(5.8) $$\bigcup_{i \in I}(B \cap A_i) = B \cap \bigcup_{i \in I} A_i,$$

(5.9) $$\bigcap_{i \in I}(B \cup A_i) = B \cup \bigcap_{i \in I} A_i.$$

Proof: We only prove (5.8).
Proof of "$\subseteq$": It follows from $B \cap A_i \subseteq A_i$ for all i that $\bigcup_i (B \cap A_i) \subseteq \bigcup_i A_i$. Moreover, $B \cap A_i \subseteq B$ for all i implies $\bigcup_i (B \cap A_i) \subseteq \bigcup_i B$ which equals B. It follows that $\bigcup_i (B \cap A_i)$ is contained in the intersection $\Big(\bigcup_i A_i\Big) \cap B$.
Proof of "$\supseteq$": Let $x \in B \cap \bigcup_i A_i$. Then $x \in B$ and $x \in A_{i^\star}$ for some $i^\star \in I$, hence $x \in B \cap A_{i^\star}$, hence $x \in \bigcup_i (B \cap A_i)$. ■

Proposition 5.2 (Rewrite unions as disjoint unions)**.** *Let $(A_j)_{j\in\mathbb{N}}$ be a sequence of sets which all are contained within the universal set Ω. For $n \in \mathbb{N}$ let $B_n := \bigcup_{j=1}^{n} A_j \;=\; A_1 \cup A_2 \cup \dots \cup A_n$*

Further, let $C_1 := A_1 = B_1$ and $C_{n+1} := A_{n+1} \setminus B_n$ ($n \in \mathbb{N}$). Then

a. *The sequence $(B_j)_j$ is increasing: $m < n \Rightarrow B_m \subseteq B_n$,*

b. *For each $n \in \mathbb{N}$, $\bigcup_{j=1}^{n} A_j \;=\; \bigcup_{j=1}^{n} B_j$,*

c. *The sets C_j are mutually disjoint and $\bigcup_{j=1}^{n} A_j \;=\; \bigcup_{j=1}^{n} B_j \;=\; \biguplus_{j=1}^{n} C_j$.*

Proof: Homework!

■

In certain situation it is advantageous to think of the subsets of a universal set Ω as "binary" functions $\Omega \to \{0,1\}$.

Definition 5.4 (indicator function for a set)**.** Let Ω be "the" universal set, i.e., we restrict our scope of interest to subsets of Ω. Let $A \subseteq \Omega$. Let $1_A : \Omega \to \{0,1\}$ be the function

$$1_A(\omega) \;:=\; \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{5.10}$$

1_A is called the **indicator function** of the set A. [51]

The above association of a subset A of Ω with its indicator function is unique:

Proposition 5.3. *Let $\mathscr{F}(\Omega,\{0,1\})$ denote the set of all functions $f : \Omega \to \{0,1\}$, i.e., all functions f with domain Ω for which the only possible function values $f(\omega)$ are zero or one.*

a. *The mapping*

$$F : 2^{\Omega} \to \mathscr{F}(\Omega,\{0,1\}), \quad \textit{defined as} \;\; F(A) := 1_A \tag{5.11}$$

which assigns to each subset of Ω its indicator function is injective.

b. *Let $f \in \mathscr{F}(\Omega,\{0,1\})$. Further, let $A := \{f = 1\} := f^{-1}(\{1\}) := \{a \in A : f(a) = 1\}$.* [52] *Then $f = 1_A$.*

c. *The function F above is bijective and its inverse function is*

$$G : \mathscr{F}(\Omega,\{0,1\}) \to 2^{\Omega}, \quad \textit{defined as} \;\; G(f) := \{f = 1\}. \tag{5.12}$$

[51] Some authors call this function the **characteristic function** of the set A and some choose to write χ_A or $\mathbb{1}_A$ instead of 1_A.

[52] The inverse image $g^{-1}(B)$ for a subset B of the codomain of a function g will be defined in def. 6.1 on p. 102

Proof of ***a***: *This follows from* ***c*** *which will be proved below.*

Proof of ***b***: *We have*

$$\begin{aligned} f(\omega) = 1 \ &\Leftrightarrow \omega \in \{f = 1\} \quad \textit{(def. of inverse image)} \\ &\Leftrightarrow \omega \in A \quad \textit{(because } A = \{f=1\}\textit{)} \\ &\Leftrightarrow 1_A(\omega) = 1 \quad \textit{(def. of indicator function).} \end{aligned}$$

It follows that $f(\omega) = 1$ *if and only if* $1_A(\omega) = 1$. *As the only other possible function value is* 0 *we conclude that* $f(\omega) = 0$ *if and only if* $1_A(\omega) = 0$. *It follows that* $f(\omega) = 1_A(\omega)$ *for all* $\omega \in \Omega$, *i.e.,* $f = 1_A$ *and this proves* ***b***.

Proof of ***c***: *It follows from* ***a*** *and* ***b*** *that* F *is bijective. According to theorem 4.1 on p.83 about the characterization of inverse functions we have a second proof if we can demonstrate that* F *and* G *are inverse to each other. To prove this it suffices to show that*

$$G \circ F \ = id_{2^\Omega} \quad \textit{and} \quad F \circ G \ = \ id_{\mathscr{F}(\Omega, \{0,1\})}. \tag{5.13}$$

Let $A \subseteq \Omega$ *and* $A \in \Omega$. *Then*

$$G \circ F(A) \ = G(1_A) \ = \ \{1_A = 1\} \ = \ \{\omega \in \Omega : 1_A(\omega) = 1\} \ = \ \{\omega \in \Omega : \omega \in A\} \ = \ A.$$

This proves $G \circ F = id_{2^\Omega}$. *Let* $\omega \in \Omega$. *Then*

$$\begin{aligned} \big(F \circ G(f)\big)(\omega) \ &= F(\{f=1\})(\omega) \ = \ 1_{\{f=1\}}(\omega) \\ &= \begin{cases} 1 & \textit{iff } \omega \in \{f=1\}, \\ 0 & \textit{iff } \omega \notin \{f=1\} \end{cases} = \begin{cases} 1 & \textit{iff } f(\omega) = 1, \\ 0 & \textit{iff } f(\omega) \neq 1 \end{cases} = \begin{cases} 1 & \textit{iff } f(\omega) = 1, \\ 0 & \textit{iff } f(\omega) = 0 \end{cases} = \ f(\omega). \end{aligned}$$

The equation next to the last results from the fact that the only possible function values for f *are* 0 *and* 1. *It follows that* $F \circ G(f) \ = \ id_{\mathscr{F}(\Omega,\{0,1\})}(f)$ *for all* $f \in \mathscr{F}(\Omega, \{0,1\})$, *hence* $F \circ G \ = \ id_{\mathscr{F}(\Omega,\{0,1\})}$. *We have proved* (5.13) *and hence* ***c***. ■

5.2 Cartesian products of more than two sets

Remark 5.1 (Associativity of cartesian products)**.** Assume we have three sets A, B and C. We can then look at

$$\begin{aligned} (A \times B) \times C \ &= \ \{((a,b),c) : a \in A, b \in B, c \in C\} \\ A \times (B \times C) \ &= \ \{(a,(b,c)) : a \in A, b \in B, c \in C\} \end{aligned}$$

The mapping

$$F : (A \times B) \times C \to A \times (B \times C), \qquad \big((a,b),c\big) \mapsto \big((a,b),c\big)$$

is bijective because it has the mapping

$$G : A \times (B \times C) \to (A \times B) \times C, \qquad \big((a,b),c\big) \mapsto \big((a,b),c\big)$$

as an inverse. For both $(A \times B) \times C$ and $A \times (B \times C)$ there are bijections to the set $\{(a,b,c) : a \in A, b \in B, c \in C\}$ of all triplets (a,b,c): the obvious bijections would be $(a,b,c) \mapsto \big((a,b),c\big)$ and $(a,b,c) \mapsto \big((a,b),c\big)$.

This remark leads us to the following definition:

Definition 5.5 (Cartesian Product of three or more sets)**.** The **cartesian product** of three sets A, B and C is defined as

$$A \times B \times C \ := \ \{(a, b, c) : a \in A, b \in B, c \in C\}$$

i.e., it consists of all pairs (a, b, c) with $a \in A$, $b \in B$ and $c \in C$.

More generally, for N sets $X_1, X_2, X_3, \dots, X_N$ ($N \in \mathbb{N}$), we define the **cartesian product** as [53]

$$X_1 \times X_2 \times X_3 \times \dots \times X_N \ := \ \{(x_1, x_2, \dots, x_N) : x_j \in X_j \text{ for all } 1 \leqq j \leqq N\}$$

Two elements $(x_1, x_2, \dots, x_N)$ and $(y_1, y_2, \dots, y_N)$ of $X_1 \times X_2 \times X_3 \times \dots \times X_N$ are called **equal** if and only if $x_j = y_j$ for all j such that $1 \leqq j \leqq N$. In this case we write $(x_1, x_2, \dots, x_N) = (y_1, y_2, \dots, y_N)$.

As a shorthand, we abbreviate $X^N := \underbrace{X \times X \times + \dots \times X}_{N \text{times}}$.

Example 5.4 (N–dimensional coordinates)**.** Here is the most important example of a cartesian product of N sets. Let $X_1 = X_2 = \dots = X_N = \mathbb{R}$. Then

$$\mathbb{R}^N \ = \ \{(x_1, x_2, \dots, x_N) : x_j \in \mathbb{R} \text{ for } 1 \leqq j \leqq N\}$$

is the set of points in N–dimensional space. You may not be familiar with what those are unless $N = 2$ (see example 4.1 above) or $N = 3$.

In the 3–dimensional case it is customary to write (x, y, z) rather than (x_1, x_2, x_3) . Each such triplet of real numbers represents a point in (ordinary 3–dimensional) space and we speak of its x–coordinate, y–coordinate and z–coordinate.
For the sake of completeness: If $N = 1$ the item $(x) \in \mathbb{R}^1$ (where $x \in \mathbb{R}$; observe the parentheses around x) is considered the same as the real number x. In other words, we "identify" $\mathbb{R}^1$ with $\mathbb{R}$. Such a "one–dimensional point" is simply a point on the x–axis.

A short word on vectors and coordinates: For $N \leqq 3$ you can visualize the following: Given a point x on the x–axis or in the plane or in 3–dimensional space, there is a unique arrow that starts at the point whose coordinates are all zero (the "origin") and ends at the location marked by the point x. Such an arrow is customarily called a vector.

Because it makes sense in dimensions $1, 2, 3$, an N**–tuple** $(x_1, x_2, \dots, x_N)$ is also called a vector of dimension N. You will read more about this in the chapter 9, p.132, on vectors and vector spaces.

This is worth while repeating: We can uniquely identify each $x \in \mathbb{R}^N$ with the corresponding vector: an arrow that starts in $\underbrace{(0, 0, \dots, 0)}_{N \text{times}}$ and ends in x.

More will be said about n-dimensional space in section 9, p.132 on Vectors and vector spaces.

[53] If $N > 3$ there are many ways to group the factors of a cartesian product. For $N = 4$ there already are 3 times as many possibilities as for $N = 3$:

$$X_1 \times (X_2 \times X_3 \times X_4), \ \ (X_1 \times X_2) \times (X_3 \times X_4), \ \ X_1 \times (X_2 \times X_3 \times X_4),$$

Actually proving that we can group the sets with parentheses any way we like is very tedious and will not be done in this document.

Example 5.5 (Parallelepipeds)**.** Let $a_1 < b_1, a_2 < b_2, a_3 < b_3$ be real numbers. Then

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \;=\; \{(x, y, z) : a_1 \leqq x \leqq b_1, a_2 \leqq y \leqq b_2, a_3 \leqq z \leqq b_3\}$$

is the **parallelepiped** (box or quad parallel to the coordinate axes) with sides $[a_1, b_1], [a_2, b_2]$ and $[a_3, b_3]$. This generalizes in an obvious manner to N dimensions:

Let $N \in \mathbb{N}$ and $a_j < b_j$ ($j \in \mathbb{N}, j \leqq N, a_j, b_j \in \mathbb{R}$). Then

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_N, b_N] \;=\; \{(x_1, x_2, \dots, x_N) : a_j \leqq x_j \leqq b_j, j \in \mathbb{N}, j \leqq N\}$$

is the parallelepiped with sides $[a_1, b_1], \dots, [a_N, b_N]$.

We now introduce cartesian products of an entire family of sets $(X_i)_{i \in I}$.

Definition 5.6 (Cartesian Product of a family of sets)**.** Let I be an arbitrary, non–empty set (the index set) and let $(X_i)_{i \in I}$ be a family of non–empty sets X_i. The **cartesian product** of the family $(X_i)_{i \in I}$ is the set

$$\prod_{i \in I} X_i \;:=\; (\prod X_i)_{i \in I} \;:=\; \{(x_i)_{i \in I} : x_i \in X_i \;\forall i \in I\}$$

of all familes $(x_i)_{i \in I}$ each of whose members x_j belongs to the corresponding set X_j. The "$\prod$" is the greek "upper case" letter "Pi" (whose lower case incarnation "π" you are probably more familiar with). As far as I know, it was chosen because it has the same starting "p" sound as the word "product" (as in cartesian product).

Two elements $(x_i)_{i \in I}$ and $(y_k)_{k \in I}$ of $\prod\limits_{i \in I} X_i$ are called **equal** if and only if $x_j = y_j$ for all $j \in I$. In this case we write $(x_i)_{i \in I} = (y_k)_{k \in I}$.

If all sets X_i are equal to one and the same set X, we abbreviate $X^I := \prod\limits_{i \in I} X := \prod\limits_{i \in I} X_i$.

Remark 5.2. We note that each element $(y_x)_{x \in X}$ of the cartesian product Y^X is the function

$$y(\cdot) : X \to Y, \qquad x \mapsto y_x$$

(see def.4.12 (indexed families) and the subsequent remarks concerining the equivalence of functions and families). In other words,

(5.14) $$Y^X \;=\; \{f : f \text{ is a function with domain } X \text{ and codomain } Y\}.$$

5.3 Countable sets

This brief chapter is not very precise in that we do not talk about an axiomatic approach to finite sets and countably infinite sets. You can find that in ch.13 of [1] (Beck/Geoghegan) but also in ch.7.1 on p.107 of this document.

Everyone understands what a finite set is: It is a set with a finite number of elements.

Definition 5.7 (Finite sets)**.** Let $n \in \mathbb{N}$. we say that a set X has **cardinality** n and we write $\text{card}(X) := |X| := n$ if there is a bijective mapping between X and the set $[n] := \{1, 2, \dots, n\}$. In other words, a set X of cardinality n is one whose elements can be enumerated as $x_1, x_2, \dots, x_n$: The cardinality of a finite set is simply the number of elements it contains.

We define the empty set $\emptyset$ to be finite and set $card(\emptyset) := 0$.

You may be surprised to hear this but there are ways to classify the degree of infinity when looking at infinite sets.

The "smallest degree of infinity" is found in sets that can be compared, in a sense, to the set $\mathbb{N}$ of all natural numbers. Look back to definition (2.11) on the principle of mathematical induction. It is based on the property of $\mathbb{N}$ that there is a starting point a_1 = 1 and from there you can progress in a sequence

$$a_2 = 2;\ a_3 = 3;\ a_4 = 4;\ \dots\ a_k = k;\ a_{k+1} = k + 1;\ \dots$$

in which no two elements a_j, a_k are the same for different j and k. We have a special name for infinite sets whose elements can be arranged into a sequence of that nature.

Definition 5.8 (Countable and countably infinite sets)**.** Let X an arbitrary set such that there is a bijection $f : \mathbb{N} \to X$. This means that all of the elements of X can be arranged in a sequence

$$X = \{x_1 = f(1), x_2 = f(2), x_3 = f(3), \dots\}.$$

which is infinite, i.e., we rule out the case of sets with finitely many members. X is called a **countably infinite** set. We call a set that is either finite or countably infinite a **countable set**. and we also say that X is countable.

A set that is neither finite nor countably infinite is called **uncountable** or **not countable**

The proofs given in the remainder of this brief chapter on cardinality are not precise as we do not try to establish, for example in the first proof below, that for any subset B of a countable set there either exists an $n \in \mathbb{N}$ and a bijection from B to $[n]$ or there exists a bijection between B and $\mathbb{N}$. You may be surprised to hear that even the fact that there is no bijection between $[m] = \{1, 2, \dots m\}$ and $[n] = \{1, 2, \dots n\}$ for $m \neq n$ needs a proof that is not entirely trivial.

Theorem 5.2 (Subsets of countable sets are countable)**.** *Any subset of a countable set is countable.*

Proof: It is obvious that any subset of a finite set is finite. So we only need to deal with the case where we take a subset B of a countably infinite set A. Because A is countably infinite, we can arrange its elements into a sequence

$$A = \{a_1,\ a_2, a_3,\ \dots\}$$

We walk along that sequence and set

$$\begin{aligned} b_1 &:= a_{j_1} \quad \textit{where } j_1 = \min\{j \geqq 1 : a_{j_1} \in B\}, \\ b_2 &:= a_{j_2} \quad \textit{where } j_2 = \min\{j > j_1 : a_{j_2} \in B\}, \\ b_3 &:= a_{j_3} \quad \textit{where } j_3 = \min\{j > j_2 : a_{j_3} \in B\}, \dots \\ &\dots \end{aligned}$$

i.e., b_j is element number j of the subset B. The sequence (b_j) contains exactly all elements of B which means that this set is either finite (in case there is an $n_0 \in \mathbb{N}$ such that b_{n_0} is the last element of that sequence) or it is countably infinite in case that there are infinitely many b_j. ■

The following proposition is proved again more exactly in a later chapter (see thm.7.1 on p.112)

Theorem 5.3 (Countable unions of countable set)**.** *The union of countably many countable sets is countable.*

Proof: In the finite case let the sets be

$$A_1,\ A_2,\ A_3,\ \dots,\ A_N.$$

In the countable case let the sets be

$$A_1,\ A_2,\ A_3,\ \dots,\ A_n,\ A_{n+1},\ \dots$$

In either case we can assume that the sets are mutually disjoint, i.e., any two different sets A_i, A_j have intersection $A_i \cap A_j = \emptyset$ (see definition (2.4) on p.12). This is just another way of saying that no two sets have any elements in common. The reason we may assume mutual disjointness is that if we substitute

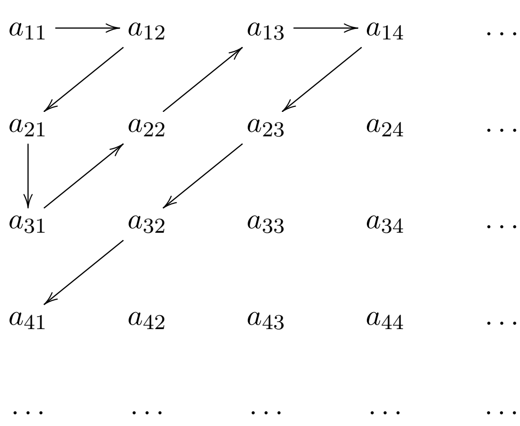
$$B_1 := A_1; \quad B_2 := A_2 \setminus B_1; \quad B_3 := A_3 \setminus B_2; \quad \dots$$

then

$$\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} B_j$$

(why?) and the B_j are mutually disjoint and also countable (thm. 5.2 on p.98). So let us assume the A_j are mutually disjoint. We write the elements of each set A_j as $a_{j1}, a_{j2}, a_{j3}, \dots$.

***A.** Let us first assume that none of those sets is finite. We start the elements of each A_j in a separate row and obtain*



Now we create a new sequence b_n by following the arrows from the start at a_{11}. We obtain

$$b_1 = a_{11};\ b_2 = a_{12};\ b_3 = a_{21};\ b_4 = a_{31};\ \dots$$

You can see that this sequence manages to collect all elements a_{ij} in that infinite two–dimensional grid and it follows that the union of the sets A_j is countable.

***B.** How do we modify this proof if some or all of the A_i are finite? We proceed as follows:*

If the predecessor A_{i-1} is finite with N_{i-1} elements, we stick the elements a_{ij} to the right of the last element $a_{i-1,N_{i-1}}$. Otherwise they start their own row.

If A_i itself is finite with N_i elements, we stick the elements $a_{i+1,j}$ to the right of the last element a_{i,N_i}. Otherwise they start their own row ...

B.1. *If an infinite number of sets has an infinite number of elements, then we have again a grid that is infinite in both horizontal and vertical directions and you create the "diagonal sequence" b_j just as before:*

Start off with the top-left element. Go one step to the right. Down–left until you hit the first column. Then down one step.

Then up–right until you hit the first row. Then one step to the right. Down–left until you hit the first column. Then down one step.

Then up–right until you hit the first row. Then one step to the right. Down–left until you hit the first column. Then down one step.

Then up–right until ...

B.2. *Otherwise, if only a finite number of sets has an infinite number of elements, then we have a grid that is infinite in only the horizontal direction. You create the "diagonal sequence" b_j almost as before.*

The exception: if you hit the bottom row, then must go one to the right rather than one down. Afterward you march again up–right until you hit the first column ... ■

Corollary 5.1 (The rational numbers are countable)**.**

Proof: Assume we can show that the set $\mathbb{Q} \cap [0,1[= \{q \in \mathbb{Q} : 0 \leqq q < 1$ is countable. Then the set

$$Q_z := \mathbb{Q} \cap [z, z+1[= \{q \in \mathbb{Q} : z \leqq q < z+1\}$$

is countable for any integer $z \in \mathbb{Z}$.

The reason: once we find a sequence b_j that runs through all elements of $\mathbb{Q} \cap [0,1[$, then the sequence $e_j := b_j + z$ runs through all elements of $\mathbb{Q}_z$.

But $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup \{-k : k \in \mathbb{N}$ is countable as a union of only three countable sets.

Observe that $\mathbb{Q} = \bigcup_{z \in \mathbb{Z}} Q_z$. It follows that $\mathbb{Q}$ is the countable union of the sets Q_z, each of which is countable.

So we are done with the proof ... except we still must prove that the set Q_0 of all rational numbers between zero and one is countable. We do that now.

Let $A_1 := 0$. *Let*

$$\begin{aligned} A_2 :=& \{z \in Q_1 : z \text{ has denominator } 2\} &=& \{\frac{0}{2}, \frac{1}{2}\} \\ A_3 :=& \{z \in Q_1 : z \text{ has denominator } 3\} &=& \{\frac{0}{3}, \frac{1}{3}, \frac{2}{3}\} \\ A_4 :=& \{z \in Q_1 : z \text{ has denominator } 4\} &=& \{\frac{0}{4}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}\} \\ & \dots \\ A_n :=& \{z \in Q_1 : z \text{ has denominator } n\} &=& \{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\} \\ & \dots \end{aligned}$$

Then each set is finite and $Q_1 = \bigcup_{k \in \mathbb{N}} A_k$ *is a countable union of countably many finite sets and hence, according to the previous theorem* (5.3), *countable. We are finished with the proof.* ■

Theorem 5.4 (The real numbers are uncountable)**.** *The real numbers are uncountable: There is no sequence* $(r_n)_{n \in \mathbb{N}}$ *such that* $\{r_n : n \in \mathbb{N}\} = \mathbb{R}$.

Proof by contradiction: This proof follows the one given in B/G [8] (thm.13.22, p.125).

According to thm.5.2 on p.98 it is enough to prove that the subset $[0,1]$ *of all decimals* $x = 0.d_1 d_2 d_3 \dots$ *where each* d_n *is a digit* $0, 1, 2, \dots, 9$ [54] *is not countable.*

Let us assume assume that $[0,1]$ *is countable, i.e., there exists a sequence of decimals* $x_n = 0.d_{n,1} d_{n,2} d_{n,3} \dots$ *as follows: for each* $x \in [0,1]$ *there exists some* $k = k(x) \in \mathbb{N}$ *such that* $x = x_k$ *and find a contradiction. For* $n \in \mathbb{N}$ *we define the digit* $\tilde{d}_n$ *as*

$$\tilde{d}_n := \begin{cases} 3 & \text{if } d_{n,n} \neq 3, \\ 4 & \text{if } d_{n,n} = 3. \end{cases}$$

Let $y := 0.\tilde{d}_1 \tilde{d}_2 \tilde{d}_3 \dots$. *Clearly* $y \in [0,1]$. *It follows from* $d_{1,1} \neq \tilde{d}_1$ *that* $y \neq x_1$. *It follows from* $d_{2,2} \neq \tilde{d}_2$ *that* $y \neq x_2$. *You should get the idea: Let* $k \in \mathbb{N}$. *It follows from* $d_{k,k} \neq \tilde{d}_k$ *that* $y \neq x_k$. *This contradicts the assumption that for each* $x \in [0,1]$ *there exists some* $k = k(x) \in \mathbb{N}$ *such that* $x = x_k$. ■

[54] We must include the number $1 = 0.\bar{9}$.

6 Sets and Functions, direct and indirect images (Study this!)

6.1 Direct images and indirect images (preimages) of a function

Definition 6.1. Let X, Y be two non-empty sets and $f : X \to Y$ be an arbitrary function with domain X and codomain Y. Let $A \subseteq X$ and $B \subseteq Y$. Let

$$f(A) := \{f(x) : x \in A\}, \tag{6.1}$$

$$f^{-1}(B) := \{x \in X : f(x) \in B\}. \tag{6.2}$$

We call $f(A)$ the **direct image** of A under f and we call We call $f^{-1}(B)$ the **indirect image** or **preimage** of B under f

Notational conveniences:

If we have a set that is written as $\{\dots\}$ then we may write $f\{\dots\}$ instead of $f(\{\dots\})$ and $f^{-1}\{\dots\}$ instead of $f^{-1}(\{\dots\})$. Specifically for $x \in X$ and $y \in Y$ we get $f^{-1}\{x\}$ and $f^{-1}\{y\}$. Many mathematicians will write $f^{-1}(y)$ instead of $f^{-1}\{y\}$ but this writer sees no advantages doing so whatsover. There seemingly are no savings with respect to time or space for writing that alternate form but we are confounding two entirely separate items: a subset $f^{-1}\{y\}$ of X v.s. the function value $f^{-1}(y)$ of $y \in Y$ which is an element of X. We can talk about the latter only in case that the inverse function f^{-1} of f exists.

In measure theory and probability theory the following notation is also very common: $\{f \in B\}$ rather than $f^{-1}(B)$ and $\{f = y\}$ rather than $f^{-1}\{y\}$

Let $a < b \in \mathbb{R}$. We write $\{a \leqq f \leqq b\}$ for $f^{-1}([a,b])$, $\{a < f < b\}$ for $f^{-1}(]a,b[)$, $\{a \leqq f < b\}$ for $f^{-1}([a,b[)$ and $\{a < f \leqq b\}$ for $f^{-1}(]a,b])$, $\{f \leqq b\}$ for $f^{-1}(]-\infty,b])$, etc.

Proposition 6.1. *Some simple properties:*

$$f(\emptyset) = f^{-1}(\emptyset) = \emptyset \tag{6.3}$$

$$A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2) \tag{6.4}$$

$$B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2) \tag{6.5}$$

$$x \in X \Rightarrow f(\{x\}) = \{f(x)\} \tag{6.6}$$

$$f(X) = Y \quad \Leftrightarrow \quad f \text{ is surjective} \tag{6.7}$$

$$f^{-1}(Y) = X \quad \text{always!} \tag{6.8}$$

Proof: Homework! ■

Proposition 6.2 (f^{-1} is compatible with all basic set ops). *In the following we assume that J is an arbitrary index set, and that $B \subseteq Y$, $B_j \subseteq Y$ for all j.*

The following all are true:

$$f^{-1}\left(\bigcap_{j\in J} B_j\right) = \bigcap_{j\in J} f^{-1}(B_j) \tag{6.9}$$

$$f^{-1}\left(\bigcup_{j\in J} B_j\right) = \bigcup_{j\in J} f^{-1}(B_j) \tag{6.10}$$

$$f^{-1}(B^{\complement}) = f^{-1}(B)^{\complement} \tag{6.11}$$

$$f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2) \tag{6.12}$$

Proof of (6.9): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}\left(\bigcap_{j\in J} B_j\right) &\Leftrightarrow f(x) \in \bigcap_{j\in J} B_j \quad (\textit{def preimage}) \\ &\Leftrightarrow \forall j\ f(x) \in B_j \quad (\textit{def} \cap) \\ &\Leftrightarrow \forall j\ x \in f^{-1}(B_j) \quad (\textit{def preimage}) \\ &\Leftrightarrow x \in \bigcap_{j\in J} f^{-1}(B_j) \quad (\textit{def} \cap) \end{aligned} \tag{6.13}$$

Proof of (6.10): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}\left(\bigcup_{j\in J} B_j\right) &\Leftrightarrow f(x) \in \bigcup_{j\in J} B_j \quad (\textit{def preimage}) \\ &\Leftrightarrow \exists j_0 :\ f(x) \in B_{j_0} \quad (\textit{def} \cup) \\ &\Leftrightarrow \exists j_0 : x \in f^{-1}(B_{j_0}) \quad (\textit{def preimage}) \\ &\Leftrightarrow x \in \bigcup_{j\in J} f^{-1}(B_j) \quad (\textit{def} \cup) \end{aligned} \tag{6.14}$$

Proof of (6.11): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}(B^{\complement}) &\Leftrightarrow f(x) \in B^{\complement} \quad (\textit{def preimage}) \\ &\Leftrightarrow f(x) \notin B \quad (\textit{def } (\cdot)^{\complement}) \\ &\Leftrightarrow x \notin f^{-1}(B) \quad (\textit{def preimage}) \\ &\Leftrightarrow x \in f^{-1}(B)^{\complement} \quad ((\cdot)^{\complement}) \end{aligned} \tag{6.15}$$

Proof of (6.12): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}(B_1 \setminus B_2) &\Leftrightarrow x \in f^{-1}(B_1 \cap B_2^{\complement}) \quad (\textit{def} \setminus) \\ &\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2^{\complement}) \quad (\textit{see } (6.9)) \\ &\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2)^{\complement} \quad (\textit{see } (6.11)) \\ &\Leftrightarrow x \in f^{-1}(B_1) \setminus f^{-1}(B_2) \quad (\textit{def} \setminus) \end{aligned} \tag{6.16}$$

■

Proposition 6.3 (Properties of the direct image)**.** *In the following we assume that J is an arbitrary index set, and that $A \subseteq X$, $A_j \subseteq X$ for all j.*
The following all are true:

$$f(\bigcap_{j\in J} A_j) \subseteq \bigcap_{j\in J} f(A_j) \tag{6.17}$$

$$f(\bigcup_{j\in J} A_j) = \bigcup_{j\in J} f(A_j) \tag{6.18}$$

Proof of (6.17): This follows from the monotonicity of the direct image (see 6.4):

$$\begin{aligned}\bigcap_{j\in J} A_j \subseteq A_i\,\forall i \in J &\Rightarrow f(\bigcap_{j\in J} A_j) \subseteq f(A_i)\,\forall i \in J \quad (\textit{see } 6.4)\\ &\Rightarrow f(\bigcap_{j\in J} A_j) \subseteq \bigcap_{i\in J} f(A_i) \quad (\textit{def} \cap)\end{aligned}$$

First proof of (6.18)) - "Expert proof":

$$y \in f(\bigcup_{j\in J} A_j) \Leftrightarrow \exists\, x \in X :\ f(x) = y \textit{ and } x \in \bigcup_{j\in J} A_j \quad (\textit{def } f(A)) \tag{6.19}$$

$$\Leftrightarrow \exists\, x \in X \textit{ and } j_0 \in J :\ f(x) = y \textit{ and } x \in A_{j_0} \quad (\textit{def} \cup) \tag{6.20}$$

$$\Leftrightarrow \exists\, x \in X \textit{ and } j_0 \in J :\ f(x) = y \textit{ and } f(x) \in f(A_{j_0}) \quad (\textit{def } f(A)) \tag{6.21}$$

$$\Leftrightarrow \exists\, j_0 \in J :\ y \in f(A_{j_0}) \quad (\textit{def } f(A)) \tag{6.22}$$

$$\Leftrightarrow y \in \bigcup_{j\in J} f(A_j) \quad (\textit{def} \cup) \tag{6.23}$$

Alternate proof of (6.18) *- Proving each inclusion separately. Unless you have a lot of practice reading and writing proofs whose subject is the equality of two sets you should write your proof the following way:*

A. Proof of "$\subseteq$":

$$y \in f(\bigcup_{j\in J} A_j) \Rightarrow \exists\, x \in X :\ f(x) = y \textit{ and } x \in \bigcup_{j\in J} A_j \quad (\textit{def } f(A)) \tag{6.24}$$

$$\Rightarrow \exists\, j_0 \in J :\ f(x) = y \textit{ and } x \in A_{j_0} \quad (\textit{def} \cup) \tag{6.25}$$

$$\Rightarrow y = f(x) \in f(A_{j_0}) (\textit{def } f(A)) \tag{6.26}$$

$$\Rightarrow y \in \bigcup_{j\in J} f(A_j) \quad (\textit{def} \cup) \tag{6.27}$$

B. Proof of "$\supseteq$":

This is a trivial consequence from the monotonicity of $A \mapsto f(A)$:

$$A_i \subseteq \bigcup_{j\in J} A_j\ \forall\, i \in J \Rightarrow f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\ \forall\, i \in J \tag{6.28}$$

$$\Rightarrow \bigcup_{i\in J} f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\ \forall\, i \in J \quad (\textit{def} \cup) \tag{6.29}$$

■

You see that the "elementary" proof is barely longer than the first one, but it is so much easier to understand!

Remark 6.1. In general you will not have equality in (6.17). Counterexample: $f(x) = x^2$ with domain $\mathbb{R}$: Let $A_1; =]-\infty, 0]$ and $A_2; = [0, \infty[$. Then $A_1 \cap A_2 = \{0\}$, hence $f(A_1 \cap A_2) = f(\{0\}) = \{0\}$. On the other hand, $f(A_1) = f(A_2) = [0, \infty]$, hence $f(A_1) \cap f(A_2) = [0, \infty]$ which is clearly bigger than $\{0\}$. ■

Proposition 6.4 (Preimage of function compositions). *Let X, Y, Z be an arbitrary, non–empty sets. Let $f : X \to Y$ and $g : Y \to Z$ and let $W \subseteq Z$. Then*

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}, \text{ i.e., } (g \circ f)^{-1}(W) = f^{-1}\big(g^{-1}(W)\big) \text{ for all } W \subseteq Z.$$

Proof:
a. "$\subseteq$": Let $W \subseteq Z$ and $x \in (g \circ f)^{-1}(W)$. Then $(g \circ f)(x) = g\big(f(x)\big) \in W$, hence $f(x) \in g^{-1}(W)$. But then $x \in f^{-1}\big(g^{-1}(W)\big)$. This proves "$\subseteq$".

b. "$\supseteq$": Let $W \subseteq Z$ and $x \in f^{-1}\big(g^{-1}(W)\big)$. Then $f(x) \in g^{-1}(W)$, hence $h(x) = g(f(x)) \in W$, hence $x \in h^{-1}(W) = (g \circ f)^{-1}(W)$. This proves "$\supseteq$". ■

Proposition 6.5 (Indirect image and fibers of f). *Let X, Y be non–empty sets and let $f : X \to Y$ be a function. We define on the domain X a relation "$\sim$" as follows:*

$$x_1 \sim x_2 \Leftrightarrow f(x_1) = f(x_2), \text{ i.e.,}. \tag{6.30}$$

***a.** Then "$\sim$" is an equivalence relation and its equivalence classes which we denote by $[x]_f$* [55] *are obtained as follows:*

$$[x]_f = \{a \in X : f(a) = f(x)\} \quad (x \in X). \tag{6.31}$$

***b.** The following is true for this equivalence relation:*

$$x \in X \Rightarrow \Big[\, [x]_f = \{a \in X : f(a) = f(x)\} = f^{-1}\{f(x)\} \,\Big] \tag{6.32}$$

$$A \subseteq X \Rightarrow f^{-1}(f(A)) = \bigcup_{a \in A} [a]_f. \tag{6.33}$$

Proof of (6.32): The equation on the left is nothing but the definition of the equivalence classes generated by an equivalence relation, the equation on the right follows from the definition of preimages.

Proof of (6.33):

As $f(A) = f(\bigcup_{x \in A}\{x\}) = \bigcup_{x \in A}\{f(x)\}$ (see 6.18), it follows that

$$f^{-1}(f(A)) = f^{-1}\Big(\bigcup_{x \in A}\{f(x)\}\Big) \tag{6.34}$$

$$= \bigcup_{x \in A} f^{-1}\{f(x)\} \quad \text{(see 6.10)} \tag{6.35}$$

$$= \bigcup_{x \in A} [x]_f \quad \text{(see 6.32)} \tag{6.36}$$

■

[55] $[x]_f$ is called the **fiber over** $f(x)$ of the function f.

Corollary 6.1.

$$A \in X \quad \Rightarrow \quad f^{-1}(f(A)) \; \supseteq \; A. \tag{6.37}$$

Proof: It follows from $x \sim x$ for all $x \in X$ that $x \in [x]_f$, i.e., $\{x\} \subseteq [x]_f$ for all $x \in X$. But then

$$A \; = \; \bigcup_{x \in A} \{x\} \; \subseteq \; \bigcup_{x \in A} [x]_f \; = \; f^{-1}(f(A)) \tag{6.38}$$

where the last equation holds because of (6.33). ■

Proposition 6.6.

$$B \; \subset Y \quad \Rightarrow \quad f(f^{-1}(B)) \; = \; B \cap f(X). \tag{6.39}$$

Proof of "⊆":

Let $y \in f(f^{-1}(B))$. There exists $x_0 \in f^{-1}(B)$ such that $f(x_0) = y$ (def direct image). We have
a. *$x_0 \in f^{-1}(B) \; \Rightarrow y = f(x_0) \in B$ (def. of preimage)*
b. *Of course $x_0 \in X$. Hence $y = f(x_0) \in f(X)$.*
***a** and **b** together imply $y \in B \cap f(X)$.*

Proof of "⊇":

Let $y \in f(X)$ and $y \in B$. We must prove that $y \in f(f^{-1}(B))$. Because $y \in f(X)$ there exists $x_0 \in X$ such that $y = f(x_0)$. Because $y = f(x_0) \in B$ we conclude that $x_0 \in f^{-1}(B)$ (def preimage).

We abbreviate $A := f^{-1}(B)$. Now it easy to see that

$$x_0 \in f^{-1}(B) = A \quad \Rightarrow \quad y \; = \; f(x_0) \in f(f^{-1}(B)). \tag{6.40}$$

We have shown that if $y \in f(X)$ and $y \in B$ then $y \in f(f^{-1}(B))$. The proof is completed. ■

Remark 6.2. Be sure to understand how the assumption $y \in f(X)$ was used.

Corollary 6.2.

$$B \in Y \quad \Rightarrow \quad f(f^{-1}(B)) \; \subseteq \; B. \tag{6.41}$$

Trivial as $f(f^{-1}(B)) \; = \; B \cap f(X) \; \subseteq \; B$. ■

7 Some miscellaneous topics

Although this chapter only contains a single topic at this time (cardinality), additional topics are planned in the future.

7.1 Cardinality - Alternate approach to Beck/Geoghegan (Study this!)

This chapter gives an alternative presentation of cardinality from that of ch.13 of B/G. It also gives a more exact presentation of the material in ch.5.3 on p.97 of this document.

Notation: *In this entire chapter on cardinality, if $n \in \mathbb{N}$, the symbol $[n]$ does not denote an equivalence class of any kind but the set $\{1, 2, \dots, n\}$ of the first n natural numbers.*

At the beginning of this chapter we look at two lemmata that let you replace bijective and surjective functions with more suitable ones that inherit bijectivity or surjectivity. This will come in handy when we prove propositions concerning cardinality.

The first lemma shows how to preserve bijectivity if two function values need to be switched around.

Lemma 7.1. *Let $X, Y \neq \emptyset$, let $f : X \to Y$ be bijective and let $x_1, x_2 \in X$. Let*

$$g(x) := \begin{cases} f(x_2) & \text{if } x = x_1, \\ f(x_1) & \text{if } x = x_2, \\ f(x) & \text{if } x \neq x_1, x_2. \end{cases} \tag{7.1}$$

(In other words, we swap two function arguments). Then $g : X \to Y$ also is bijective.

Proof: *Let $y_1 := f(x_1)$ and $y_2 := f(x_2)$. Let $f^{-1} : Y \to X$ be the inverse function of f and define $G : Y \to X$ as follows*

$$G(y) := \begin{cases} f^{-1}(y_2) & \text{if } y = y_1, \\ f^{-1}(y_1) & \text{if } y = y_2, \\ f^{-1}(y) & \text{if } y \neq y_1, y_2. \end{cases} \tag{7.2}$$

We show that G satisfies $G \circ g = id_X$ and $g \circ G = id_Y$, i.e., g has G as its inverse. This suffices to prove bijectivity of g.

$$\begin{aligned} y \neq y_1, y_2 \quad \Rightarrow \quad g \circ G(y) &= g(f^{-1}(y)) = f(f^{-1}(y)) = y \text{ as } f^{-1}(y) \neq x_1, x_2, \\ g \circ G(y_1) &= g(f^{-1}(y_2)) = g(x_2) = f(x_1) = y_1 \text{ as } f^{-1}(y_2) = x_2, \\ g \circ G(y_2) &= g(f^{-1}(y_1)) = g(x_1) = f(x_2) = y_2 \text{ as } f^{-1}(y_1) = x_1. \end{aligned}$$

Further,

$$\begin{aligned} x \neq x_1, x_2 \quad \Rightarrow \quad G \circ g(x) &= G(f(x)) = f^{-1}f(x)) = y \text{ as } f(x) \neq y_1, y_2, \\ G \circ g(x_1) &= G(f(x_2)) = G(y_2) = f^{-1}(y_1)) = x_1 \text{ as } f(x_1) = y_1, \\ G \circ g(x_2) &= G(f(x_1)) = G(y_1) = f^{-1}(y_2)) = x_2 \text{ as } f(x_2) = y_2. \end{aligned}$$

We have proved that g has an inverse, the function G. ■

Note that the validity of $G \circ g \;=\; id_X$ and $g \circ G \;=\; id_Y$ is obvious without the use of any formalism: g differs from f only in that it switches around the function values $f(x_1)$ and $f(x_2)$. and G differs from f^{-1} only in that this switch is reverted.

A more general version of the above shows how to preserve surjectivity if two function values need to be switched around.

Lemma 7.2. *Let $X, Y \neq \emptyset$ and assume that Y contains at least two elements y_1 and y_2. Let $f : X \to Y$ be surjective.*

Let $A_1 \;:=\; f^{-1}\{y_1\}$, $A_2 \;:=\; f^{-1}\{y_2\}$, and $B \;:=\; X \setminus (A_1 \cup A_2)$. Let

$$g(x) \;:=\; \begin{cases} y_2 & \text{if } x \in A_1, \\ y_1 & \text{if } x \in A_2, \\ f(x) & \text{if } x \in B. \end{cases} \tag{7.3}$$

In other words, everything that f maps to y_1 is now mapped to y_2 and everything that f maps to y_2 is now mapped to y_1. Then $g : X \to Y$ also is surjective.

Proof:

We notice that A_1, A_2, B partition X into three mutually exclusive parts: $X = B \uplus A_1 \uplus A_2$

and that the sets $f(A_1) = \{y_1\}$, $f(A_2) = \{y_2\}$, $f(B) = Y \setminus \{y_1, y_2\}$

partition Y into $Y = f(B) \uplus f(A_1) \uplus f(A_2)$. (Do you see why $f(B) = Y \setminus \{y_1, y_2\}$?)

B and hence $f(B)$ might be empty but none of the other four sets are.

It follows that there is indeed a function value $g(x)$ for each $x \in X$ and there is exactly one such value, i.e., g in fact defines a mapping from X to Y.

The surjectivity of g follows from that of f and the fact that

$$Y \;=\; f(B) \cup f(A_1) \cup f(A_2) \;=\; g(B) \cup g(A_2) \cup g(A_1) \tag{7.4}$$

(see (6.18) on p. 104 in prop. 6.3 (Properties of the direct image)). ■

The definitions of finite, countable, countably infinite and uncountable sets were given at the beginning of ch. 5.3: "Countable sets" on p. 97.

Definition 7.1 (cardinality comparisons)**.** Given two arbitrary sets X and Y we say **cardinality of** $X \;\leqq$ **cardinality of** Y and we write **card**$(X) \leqq$ **card**(Y) if there is an injective mapping $f : X \to Y$.

We say X, Y **have same cardinality** and we write **card**$(X) =$ **card**(Y) if both $\text{card}(X) \leqq \text{card}(Y)$ and $\text{card}(Y) \leqq \text{card}(X)$, i.e., there is a bijective mapping $f : X \xrightarrow{\sim} Y$.

Finally we say **cardinality of** $X \;<$ **cardinality of** Y and we write **card**$(X) <$ **card**(Y) if both $\text{card}(X) \leqq \text{card}(Y)$ and $\text{card}(Y) \neq \text{card}(X)$, i.e., there is an injective mapping but not a bijection $f : X \to Y$.

Proposition 7.1. *Let $m, n \in \mathbb{N}$. Let $\emptyset \neq A \subseteq [m]$. If $m < n$ then there is no surjection from A to $[n]$.*

Proof by induction on n:

Base case: Let $n = 2$. This implies $m = 1$ and $A = [1]$ (no other non-empty subset of $[1]$). For an arbitrary function $f : A \to [2]$ we have either $f(1) = 1$ in which case $2 \notin f(A)$ or $f(1) = 2$ in which case $1 \notin f(A)$. This proves the base case.

Induction assumption: Fix $n \in \mathbb{N}$ and assume that for any $\tilde{m} < n$ and non-empty $\tilde{A} \subseteq [\tilde{m}]$ there is no surjective $\tilde{f} : \tilde{A} \to [n]$.

We must prove the following: Let $m \in \mathbb{N}$ and $\emptyset \neq A \subseteq [m]$. If $m < n+1$ then there is no surjection from A to $[n+1]$.

We now assume to the contrary that a surjective $f : A \to [n+1]$ exists.

***case 1:** $n \notin A$:*

As $m < n+1$ this implies both $n, n+1 \notin A$, hence $A \subseteq [n-1]$.

Let $\tilde{A} := A \setminus f^{-1}\{n+1\}$. Then $\tilde{A} \subseteq A \subseteq [n-1]$

and, because the surjective f "hits" every integer between 1 and $n+1$ and we only removed those $a \in A$ which map to $n+1$, the restriction $\tilde{f}$ of f to $\tilde{A}$ covers any integer between 1 and n.

In other words, $\tilde{f} : \tilde{A} \to [n]$ is surjective, contradictory to our induction assumption.

***case 2:** $n \in A$ **and** $f(n) = n+1$: As in case 1, let $\tilde{A} := A \setminus f^{-1}\{n+1\}$.*

Then $\tilde{A} \subseteq [n-1]$ because n was discarded from A as an element of $f^{-1}\{n+1\}$.

Again, the surjective f "hits" every integer between 1 and $n+1$ and again, we only removed those $a \in A$ which map to $n+1$.

It follows as in case 1 that $\tilde{f} : \tilde{A} \to [n]$ is surjective, contradictory to our induction assumption.

***case 3:** $n \in A$ **and** $f(n) \neq n+1$: According to lemma 7.1 on p.109 we can replace f by a surjective function g which maps n to $n+1$.*

This function satisfies the conditions of case 2 above, for which it was already proved that no surjective mapping from A to $[n+1]$ exists. We have reached a contradiction. ■

Corollary 7.1 (No bijection from $[m]$ to $[n]$ exists). *B/G Thm.13.4: Let $m, n \in \mathbb{N}$. If $m \neq n$ then there is no bijective $f : [m] \xrightarrow{\sim} [n]$.*

Proof: We may assume $m < n$ and can now apply prop. 7.1 with $A := [m]$. ■

Corollary 7.2 (Pigeonhole Principle). *B/G Prop.13.5: Let $m, n \in \mathbb{N}$. If $m < n$ then there is no injective $f : [n] \to [m]$.*

Proof: Otherwise g would have a (surjective) left inverse $g : [m] \to [n]$ in contradiction to the preceding proposition. ■

Proposition 7.2 (B/G Prop.13.6, p.122: Subsets of finite sets are finite)**.** *Let $\emptyset \neq B \subseteq A$ and let A be finite. Then B is finite.*

Proof: Done by induction on the cardinality n of sets:

Base case: $n = 1$ or $n = 2$: Proof obvious.

Induction assumption: Assume that all subsets of sets of cardinality less than n are finite.

Now let A be a set with cardinality card$(A) = n$. there is a bijection $a(\cdot) : [n] \xrightarrow{\sim} A$. Let $B \subseteq A$.

***Case 1:** $a(n) \in B$**:** Let $B_n := B \setminus \{a(n)\}$ and $A_n := A \setminus \{a(n)\}$.*

Then the restriction of $a(\cdot)$ to $[n-1]$ is a bijection $[n-1] \xrightarrow{\sim} A_n$ according to B/G prop.13.2.

As card$(A_n) = n-1$ and $B_n \subseteq A_n$ it follows from the induction assumption that B_n is finite: there exists $m \in \mathbb{N}$ and a bijection $b(\cdot) : [m] \xrightarrow{\sim} B_n$.

We now extend $b(\cdot)$ to $[m+1]$ by defining $b(m+1) := a(n)$. It follows that this extension remains injective and it is also surjective if we choose as codomain $B_n \cup \{a(n)\} = B$.

It follows that B is finite.

***Case 2:** $a(n) \notin B$**:** We pick an arbitrary $b \in B$. Let $j := a^{-1}(b)$. Clearly $j \in [n]$.*

Now we modify the mapping $a(\cdot)$ by switching the function values for j and n. We obtain another bijection $f : [n] \xrightarrow{\sim} A$ (see lemma 7.1 on p. 107) for which $f(n) = a(j) = b \in B$.

We now can apply what was proved in case 1 and obtain that B is finite. ■

Proposition 7.3 (B/G Cor.13.16, p.122)**.** *$\mathbb{N}^2$ is countable.*

Proof: [56] *Done by directly specifiying a bijection $F : \mathbb{N}^2 \xrightarrow{\sim} \mathbb{N}$.*

The following definitions and observations will make it easier to understand this proof. Let

$$s_0 := 0; \qquad s_n := \sum_{j=1}^{n} j = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

For the last equality see [1] B/G prop.4.11 on p.37. We note that

$$s_{n-1} + n = s_n, \tag{7.5}$$

hence

$$A_n := \{j \in \mathbb{N} : s_{n-1} < j \leqq s_n\} = \{s_{n-1}+1, s_{n-1}+2, \dots, s_{n-1}+n\} \quad (n \in \mathbb{N}) \tag{7.6}$$

*is a partition of $\mathbb{N}$ (see example 5.3.**d** on p.92).*

For $n \in \mathbb{N} \setminus \{1\} = \{2, 3, 4, \dots\}$ let

$$D_n := \{(i,j) \in \mathbb{N}^2 : i + j = n\}$$

[56] Understanding this proof is not very important and you will understand the essence of it if you read instead the subsequent remark 7.1.

be the set of all pairs of natural numbers whose sum equals n*. Clearly,* $(D_{n+1})_{n\geqq 2}$ *is a partition of* $\mathbb{N}^2$*.*

For $n \in \mathbb{N}$ *let us look at the mapping*

$$f_n : A_n \to D_{n+1}, \qquad s_{n-1}+k \;\mapsto\; f_n(s_{n-1}+k) := (n+1-k, k). \;\; (k \in \mathbb{N}, 1 \leqq k \leqq n) \tag{7.7}$$

We see from the second equality in (7.6) *that the argument values* $s_{n-1}+k$ $(1 \leqq k \leqq n)$ *in fact coincide with the domain* A_n *and it follows that* (7.7) *indeed defines a function* $A_n \to \mathbb{N}^2$*.*

It is immediate that $D_{n+1} = \{(n+1-k, k) : 1 \leqq k \leqq n\}$*. We conclude that* $f_n(A_n)) = D_{n+1}$ *and we have proved surjectivity of* f_n*.*

Finally we observe that if $1 \leqq k, k' \leqq n$ *and* $k \neq k'$ *then*

$$f_n(s_{n-1}+k) = (n+1-k, k) \neq (n+1-k', k') = f_n(s_{n-1}+k')$$

and this proves injectivity of f_n*.*

We now "glue together" the functions f_n *to obtain*

a function f *with domain* $\bigcup\Big[A_n : n \in \mathbb{N}\Big] = \mathbb{N}$ *and codomain* $\bigcup\Big[D_{n+1} : n \in \mathbb{N}\Big] = \mathbb{N}^2$ *as follows:*

$f(m) := f_n(m)$ *for* $m \in A_n$*, i.e.,*

$$f(s_{n-1}+k) \;=\; f_n(s_{n-1}+k) = (n+1-k, k) \;\text{ for } k \in A_n.$$

f *inherits injectivity from the individual* f_n *as the ranges* $f_n(A_n) = D_{n+1}$ *are mutually disjoint for different values of* n *and* f *inherits surjectivity from the* f_n *as*

$$f(\mathbb{N}) \;=\; \bigcup\Big[f_n(A_n) : n \in \mathbb{N}\Big] = \mathbb{N}^2 \;=\; \bigcup\Big[f(A_n) : n \in \mathbb{N}\Big] = \mathbb{N}^2$$

To summarize, we have proved that f *is a bijective mapping between* $\mathbb{N}$ *and* $\mathbb{N}^2$ *and this proves that* $\mathbb{N}^2$ *is countable.* ■

Remark 7.1. The following will help to visualize the proof just given. We think of $\mathbb{N}^2$ as a matrix with "infinitely many rows and columns"

$$(1,1)\ (1,2)\ (1,3)\ \dots \tag{7.8}$$
$$(2,1)\ (2,2)\ (2,3)\ \dots \tag{7.9}$$
$$(3,1)\ (3,2)\ (3,3)\ \dots \tag{7.10}$$

We reorganize this matrix into an ordinary sequence $(f(j))_{j\in\mathbb{N}}$ as follows:

$$f(1) = f_1(1) = (1,1), \tag{7.11}$$
$$f(2) = f_2(2) = (1,2),\ f(3) = f_2(3) = (2,1), \tag{7.12}$$
$$f(4) = f_3(4) = (1,3),\ f(5) = f_3(5) = (2,2),\ f(6) = f_3(6) = (3,1), \tag{7.13}$$
$$f(7) = f_4(7) = (1,4),\ f(8) = f_4(8) = (2,3),\ f(9) = f_4(9) = (3,2), f(10) = f_4(10) = (4,1), \tag{7.14}$$
$$\dots \tag{7.15}$$

In other words, we traverse first D_2, then D_3, then D_4, ... starting for each D_n at the upper right $(1, n-1)$ and ending at the lower left $(n-1, 1)$).

Lemma 7.3. *Author's note: DELETED: This lemma is covered by prop.5.2 (Rewrite unions as disjoint unions) on p.93.*

The following proposition was proved informally before (see thm.5.3 on p.99)

Theorem 7.1 (B/G prop.13.6: Countable unions of countable sets)**.** *The union of countably many countable sets is countable.*

Proof: Let the sets

$$A_1,\ A_2,\ A_3,\ \dots \ \text{be countable and let}\ A := \bigcup_{n\in\mathbb{N}} A_i.$$

We assume that at least one of those sets is not empty: otherwise their union is empty, hence finite, hence countable and we are done.

We may assume, on account of prop.5.2 that the sets are mutually disjoint, i.e., any two different sets A_i, A_j have intersection $A_i \cap A_j = \emptyset$ (see definition (2.4) on p.12).

***A.** As each of the non-empty A_i is countable, either A is finite and we have an $N_i \in \mathbb{N}$ and a bijective mapping $a_i(\cdot) : A_i \xrightarrow{\sim} [N_i]$, or A_i is countably infinite and we have a bijective mapping $a_i(\cdot) : A_i \xrightarrow{\sim} \mathbb{N}$.*

We now define the mapping $f : A \to \mathbb{N}^2$ as follows: Let $a \in A$. As the A_j are disjoint there is a unique index i such that $a \in A_i$ and, as sets do not contain duplicates of their elements, there is a unique index j such that $a = a_i(j)$.

In other words, for any $a \in A$ there exists a unique pair $(i_a, j_a) \in \mathbb{N}^2$ such that $a = a_{i_a}(j_a)$ and the assignment $a \mapsto (i_a, j_a)$ defines an injective function $f : A \to \mathbb{N}^2$.

But then this same assignment gives us a bijective function $F : A \xrightarrow{\sim} f(A)$.

$f(A)$ is countable as a subset of the countable set $\mathbb{N}^2$ and this proves the theorem as any subset of a countable set is countable (see B/G prop.13.10). ■

Corollary 7.3. *Let the set X not be countable and let $A \subseteq X$ be countable. Then its complement $A^{\complement}$ is not countable.*

Proof: Left to the reader.

Definition 7.2 (algebraic numbers)**.** Let $x \in \mathbb{R}$ be the root (zero) of a polynomial with integer coefficients. We call such x an **algebraic number** and we call any real number that is not algebraic a **transcendental number**

Proposition 7.4 (B/G Prop.13.21, p.125: All algebraic numbers are countable)**.** *All algebraic numbers are countable.*

Proof: Let P be the set of all integer polynomials and Z the set of zeroes for all such polynomials. Let

$$P_n := \{\text{polynomials } p(x) = \sum_{j=0}^{k} a_j x^j : a_j \in \mathbb{Z} \text{ and } -n \leqq a_j \leqq n\}. \tag{7.16}$$

Then P_n is finite and

(7.17) $$Z_n := \{x \in \mathbb{R} : p(x) = 0 \text{ for some } p \in P_n\}$$

also is finite as a polynomial of degree n has at most n zeroes.

But Z is the countable union of the sets Z_n. It follows that Z is countable. ■

Here are some trivial consequences of the fact that $\mathbb{R}$ is not countable (see thm. 5.4, p.5.4 and B/G Thm.13.22).

Proposition 7.5. *All transcendental numbers are not countable.*

Proof: the uncountable real numbers are the disjoint union of the countable algebraic numbers with the transcendentals. The assertion follows from cor.7.3. ■

7.2 Addenda to chapter 7 (Some miscellaneous topics)

7.2.1 More on Cardinality

Theorem 7.2 (Finite Cartesians of countable sets are countable)**.** *The Cartesian product of finitely many countable sets is countable.*

Proof by induction: Let $X := X_1 \times \cdots \times X_n$ We may assume that none of the factor sets X_j is empty: Otherwise the Cartesian is empty too and there is nothing to prove.

We could choose $k = 1$ for which the proof is a triviality as the base case, but it is more instructive to choose $k = 2$ instead.

So let X_1, X_2 be two nonempty countable sets. We now prove that $X_1 \times X_2$ is countable.

For fixed $x_1 \in X_1$ the function $F_2 : X_2 \to \{x_1\} \times X_2; \quad x_2 \mapsto (x_1, x_2)$ is bijective because it has as an inverse the function $G_2 : \{x_1\} \times X_2 \to X_2; \quad (x_1, x_2) \mapsto x_2$. It follows that $\{x_1\} \times X_2$ is countable.

But then $X_1 \times X_2 = \bigcup_{x \in X_1} \{x_1\} \times X_2$ is countable according to thm.7.1 on p.112. We have proved the base case.

Our induction assumption is that $X_1 \times \cdots \times X_k$ is countable. We must prove that $X_1 \times \cdots \times X_{k+1}$ is countable. We can "identify"

(7.18) $$X_1 \times \cdots \times X_{k+1} = (X_1 \times \cdots \times X_k) \times X_{k+1}$$

by means of the bijection $(x_1, \dots, x_n, x_{n+1}) \mapsto \big((x_1, \dots, x_n), x_{n+1}\big)$. According to the induction assumption the set $X_1 \times \cdots \times X_k$ is countable.

The proof for the base case shows that $X_1 \times \cdots \times X_{k+1}$ as the Cartesian product of the two sets $X_1 \times \cdots \times X_k$ and X_{k+1} is countable. This finishes the proof of the induction step. ■

Corollary 7.4. *Let $n \in \mathbb{N}$. The sets $\mathbb{Q}$ and $\mathbb{Z}$ are countable.*

Proof: This follows from the preceding theorem because the sets $\mathbb{Q}$ and $\mathbb{Z}$ are countable. ■

8 Real functions (Understand this!)

8.1 Operations on real functions

Definition 8.1 (real functions)**.** Let X be an arbitrary, nonempty set. If the codomain Y of a mapping

$$f(\cdot): \quad X \longrightarrow Y \qquad x \longmapsto f(x)$$

is a subset of $\mathbb{R}$, then we call $f(\cdot)$ a **real function** or **real valued function**.

Remember that this definition does not exclude the case $Y = \mathbb{R}$ because $Y \subseteq \mathbb{R}$ is in particular true if both sets are equal.

Real functions are a pleasure to work with because, given any fixed argument x_0, the object $f(x_0)$ is just an ordinary number. In particular you can add, subtract, multiply and divide real functions. Of course, division by zero is not allowed:

Definition 8.2 (Operations on real functions)**.** Let X an arbitrary non-empty set.
Given are two real functions $f(\cdot), g(\cdot) : X \to (R)$ and a real number α. The **sum** $f + g$, **difference** $f - g$, **product** fg or $f \cdot g$, **quotient** f/g or $\frac{f}{g}$, and **scalar product** αf are defined by doing the operation in question with the numbers $f(x)$ and $g(x)$ for each $x \in X$. In other words these items are defined by the following equations:

$$\begin{aligned}
(f+g)(x) &:= f(x) + g(x), \\
(f-g)(x) &:= f(x) - g(x), \\
(fg)(x) &:= f(x)g(x), \\
(f/g)(x) &:= f(x)/g(x) \qquad \text{for all } x \in X \text{ where } g(x) \neq 0, \\
(\alpha f)(x) &:= \alpha \cdot g(x).
\end{aligned} \tag{8.1}$$

Before we list some basic properties of addition and scalar multiplication of functions (the operations that interest us the most), let us have a quick look at constant functions.

Definition 8.3 (Constant functions)**.** Let a be an ordinary real number. You can think of a as a function from any non-empty set X to $\mathbb{R}$ as follows:

$$a(\cdot): \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto a$$

In other words, the function $a(\cdot)$ assigns to each $x \in X$ one and the same value a. We call such a function a **constant real function**.

The most important constant function is the **zero function** $0(\cdot)$ which maps any $x \in X$ to the number zero. We usually just write 0 for this function unless doing so would confuse the reader. Note that scalar multiplication $(\alpha f)(x) = \alpha \cdot g(x)$ is a special case of multiplying two functions $(gf)(x) = g(x)f(x)$: Let $g(x) = \alpha$ for all $x \in X$ (constant function α).

The concept of a constant function makes sense for an arbitrary, nonempty codomain Y (i.e., Y need not be a set of real numbers):

We call any mapping f from X to Y a **constant function**. if its image $f(X) \subseteq Y$ is a singleton, i.e, it consists of exactly one element.

One last definition before we finally get so see some examples:

Definition 8.4 (Negative function)**.** Let X be an arbitrary, non-empty set and let

$$f(\cdot): \ X \longrightarrow \mathbb{R} \qquad x \longmapsto f(x)$$

be a real function on X. The function

$$-f(\cdot): \ X \longrightarrow \mathbb{R} \qquad x \longmapsto -f(x)$$

which assigns to each $x \in X$ the value $-f(x)$ is called **negative** f or **minus** f. Sometimes we write $-f$ rather than $-f(\cdot)$.

All those last definitions about sums, products, scalar products, ... of real functions are very easy to understand if you remember that, for any fixed $x \in X$, you just deal with ordinary numbers!

Example 8.1 (Arithmetic operations on real functions)**.** // For simplicity, we set $X := \mathbb{R}_+ = \{x \in \mathbb{R} : x \geqq 0\}$. Let

$$f(\cdot): \ \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto (x-1)(x+1)$$
$$g(\cdot): \ \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto x-1$$
$$h(\cdot): \ \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto x+1$$

Then

$$\begin{aligned}
(f+h)(x) &= (x-1)(x+1)+x+1 &&= x^2-1+x+1 = x(x+1) && \forall x \in \mathbb{R}_+\\
(f-g)(x) &= (x-1)(x+1)-(x-1) &&= x^2-1-x+1 = x(x-1) && \forall x \in \mathbb{R}_+\\
(gh)(x) &= (x-1)(x+1) &&= f(x) && \forall x \in \mathbb{R}_+\\
(f/h)(x) &= (x-1)(x+1)/(x+1) &&= x-1 = g(x) && \forall x \in \mathbb{R}_+\\
(f/g)(x) &= (x-1)(x+1)/(x-1) &&= x+1 = h(x) && \forall x \in \mathbb{R}_+ \setminus \{1\}
\end{aligned}$$

It is really, really important to understand that $f/g(\cdot)$ and $h(\cdot)$ are **not the same functions** on $\mathbb{R}_+$. Matter of fact, $f/g(\cdot)$ is not defined for all $x \in \mathbb{R}_+$ because for $x = 1$ you obtain $\frac{(1-1)(1+1)}{1-1} = 0/0$. The domain of f/g is different from that of h and both functions thus are different.

8.2 Maxima, suprema, limsup ... (Study this!)

Definition 8.5 (Upper and lower bounds, maxima and minima)**.** [57] Let $A \subseteq \mathbb{R}$. Let $l, u \in \mathbb{R}$. We call l a **lower bound** of A if $l \leqq a$ for all $a \in A$. We call u an **upper bound** of A if $u \geqq a$ for all $a \in A$.

We call A **bounded above** if this set has an upper bound and we call A **bounded below** if A has a lower bound. We call A **bounded** if A is both bounded above and bounded below.

[57] The definitions here were previously given for $\mathbb{Z}$. See def.2.12 on p.18.

A **minimum** (min) of A is a lower bound l of A such that $l \in A$. A **maximum** (max) of A is an upper bound u of A such that $u \in A$.

The next proposition will show that min and max are unique if they exist. This makes it possible to write $\min(A)$ or $\min A$ for the minimum of A and $\max(A)$ or $\max A$ for the maximum of A.

Proposition 8.1. *Let $A \subseteq \mathbb{R}$. If A has a maximum then it is unique. If A has a minimum then it is unique.*

Proof for maxima: Let u_1 and u_2 be two maxima of A: both are upper bounds of A and both belong to A. As u_1 is an upper bound, it follows that $a \leqq u_1$ for all $a \in A$. Hence $u_2 \leqq u_1$. As u_2 is an upper bound, it follows that $u_1 \leqq u_2$ and we have equality $u_1 = u_2$. The proof for minima is similar. ■

Definition 8.6. Given $A \subseteq \mathbb{R}$ we define

$$\begin{aligned} A_{\ell ow b} &:= \{l \in \mathbb{R} : l \text{ is lower bound of } A\} \\ A_{upp b} &:= \{u \in \mathbb{R} : u \text{ is upper bound of } A\}. \end{aligned} \tag{8.2}$$

We say that A is **bounded above** if $A_{upp b} \neq \emptyset$ and we say that A is **bounded below** if $A_{\ell ow b} \neq \emptyset$.

Axiom 8.1. (see [1] B/G axiom 8.52, p.83).

> **Completeness axiom for $\mathbb{R}$:** Let $A \subseteq \mathbb{R}$. If its set of lower bounds $A_{upp b}$ is not empty then $A_{upp b}$ has a minimum.

The above has the status of an axiom due to the fact that the real numbers usually are given axiomatically as an "archimedian ordered field" which satisfies the completeness axiom just stated.

Remark 8.1. $A_{\ell ow b}$ and/or $A_{upp b}$ may be empty: $A = \mathbb{R}$, $A = \mathbb{R}_{>0}$, $A = \mathbb{R}_{<0}$.

Definition 8.7. Let $A \subseteq \mathbb{R}$. If $A_{upp b}$ is not empty then $\min(A_{upp b})$ exists by axiom 8.1 and it is unique by prop. 8.1. We write $\sup(A)$ or $\text{l.u.b.}(A)$ for $\min(A_{upp b})$ and call it the **supremum** or **least upper bound** of A.

We will see in cor.8.1 that, if $A_{\ell ow b}$ is not empty, then $\max(A_{\ell ow b})$ exists and it is unique by prop. 8.1. We write $\inf(A)$ or $\text{g.l.b.}(A)$ for $\max(A_{\ell ow b})$ and call it the **infimum** or **greatest lower bound** of A.

Proposition 8.2 (Duality of upper and lower bounds, min and max, inf and sup)**.** *Let $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$. Then the following is true for $-x$ and $-A = \{-y : y \in A\}$:*

$$\begin{aligned} &-x \text{ is a lower bound of } A \iff x \text{ is an upper bound of -}A \quad \text{and vice versa}, \\ &-x \in A_{upp b} \iff x \in (-A)_{\ell ow b} \quad \text{and vice versa}, \\ &-x = \sup(A) \iff x = \inf(-A) \quad \text{and vice versa}, \\ &-x = \max(A) \iff x = \min(-A) \quad \text{and vice versa}. \end{aligned} \tag{8.3}$$

Proof: A simple consequence of

$$-x \leq y \iff x \geqq -y \quad \text{and} \quad -x \geqq y \iff x \leqq -y. \qquad ■$$

Corollary 8.1. *Let $A \subseteq \mathbb{R}$. If A has lower bounds then* $\inf(A)$ *exists.*

Proof: According to the duality proposition prop.8.2, if A has lower bounds then $(-A)$ has upper bounds. It follows from the completeness axiom that $\sup(-A)$ *exists. We apply once more prop.8.2 to prove that* $\inf(A)$ *exists:* $\inf(A) = \sup(-A)$.

Here are some examples. We define ***for all three*** *of them $f(x) := -x$ and $g(x) := x$.*

Example 8.2 (Example a: Maximum exists)**.** Let $X_1 := \{t \in \mathbb{R} : 0 \leqq t \leqq 1\}$.
For each $x \in X_1$ we have $|f(x) - g(x)| = g(x) - f(x) = 2x$ and the biggest possible such difference is $g(1) - f(1) = 2$, . So $\max(X_1)$ exists and equals $\max(X_1) = 2$.

Example 8.3 (Example b: Supremum is finite)**.** Let $X_2 := \{t \in \mathbb{R} : 0 \leqq t < 1\}$, i.e., we now exclude the right end point 1 at which the maximum difference was attained. For each $x \in X$ we have

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

and the biggest possible such difference is certainly bigger than

$$g(0.9999999999) - f(0.9999999999) = 1.9999999998.$$

If you keep adding $5,000$ 9s to the right of the argument x, then you get the same amount of 9s inserted into the result $2x$, so $2x$ comes closer than anything you can imagine to the number 2, without actually being allowed to reach it.

The supremum is still considered in a case like this to be 2. This precisely is the difference in behavior between the supremum $s := \sup(A)$ and the **maximum** $m := \max(A)$ of a set $A \subseteq \mathbb{R}$ of real numbers: For the maximum there must actually be at least one element $a \in A$ so that $a = \max(A)$.

For the supremum it is sufficient that there is a sequence $a_1 \leqq a_2 \leqq \dots$ which approximates s from below in the sense that the difference $s - a_n$ "drops down to zero" as n approaches infinity. We will not be more exact than this because doing so would require us to delve into the concepts of convergence and contact points.

Example 8.4 (Example c: Supremum is infinite)**.** Let $X_3 := \mathbb{R}_{\geq 0} = \{t \in \mathbb{R} : 0 \leqq t\}$. For each $x \in X_3$ we have again $|f(x) - g(x)| = g(x) - f(x) = 2x$. But there is no more limit to the right for the values of x. The difference $2x$ will exceed all bounds and that means that the only reasonable value for $\sup\{|f(x) - g(x)| \,:\, x \in X_3\}$ is $+\infty$.

As in case b above, the $\max$ does not exist because there is no $x_0 \in X_3$ such that $|f(x_0) - g(x_0)|$ attains the highest possible value among all $x \in X_3$.

You should understand that even though $\sup(A)$ as best approximation of the largest value of $A \subseteq \mathbb{R}$ is allowed to take the "value" $+\infty$ or $-\infty$ this cannot be allowed for $\max(A)$.

How so? The infinity values are not real numbers, but, by definition of the maximum, if $\alpha := \max(A)$ exists, then $\alpha \in A$. In particular, the max must be a real number.

That last example motivates the following definition.

Definition 8.8 (Supremum and Infimum of unbounded and empty sets)**.** If A is not bounded from above then we define

(8.4)
$$\sup A = \infty$$

If A is not bounded from below then we define

(8.5)
$$\inf A = -\infty$$

Finally we define

(8.6)
$$\sup \emptyset = -\infty, \qquad \inf \emptyset = +\infty$$

Note that we have defined infimum and supremum for any kind of set: empty or not, bounded above or below or not. We use those definitions to define infimum and supremum for functions, sequences and indexed families.

Definition 8.9 (supremum and infimum of functions)**.** Let X be an arbitrary set, $A \subseteq X$ a subset of X, $f : X \to \mathbb{R}$ a real function on X. Look at the set $f(A) = \{f(x) : x \in A\}$, i.e., the image of A under $f(\cdot)$.

The **supremum of** $f(\cdot)$ **on** A is then defined as

(8.7)
$$\sup_A f \;:=\; \sup_{x \in A} f(x) \;:=\; \sup\left(f(A)\right)$$

The **infimum of** $f(\cdot)$ **on** A is then defined as

(8.8)
$$\inf_A f \;:=\; \inf_{x \in A} f(x) \;:=\; \inf\left(f(A)\right)$$

Definition 8.10 (supremum and infimum of families)**.** Let $(x_i)_{i \in I}$ be an indexed family of real numbers x_i.

The **supremum of** $(x_i)_{i \in I}$ is then defined as

(8.9)
$$\sup\,(x_i) \;:=\; \sup_i\,(x_i) \;:=\; \sup\,(x_i)_i \;:=\; \sup\,(x_i)_{i \in I} \;:=\; \sup_{i \in I} x_i \;:=\; \sup\,\{x_i : i \in I\}$$

The **infimum of** $(x_i)_{i \in I}$ is then defined as

(8.10)
$$\inf\,(x_i) \;:=\; \inf_i\,(x_i) \;:=\; \inf\,(x_i)_i \;:=\; \inf\,(x_i)_{i \in I} \;:=\; \inf_{i \in I} x_i \;:=\; \inf\,\{x_i : i \in I\}$$

The definition above for families is consistent with the one given earlier for sequences (the special case of $I = \{k \in \mathbb{Z} : k \geq k_0$ for some $k_0 \in \mathbb{Z})$. We repeat it here for your convenience.

Definition 8.11 (supremum and infimum of sequences)**.** Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers x_n. The **supremum of** $(x_n)_{n\in\mathbb{N}}$ is then defined as

$$\sup(x_n) := \sup(x_n)_{n\in\mathbb{N}} := \sup_{n\in\mathbb{N}} x_n = \sup\{x_n : n\in\mathbb{N}\} \tag{8.11}$$

The **infimum of** $(x_n)_{n\in\mathbb{N}}$ is then defined as

$$\inf(x_n) := \inf(x_n)_{n\in\mathbb{N}} := \inf_{n\in\mathbb{N}} x_n = \inf\{x_n : n\in\mathbb{N}\} \tag{8.12}$$

We note that the "duality principle" for min and max, sup and inf is true in all cases above: You flip the sign of the items you examine and the sup/max of one becomes the inf/min of the other and vice versa.

Proposition 8.3. *X be a nonempty set and $\varphi, \psi : X \to \mathbb{R}$ be two real valued functions on X. Let $A \subseteq X$. Then*

$$\sup\{\varphi(x)+\psi(x) : x\in A\} \leqq \sup\{\varphi(y) : y\in A\} + \sup\{\psi(z) : z\in A\}, \tag{8.13}$$

$$\inf\{\varphi(x)+\psi(x) : x\in A\} \geqq \inf\{\varphi(y) : y\in A\} + \inf\{\psi(z) : z\in A\}. \tag{8.14}$$

Proof:
We only prove (8.13)*. The proof of* (8.14) *is similar.* [58]

Let $U := \{\varphi(x)+\psi(x) : x\in A\}$, $V := \{\varphi(y) : y\in A\}$, $W := \{\psi(z) : z\in A\}$. *Let* $x\in A$.

Then $\sup(V)$ *is an upper bound of* $\varphi(x)$ *and* $\sup(W)$ *is an upper bound of* $\psi(x)$,

hence $\sup(V)+\sup(W) \geqq \varphi(x)+\psi(x)$.

This is true for all $x\in A$, *hence* $\sup(V)+\sup(W)$ *is an upper bound of* U.

It follows that $\sup(V)+\sup(W)$ *dominates the least upper bound* $\sup(U)$ *of* U *and this proves* (8.13). ■

Definition 8.12 (Tail sets of a sequence)**.** Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Let

$$T_n := \{x_j : j \geqq n\} = \{x_n, x_{n+1}, x_{n+2}, x_{n+3}, \dots\} \tag{8.15}$$

be what remains in the sequence after we discard the first $n-1$ elements. We call $(T_n)_{n\in\mathbb{N}}$ the **tail set** for n of the given sequence $(x_k)_{k\in\mathbb{N}}$.

Remark 8.2. Some simple properties of tail sets:

a. We deal with sets and not with sequences T_n: If, e.g., $x_n = (-1)^n$ then each $T_n = \{-1, 1\}$ only contains two items and not infinitely many.

b. The tail set sequence $(T_n)_{n\in\mathbb{N}}$ is "decreasing": If $m < n$ then $T_m \supseteq T_n$.

We use the following result without a proof as it requires the definition of limits of sequences of real numbers. This document is written for an audience that knows calculus. [59] You should know enough about limits to understand the following:

[58] (8.14)can also be deduced from (8.13) and the fact that $\inf\{\varphi(u) : u\in A\} = -\sup\{-\varphi(v) : v\in A\}$.
[59] on the level of (see [11] Stewart, J: Single Variable Calculus)

Let x_n be a sequence of real numbers that is non-decreasing, i.e., $x_n \leqq x_{n+1}$ for all n (see def. 16.1 on p.246), and bounded above. Then $\lim\limits_{n\to\infty} x_n$ exists and coincides with $\sup\{x_n : n \in \mathbb{N}\}$ (see the proof of [1] B/G thm 10.19, p.101).
Further, if y_n is a sequence of real numbers that is non-increasing, i.e., $y_n \geqq y_{n+1}$ for all n, and bounded below, the analogous result is that $\lim\limits_{n\to\infty} y_n$ exists and coincides with $\inf\{y_n : n \in \mathbb{N}\}$.

We do not use the above except to justify the following notation.

(8.16)
$$\begin{aligned}\lim_{n\to\infty}\big(\sup\{x_j : j\in\mathbb{N}, j\geqq n\}\big) &:= \lim_{n\to\infty}\big(\sup(T_n)\big) := \inf\big(\{\sup(T_n) : n\in\mathbb{N}\}\big),\\ \lim_{n\to\infty}\big(\inf\{x_j : j\in\mathbb{N}, j\geqq n\}\big) &:= \lim_{n\to\infty}\big(\inf(T_n)\big) := \sup\big(\{\inf(T_n) : n\in\mathbb{N}\}\big).\end{aligned}$$

An expression like $\sup\{x_j : j\in\mathbb{N}, j\geqq n\}$ can be written more compactly as $\sup\limits_{j\in\mathbb{N}, j\geqq n}\{x_j\}$. Moreover, when dealing with sequences (x_n), it is understood in most cases that $n\in\mathbb{N}$ or $n\in\mathbb{Z}_{\geqq 0}$ and the last expression simplifies to $\sup\limits_{j\geqq n}\{x_j\}$. This can also be written as $\sup\limits_{j\geqq n}(x_j)$ or $\sup\limits_{j\geqq n} x_j$.

In other words, (8.16) becomes

(8.17)
$$\begin{aligned}\inf_{n\in\mathbb{N}}\big(\sup_{j\geqq n} x_j\big) &= \inf\big(\{\sup(T_n) : n\in\mathbb{N}\}\big) = \lim_{n\to\infty}\big(\sup(T_n)\big) = \lim_{n\to\infty}\big(\sup_{j\geqq n} x_j\big),\\ \sup_{n\in\mathbb{N}}\big(\inf_{j\geqq n} x_j\big) &= \sup\big(\{\inf(T_n) : n\in\mathbb{N}\}\big) = \lim_{n\to\infty}\big(\inf(T_n)\big) = \lim_{n\to\infty}\big(\inf_{j\geqq n} x_j\big).\end{aligned}$$

The above justifies the following definition:

Definition 8.13. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and let $T_n = \{x_j : j\in\mathbb{R}, j\geqq n\}$ be the tail set for x_n. Assume that T_n is bounded above for some $n_0\in\mathbb{N}$ (and hence for all $n\geqq n_0$). We call

$$\limsup_{n\to\infty} x_j := \lim_{n\to\infty}\big(\sup_{j\geqq n} x_j\big) = \inf_{n\in\mathbb{N}}\big(\sup_{j\geqq n} x_j\big) = \inf_{n\in\mathbb{N}}\big(\sup(T_n)\big)$$

the **lim sup** or **limit superior** of the sequence (x_n).

If, for each n, T_n is not bounded above then we say $\limsup\limits_{n\to\infty} x_j = \infty$.

Assume that T_n is bounded below for some n_0 (and hence for all $n\geqq n_0$). We call

$$\liminf_{n\to\infty} x_j := \lim_{n\to\infty}\big(\inf_{j\geqq n} x_j\big) = \sup_{n\in\mathbb{N}}\big(\inf_{j\geqq n} x_j\big) = \sup_{n\in\mathbb{N}}\big(\inf(T_n)\big)$$

the **lim inf** or **limit inferior** of the sequence (x_n).

If, for each n, T_n is not bounded below then we say $\liminf\limits_{n\to\infty} x_j = -\infty$.

Proposition 8.4. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded above with tail sets T_n.*

A. *Let*

$$(8.18)\qquad \begin{aligned} \mathscr{U} &:= \{y \in \mathbb{R} : T_n \cap [y,\infty[\neq \emptyset \text{ for all } n \in \mathbb{N}\},\\ \mathscr{U}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \geqq y\},\\ \mathscr{U}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \cdots \in \mathbb{N} \text{ such that } x_{n_j} \geqq y \text{ for all } j \in \mathbb{N}\},\\ \mathscr{U}_3 &:= \{y \in \mathbb{R} : x_n \geqq y \text{ for infinitely many } n \in \mathbb{N}\}. \end{aligned}$$

Then $\mathscr{U} = \mathscr{U}_1 = \mathscr{U}_2 = \mathscr{U}_3$.

B. *There exists* $z = z(\mathscr{U}) \in \mathbb{R}$ *such that* $\mathscr{U}$ *is either an interval* $]-\infty, z]$ *or an interval* $]-\infty, z[$.

C. *Let* $u := \sup(\mathscr{U})$. *Then* $u = z = z(\mathscr{U})$ *as defined in part B. Further,* u *is the only real number such that*

C1. (8.19) $$u - \varepsilon \in \mathscr{U} \quad \text{and} \quad u + \varepsilon \notin \mathscr{U} \quad \text{for all } \varepsilon > 0.$$

C2. *There exists a subsequence* $(n_j)_{j\in\mathbb{N}}$ *of integers such that* $u = \lim\limits_{j\to\infty} x_{n_j}$ *and* u *is the largest real number for which such a subsequence exists.*

*Proof of **A**:*

A.1 - $\mathscr{U} = \mathscr{U}_1$: *This equality is valid by definition of tailsets of a sequence:*

$$x \in T_n \Leftrightarrow x = x_j \text{ for some } j \geqq n \Leftrightarrow x = x_{n+k} \text{ for some } k \in \mathbb{Z}_{\geqq 0}$$

from which it follows that $x \in T_n \cap [y,\infty[\Leftrightarrow x = x_{n+k}$ *for some* $k \geqq 0$ *and* $x_{n+k} \geqq y$.

A.2 - $\mathscr{U}_1 \subseteq \mathscr{U}_2$:

Let $y \in \mathscr{U}_1$ *and* $n \in \mathbb{N}$. *We prove the existence of* $(n_j)_j$ *by induction on* j.

Base case $j = 1$: *As* $T_2 \cap [y,\infty[\neq \emptyset$ *there is some* $x \in T_2$ *such that* $y \leqq x < \infty$, *i.e.,* $x \geqq y$. *Because* $x \in T_2 = \{x_2, x_3, \dots\}$ *we have* $x = x_{n_1}$ *for some integer* $n_1 > 1$ *and we have proved the existence of* n_1.

Induction assumption: Assume that $n_1 < n_2 < \cdots < n_{j_0}$ *have already been picked.*

Induction step: Let $n = n_{j_0}$. *As* $y \in \mathscr{U}_1$ *there is* $k \in \mathbb{N}$ *such that* $x_{n_{j_0}+k} \geqq y$. *We set* $n_{j_0+1} := n_{j_0} + k$. *As this index is strictly larger than* n_{j_0}, *the induction step has been proved.*

A.3 - $\mathscr{U}_2 \subseteq \mathscr{U}_3$: *This is trivial: Let* $y \in \mathscr{U}_2$. *The strictly increasing subsequence* $n_1 < n_2 < n_3 < \cdots \in \mathbb{N}$ *constitutes the infinite set of indices that is required to grant* y *membership in* $\mathscr{U}_3$.

A.4 - $\mathscr{U}_3 \subseteq \mathscr{U}$: *Let* $y \in \mathscr{U}_3$. *Fix some* $n \in \mathbb{N}$.

Let $J = J(y) \subseteq \mathbb{N}$ *be the infinite set of indices* j *for which* $x_j \geqq y$. *At most finitely many of those* j *can be less than that given* n *and there must be (infinitely many)* $j \in J$ *such that* $j \geqq n$

Pick any one of those, say j'. *Then* $x_{j'} \in T_n$ *and* $x_{j'} \geqq y$. *It follows that* $y \in \mathscr{U}$

We have shown the following sequence of inclusions:

$$\mathscr{U} = \mathscr{U}_1 \subseteq \mathscr{U}_2 \subseteq \mathscr{U}_3 \subseteq \mathscr{U}$$

It follows that all four sets are equal and part A of the proposition has been proved.

*Proof of **B**: Let $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $y_2 \in \mathscr{U}$.*

It follows from $[y_2, \infty[\subseteq [y_1, \infty[$ that, because $T_n \cap [y_2, \infty[\neq \emptyset$ for all $n \in \mathbb{N}$, we must have $T_n \cap [y_1, \infty[\neq \emptyset$ for all $n \in \mathbb{N}$, i.e., $y_1 \in \mathscr{U}$.

But that means that $\mathscr{U}$ must be an interval of the form $]-\infty, z]$ or $]-\infty, z[$ for some $z \in \mathbb{R}$.

*Proof of **C**: Let $z = z(\mathscr{U})$ as defined in part B and $u := \sup(\mathscr{U})$.*

*Proof of **C.1 - (8.19) part 1**, $u - \varepsilon \in \mathscr{U}$:*

As $u - \varepsilon$ is smaller than the least upper bound u of $\mathscr{U}$, $u - \varepsilon$ is not an upper bound of $\mathscr{U}$. Hence there is $y > u - \varepsilon$ such that $y \in \mathscr{U}$. It follows from part B that $u - \varepsilon \in \mathscr{U}$.✓

*Proof of **C.1 - (8.19) part 2**, $u + \varepsilon \notin \mathscr{U}$:*

This is trivial as $u + \varepsilon > u = \sup(\mathscr{U})$ implies that $y \leqq u < u + \varepsilon$ for all $y \in \mathscr{U}$.

But then $y \neq u$ for all $y \in \mathscr{U}$, i.e., $u \notin \mathscr{U}$. This proves $u + \varepsilon \notin \mathscr{U}$.

*Proof of **C.2**: We construct by induction a sequence $n_1 < n_2 < \dots$ of natural numbers such that*

$$u - 1/j \leq x_{n_j} \leqq u + 1/j. \tag{8.20}$$

*Base case: We have proved as part of **C.1** that $x_n \geqq u + 1$ for at most finitely many indices n. Let K be the largest of those.*

As $u - 1 \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geqq u - 1$. Infinitely many of those n must exceed K. We pick one of them and that will be n_1. Clearly, n_1 satisfies (8.20) and this proves the base case.

Induction step: Let us now assume that $n_1 < n_2 < \cdots < n_k$ satisfying (8.20) have been constructed. $x_n \geqq u + 1/(k+1)$ is possible for at most finitely many indices n. Let K be the largest of those.

As $u - 1/(k+1) \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geqq u - 1/(k+1)$. Infinitely many of those n must exceed $\max(K, n_k)$. We pick one of them and that will be n_{k+1}. Clearly, n_{k+1} satisfies (8.20) and this finishes the proof by induction.

We now show that $\lim\limits_{j \to \infty} x_{n_j} = u$. Given $\varepsilon > 0$ there is $N = N(\varepsilon)$ such that $1/N < \varepsilon$. It follows from (8.20) that $|x_{n_j} - u| \leqq 1/j < 1/N < \varepsilon$ for all $j \geqq n$ and this proves that $x_{n_j} \to u$ as $j \to \infty$.

*We will be finished with the proof of **C.2** if we can show that if $w > u$ then there is no sequence $n_1 < n_2 < \dots$ such that $x_{n_j} \to w$ as $j \to \infty$.*

Let $\varepsilon := (w - u)/2$. According to (8.19), $u + \varepsilon \notin \mathscr{U}$. But then, by definition of $\mathscr{U}$, there is $n \in \mathbb{N}$ such that $T_n \cap [u + \varepsilon, \infty[= \emptyset$.

But $u + \varepsilon = w - \varepsilon$ and we have $T_n \cap [w - \varepsilon, \infty[= \emptyset$. This implies that $|w - x_j| \geqq \varepsilon$ for all $j \geqq n$ and that rules out the possibility of finding n_j such that $\lim\limits_{j \to \infty} x_{n_j} = w$. ■

Corollary 8.2. *As in prop.8.4, let $u := \sup(\mathscr{U})$. Then $\mathscr{U} =]-\infty, u]$ or $\mathscr{U} =]-\infty, u[$.*

Further, u is determined by the following property: For any $\varepsilon > 0$, $x_n > u - \varepsilon$ for infinitely many n and

$x_n > u + \varepsilon$ for at most finitely many n.

Proof: This follows from $\mathscr{U} = \mathscr{U}_3$ and parts B and C of prop.8.4. ■

When we form the sequence $y_n = -x_n$ then the roles of upper bounds and lower bounds, max and min, inf and sup will be reversed. Example: x is an upper bound for $\{x_j : j \geqq n$ if and only if $-x$ is a lower bound for $\{y_j : j \geqq n$.

The following "dual" version of prop. 8.4 is a direct consequence of the duality of upper/lower bounds, min/max, inf/sup proposition prop.8.2, p.116.

Proposition 8.5. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded below with tail sets T_n.*

***A.** Let*

(8.21)
$$\begin{aligned}\mathscr{L} &:= \{y \in \mathbb{R} : T_n \cap]-\infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\\ \mathscr{L}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \leqq y\},\\ \mathscr{L}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \text{ such that } x_{n_j} \leqq y \text{ for all } j \in \mathbb{N}\},\\ \mathscr{L}_3 &:= \{y \in \mathbb{R} : x_n \leqq y \text{ for infinitely many } n \in \mathbb{N}\}.\end{aligned}$$

Then $\mathscr{L} = \mathscr{L}_1 = \mathscr{L}_2 = \mathscr{L}_3$.

***B.** There exists $z = z(\mathscr{L}) \in \mathbb{R}$ such that $\mathscr{L}$ is either an interval $[z, \infty[$ or an interval $]z, \infty[$.*

***C.** Let $l := \inf(\mathscr{L})$. Then $l = z = z(\mathscr{L})$ as defined in part B. Further, l is the only real number such that*

- ***C1.*** (8.22) $$l + \varepsilon \in \mathscr{L} \quad \text{and} \quad l - \varepsilon \notin \mathscr{L}$$
- ***C2.*** *There exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $l = \lim\limits_{j\to\infty} x_{n_j}$ and l is the smallest real number for which such a subsequence exists.*

Proof: Let $y_n = -x_n$ and apply prop.8.4. ■

Proposition 8.6. *Let (x_n) be a bounded sequence of real numbers. As in prop. 8.4 and prop 8.5, let*

(8.23)
$$\begin{aligned} u &= \sup(\mathscr{U}) = \sup\{y \in \mathbb{R} : T_n \cap [y, \infty[\neq \emptyset \text{ for all } n \in \mathbb{N}\},\\ l &= \inf(\mathscr{L}) = \inf\{y \in \mathbb{R} : T_n \cap]-\infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\end{aligned}$$

Then

$$u = \limsup_{n\to\infty} x_j \quad \text{and} \quad l = \liminf_{n\to\infty} x_j.$$

Proof that $u = \limsup\limits_{n\to\infty} x_j$: Let

(8.24)
$$\beta_n := \sup_{j\geqq n} x_j, \quad \beta := \inf_n \beta_n = \limsup_{n\to\infty} x_n.$$

We will prove that β has the properties listed in prop.8.4.C that uniquely characterize u: For any $\varepsilon > 0$, we have

$$\beta - \varepsilon \in \mathscr{U} \quad \text{and} \quad \beta + \varepsilon \notin \mathscr{U}$$

Another way of saying this is that

$$b \in \mathscr{U} \text{ for } b < \beta \quad \text{and} \quad a \notin \mathscr{U} \text{ for } a > \beta. \tag{8.25}$$

We now prove the latter characterization.

Let $a \in \mathbb{R}, a > \beta = \inf\{\beta_n : n \in \mathbb{N}\}$. Then a is not a lower bound of the β_n: $\beta_{n_0} < a$ for some $n_0 \in \mathbb{N}$.

As the β_n are not increasing in n, this implies strict inequality $\beta_j < a$ for all $j \geqq n_0$. By definition, β_j is the least upper bound (hence an upper bound) of the tail set T_j. We conclude that $x_j < a$ for all $j \geqq n_0$.

From that we conclude that $T_n \cap [a, \infty[= \emptyset$ for all $j \geqq n_0$. It follows that $a \notin \mathscr{U}$.

Now let $b \in \mathbb{R}, b < \beta = g.l.b\{\beta_n : n \in \mathbb{N}\}$. As $\beta \leqq \beta_n$ we obtain $b < \beta_n$ for all n.

In other words, $b < \sup(T_n)$ for all n: It is possible to pick some $x_k \in T_n$ such that $b < x_k$.

But then $T_n \cap [b, \infty[\neq \emptyset$ for all n and we conclude that $b \in \mathscr{U}$.

We put everything together and see that β has the properties listed in (8.25). This finishes the proof that $u = \limsup\limits_{n\to\infty} x_j$. The proof that $l = \liminf\limits_{n\to\infty} x_j$ follows again by applying what has already been proved to the sequence $(-x_n)$. ■

We have collected everything to prove

Theorem 8.1 (Characterization of limsup and liminf). *Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}$. Then*

- ***a1.*** $\limsup\limits_{n\to\infty} x_n$ *is the largest of all real numbers x for which a sequence $n_1 < n_2 < \cdots \in \mathbb{N}$ can be found such that $x = \lim\limits_{j\to\infty} x_{n_j}$.*
- ***a2.*** $\limsup\limits_{n\to\infty} x_n$ *is the only real number u such that, for all $\varepsilon > 0$, the following is true:*
  $x_n > u + \varepsilon$ for at most finitely many n and $x_n > u - \varepsilon$ for infinitely many n.
- ***b1.*** $\liminf\limits_{n\to\infty} x_n$ *is the smallest of all real numbers x for which a sequence $n_1 < n_2 < \cdots \in \mathbb{N}$ can be found such that $x = \lim\limits_{j\to\infty} x_{n_j}$.*
- ***b2.*** $\liminf\limits_{n\to\infty} x_n$ *is the only real number l such that, for all $\varepsilon > 0$, the following is true:*
  $x_n < l - \varepsilon$ for at most finitely many n and $x_n < l + \varepsilon$ for infinitely many n.

*Proof: We know from prop.8.6 on p.123 that $\limsup\limits_{n\to\infty} x_n$ is the unique number u described in part **C** of prop.8.4, p.120:*

$$u - \varepsilon \in \mathscr{U} \quad \text{and} \quad u + \varepsilon \notin \mathscr{U} \quad \text{for all } \varepsilon > 0$$

and u is the largest real number for which there exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $u = \lim\limits_{j\to\infty} x_{n_j}$.

*$u-\varepsilon \in \mathscr{U} = \mathscr{U}_3$ (see part A of prop.8.6) means that there are infinitely many n such that $x_n \geqq u-\varepsilon$ and $u+\varepsilon \notin \mathscr{U} = \mathscr{U}_3$ means that there are at most finitely many n such that $x_n \geqq u+\varepsilon$. This proves **a1** and **a2**.*

*We also know from prop.8.6 that $\liminf\limits_{n\to\infty} x_n$ is the unique number l described in part **C** of prop.8.5, p.123: $l+\varepsilon \in \mathscr{L}$ and $l-\varepsilon \notin \mathscr{L}$ for all $\varepsilon > 0$ and l is the smallest real number for which there exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $u = \lim\limits_{j\to\infty} x_{n_j}$.*

*$l+\varepsilon \in \mathscr{L} = \mathscr{L}_3$ (see part **A** of prop.8.6) means that there are infinitely many n such that $x_n \leqq l+\varepsilon$ and $l-\varepsilon \notin \mathscr{L} = \mathscr{L}_3$ means that there are at most finitely many n such that $x_n \leqq l-\varepsilon$. This proves **b1** and **b2**.*
■

*Proof of thm.8.1 **without the use of prop.8.6, prop.8.4 and the dual propositions for the liminf**.*

Step 1:

*Let $\varepsilon > 0$. It follows from $\beta_n = \sup(T_n) = \sup\{x_j : j \geqq n\}$ and $\beta_n \searrow \beta = \limsup_n x_n$ that $\beta_n < \beta+\varepsilon$ for all $n \geqq N$ for a suitable $N = N(\varepsilon) \in \mathbb{N}$. But then $\beta+\varepsilon$ exceeds the upper bound β_N of T_N and follows that all of its elements, i.e., all x_n with $n \geqq N$, satisfy $x_n < \beta+\varepsilon$. Hence only some or all of the finitely many $x_1, x_2, \dots x_{N-1}$ can exceed $\beta+\varepsilon$. It follows that β satisfies the first half of **a1** of thm.8.1.*

Step 2: We create a subsequence $(x_{n_j})_j$ such that

$$\beta_{n_j} \ \geqq\ x_{n_j} \ >\ \beta_{n_j} - 1/j \tag{8.26}$$

for all $j \in \mathbb{N}$ as follows.

$\beta_1 = \sup(T_1)$ is the smallest upper bound for T_1, hence $\beta_1 - 1$ is not an upper bound and we can find some $k \in \mathbb{N}$ such that $\beta_1 \geqq x_k > \beta_1 - 1$. We set $n_1 := k$.

Having constructed $n_1 < n_2 < \dots < n_k$ such that $\beta_{n_j} \ \geqq\ x_{n_j} \ >\ \beta_{n_j} - 1/j$ for all $j \leqq k$ we now find $x_{n_{k+1}}$ with an index $n_{k+1} > n_k$ as follows.

$\beta_{n_k+1} - \frac{1}{k+1}$ is not an upper bound for T_{n_k+1}, hence there exists some $i \in \mathbb{N}$ such that x_{n_k+i} (which belongs to T_{n_k+1}) satisfies

$$x_{n_k+i} \ >\ \beta_{n_k+1} - \frac{1}{k+1}. \tag{8.27}$$

We set $n_{k+1} := n_k + i$. The sequence β_n non–increasing (i.e., decreasing) and it follows from $n_{k+1} = n_k + i \geqq n_k + 1$ that $\beta_{n_{k+1}} \leqq \beta_{n_k+1}$. But then (8.27) implies that $x_{n_{k+1}} > \beta_{n_{k+1}} - \frac{1}{k+1}$. We note that $x_{n_{k+1}} \leqq \beta_{n_{k+1}}$ because $x_{n_{k+1}} \in T_{n_{k+1}}$ and $\beta_{n_{k+1}} = \sup(T_{n_{k+1}})$ is an upper bound for all elements of $T_{n_{k+1}}$. Together with (8.27) we have

$$\beta_{n_{k+1}} \ \geqq\ x_{n_{k+1}} \ >\ \beta_{n_k+1} - \frac{1}{k+1}. \tag{8.28}$$

It follows that $x_{n_{k+1}}$ satisfies (8.26) and the the proof of step 1 is completed.

Step 3: The sequence x_{n_j} we constructed in step 2 converges to $\beta = \limsup_n x_n$. This is true because $\lim_k \beta_{n_k} = \beta$, $\lim_k \beta_{n_k} - 1/k = \lim_k \beta_{n_k} - \lim_k 1/k = \beta - 0 = \beta$ and x_{n_j} is "sandwiched" between two sequences which both converge to the same limit β.

It follows from step 1 that no subsequence of (x_n) *can converge to a number* u *bigger than* β*: Let* $\varepsilon := 1/2(u-\beta)$*. Then all but finitely many* x_j *satisfy* $x_j \leqq \beta+\varepsilon$*, hence* $x_j \leqq u-\varepsilon$ *and it follows that the distance* $d(x_j, u)$ *exceeds* ε *for* $j \geqq N$ *and no subsequence converging to* u *can be extracted. This proves* ***a1*** *of thm.8.1.*

*Step 4. We still must prove the missing half of thm.8.1.**a2**:* $x_n > \beta - \varepsilon$ *for infinitely many* n*.*

Let $\varepsilon > 0$*. and let* $j \in \mathbb{N}$ *be so big that* $1/j < \varepsilon$*. Let* x_{n_j} *be again the subsequence constructed in step 2. It follows from (8.26) and* $\beta_{n_j} \geqq \beta$ *and* $1/j < \varepsilon$ *that* $x_{n_j} > \beta - \varepsilon$*. This proves the missing half of thm.8.1.**a2**.*

Uniqueness of β*: If we have some* $v > \beta$ *tnen we set* $\varepsilon := (v-\beta)/3$*. Because* $v-\varepsilon > \beta+\varepsilon$*, at most finitely many* x_n *satisfy* $x_n > v - \varepsilon$*. It follows that* v *does not satisfy part 2 of thm.8.1.**a2**.*

Finally let $v < \beta$*. Let* $\varepsilon := (\beta - v)/3$*. Because* $\beta - \varepsilon > v + \varepsilon$*, infinitely many* x_n *satisfy* $x_n > v + \varepsilon$*. It follows that* v *does not satisfy part 1 of thm.8.1.**a2**. We have proved that* $\limsup_n x_n$ *is uniquely determined by the inequalities of thm.8.1.**a2** and we have shown both **a1** and **a2** of thm.8.1.*

*Parts **b1** and **b2** of thm.8.1 follow now easily from applying parts **a1** and **a2** to the sequence* $y_n := -x_n$*.* ■

Theorem 8.2 (Characterization of limits via limsup and liminf)**.** *Let* $(x_n)_{n\in\mathbb{N}}$ *be a bounded sequence in* $\mathbb{R}$*. Then* (x_n) *converges to a real number if and only if liminf and limsup for that sequence coincide and we have*

$$\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n. \tag{8.29}$$

Proof of "⇒": Let $L := \lim\limits_{n\to\infty} x_n$*. Let* $\varepsilon > 0$*. There is* $N = N(\varepsilon) \in \mathbb{N}$ *such that* $T_k \subseteq \,]\, L-\varepsilon, L+\varepsilon\, [$ *for all* $k \geqq N$*. But then*

$$L - \varepsilon \leqq \alpha_k := \inf(T_k) \leq \beta_k := \sup(T_k) \leqq L + \varepsilon \ \text{ for all } k \geqq N.$$

It follows from $T_j \subseteq T_k$ *for all* $j \geqq k$ *that*

$$\begin{aligned} L-\varepsilon &\leq \alpha_k \leqq \alpha_j \leqq \beta_j \leqq \beta_k \leqq L+\varepsilon, \quad \text{hence} \\ L-\varepsilon &\leq \lim_{k\to\infty} \alpha_k = \liminf_{k\to\infty} x_k \leqq \limsup_{k\to\infty} x_k = \lim_{k\to\infty} \beta_k \leqq L+\varepsilon. \end{aligned}$$

The equalities above result from prop.8.6. We have shown that, for any $\varepsilon > 0$*,* $\liminf\limits_{k\to\infty} x_k$ *and* $\limsup\limits_{k\to\infty} x_k$ *differ by at most* 2ε*, hence they are equal.*

Proof of "⇐": Let $L := \liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$*. Let* $\varepsilon > 0$*. We know from (8.19), p.121 and (8.22), p.123 that* $L + \varepsilon/2 \notin \mathscr{U}$ *and* $L - \varepsilon/2 \notin \mathscr{L}$ *But then there are at most finitely many* n *for which* x_n *has a distance from* L *which exceeds* $\varepsilon/2$*. Let* N *be the maximum of those* n*. It follows that* $|x_n - L| < \varepsilon$ *for all* $n > N$*, hence* $L = \lim\limits_{n\to\infty} x_n$*.* ■

8.3 Sequences of sets and indicator functions and their liminf and limsup

Let Ω *be a non-empty set and let* $f_n : \Omega \to \mathbb{R}$ *be a sequence of real-valued functions. Let* $\omega \in \Omega$*. Then* $\big(f_n(\omega)\big)_{n\in\mathbb{N}}$ *is a sequence of real numbers for which we can examine* $\liminf_n f_n(\omega)$ *and* $\limsup_n f_n(\omega)$*. We will look at those two expressions as functions of* ω*.*

Definition 8.14 (limsup and liminf of a sequence of real functions)**.** Let Ω be a non-empty set and let $f_n : \Omega \to \mathbb{R}$ be a sequence of real-valued functions such that $f_n(\omega)$ is bounded for all $\omega \in \Omega$. [60]

We define

$$\liminf_{n\to\infty} f_n \ :\ \Omega \to \mathbb{R} \quad \text{as follows: } \omega \mapsto \liminf_{n\to\infty} f_n(\omega), \tag{8.31}$$

$$\limsup_{n\to\infty} f_n \ :\ \Omega \to \mathbb{R} \quad \text{as follows: } \omega \mapsto \limsup_{n\to\infty} f_n(\omega). \tag{8.32}$$

Remark 8.3. We recall from thm.8.1 (Characterization of limsup and liminf) on p.124 that

$$\liminf_{n\to\infty} f_n(\omega) \ = \inf\{\alpha \in \mathbb{R} : \lim_{j\to\infty} f_{n_j}(\omega) \ = \ \alpha \text{ for some subsequence } n_1 < n_2 < \dots\} \tag{8.33}$$

$$\limsup_{n\to\infty} f_n(\omega) \ = \sup\{\beta \in \mathbb{R} : \lim_{j\to\infty} f_{n_j}(\omega) \ = \ \beta \text{ for some subsequence } n_1 < n_2 < \dots\} \tag{8.34}$$

We now characterize $\liminf_n f_n$ and $\limsup_n f_n$ for functions f_n such that $f_n(\omega)$ is either zero or one. We have seen in prop.5.3 on p.94 that any such function is the indicator function 1_A of the set

$$A \ := \ \{f = 1\} \ = \ f^{-1}(\{1\}) \ = \ \{\omega \in \Omega : f(\omega) = 1\} \ \subseteq \ \Omega.$$

Proposition 8.7 (liminf and limsup of binary functions)**.** *Let $\Omega \neq \emptyset$ and $f_n : \Omega \to \{0, 1\}$. Let $\omega \in \Omega$. Then either $\liminf_n f_n(\omega) \ = 1$ or $\liminf_n f_n(\omega) \ = 0$ and either $\limsup_n f_n(\omega) \ = 1$ or $\limsup_n f_n(\omega) \ = 0$. Further*

$$\liminf_{n\to\infty} f_n(\omega) \ = 1 \ \Leftrightarrow f_n(\omega) \ = 1 \textit{ \textbf{except for at most finitely many} } n \in \mathbb{N} \tag{8.35}$$

$$\limsup_{n\to\infty} f_n(\omega) \ = 1 \ \Leftrightarrow f_n(\omega) \ = 1 \textit{ for infinitely many } n \in \mathbb{N} \tag{8.36}$$

Proof: It follows from (8.33), (8.34) and $0 \leqq f_n(\omega) \leqq 1$ that $0 \leqq \liminf_n f_n(\omega) \leqq \limsup_n f_n(\omega) \leqq 1$.

We conclude from (8.33) that $\liminf_n f_n(\omega) = 0$ if a subsequence $n_1 < n_2 < \dots$ can be found such that $f_{n_j}(\omega) = 0$ for all j and that $\liminf_n f_n(\omega) = 1$ if no such subsequence exists, i.e., if $f_n(\omega) = 1$ for all except at most finitely many n. This proves not only (both directions(!) of) (8.35) but also that either $\liminf_n f_n(\omega) \ = 1$ or $\liminf_n f_n(\omega) \ = 0$

We conclude from (8.34) that $\limsup_n f_n(\omega) = 1$ if a subsequence $n_1 < n_2 < \dots$ can be found such that $f_{n_j}(\omega) = 1$ for all j and that $\limsup_n f_n(\omega) = 0$ if no such subsequence exists, i.e., if $f_n(\omega) = 0$ for

[60] In more advanced texts you will find the following

Definition 8.15 (Extended real functions)**.** The set $\overline{\mathbb{R}} \ := \ \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ is called the **extended real numbers line**. A mapping F whose codomain is a subset of $\overline{\mathbb{R}}$ is called an **extended real function**.

The above allows to define the functions $\liminf_n f_n$ and $\limsup_n f_n$ even if there are arguments ω for which $\liminf_n f_n(\omega)$ and/or $\limsup_n f_n(\omega)$ assumes one of the values $\pm\infty$. There are many issues with functions that allow some arguments to have infinite value (hint: if $F(x) = \infty$ and $F(y) = \infty$, what is $F(x) - F(y)$?)

We only list the following rule which might come unexpected to you:

$$0 \cdot \pm\infty \ = \ \pm\infty \cdot 0 \ = \ 0 \tag{8.30}$$

all except at most finitely many n. This proves not only (both directions(!) of) (8.36) but also that either $\limsup_n f_n(\omega) = 1$ *or* $\limsup_n f_n(\omega) = 0$. ■

We now look at indicator functions 1_{A_n} of a sequence of sets $A_n \subseteq \Omega$. For such a sequence we define

$$A_\star := \bigcup_{n\in\mathbb{N}} \bigcap_{j\geqq n} A_j, \qquad A^\star := \bigcap_{n\in\mathbb{N}} \bigcup_{j\geqq n} A_j. \tag{8.37}$$

Proposition 8.8. *Let $\omega \in \Omega$. Then*

$$\omega \in A_\star \Leftrightarrow \omega \in A_n \text{ for all except at most finitely many } n \in \mathbb{N}. \tag{8.38}$$

$$\omega \in A^\star \Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}, \tag{8.39}$$

a. *Proof that $\omega \in A_\star \Rightarrow \omega \in A_n$ for all except at most finitely many $n \in \mathbb{N}$:*
We will prove the contrapositive: Assume that there exists $1 \leqq n_1 < n_2 < \dots$ such that $\omega \notin A_{n_j}$ for all $j \in \mathbb{N}$. We must show that $\omega \notin A_\star$.

Let $k \in \mathbb{N}$. Then $k \leqq n_k$ (think!) and it follows from $\omega \notin A_{n_k}$ and $A_{n_k} \supseteq \bigcap\limits_{j\geqq n_k} A_j \supseteq \bigcap\limits_{j\geqq k} A_j$ that there is no $k \in \mathbb{N}$ such that $\omega \in \bigcap\limits_{j\geqq k} A_j$.

*But then $\omega \notin \bigcup\limits_{k} \bigcap\limits_{j\geqq k} A_j = A_\star$ and we are done with the proof of **a**.*

b. *Proof that $\omega \in A_n$ for all except at most finitely many $n \in \mathbb{N} \Rightarrow \omega \in A_\star$:*
By assumption there exists some $N \in \mathbb{N}$ such that $\omega \in A_n$ for all $n \geqq N$.

*It follows that $\omega \in \bigcap_{n\geqq N} A_n \subseteq \bigcup_{m\in\mathbb{N}} \bigcap_{n\geqq m} A_n = A_\star$ and **b** has been proved.*

c. *Proof that $\omega \in A^\star \Rightarrow \omega \in A_n$ for infinitely many $n \in \mathbb{N}$:*
Let $\omega \in A^\star$. We will recursively construct $1 \leqq n_1 < n_2 < \dots$ such that $\omega \in A_{n_j}$ for all $j \in \mathbb{N}$.

We observe that $\omega \in \bigcup\limits_{j\geqq n} A_j$ for all $n \in \mathbb{N}$. As $\omega \in \bigcup_{j\geq 1} A_j$ there exists $n_1 \geqq 1$ such that $\omega \in A_{n_1}$ and we have constructed the base case.

*Let $k \in \mathbb{N}$. If we already have found $n_1 < n_2 < \dots n_k$ such that $\omega \in A_{n_j}$ for $1 \leqq j \leqq k$ then we find n_{k+1} as follows: As $\omega \in \bigcup_{j\geqq n_k+1} A_j$ there exists $n_{k+1} \geqq n_k + 1$ such that $\omega \in A_{n_{k+1}}$. We have constructed our infinite sequence and this finishes the proof of **c**.*

d. *Proof that if $\omega \in A_n$ for infinitely many $n \in \mathbb{N} \Rightarrow \omega \in A^\star$:*
For $n \in \mathbb{N}$ we abbreviate $\Gamma_n := \bigcup_{j\geqq n} A_j$.

Let $1 \leqq n_1 < n_2 < \dots$ such that $\omega \in A_{n_j}$ for all $j \in \mathbb{N}$. Let $k \in \mathbb{N}$.

*Then $n_k \geqq k$, hence $\omega \in A_{n_k} \in \Gamma_{n_k} \subseteq \Gamma_k$ for all $k \in \mathbb{N}$, hence $\omega \in \bigcap_{k\in\mathbb{N}} \Gamma_k = A^\star$. We have proved **d**.* ■

Proposition 8.9 (liminf and limsup of indicator functions)**.**

$$1_{A_\star} = \liminf_{n\to\infty} 1_{A_n} \quad \text{and} \quad 1_{A^\star} = \limsup_{n\to\infty} 1_{A_n} \tag{8.40}$$

Proof: Let $\omega \in \Omega$. Then

$$1_{A_\star}(\omega) = 1 \;\Leftrightarrow\; \omega \in A_\star \;\Leftrightarrow \omega \in A_n \text{ for all except at most finitely many } n \in \mathbb{N} \tag{8.41}$$
$$\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for all except at most finitely many } n \in \mathbb{N} \tag{8.42}$$
$$\Leftrightarrow \liminf_n 1_{A_n}(\omega) = 1 \tag{8.43}$$

The second equivalence follows from prop.8.8 and the last equivalence follows from prop.8.7 and this proves the first equation. Similarly we have

$$1_{A^\star}(\omega) = 1 \;\Leftrightarrow\; \omega \in A^\star \;\Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N} \tag{8.44}$$
$$\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for infinitely many } n \in \mathbb{N} \tag{8.45}$$
$$\Leftrightarrow \limsup_n 1_{A_n}(\omega) = 1 \tag{8.46}$$

Again the second equivalence follows from prop.8.8 and the last equivalence follows from prop.8.7. ■

This last proposition is the reason for the following definition.

Definition 8.16 (limsup and liminf of a sequence of real functions)**.** Let Ω be a non-empty set and let $A_n \subseteq \Omega$ $(n \in \mathbb{N})$. We define

$$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{j\geqq n} A_j, \tag{8.47}$$

(8.48)

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{j\geqq n} A_j. \tag{8.49}$$

We call $\liminf\limits_{n\to\infty} A_n$ the **limit inferior** and $\limsup\limits_{n\to\infty} A_n$ the **limit superior** of the sequence A_n.

We note that $\liminf\limits_{n\to\infty} A_n = \limsup\limits_{n\to\infty} A_n$ if and only if the functions $\liminf\limits_{n\to\infty} 1_{A_n}$ and $\limsup\limits_{n\to\infty} 1_{A_n}$ coincide (prop. 8.40) which is true if and only if the sequence $1_{A_n}(\omega)$ has a limit for all $\omega \in \Omega$ (thm.8.2 on p.126). In this case we define

$$\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n \tag{8.50}$$

and we call this set the **limit** of the sequence A_n.

Note 8.1 (Notation for limits of monotone sequences of sets)**.** Let (A_n) be a non-decreasing sequence of sets, i.e., $A_1 \subseteq A_2 \subseteq \dots$ and let $A := \bigcup_n A_n$. Further let B_n be a non-increasing sequence of sets, i.e., $B_1 \supseteq B_2 \supseteq \dots$ and let $B := \bigcap_n B_n$. We write suggestively [61]

$$A_n \nearrow A \quad (n \to \infty), \qquad B_n \searrow B \quad (n \to \infty).$$

Example 8.5. Let $A_n \subseteq \Omega$.

[61] See note 10.1 on p.159.

(8.51) **a.** If $A_n \nearrow$ then $\liminf\limits_{n\to\infty} A_n = \limsup\limits_{n\to\infty} A_n = \bigcup\limits_{n\in\mathbb{N}} A_n.$

(8.52) **b.** If $A_n \searrow$ then $\liminf\limits_{n\to\infty} A_n = \limsup\limits_{n\to\infty} A_n = \bigcap\limits_{n\in\mathbb{N}} A_n.$

Exercise 8.1. Prove the assertions of example 8.5 above.

Note 8.2 (Liminf and limsup of number sequences vs their tail sets)**.** Let $x_n \in \mathbb{R}$ be a sequence of real numbers. We then can associate with this sequence that of its tail sets $T_n := \{x_j : j \geqq n\}$.

Do not confuse $\liminf_n x_n = \sup_n \big(\inf(T_n)\big)$ with $\liminf_n T_n = \bigcup_n \big(\bigcap_{k\geq n} T_k\big)$ and do not confuse $\limsup_n x_n = \inf_n \big(\sup(T_n)\big)$ with $\limsup_n T_n = \bigcap_n \big(\bigcup_{k\geq n} T_k\big)$:

$\liminf\limits_n x_n$ ($\limsup\limits_n x_n$) is a number: it is the lowest possible (highest possible) limit of a convergent subsequence $(x_{n_j})_{j\in\mathbb{N}}$. On the other hand we deal with a set(!) $\liminf_n T_n = \limsup_n T_n = \bigcap_n T_n$. The last equality follows from example 8.5 and the fact the the sequence of tailsets T_n is always non-increasing.

8.4 Addenda to chapter 8 (Real functions)

8.4.1 Sequences that enumerate parts of $\mathbb{Q}$

We informally defined the real numbers in ch.2.2 (Numbers) on p.14 as the set of all decimals, i.e., all numbers x which can be written as

(8.53) $$x = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad \text{where } d_j \text{ is a digit, i.e., } d_j = 0, 1, 2, \dots, 9,$$

(8.54) $$\text{i.e., } x = \lim_{k\to\infty} x_k \quad \text{where } x_k = m + \sum_{j=1}^{k} d_j 10^{-j}.$$

Each x_k is a (finite) sum of fractions, hence $x_k \in \mathbb{Q}$.

We proved in ch.5.3 (Countable sets) on p.97 that $\mathbb{Q}$ and hence all of its subsets can be sequenced: If $A \subseteq \mathbb{Q}$ there is a sequence $(q_n)_n$ of fractions such that $A = \{q_n : n \subseteq \mathbb{N}\}$. We apply this to $A := \mathbb{Q}$ and find that for each $k \in \mathbb{N}$ there is some $n \in \mathbb{N}$ such that $x_k = q_n$. Of course n depends on k, i.e., we have a functional dependency $n = n(k) = n_k$.

It follows from (8.54) that $q_{n_k} \to x$ as $k \to \infty$. In other words, we have proved the following

Theorem 8.3 (Universal sequence of rational numbers with convergent subsequences to any real number)**.**

There is a sequence $(q_n)_{n\in\mathbb{N}}$ of fractions which satisfies the following: For any $x \in \mathbb{R}$ there is a sequence $n_1, n_2, n_3, \dots,$ of natural numbers such that $x = \lim\limits_{k\to\infty} q_{n_k}$.

Remark 8.4.

a. The above theorem can be phrased as follows: There is a sequence $(q_n)_{n\in\mathbb{N}}$ of fractions such that for any $x \in \mathbb{R}$ one can find a subsequence $(q_{n_j})_{j\in\mathbb{N}}$ of $(q_n)_n$ which converges to x.

b. What is remarkable about thm.8.3: A **single** sequence $(q_n)_n$ is so rich that its ingredients can be used to approximate any item in the uncountable**!** set $\mathbb{R}$

c. Let $A := \{x \in \mathbb{R} : x^2 \leqq 2\} \ = [\,-\sqrt{2}, \sqrt{2}\,]$ and let $A_{\mathbb{Q}} := A \cap \mathbb{Q} = \{q \in \mathbb{Q} : q^2 \leqq 2\}$. A is of such a shape that for any $x \in A$ the partial sums $x_k \ = \ m \ + \ \displaystyle\sum_{j=1}^{k} d_j 10^{-j}$ which converge to x belong to $A_{\mathbb{Q}}$. (Why? Especially, why also for $x = \pm\sqrt{2}$?)

9 Vectors and vector spaces (Understand this!)

9.1 $\mathbb{R}^N$: Euclidean space

Most if not all of the material of this chapter is familiar to those of you who took a linear algebra course and those of you who took a Calc 3 course should know this from two or three dimensional space.

9.1.1 N–dimensional Vectors

This following definition of a vector is much more specialized than what is usually understood among mathematicians. For them, a vector is an element of a ***"vector space"*** *. You can find later in the document the definition of a vector space ((9.4) on p.138) What you see here is a definition of vectors of "finite dimension".*

Definition 9.1 (N–dimensional vectors)**.** A **vector** is a finite, ordered collection $\vec{v} = (x_1, x_2, x_3, \dots, x_N)$ of real numbers $x_1, x_2, x_3, \dots, x_N$. "Ordered" means that it matters which number comes first, second third, ... If the vector has N elements then we say that it is N**–dimensional** . The set of all N–dimensional vectors is written as $\mathbb{R}^N$.

You are encouraged to go back to the section on cartesian products (5.5 on p.96) to review what was said there about $\mathbb{R}^N = \underbrace{\mathbb{R} \times \mathbb{R} \times + \cdots \times \mathbb{R}}_{N times}$. *Here are some examples of vectors:*

Example 9.1 (Two–dimensional vectors)**.** The two–dimensional vector with coordinates $x = -1.5$ and $y = \sqrt{2}$ is written $(-1.5, \sqrt{2})$ and we have $(-1.5, \sqrt{2}) \in \mathbb{R}^2$. Order matters, so this vector is different from $(\sqrt{2}, -1.5) \in \mathbb{R}^2$.

Example 9.2 (Three–dimensional vectors). $\vec{v_t} = (3-t, 15, \sqrt{5t^2 + \frac{22}{7}}) \in \mathbb{R}^3$ with coordinates $x = 3-t$, $y = 15$ and $z = \sqrt{5t^3 + \frac{22}{7}}$ is an example of a parametrized vector (parametrized by t). Each specific value of t defines an element of $\in \mathbb{R}^3$, e.g., $\vec{v}_{-2} = (5, 15, \sqrt{20 + \frac{22}{7}})$. Note that

$$F : \mathbb{R} \to \mathbb{R}^3 \qquad t \mapsto F(t) = \vec{v_t}$$

defines a mapping from $\mathbb{R}$ into $\mathbb{R}^3$ in the sense of definition (4.6) on p.74. Each argument s has assigned to it one and only one argument $\vec{v_s} = (3 - s, 15, \sqrt{5s^2 + \frac{22}{7}}) \in \mathbb{R}^3$.
Or, is it rather that we have three functions

$$x(\cdot) : \mathbb{R} \to \mathbb{R} \qquad t \to x(t) = 3 - t.$$
$$y(\cdot) : \mathbb{R} \to \mathbb{R} \qquad t \to y(t) = 15.$$
$$z(\cdot) : \mathbb{R} \to \mathbb{R} \qquad t \to z(t) = \sqrt{5t^2 + \frac{22}{7}}$$

and $t \to \vec{v_t} = (x(t), y(t), z(t))$ is a vector of three real valued functions $x(\cdot), y(\cdot), z(\cdot)$?

Both points of view are correct and it depends on the specific circumstances how you want to interpret $\vec{v_t}$

Example 9.3 (One–dimensional vectors)**.** Let us not forget about the one–dimensional case: A one-dimensional vector has a single coordinate.

For example, $\vec{w_1} = (-3) \in \mathbb{R}^1$ with coordinate $x = -3 \in \mathbb{R}$ and $\vec{w_2} = (5.7a) \in \mathbb{R}^1$ with coordinate $x = 5.7a \in \mathbb{R}$ are one–dimensional vectors. $\vec{w_2}$ is not a fixed number but parametrized by a.

Mathematicians do not distinguish between the one–dimensional vector (x) and its coordinate value, the real number x. For brevity, they will simply write $\vec{w_1} = -3$ and $\vec{w_2} = 5.7a$.

Example 9.4 (Vectors as functions)**.** An N–dimensional vector $\vec{x} = (x_1, x_2, x_3, \cdots, x_N)$ can be interpreted as a real function (remember: a real function is one which maps it arguments into $\mathbb{R}$)

$$\begin{aligned} &f_{\vec{x}}(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R} \qquad m \mapsto x_m \\ &f_{\vec{x}}(1) = x_1,\ f_{\vec{x}}(2) = x_2,\ \cdots,\ f_{\vec{x}}(N) = x_N, \end{aligned} \tag{9.1}$$

i.e., as a real function whose domain is the natural numbers $1, 2, 3, \cdots, N$. This goes also the other way around: given a real function $f(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R}$ we can associate with it the vector

$$\begin{aligned} &\vec{v}_{f(\cdot)} := (f(1), f(2), f(3), \cdots, f(N)) \\ &\vec{v}_{f_1} = f(1),\ \vec{v}_{f_2} = f(2),\ , \cdots, \vec{v}_{f_N} = f(N) \end{aligned} \tag{9.2}$$

9.1.2 Addition and scalar multiplication for N–dimensional vectors

Definition 9.2 (Addition and scalar multiplication in $\mathbb{R}^N$)**.** Given are two N–dimensional vectors $\vec{x} = (x_1, x_2, \ldots, x_N)$ and $\vec{y} = (y_1, y_2, \ldots, y_N)$ and a real number α.

We define the **sum** $\vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ as the vector $\vec{z}$ with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \ldots; \quad z_N = x_N + y_N; \tag{9.3}$$

We define the **scalar product** $\alpha\vec{x}$ of α and $\vec{x}$ as the vector $\vec{w}$ with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \ldots; \quad w_N = \alpha x_N; \tag{9.4}$$

Figure 9.1 below describes vector addition.

Adding two vectors $\vec{v}$ and $\vec{w}$ means that you take one of them, say $\vec{v}$, and shift it in parallel (without rotating it in any way or flipping its direction), so that its starting point moves from the origin to the endpoint of the other vector $\vec{w}$. Look at the picture and you see that the vectors $\vec{v}$, $\vec{w}$ and $\vec{v}$ shifted form three pages of a parallelogram. $\vec{v} + \vec{w}$ is then the diagonal of this parallelogram which starts at the origin and ends at the endpoint of $\vec{v}$ shifted.

9.1.3 Length of N–dimensional vectors, the Euclidean Norm

*It is customary to write $\|\vec{v}\|_2$ for the length, sometimes also called the **Euclidean norm** of the vector $\vec{v}$.*

Example 9.5 (Length of one–dimensional vectors)**.** For a vector $\vec{v} = x \in \mathbb{R}$ its length is its absolute value $\|\vec{v}\|_2 = |x|$. This means that $\|-3.57\|_2 = |-3.57| = 3.57$ and $\|\sqrt{2}\|_2 = |\sqrt{2}| \approx 1.414$.

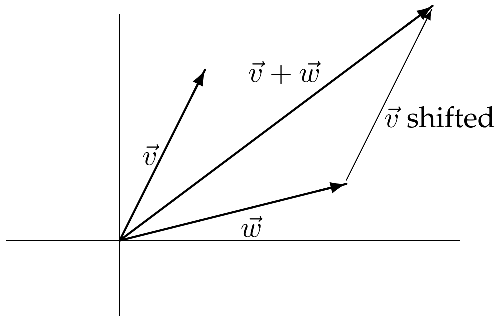


Figure 9.1: Adding two vectors.

Example 9.6 (Length of two–dimensional vectors)**.** We start with an example. Look at $\vec{v} = (4, -3)$. Think of an xy-coordinate system with origin (the spot where x-axis and y-axis intersect) $(0, 0)$. Then $\vec{v}$ is represented by an arrow which starts at the origin and ends at the point with coordinates $x = 4$ and $y = -3$ (see figure 9.2). How long is that arrow?

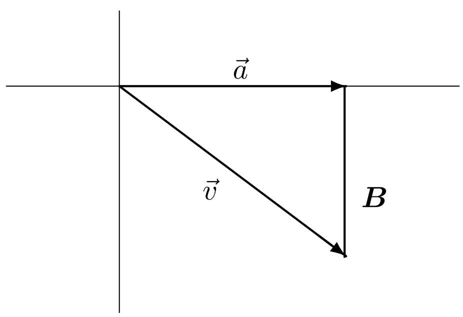


Figure 9.2: Length of a 2–dimensional vectors.

Think of it as the hypothenuse of a right angle triangle whose two other sides are the horizontal arrow from $(0, 0)$ to $(4, 0)$ (the vector $\vec{a} = (4, 0)$) and the vertical line $\boldsymbol{B}$ between $(4, 0)$ and $(4, -3)$. Note that $\boldsymbol{B}$ is not a vector because it does not start at the origin! Obviously (I hope this is obvious) we have $\|\vec{a}\|_2 = 4$ and length–of$(\boldsymbol{B}) = 3$. Pythagoras tells us that

$$\|\vec{v}\|_2^2 \;=\; \|\vec{a}\|_2^2 + \big(\text{length–of–}\boldsymbol{B}\big)^2$$

and we obtain for the vector $(4, -3)$ that $\|\vec{v}\|_2 = \sqrt{16 + 9} = 5$.

The above argument holds for any vector $\vec{v} = (x, y)$ with arbitrary $x, y \in \mathbb{R}$. The horizontal leg on the x-axis is then $\vec{a} = (x, 0)$ with length $|x| = \sqrt{x^2}$ and the vertical leg on the y-axis is a line equal in length to $\vec{b} = (0, y)$ the length of which is $|y| = \sqrt{y^2}$ The theorem of Pythagoras yields $\|(x,y)\|_2^2 = x^2 + y^2$ which becomes, after taking square roots on both sides,

$$\|(x,y)\|_2 = \sqrt{x^2 + y^2} \tag{9.5}$$

Example 9.7 (Length of three–dimensional vectors)**.** This is not so different from the two–dimensional case above. We build on the previous example. Let $\vec{v} = (4, -3, 12)$. Think of an xyz-coordinate system with origin (the spot where x-axis, y-axis and z-axis intersect) $(0,0,0)$. Then $\vec{v}$ is represented by an arrow which starts at the origin and ends at the point with coordinates $x = 4$, $y = -3$ and $z = 12$. How long is that arrow?

Remember what the standard 3–dimensional coordinate system looks like: The x-axis goes from west to east, the y-axis goes from south to north and the z-axis goes vertically from down below to the sky. Now drop a vertical line $\boldsymbol{B}$ from the point with coordinates $(4, -3, 12)$ to the xy–plane which is "spanned" by the x-axis and y-axis. This line will intersect the xy–plane at the point with coordinates $x = 4$ and $y = -3$ (and $z = 0$. Why?)

Note that $\boldsymbol{B}$ is not a vector because it does not start at the origin! It should be clear that length–of($\boldsymbol{B}$) $= |z| = 12$.

Now we connect the origin $(0,0,0)$ with the point $(4,-3,0)$ in the xy–plane which is the endpoint of $\boldsymbol{B}$.

We can forget about the z–dimension because this arrow is entirely contained in the xy–plane. Matter of fact, it is a genuine two–dimensional vector $\vec{a} = (4, -3)$ because it starts in the origin. Observe that $\vec{a}$ has the same values 4 and -3 for its x– and y–coordinates as the original vector $\vec{v}$. [62] We know from the previous example about two–dimensional vectors that

$$\|\vec{a}\|_2^2 = \|(x,y)\|_2^2 = x^2 + y^2 = 16 + 9 = 25.$$

At this point we have constructed a right angle triangle with a) hypothenuse $\vec{v} = (x, y, z)$ where we have $x = 4$, $y = -3$ and $z = 12$, b) a vertical leg with length $|z| = 12$ and c) a horizontal leg with length $\sqrt{x^2 + y^2}$ = 5. Pythagoras tells us that

$$\|\vec{v}\|_2^2 = z^2 + \|(x,y)\|_2^2 \;=\; 144 + 25 = 169 \quad \text{or} \quad \|\vec{v}\|_2 = 13.$$

None of what we just did depended on the specific values 4, -3 and 12. Any vector $(x, y, z) \in \mathbb{R}^3$ is the hypothenuse of a right triangle where the square lengths of the legs are z^2 and $x^2 + y^2$. This means we have proved the general formula $\|(x,y,z)\|^2 = x^2 + y^2 + z^2$ or

$$\|(x,y,z)\| = \sqrt{x^2 + y^2 + z^2} \tag{9.6}$$

The previous examples show how to extend the concept of "length" to vector spaces of any finite dimension:

Definition 9.3 (Euclidean norm)**.** Let $n \in \mathbb{N}$ and $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ be an n–dimension vector. The **Euclidean norm** $\|\vec{v}\|_2$ of $\vec{v}$ is defined as follows:

$$\|\vec{v}\|_2 \quad = \quad \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2} \quad = \quad \sqrt{\sum_{j=1}^{n} {x_j}^2} \tag{9.7}$$

[62] You will learn in the chapter on vector spaces that the vector $\vec{a} = (4, -3)$ is the projection on the xy–coordinates $\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2 \qquad (x,y,z) \mapsto (x,y)$ of the vector $\vec{v} = (4, -3, 12)$. (see Example 9.19) on p.143)

This definition is important enough to write the special cases for $n = 1, 2, 3$ where $\|\vec{v}\|_2$ coincides with the length of $\vec{v}$:

$$(9.8) \qquad \begin{aligned} 1-dim: &\quad \|(x)\|_2 \;=\; \sqrt{x^2} \;=\; |x| \\ 2-dim: &\quad \|(x,y)\|_2 \;=\; \sqrt{x^2+y^2} \\ 3-dim: &\quad \|(x,y,z)\|_2 \;=\; \sqrt{x^2+y^2+z^2} \end{aligned}$$

Proposition 9.1 (Properties of the Euclidean norm)**.** *Let $n \in \mathbb{N}$. Then the Euclidean norm, viewed as a function*

$$\|\cdot\|_2 : \;\mathbb{R}^n \longrightarrow \mathbb{R} \qquad \vec{v} = (x_1, x_2, \dots, x_n) \longmapsto \|\vec{v}\|_2 = \sqrt{\sum_{j=1}^{n} {x_j}^2}$$

has the following three properties:

$$(9.9a) \qquad \|\vec{v}\|_2 \geqq 0 \quad \forall \vec{v} \in \mathbb{R}^n \quad and \quad \|\vec{v}\|_2 = 0 \;\Leftrightarrow\; \vec{v} = 0 \quad \textit{positive definite}$$

$$(9.9b) \qquad \|\alpha\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R} \quad \textit{homogeneity}$$

$$(9.9c) \qquad \|\vec{v} + \vec{w}\|_2 \leqq \|\vec{v}\|_2 + \|\vec{w}\|_2 \quad \forall\, \vec{v}, \vec{w} \in \mathbb{R}^n \quad \textit{triangle inequality}$$

Proof:
a. *It is certainly true that $\|\vec{v}\|_2 \geqq 0$ for any n–dimensional vector $\vec{v}$ because it is defined as $+\sqrt{K}$ where the quantity K is, as a sum of squares, non–negative. If 0 is the zero vector with coordinates $x_1 = x_2 = \ldots = x_n = 0$ then obviously $\|0\|_2 = \sqrt{0 + \ldots + 0} = 0$. Conversely, let $\vec{v} = (x_1, x_2, \dots, x_n)$ be a vector in $\mathbb{R}^n$ such that $\|\vec{v}\|_2 = 0$. This means that $\sqrt{\sum_{j=1}^{n} {x_j}^2} = 0$ which is only possible if everyone of the non–negative x_j is zero. In other words, $\vec{v}$ must be the zero vector 0.*

b. *Let $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then*

$$\begin{aligned} \|\alpha\vec{v}\|_2 \;&=\; \sqrt{\sum_{j=1}^{n} (\alpha x_j)^2} \;=\; \sqrt{\sum_{j=1}^{n} \alpha^2 \alpha {x_j}^2} \;=\; \sqrt{\alpha^2 \sum_{j=1}^{n} \alpha {x_j}^2} \;=\; \sqrt{\alpha^2}\sqrt{\sum_{j=1}^{n} \alpha {x_j}^2} \\ &= \sqrt{\alpha^2}\|\vec{v}\|_2 = \; |\alpha| \cdot \|\vec{v}\|_2 \end{aligned}$$

because it is true that $\sqrt{\alpha^2} \;=\; |\alpha|$ for any real number α (see assumption 2.1 on p.16).

c. *The proof will only be given for $n = 1, 2, 3$.*
$\boldsymbol{n = 1}$ *: (9.9.c) simply is the triangle inequality for real numbers (see (2.2) on 16) and we are done.*
$\boldsymbol{n = 2, 3}$ *: Look back at the picture about addition of vectors in the plane or in space (see p.134). Remember that for any two vectors $\vec{v}$ and $\vec{w}$ you can always build a triangle whose sides have length $\|\vec{v}\|_2$, $\|\vec{w}\|_2$ and $\|\vec{v} + \vec{w}\|_2$. It is clear that the length of any one side cannot exceed the sum of the lengths of the other two sides, so we get specifically $\|\vec{v} + \vec{w}\|_2 \leqq \|\vec{v}\|_2 + \|\vec{w}\|_2$ and we are done.*

The geometric argument is not exactly an exact proof but I used it nevertheless because it shows the origin of the term "triangle inequality" for property (9.9.c). An exact proof will be given for arbitrary $n \in \mathbb{N}$ as a consequence of the so–called Cauchy–Schwartz inequality (cor.9.1). The inequality itself is stated and proved in prop.9.9 on p.146 in the section which discusses inner products (dot products) on vector spaces. ■

9.2 General vector spaces

9.2.1 Vector spaces: Definition and examples

Part of this follows [3] Brin, Matthew and Marchesi, Gerald: Linear Algebra, a text for Math 304, Spring 2016.

Mathematicians are very fond of looking at very different objects and figuring out what they have in common. They then create an abstract concept whose items have those properties and examine what they can conclude. For those of you who have had some exposure to object oriented programming: It's like defining a base class, e.g., "mammal", that possesses the core properties of several concrete items such as "horse", "pig", "whale" (sorry – can't require that all mammals have legs). We have looked at the following items that seem to be quite different:

real numbers
N–dimensional vectors
real functions

Well, that was disingenuous. I took great pains to explain that real numbers and one–dimensional vectors are sort of the same (see 9.3 on p.133). Besides I also explained that N–dimensional vectors can be thought of as real functions on the domain $X = \{1, 2, 3, \cdots, N\{$. (see 9.4 on p.133). Never mind, I'll introduce you now to vector spaces as sets of objects which you can "add" and multiply with real numbers according to rules which are guided by those that apply to addition and multiplication of ordinary numbers.

Here is quick reminder on how we add N–dimensional vectors and multiply them with scalars (real numbers) (see (9.1.2) on p.133). Given are two N–dimensional vectors
$\vec{x} = (x_1, x_2, \dots, x_N)$ and $\vec{y} = (y_1, y_2, \dots, y_N)$ and a real number α. Then the sum $\vec{z} = \vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ is the vector with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_N = x_N + y_N;$$

and the scalar product $\vec{w} = \alpha\vec{x}$ of α and $\vec{x}$ is the vector with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_N = \alpha x_N;$$

Example 9.8 (Vector addition and scalar multiplication)**.** We use $N = 2$ in this example:
Let $a = (-3, 1/5)$, $b = (5, \sqrt{2})$ We add those vectors by adding each of the coordinates separately:

$$a + b = (2, 1/5 + \sqrt{2})$$

and we multiply a with a scalar $\lambda \in \mathbb{R}$, e.g. $\lambda = 100$, by multiplying each coordinate with λ:

$$100a = 100(-3, 1/5) = (-300, 20).$$

In the last example I have avoided using the notation "$\vec{x}$" with the cute little arrows on top for vectors. I did that on purpose because this notation is not all that popular in Math even for N–dimensional vectors and definitely not for the more abstract vectors as elements of a vector space. Here now is the definition of a vector space, taken almost word for word from the book "Introductory Real Analysis" (Kolmogorov/Fomin [8]). This definition is quite lengthy because a set needs to satisfy many rules to be a vector space.

Definition 9.4 (Vector spaces (linear spaces))**.** A non–empty set V of elements $x, y, z, \dots$ is called a **vector space** or **linear space** if it satisfies the following:

A. Any two elements $x, y \in V$ uniquely determine a third element $x + y \in V$, called the **sum** of x and y with the following properties:

1. $x + y \;=\; y + x$ (**commutativity**);
2. $(x + y) + z \;=\; x + (y + z)$ (**associativity**);
3. There exists an element $0 \in V$, called the **zero element**, or **zero vector**, or **null vector**, with the property that $x + 0 = x$ for each $x \in V$;
4. For every $x \in V$, there exists an element $-x \in V$, called the **negative** of x, with the property that $x + (-x) = 0$ for each $x \in V$. When adding negatives, then there is a convenient short form. We write $x - y$ as an abbreviation for $x + (-y)$;

B. Any real number α and element $x \in V$ together uniquely determine an element $\alpha x \in V$ (sometimes also written $\alpha \cdot x$), called the **scalar product** of α and x. It has the following properties:

1. $\alpha(\beta x) \;=\; (\alpha\beta)x$;
2. $1x \;=\; x$;

C. The operations of addition and scalar multiplication obey the two **distributive laws**

1. $(\alpha + \beta)x \;=\; \alpha x + \beta x$;
2. $\alpha(x + y) \;=\; \alpha x + \alpha y$;

The elements of a vector space are called **vectors**

Definition 9.5 (Subspaces of vector spaces)**.** Let V be a vector space and let $A \subseteq V$ be a non–empty subset of V with the following property: For any $x, y \in A$ and $\alpha \in \mathbb{R}$ the sum $x + y$ and the scalar product αx also belong to A. Then A is called a **subspace** of V.

Note that if $\alpha = 0$ then $\alpha x = 0$ and it follows that the null vector belongs to A.

We ruled out the case $A = \emptyset$ but did not ask that A be a strict subset of V ((2.3) on p.11). In other words, V is a subspace of itself.

The set $\{0\}$ which contains the null vector 0 of V as its single element also is a subspace, the so called **nullspace**

Proposition 9.2 (Subspaces are vector spaces)**.** *A subspace of a vector space is a vector space, i.e., it satisfies all requirements of definition (9.4).*

Proof: None of the equalities that are part of the definition of a vector space magically ceases to be valid just because we look at a subset. The only thing that could go wrong is that some of the expressions might not *belong to A anymore. I'll leave it to you to figure out why this won't be the case, but I'll show you the proof for the second distributive law of part C.*

We must prove that for any $x, y \in A$ and $\lambda \in \mathbb{R}$

$$\lambda(x+y) = \lambda x + \lambda y :$$

First, $x+y \in A$ because a subspace contains the sum of any two of its elements. It follows that $\lambda(x+y)$ as product of a real number with an element of A again belongs to A because it is a subspace. Hence the left hand side of the equation belongs to A.

Second, both λx and λy belong to A because each is the scalar product of λ with an element of A and this set is a subspace. It follows for the same reason that the right hand side of the equation as the sum of two elements of the subspace A belongs to A.

Equality of $\lambda(x+y)$ and $\lambda x + \lambda y$ is true because it is true if we look at x and y as elements of V. ■

Remark 9.1 (Closure properties)**.** If a subset B of a larger set X has the property that certain operations on members of B will always yield elements of B, then we say that B is **closed** with respect to those operations.

A subspace is a subset of a vector space which is closed with respect to vector addition and scalar multiplication.

You have already encountered the following examples of vector spaces:

Example 9.9 (Vector space $\mathbb{R}$)**.** The real numbers $\mathbb{R}$ are a vector space if you take the ordinary addition of numbers as "+" and the ordinary multiplication of numbers as scalar multiplication.

Example 9.10 (Vector space $\mathbb{R}^n$)**.** More general, the sets $\mathbb{R}^n$ of n–dimensional vectors are vector spaces when you define addition and scalar multiplication as in (9.2) on p.133. To see why, just look at each component (coordinate) separately and you just deal with ordinary real numbers.

Example 9.11 (Vector space of real functions)**.** The set

$$\mathscr{F}(X,\mathbb{R}) := \{f(\cdot) : f(\cdot) \text{ is a real function on } X\} \tag{9.10}$$

of all real functions on an arbitrary non–empty set X [63] is a vector space if you define addition and scalar multiplication as in (8.2) on p.114. The reason is that you can verify the properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ and again, you just deal with ordinary real numbers. The "sup–norm" [64]

$$\|f(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\}$$

is **not a real function** on all of $\mathscr{F}(X,\mathbb{R})$ because $\|f(\cdot)\|_\infty = +\infty$ for any unbounded $f(\cdot) \in \mathscr{F}(X,\mathbb{R})$.

The subset

$$\mathscr{B}(X,\mathbb{R}) = \{h(\cdot) : h(\cdot) \text{ is a bounded real function on } X\}$$

(see (10.1) on p. 154) is a subspace of the vector space of all real functions on X. On this subspace the sup–norm truly is a real function because $\|f(\cdot)\|_\infty < \infty$.

[63] Note that $\mathscr{F}(X,\mathbb{R}) = \mathbb{R}^X$ (see remark 5.2, p.97 which follows def.5.6 of the Cartesian Product of a family of sets.)

[64] Norms in general and $\|\cdot\|_\infty$ in particular will be discussed in ch.10 on p.150.

And here are some more examples:

Example 9.12 (Subspace $\{(x,y) : x = y\}$)**.** The set $V := \{(x,x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ of all vectors in the plane with equal x and y coordinates has the following property: For any two vectors $\vec{x} = (a,a)$ and $\vec{y} = (b,b) \in V$ $(a, b \in \mathbb{R})$ and real number α the sum $\vec{x} + \vec{y} = (a+b, a+b)$ and the scalar product $\alpha\vec{x} = (\alpha a, \alpha a)$ have equal x–and y–coordinates, i.e., they again belong to V. Moreover the zero–vector 0 with coordinates $(0,0)$ belongs to V. It follows that the subset L of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (see (9.5) on p.138).

A proof for the following is omitted even though it is not difficult:

Example 9.13 (Subspace $\{(x,y) : y = \alpha x\}$)**.** Any subset of the form

$$V_\alpha := \{(x,y) \in \mathbb{R}^2 : y = \alpha x\}$$

is a subspace of $\mathbb{R}^2$ ($\alpha \in \mathbb{R}$). Draw a picture: V_α is the straight line through the origin in the xy–plane with slope α.

Example 9.14 (Embedding of linear subspaces)**.** The last example was about the subspace of a bigger space. Now we switch to the opposite concept, the **embedding** of a smaller space into a bigger space. We can think of the real numbers $\mathbb{R}$ as a part of the xy–plane $\mathbb{R}^2$ or even 3–dimensional space $\mathbb{R}^3$ by identifying a number a with the two-dimensional vector $(a,0)$ or the three-dimensional vector $(a,0,0)$. Let $M < N$. It is not a big step from here that the most natural way to uniquely associate an N–dimensional vector with an M–dimensional vector $\vec{x} := (x_1, x_2, \dots, x_M)$ by adding zero-coordinates to the right:

$$\vec{x} := (x_1, x_2, \dots, x_M, \underbrace{0, 0, \dots, 0}_{N-M \text{ times}})$$

Example 9.15 (All finite–dimensional vectors)**.** Let

$$\mathfrak{S} \ := \ \bigcup_{n \in \mathbb{N}} \mathbb{R}^n \ = \ \mathbb{R}^1 \cup \mathbb{R}^2 \cup \dots \cap \mathbb{R}^n \cup \dots$$

be the set of all vectors of finite (but unspecified) dimension.

We can define addition for any two elements $\vec{x}, \vec{y} \in \mathfrak{S}$ as follows: If $\vec{x}$ and $\vec{y}$ both happen to have the same dimension N then we add them as usual: the sum will be $x_1 + y_1, x_2 + y_2, \dots, x_N + y_N,$. If not, then one of them, say $\vec{x}$ will have dimension M smaller than the dimension N of $\vec{y}$. We now define the sum $\vec{x} + \vec{y}$ as the vector

$$\vec{z} := (x_1 + y_1, x_2 + y_2, \dots, x_M + y_M, y_{M+1}, y_{M+2}, \dots, y_N)$$

which is hopefully what you expected to happen.

Example 9.16 (All sequences of real numbers)**.** Let $\mathbb{R}^{\mathbb{N}} = \prod_{j \in \mathbb{N}} \mathbb{R}$ (see (5.6) on p.97). Is this the same set as $\mathfrak{S}$ from the previous example? The answer is No. Can you see why? I would be surprised if you do, so let me give you the answer: Each element $x \in \mathfrak{S}$ is of some finite dimension, say N, meaning that that it has no more than N coordinates. Each element $y \in \mathbb{R}^{\mathbb{N}}$ is a collection of numbers $y_1, y_2, \dots$ none of which need to be zero. In fact, $\mathbb{R}^{\mathbb{N}}$ is the vector space of all sequences of real numbers. Addition is of course done coordinate by coordinate and scalar multiplication with $\alpha \in \mathbb{R}$ is done by multiplying each coordinate with α.

There is again a natural way to embed $\mathfrak{S}$ into $\mathbb{R}^{\mathbb{N}}$ as follows: We transform an N–dimensional vector $(a_1, a_2, \dots, a_N)$ into an element of $\mathbb{R}^{\mathbb{N}}$ (a sequence $(a_j)_{j \in \mathbb{N}}$) by setting $a_j = 0$ for $j > N$.

Definition 9.6 (linear combinations)**.** Let V be a vector space and let $x_1, x_2, x_3, \dots, x_n \in V$ be a finite number of vectors in V. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. We call the finite sum

$$\sum_{j=0}^{n} \alpha_j x_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_n x_n \tag{9.11}$$

a **linear combination** of the vectors x_j . The multipliers $\alpha_1, \alpha_2, \dots$ are called **scalars** in this context.

In other words, linear combinations are sums of scalar multiples of vectors. You should understand that the expression in (9.11) always is an element of V, no matter how big $n \in \mathbb{N}$ was chosen:

Proposition 9.3 (Vector spaces are closed w.r.t. linear combinations)**.** *Let V be a vector space and let $x_1, x_2, x_3, \dots, x_n \in V$ be a finite number of vectors in V. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. Then the linear combination $\sum_{j=0}^{n} \alpha_j x_j$ also belongs to V. Note that this is also true for subspaces because those are vector spaces, too.*

Proof: Trivial. ■

Proposition 9.4. *Let V be a vector space and let $(W_i)_{i \in I}$ be a family of subspaces of V. Let $W := \bigcap \left[W_i : i \in I \right]$. Then W is a subspace of V.*

Proof: It suffices to show that W is not empty and that any linear combination of items in W belongs to W. As $0 \in W_i$ for each $i \in I$, it follows that $0 \in W$, hence $W \neq \emptyset$.

Let $x_1, x_2, \dots x_k \in W$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R} (k \in \mathbb{N})$. Let $x := \sum_{j=1}^{k} \alpha_j x_j$. Then $x \in W_i$ for all i because each W_i is a vector space, hence $x \in W$. ■

Definition 9.7 (Linear span)**.** Let V be a vector space and $A \subseteq V$. Then $span(A)$:= the set of all linear combinations of vectors in A is called the **span** or **linear span** of A. In other words,

$$span(A) = \{ \sum_{j=1}^{k} \alpha_j x_j : k \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in A \; (1 \leqq j \leqq k) \; \}. \tag{9.12}$$

Proposition 9.5. *Let V be a vector space and $A \subseteq V$. Then $span(A)$ is a subspace of V.*

Proof: Let $y_j \in span(A)$ for $j = 1, 2, \dots k$, i.e. y_j is a linear combination of vectors $x_{j,1}, x_{j,2}, \dots x_{j,n_j} \in A$. But then any linear combination of $y_1, y_2, \dots y_k$ is a linear combination of the vectors

$(x_{1,1}, x_{1,2}, \dots x_{1,n_1}),\ (x_{2,1}, x_{2,2}, \dots x_{2,n_2}), \dots, (x_{k,1}, x_{k,2}, \dots x_{k,n_k})$. ■

Theorem 9.1. *Let V be a vector space and $A \subseteq V$. Let $\mathfrak{V} := \{ W \subseteq V : W \supseteq A$ and W is a subspace of $V \}$. Then $span(A) = \bigcap \left[W : W \in \mathfrak{V} \right]$.*

Clearly, $span(A) \supseteq A$ It follows from prop.9.5 that $span(A) \in \mathfrak{V}$, hence $span(A) \supseteq \bigcap [W : W \in \mathfrak{V}]$.

On the other hand, Any subspace W of V that contains A also contains all its linear combinations, hence $span(A) \subseteq W$ for all $W \in \mathfrak{V}$. But then $span(A) \subseteq \bigcap [W : W \in \mathfrak{V}]$. ■

Remark 9.2 (Linear span(A) = subspace generated by A)**.** Let V be a vector space and $A \subseteq V$. Theorem 9.1 justifies to call $span(A) :=$ the **subspace generated by** A.

Definition 9.8 (linear mappings)**.** Let V_1, V_2 be two vector spaces.

Let $f(\cdot) : V_1 \to V_2$ be a mapping with the following properties:

(9.13a) $\quad f(x+y) = f(x) + f(y) \quad \forall x, y \in V_1$ $\quad$ **additivity**

(9.13b) $\quad f(\alpha x) = \alpha f(x) \quad \forall x \in V_1, \; \forall \alpha \in \mathbb{R}$ $\quad$ **homogeneity**

Then we call $f(\cdot)$ a **linear mapping**.

Note 9.1 (Note on homogeneity)**.** We encountered homogeneity when looking at the properties of the Euclidean norm ((9.9) on p.136), but homogeneity is defined differently there in that you had to take the absolute value $|\alpha|$ instead of α.

Remark 9.3 (Linear mappings are compatible with linear combinations)**.** We saw in the last proposition that vector spaces are closed with respect to linear combinations. Linear mappings and linear combinations go together very well in the following sense:

Remember that for any kind of mapping $x \mapsto f(x)$, $f(x)$ was called the image of x. Now we can express what linear mappings are about like this:

A: The image of the sum is the sum of the image

B: The image of the scalar multiple is the scalar multiple of the image

C: The image of the linear combination is the linear combination of the images

Mathematicians express this by saying that linear mappings **preserve** or are **compatible with** linear combinations.

Proposition 9.6 (Linear mappings preserve linear combinations)**.** *Let V_1, V_2 be two vector spaces.*

Let $f(\cdot) : V_1 \to V_2$ be a linear map and let $x_1, x_2, x_3, \dots, x_n \in V_1$ be a finite number of vectors in the domain V_1 of $f(\cdot)$. Let $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n \in \mathbb{R}$. Then $f(\cdot)$ preserves any such linear combination:

$$f(\sum_{j=0}^{n} \lambda_j x_j) = \sum_{j=0}^{n} \lambda_j f(x_j). \tag{9.14}$$

Proof:
First we note that $f(\lambda_j x_j) = \lambda_j f(x_j)$ for all j because linear mappings preserve scalar multiples and the proof is done for $n = 1$. Because they also preserve the addition of any two elements, the proposition holds for $n = 2$. We prove the general case by induction (see (2.11) on p.17). Our induction assumption is

$$f(\sum_{j=0}^{n-1} \lambda_j x_j) = \sum_{j=0}^{n-1} \lambda_j f(x_j).$$

We use it in the third equality of the following:

$$f(\sum_{j=0}^{n}\lambda_j x_j) \;=\; f(\sum_{j=0}^{n-1}\lambda_j x_j+\lambda_n x_n) = \; f(\sum_{j=0}^{n-1}\lambda_j x_j)+f(\lambda_n x_n) = \; \sum_{j=0}^{n-1}\lambda_j f(x_j)+f(\lambda_n x_n) = \; \sum_{j=0}^{n}\lambda_j f(x_j)$$

■

Here are some examples of linear mappings.

Example 9.17 (Projection on the first coordinate)**.** Let $N \in \mathbb{N}$. The map

$$\pi_1(\cdot) : \mathbb{R}^N \to \mathbb{R} \qquad (x_1, x_2, \dots, x_N) \mapsto x_1$$

is called the **projection** on the first coordinate or the first coordinate function.

Example 9.18 (Projections on any coordinate)**.** More generally, let $N \in \mathbb{N}$ and $1 \leqq j \leqq N$. The map

$$\pi_j(\cdot) : \mathbb{R}^N \to \mathbb{R} \qquad (x_1, x_2, \dots, x_N) \mapsto x_j$$

is called the **projection** on the jth coordinate or the j**th coordinate function**.

It is easy to see what that means if you set $N = 2$: For the two–dimensional vector $\vec{v} := (3.5, -2) \in \mathbb{R}^2$ you get $\pi_1(\vec{v}) = 3.5$ and $\pi_2(\vec{v}) = -2$.

Example 9.19 (Projections on any lower dimensional space)**.** In the last two examples we projected $\mathbb{R}^N$ onto a one–dimensional space. More generally, we can project $\mathbb{R}^N$ onto a vector space $\mathbb{R}^M$ of lower dimension M (i.e., we assume $M < N$) by keeping M of the coordinates and throwing away the remaining $N - M$. Mathematicians express this as follows:

Let $M, N, i_1, i_2, \dots, i_M \in \mathbb{N}$ such that $M < N$ and $1 \leqq i_1 < i_2 < \dots < i_M \leqq N$. The map

$$\pi_{i_1,i_2,\dots,i_M}(\cdot) : \mathbb{R}^N \to \mathbb{R}^M \qquad (x_1, x_2, \dots, x_N) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_M}) \tag{9.15}$$

is called the **projection** on the coordinates $i_1, i_2, \dots, i_M$.[65]

Example 9.20. Let $x_0 \in A$. The mapping

$$\varepsilon_{x_0} : \mathscr{F}(A, \mathbb{R}) \to \mathbb{R}; \qquad f(\cdot) \mapsto f(x_0) \tag{9.16}$$

which assigns to any real function on A its value at the specific point x_0 is a linear mapping because if $h(\cdot) = \sum\limits_{j=0}^{n} a_j f_j(\cdot)$ then

$$\varepsilon_{x_0}(\sum_{j=0}^{n} a_j f_j(\cdot)) \;= \varepsilon_{x_0}(h(\cdot)) \;=\; h(x_0) \;=\; \sum_{j=0}^{n} a_j f_j(x_0) \;=\; \sum_{j=0}^{n} a_j \varepsilon_{x_0}(f_j(\cdot))$$

and this proves the linearity of the mapping $\varepsilon_{x_0}(\cdot)$. The mapping $\varepsilon_{x_0}(\cdot)$ is called the **abstract integral** with respect to point mass at x_0.

[65] You previously encountered an example where we made use of the projection

$$\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2 \qquad (x, y, z) \mapsto (x, y).$$

This was in the course of computing the length of a 3–dimensional vector (see (9.5) on p.133).

Lemma 9.1 ($F \circ \mathit{span} = \mathit{span} \circ F$)**.** *[3] Brin/Marchesi Linear Algebra, general lemma 4.1.7: Let V, W be two vector spaces and $F : V \to W$ a linear mapping from V to W. Let $A \subseteq V$. Then*

$$F(\mathit{span}(A)) = \mathit{span}(F(A)). \tag{9.17}$$

Proof: *See Brin/Marchesi Linear Algebra, general lemma 4.1.7.* ■

Definition 9.9 (Linear dependence and independence)**.** Let V be a vector space and $A \subseteq V$

- **a.** A is called **linearly dependent** if the following is true: There exist distinct vectors $x_1, x_2, \dots x_k \in A$ and scalars $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$) such that not all scalars α_j are zero ($1 \leqq j \leqq k$) and $\sum_{j=1}^{k} \alpha_j x_j = 0$.
- **b.** A is called **linearly independent** if A is not linearly dependent, i.e., if the following is true: Let $x_1, x_2, \dots x_k \in A$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$). If $\sum_{j=1}^{k} \alpha_j x_j = 0$ then $\alpha_j = 0$ for all $1 \leqq j \leqq k$.

Definition 9.10 (Basis of a vector space)**.** Let V be a vector space and $B \subseteq V$. B is called a **basis** of V if **a.** B is linearly independent and **b.** $\mathit{span}(B) = V$.

Lemma 9.2. *Let V be a vector space and $A \subseteq V$ linearly independent. If $\mathit{span}(A) \subsetneq V$ and $y \in \mathit{span}(A)^{\complement}$ then $A' := A \cup \{y\}$ is linearly independent.*

Proof:
Let $x_1, x_2, \dots x_k \in A'$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$) such that

$$\sum_{j=1}^{k} \alpha_j x_j = 0 \tag{9.18}$$

We must show that each α_j is zero.

Case 1: *$y \neq x_j$ for all j:*
Then $y \in A$ and it follows from the linear independence of A that each α_j is zero.

Case 2: *$y = x_{j_0}$ for some $1 \leq j_0 \leqq k$:*
We first show that $\alpha_{j_0} = 0$: Otherwise

$$x_{j_0} = \sum_{j \neq j_0} \frac{-\alpha_j}{\alpha_{j_0}} x_j \tag{9.19}$$

is a linear combination of elements of A, contrary to the assumption that $x_{j_0} = y \in \mathit{span}(A)^{\complement}$ and we have shown that $\alpha_{j_0} = 0$.

It follows from (9.18) that

$$\sum_{j \neq j_0} \alpha_j x_j = 0 \tag{9.20}$$

and It follows as in case 1 from the linear independence of A that if $j \neq j_0$ then α_j also is zero. ■

9.2.2 Normed vector spaces (Study this!)

The following definition of inner products and proof of the Cauchy–Schwartz inequality were taken from "Calculus of Vector Functions" (Williamson/Crowell/Trotter [13]).

Definition 9.11 (Inner products)**.** Let V be a vector space with a function

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y := \bullet(x,y)$$

which satisfies the following properties:

(9.21a) $x \bullet x \geqq 0 \quad \forall x \in V$ and $x \bullet x = 0 \;\Leftrightarrow\; x = 0$ **positive definite**

(9.21b) $x \bullet y = y \bullet x \quad \forall x, y \in V$ **symmetry**

(9.21c) $(x+y) \bullet z = x \bullet z + y \bullet z \quad \forall\, x, y, z \in V$ **additivity**

(9.21d) $(\lambda x) \bullet y = \lambda(x \bullet y) \quad \forall\, x, y \in V \quad \forall\, \lambda \in \mathbb{R}$ **homogeneity**

We call such a function an **inner product**. [66]

Note that additivity and homogeneity of the mapping $x \mapsto x \bullet y$ for a fixed $y \in V$ imply linearity of that mapping and the symmetry property implies that the mapping $y \mapsto x \bullet y$ for a fixed $x \in V$ is linear too. In other words, an inner product is binear in the following sense:

Definition 9.12 (Bilinearity)**.** Let V be a vector space with a function

$$F(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto F(x,y).$$

$F(\cdot,\cdot)$ is called **bilinear** if it is linear in each component, i.e., the mappings

$$F_1 : V \to \mathbb{R}; \quad x \mapsto F(x,y)$$
$$F_2 : V \to \mathbb{R}; \quad y \mapsto F(x,y)$$

are both linear.

Proposition 9.7 (Algebraic properties of the inner product)**.** *Let V be a vector space with inner product $\bullet(\cdot,\cdot)$. Let $a, b, x, y \in V$. Then*

(9.22a) $(a+b) \bullet (x+y) = a \bullet x \;+\; b \bullet x \;+\; a \bullet y \;+\; b \bullet y$

(9.22b) $(x+y) \bullet (x+y) = x \bullet x \;+\; 2(x \bullet y) \;+\; y \bullet y$

(9.22c) $(x-y) \bullet (x-y) = x \bullet x \;-\; 2(x \bullet y) \;+\; y \bullet y$

*Proof of **a**:*

$$\begin{aligned}(a+b) \bullet (x+y) &= (a+b) \bullet x \;+\; (a+b) \bullet y \\ &= a \bullet x \;+\; b \bullet x \;+ a \bullet y \;+\; b \bullet y.\end{aligned}$$

We used linearity in the second argument for the first equality and linearity in the first argument for the second equality.

*Proof of **b** and **c**: The proof is required as part of an upcoming homework.*

The following is the most important example of an inner product.

[66] also called **dot product**, e.g., in [3] Brin/Marchesi Linear Algebra, ch.6, Orthogonality.

Proposition 9.8 (Inner product on $\mathbb{R}^N$)). *Let $N \in \mathbb{N}$. Then the real function*

$$(9.23) \qquad (\vec{v}, \vec{w}) \ \mapsto \ x_1 y_1 + x_2 y_2 + \ldots + x_N y_N \ = \ \sum_{j=1}^{N} x_j y_j$$

is an inner product on $\mathbb{R}^N \times \mathbb{R}^N$.

Proof:
a : *For $\vec{v} = \vec{w}$ we obtain $(\vec{v} \bullet \vec{v}) = \|\vec{v}\|$ and positive definiteness of the inner product follows from that of the Euclidean norm.*
b : *Symmetry is clear because $x_j y_j = y_j x_j$.*
c : *Additivity follows from the fact that $(x_j + y_j) z_j = x_j z_j + y_j z_j$.*
d : *Homogeneity follows from the fact that $(\lambda x_j) y_j = \lambda(x_j y_j)$.* ■

Proposition 9.9 (Cauchy–Schwartz inequality for inner products). *Let V be a vector space with an inner product*

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x, y) \mapsto x \bullet y := \bullet(x, y)$$

Then

$$(x \bullet y)^2 \ \leqq \ (x \bullet x)\,(y \bullet y).$$

Proof:
Step 1: *We assume first that $x \bullet x = y \bullet y = 1$. Then*

$$\begin{aligned} 0 \ &\leqq (x - y \bullet x - y) \\ &= x \bullet x - 2x \bullet y + y \bullet y = 2 - 2(x \bullet y) \end{aligned}$$

where the first equality follows from proposition (9.7) on p.145.
This means $2(x \bullet y) \ \leqq \ 2$, i.e., $x \bullet y \ \leqq \ 1 \ = \ (x \bullet x)\,(y \bullet y)$ where the last equality is true because we had assumed $x \bullet x = y \bullet y = 1$. The Cauchy–Schwartz inequality is thus true under that special assumption.

Step 2: *General case: We do not assume anymore that $x \bullet x = y \bullet y = 1$. If x or y is zero then the Cauchy–Schwartz inequality is trivially true because, say if $x = 0$ then the left hand side becomes*

$$(x \bullet y)^2 = (0x \bullet y)^2 = 0(x \bullet y)^2 = 0$$

whereas the right hand side is, as the product of two non–negative numbers $x \bullet x$ and $y \bullet y$, non–negative.

So we can assume that x and y are not zero. On account of the positive definiteness we have $x \bullet x > 0$ and $y \bullet y > 0$. This allows us to define $u := x/\sqrt{x \bullet x}$ and $v := y/\sqrt{y \bullet y}$. But then

$$\begin{aligned} u \bullet u \ &= (x \bullet x)/\sqrt{x \bullet x}^2 \ = 1 \\ v \bullet v \ &= (y \bullet y)/\sqrt{y \bullet y}^2 \ = 1. \end{aligned}$$

We have already seen in step 1 that $u \bullet v \ \leqq \ 1$. It follows that

$$(x \bullet y)/(\sqrt{x \bullet x}\sqrt{y \bullet y}) \ = \ (x/\sqrt{x \bullet x}) \bullet (y/\sqrt{y \bullet y}) \ \leqq \ 1$$

We multiply both sides with $\sqrt{x \bullet x}\sqrt{y \bullet y}$,

$$x \bullet y \ \leqq \ \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We replace x by $-x$ and obtain

$$-(x \bullet y) \leqq \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

Think for a moment about the meaning of the absolute value and it is clear that the last two inequalities together prove that

$$|x \bullet y| \leqq \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We square this and obtain

$$(x \bullet y)^2 \leqq (x \bullet x)(y \bullet y)$$

and the Cauchy–Schwartz inequality is proved. ■

Note 9.2. We previously discussed the sup–norm

$$\|f(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\} \tag{9.24}$$

for real functions on some non–empty set X and the Euclidean norm

$$\|\vec{x}\|_2 = \sqrt{\sum_{j=1}^{n} {x_j}^2} \tag{9.25}$$

for n–dimensional vectors $\vec{x} = (x_1, x_2, \dots, x_n)$. You saw that either one satisfies positive definiteness, homogeneity and the triangle inequality (see (10.1) on p.154 and (9.1) on p.136). As previously mentioned, mathematicians like to define new objects that are characterized by a given set of properties. As an example we had the definition of a vector space which encompasses objects as different as finite–dimensional vectors and real functions.

The sup norm $\|f(\cdot)\|_\infty$ and the Euclidean norm $\|\vec{x}\|_2$ are special instances of the concept of length or size of a vector (remember from example 9.11 on p.139 (Vector space of real functions) that bounded real valued functions $f : X \to \mathbb{R}$ are vectors when considered elements of the vector space $\mathscr{B}(X, \mathbb{R})$). Both norms have in common the three characteristics of positive definiteness, homogeneity and the triangle inequality and a very rich mathematical theory can be developed for a generalized definition of length which is based just on those properties.

Accordingly we now define a norm on a vector space by demanding the three characteristics of positive definiteness, homogeneity and the triangle inequality.

Definition 9.13 (Normed vector spaces)**.** Let V be a vector space. A **norm** on V is a real function

$$\|\cdot\| : V \to \mathbb{R} \qquad x \mapsto \|x\|$$

with the following three properties: [67]

$$\|x\| \geqq 0 \quad \forall x \in V \quad \text{and} \quad \|x\| = 0 \Leftrightarrow x = 0 \quad \textbf{positive definite} \tag{9.26a}$$

$$\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in V, \forall \alpha \in \mathbb{R} \quad \textbf{(absolute) homogeneity} \tag{9.26b}$$

$$\|x + y\| \leqq \|x\| + \|y\| \quad \forall\, x, y \in V \quad \textbf{triangle inequality} \tag{9.26c}$$

We call V a **normed vector space** and we write $(V, \|\cdot\|)$ instead of V when we wish to emphasize what norm on V we are dealing with.

[67] "absolute homogeneity" seems to be preferred to "homogeneity" nowadays.

Remark 9.4. A vector space may be endowed with more than one norm. Here are two examples.

a. It is shown in the next proposition that if $x \mapsto \|x\|$ is a norm on a vector space V and $\beta > 0$ then $x \mapsto \beta \cdot \|x\|$ also is a norm on V.

b. Here is a non-trivial example for the vector space $\mathbb{R}^n$:

Let $p \geqq 1$. Then

(9.27) $$x \mapsto \|x\|_p \; := \Big(\sum_{j=1}^{n} {x_j}^p \Big)^{1/p}$$

is a norm on $\mathbb{R}^n$). This norm is called the p**-norm** .

The Euclidean norm is a p–norm; it is the 2-norm. A proof that $\|\cdot\|_p$ is in fact a norm is not given in this document except for $p = 2$ (see cor.9.1 on p.149).

Proposition 9.10. *Let $(V, \|\cdot\|)$ be a normed vector space and let $\gamma > 0$. Let $p : \quad V \to \mathbb{R}$ be defined as $p(x) := \gamma\|x\|$. Then p also is a norm.*

Proof: The proof is required as part of an upcoming homework.

Definition 9.14 (Norm for an inner product)**.** Let V be a vector space with an inner product

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x, y) \mapsto x \bullet y$$

Then

(9.28) $$\|\cdot\| : x \;\mapsto\; \|x\| \;:=\; \sqrt{(x \bullet x)}$$

is called the **norm associated with the inner product** $\bullet(\cdot,\cdot)$.

The following theorem shows that it is justified to call $\|x\| \;:=\; \sqrt{(x \bullet x)}$ a norm.

Theorem 9.2 (Inner products define norms)**.** *Let V be a vector space with an inner product*

$$\bullet(\cdot,\cdot) : V \times V \to \mathbb{R}; \qquad (x, y) \mapsto x \bullet y$$

Then

$$\|\cdot\| : x \;\mapsto\; \|x\| \;:=\; \sqrt{(x \bullet x)}$$

defines a norm on V

Proof:
Positive definiteness *: follows immediately from that of the inner product.*

Homogeneity *: Let $x \in V$ and $\lambda \in \mathbb{R}$. Then*

$$\|\lambda x\| \;=\sqrt{(\lambda x) \bullet (\lambda x)} \;=\; \sqrt{\lambda\lambda(x \bullet x))} \;=\; |\lambda|\sqrt{x \bullet x} \;=\; |\lambda|\|x\|.$$

Triangle inequality *: Let $x, y \in V$. Then*

$$
\begin{aligned}
\|x+y\|^2 &= (x+y) \bullet (x+y) \\
&= x \bullet x \;+\; 2(x \bullet y) \;+\; y \bullet y \\
&\leqq x \bullet x \;+\; 2|x \bullet y| \;+\; y \bullet y \\
&\leqq x \bullet x \;+\; 2\sqrt{x \bullet x}\sqrt{y \bullet y} \;+\; y \bullet y \\
&= \|x\|^2 \;+\; 2\|x\|\,\|y\| \;+\; \|y\|^2 \\
&= \big(\|x\| \;+\; \|y\|\big)^2.
\end{aligned}
$$

The second equation uses bilinearity and symmetry of the inner product. The first inequality expresses the simple fact that $\alpha \leqq |\alpha|$ for any number α. The second inequality uses Cauchy–Schwartz. The next equality just substitutes the definition $\|x\| = \sqrt{(x \bullet x)}$ of the norm. The next and last equality is the binomial expansion $(a+b)^2 = a^2 + 2ab + b^2$ for the ordinary real numbers $a = \|x\|$ and $b = \|y\|$.

We take square roots in the above inequality $\|x+y\|^2 \leqq \big(\|x\| + \|y\|\big)^2$ and obtain $\|x+y\| \leqq \|x\| + \|y\|$, the triangle inequality we set out to prove. ■

It was stated in prop.9.1 on p. 136 that the Euclidean norm is in fact a norm but only positive definiteness and homogeneity were proved. We now can easily complete the proof.

Corollary 9.1. *The Euclidean norm in $\mathbb{R}^n$ defined as $\|(x_1, x_2, \dots, x_n)\|_2 = \sqrt{\sum\limits_{j=1}^{n} {x_j}^2}$ (see def. 9.3 on p.135) is a norm.*

Proof: This follows from the fact that

$$\vec{x} \bullet \vec{y} = \sum_{j=1}^{n} x_j y_j \qquad \textit{where } \vec{x} = (x_1, \dots, x_n) \textit{ and } \vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$$

defines an inner product on $\mathbb{R}^n \times \mathbb{R}^n$. (see prop.9.8) for which $\|(x_1, x_2, \dots, x_n)\|_2$ is the associated norm.
■

In chapter (10) on the topology of real numbers (p. 150) you will learn about metric spaces as a concept that generalizes the measurement of distance (or closeness, if you prefer) for the elements of a non–empty set.

9.3 Addenda to chapter 9 (Vectors and vector spaces)

9.3.1 More on norms

Exercise 9.1. Prove that the p–norm (see (9.27) of rem.9.4) is a norm on $\mathbb{R}^n$ for the special case $p = 1$:

$$\|\vec{x}\|_1 = \sum_{j=1}^{n} |x_j|$$

10 Metric Spaces

There is a branch of Mathematics, called topology, which deals with the concept of closeness. The concept of limits of a sequence $(x_n)_n$ *is based on closeness: The points of the sequence must get "arbitrarily close" to its limit as* $n \to \infty$. *Continuity of functions also can be phrased in terms of closeness: They map arbitrarily close elements of the domain to arbitrarily close elements of the codomain. In the most general setting Topology is about neighborhoods of a point without having the concept of measuring the distance of two points. We mostly won't deal with such a level of generality in this document. Instead we'll we'll focus on sets* X *that are equipped with a distance function.*

10.1 The topology of metric spaces (Study this!)

A metric is a real function of two arguments which associates with any two points $x, y \in X$ *their "distance"* $d(x, y)$.

It is clear how you measure the distance (or closeness, depending on your point of view) of two numbers x *and* y*: you plot them on an* x*–axis where the distance between two consecutive integers is exactly one inch, grab a ruler and see what you get. Alternate approach: you compute the difference. For example, the distance between* $x = 12.3$ *and* $y = 15$ *is* $x - y = 12.3 - 15 = -2.7$. *Actually, we have a problem: There are situations where direction matters and a negative distance is one that goes into the opposite direction of a positive distance, but we do not want that in this context and understand the distance to be always non–negative, i.e.,*

$$dist(x, y) \ = \ |y - x| \ = \ |x - y|$$

More importantly, you must forget what you learned in your in your science classes: "Never ever talk about a measure (such as distance or speed or volume) without clarifying its dimension". Is the speed measured in miles per hour our inches per second? Is the distance measured in inches or miles or micrometers? In the context of metric spaces we measure distance simply as a number, without any dimension attached to it. For the above example, you get

$$dist(12.3, 15) \ = \ |12.3 - 15| \ = 2.7.$$

In section *9.1.3 on p.133* *it is shown in great detail that the distance between two two–dimensional vectors* $\vec{v} = (v_1, v_2)$ *and* $\vec{w} = (w_1, w_2)$ *is*

$$dist(\vec{v}, \vec{w}) \ = \ \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2}$$

and the distance between two three–dimensional vectors $\vec{v} = (v_1, v_2, v_3)$ *and* $\vec{w} = (w_1, w_2, w_3)$ *is*

$$dist(\vec{v}, \vec{w}) \ = \ \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + (w_3 - v_3)^2}.$$

We will see in thm *10.1 on p.151* *that this distance function is a metric according to the next definition:*

Definition 10.1 (Metric spaces)**.** Let X be an arbitrary, non–empty set.
A **metric** on X is a real function of two arguments

$$d(\cdot, \cdot) : X \times X \to \mathbb{R}, \qquad (x, y) \mapsto d(x, y)$$

with the following three properties: [68]

(10.1a) $d(x,y) \geqq 0 \quad \forall x,y \in X$ and $d(x,y) = 0 \;\Leftrightarrow\; x = y$ **positive definite**

(10.1b) $d(x,y) = d(y,x) \quad \forall x,y \in X$ **symmetry**

(10.1c) $d(x,z) \leqq d(x,y) + d(y,z) \quad \forall\, x,y,z \in X$ **triangle inequality**

The pair $(X, d(\cdot,\cdot))$, usually just written as (X,d), is called a **metric space**. We'll write X for short if it is clear which metric we are talking about.

To appreciate that last sentence, you must understand that there can be more than one metric on X. See the examples below.

Remark 10.1 (Metric properties)**.** Let us quickly examine what those properties mean.

"Positive definite":	The distance is never negative and two items x and y have distance zero if and only if they are equal.
"symmetry":	the distance from x to y is no different to that from y to x. That may come as a surprise to you if you have learned in Physics about the distance from point a to point b being the vector $\vec{v}$ that starts in a and ends in b and which is the opposite of the vector $\vec{w}$ that starts in b and ends in a, i.e., $\vec{v} = -\vec{w}$. In this document we care only about size and not about direction.
"Triangle inequality":	If you directly walk from x to z then this will take less time than if you make a stopover at an intermediary y.

Before we give some examples of metric spaces, here is a theorem that tells you that a vector space with a norm, i.e., a function with the three properties of the Euclidean norm (see 9.1 on p.136), becomes a metric space as follows:

Theorem 10.1 (Norms define metric spaces)**.** *A **norm** on a vector space L is a real function* [69]

$$\|\cdot\| : L \to \mathbb{R}_+; \qquad x \mapsto \|x\|$$

such that

$$\begin{aligned} &\|x\| \geqq 0 \quad \forall x \in L \quad \textit{and} \quad \|x\| = 0 \;\Leftrightarrow\; x = 0 \quad \textbf{\textit{positive definite}} \\ &\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in L, \forall \alpha \in \mathbb{R} \quad \textbf{\textit{homogeneity}} \\ &\|x+y\| \leqq \|x\| + \|y\| \quad \forall\, x,y \in L \quad \textbf{\textit{triangle inequality}} \end{aligned} \tag{10.2}$$

The following is true:

$$d_{\|\cdot\|}(\cdot,\cdot) : (x,y) \mapsto \|y - x\|$$

defines a metric space $(L, d_{\|\cdot\|})$

[68] If you forgot the meaning of $X \times X$, it's time to review [1] B/G (Beck/Geoghegan) ch.5.3 on cartesian products.
[69] This definition was already given in the section on abstract vector spaces (def.9.13, p.147).

Proof: The proof is required as part of an upcoming homework and will not be given here. It is really simple, even the triangle inequality for the metric $d(x,y) = \|x-y\|$ follows easily from the triangle inequality for the norm.

■

Here are some examples of metric spaces.

Example 10.1 ($\mathbb{R} : d(a,b) = |b-a|$)**.** According to thm.10.1 (Norms define metric spaces) on p.151 this is a metric space because the Euclidean norm $|\cdot|$ is a norm on $\mathbb{R} = \mathbb{R}^1$.

Here is a direct proof; It is obvious that if x, y are real numbers then the difference $x - y$, and hence its absolute value, is zero if and only if $x = y$ and that proves positive definiteness.
Symmetry follows from the fact that

$$d(x,y) \;=\; |x-y| \;=\; |-(y-x)| \;=\; |y-x| \;=\; d(y,x).$$

The triangle inequality follows from the one which says that

$$|a+b| \leqq |a| + |b|$$

(see prop.2.2 The Triangle Inequality for real numbers on p.16.) as follows:

$$d(x,z) \;=\; |x-z| \;=\; |(x-y)-(z-y)| \leqq |x-y| + |z-y| \;=\; d(x,y) + d(z,y) \;=\; d(x,y) + d(y,z).$$

Example 10.2 (bounded real functions with $\boldsymbol{d(f,g) =}$ **sup**–norm of $\boldsymbol{g(\cdot) - f(\cdot)}$)**.**

$$d(f,g) = \sup\{|g(x) - f(x)| : x \in X\}$$

is a metric on the set $\mathscr{B}(X,\mathbb{R})$ of all bounded real functions on X.

This follows from the fact that $f \mapsto \sup\{|f(x)| : x \in X\}$ *is a norm on the vector space* $\mathscr{B}(X,\mathbb{R})$ *(see prop.10.1 on p.154) and from thm.10.1 (Norms define metric spaces) on p.151.*

Example 10.3 ($\mathbb{R}^N : d(\vec{x},\vec{y}) =$ Euclidean norm)**.**

$$d(\vec{x},\vec{y}) \;=\; \sqrt{(y_1-x_1)^2 + (y_2-x_2)^2 + \ldots + (y_N-x_N)^2} \;=\; \sqrt{\sum_{j=1}^{N} (y_j - x_j)^2}$$

This follows from the fact that the Euclidean norm is a norm on the vector space $\mathbb{R}^N$ *(see (9.1) on p.136).*

Just in case you think that all metrics are derived from norms, this one will set you straight.

Example 10.4 (Discrete metric)**.** Let X be non–empty. Then the function

$$d(x,y) = \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y \end{cases}$$

on $X \times X$ defines a metric.

Proof: Obviously the function is non–negative and it is zero if and only if $x = y$. Symmetry is obvious too.

The triangle inequality $d(x,z) = d(x,y) + d(y,z)$ is clear in the special case $x = z$. (Why?)

So let us assume $x \neq z$. But then $x \neq y$ or $y \neq z$ or both must be true. (Why?) That means that

$$d(x,z) = 1 \leqq d(x,y) + d(y,z)$$

and this proves the triangle inequality. ■

10.1.1 Measuring the distance of real functions

How do we compare two functions? Let us make our lives easier: How do we compare two real functions $f(\cdot)$ and $g(\cdot)$? One answer is to look at a picture with the graphs of $f(\cdot)$ and $g(\cdot)$ and look at the shortest distance $|f(x) - g(x)|$ as you run through all x. That means that the distance between the functions $f(x) = x$ and $g(x) = x^2$ is zero because $f(1) = g(1) = 1$. The distance between $f(x) = x + 1$ and $g(x) = 0$ (the x–axis) is also zero because $f(-1) = g(-1) = 0$.

Do you really think this is a good way to measure closeness? You really do not want two items to have zero distance unless they coincide. It's a lot better to look for an argument x where the value $|f(x) - g(x)|$ is largest rather than smallest. Now we are ready for a proper definition.

Definition 10.2 (Distance between real functions)**.** Let X be an arbitrary, non-empty set and let $f(\cdot), g(\cdot) : X \to \mathbb{R}$ be two real functions on X. We define the distance between $f(\cdot)$ and $g(\cdot)$ as

$$d(f,g) \ := \ d(f(\cdot), g(\cdot)) \ := \ \sup\{|f(x) - g(x)| : x \in X\} \ \ (\delta > 0) \tag{10.3}$$

The following picture illustrates this definition in the special case that the argument x is a real number, i.e., $X \subseteq \mathbb{R}$

Plot the graphs of f and g as usual and find the the spot x_0 on the x–axis for which the difference $|f(x_0) - g(x_0)|$ (the length of the vertical line that connects the two points with coordinates $(x_0, f(x_0))$ and $(x_0, g(x_0))$) has the largest possible value. The domain of f and g is the subset of $\mathbb{R}$ that corresponds to the thick portion of the x–axis.

Now that you know how to measure the distance $d(f(\cdot), g(\cdot))$ between two real functions $f(\cdot), g(\cdot)$, the next picture shows you how to visualize for a given $\delta > 0$ and $f : X \to \mathbb{R}$ the "δ–neighborhood" of f

$$N_\delta(f) \ := \ \{g(\cdot) : X \to \mathbb{R} : d(f,g) < \delta\} \ = \ \{g(\cdot) : X \to \mathbb{R} : \sup_{x \in X}|f(x) - g(x)| < \delta\} \tag{10.4}$$

If X is a subset of $\mathbb{R}$, you draw the graph of $f(\cdot) + \delta$ (the graph of $f(\cdot)$ shifted up north by the amount of δ) and the graph of $f(\cdot) - \delta$ (the graph of $f(\cdot)$ shifted down south by the amount of δ). Any function $g(\cdot)$ which stays completely inside this band, without actually touching it, belongs to the δ–neighborhood of $f(\cdot)$.

In other words assuming that the domain A is a single, connected chunk and not a collection of more than one separate intervals, the δ–neighborhood of $f(\cdot)$ is a "band" whose contours are made up on the left and right by two vertical lines and on the top and bottom by two lines that look like the graph of $f(\cdot)$ itself but have been shifted up and down by the amount of δ.

The distance of a real function $f(\cdot)$ to the zero function (see 8.3 on 114) has a special notation.

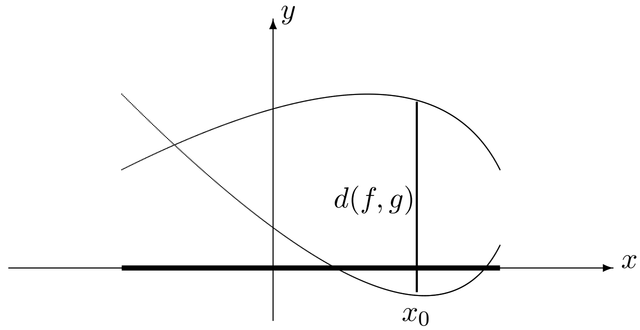


Figure 10.1: Distance of two real functions.

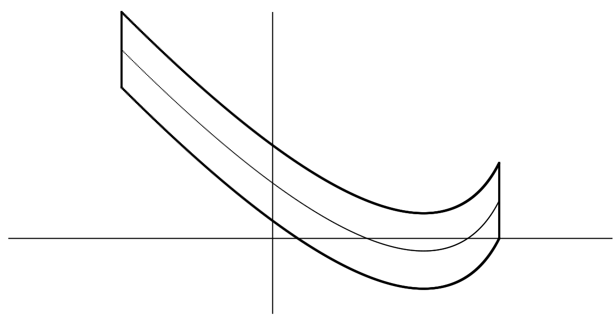


Figure 10.2: δ–neighborhood of a real function.

Definition 10.3 (Norm of bounded real functions)**.** Let X be an arbitrary, non-empty set. Let $f(\cdot) : X \to \mathbb{R}$ be a bounded real function on X, i.e., there exists a (possibly very large) number K such that $|f(x)| \leqq K$ for all $x \in X$. We define

$$\|f(\cdot)\|_\infty \; := \; \sup\{|f(x)| : x \in X\}$$

You can see that for any two bounded real functions $f(\cdot), g(\cdot)$ we have

$$\|f - g\|_\infty \; = \; \sup\{|f(x) - g(x)| : x \in X\} = d(f, g).$$

Proposition 10.1 (Properties of the sup norm of a real function)**.** *Let X be an arbitrary, non–empty set. Let*

$$\mathscr{B}(X, \mathbb{R}) := \{h(\cdot) : h(\cdot) \textit{ is a bounded real function on } X\}$$

Then the norm function

$$\|\cdot\|_\infty : \;\; \mathscr{B}(X, \mathbb{R}) \longrightarrow \mathbb{R}_+ \qquad\qquad h(\cdot) \longmapsto \|h(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\}$$

satisfies the three properties of a norm (see (10.2), p.151):

(10.5a) $\|f\|_\infty \geqq 0 \quad \forall f \in \mathscr{B}(X,\mathbb{R}) \quad and \quad \|f\|_\infty = 0 \;\Leftrightarrow\; f(\cdot) = 0$ ***positive definite***

(10.5b) $\|\alpha f(\cdot)\|_\infty = |\alpha| \cdot \|f(\cdot)\|_\infty \quad \forall f \in \mathscr{B}(X,\mathbb{R}), \forall \alpha \in \mathbb{R}$ ***homogeneity***

(10.5c) $\|f(\cdot) + g(\cdot)\|_\infty \leqq \|f(\cdot)\|_\infty + \|g(\cdot)\|_\infty \quad \forall\, f, g \in \mathscr{B}(X,\mathbb{R})$ ***triangle inequality***

Proof The proof is required as part of an upcoming homework.

Proposition 10.2 (Metric properties of the distance between real functions)**.** *Let X be an arbitrary, non–empty set.*
Let $\mathscr{B}(X,\mathbb{R}) := \{h(\cdot) : h(\cdot)$ is a bounded real function on $X\}$.
Let $f(\cdot), g(\cdot), h(\cdot) \in \mathscr{B}(X,\mathbb{R})$ Then the distance function

$$d(\cdot): \; \mathscr{B}(X,\mathbb{R}) \times \mathscr{B}(X,\mathbb{R}) \longrightarrow \mathbb{R}_+ \qquad (h_1,h_2) \longmapsto d(h_1,h_2) := \|h_1 - h_2\|_\infty$$

is a metric on $\mathscr{B}(X,\mathbb{R})$, i.e., it has the following three properties: [70]

(10.6a) $d(f,g) \geqq 0 \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R}) \quad and \quad d(f,g) = 0 \;\Leftrightarrow\; f(\cdot) = g(\cdot)$ ***positive definite***

(10.6b) $d(f,g) = d(g,f) \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R})$ ***symmetry***

(10.6c) $d(f,h) \leqq d(f,g) + d(g,h) \quad \forall\, f, g, h \in \mathscr{B}(X,\mathbb{R})$ ***triangle inequality***

We have seen in other contexts what those properties mean:
"Positive definite": The distance is never negative and two functions $f(\cdot)$ and $g(\cdot)$ have distance zero if and only if they are equal, i.e., if and only if $f(x) = g(x)$ for each argument $x \in X$.
*"symmetry": the distance from $f(\cdot)$ to $g(\cdot)$ is no different than that from $g(\cdot)$ to $f(\cdot)$. Symmetry implies that you do **not** obtain a negative distance if you walk in the opposite direction.*
"Triangle inequality": If you directly compare the maximum deviation between two functions $f(\cdot)$ and $h(\cdot)$ then this will never be more than than using an intermediary function $g(\cdot)$ and adding the distance between $f(\cdot) and g(\cdot)$ to that between $g(\cdot)$ and $h(\cdot)$.

Proof: This follows from prop.prop:norm-uc-f (Properties of the sup norm of a real function) on p.154 and thm.thm-x:norm-def-metric-spc (Norms define metric spaces) on p.151.

10.1.2 Neighborhoods and open sets

A. *Given a point $x_0 \in \mathbb{R}$ (a real number), we can look at*

(10.7)
$$\begin{aligned} N_\varepsilon(x_0) = (x_0 - \varepsilon, x_0 + \varepsilon) \;&= \{x \in \mathbb{R} : x_0 - \varepsilon < x < x_0 + \varepsilon\} \\ &= \{x \in \mathbb{R} : d(x, x_0) = |x - x_0| < \varepsilon\} \end{aligned}$$

which is the set of all real numbers x with a distance to x_0 of strictly less than a number ε (the open interval with end points $x_0 - \varepsilon$ and $x_0 + \varepsilon$). (see example (10.1) on p.152).

B. *Given a point $\vec{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ (a point in the xy–plane), we can look at*

(10.8)
$$\begin{aligned} N_\varepsilon(\vec{x}_0) \;&= \{\vec{x} \in \mathbb{R}^2 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} \\ &= \{(x,y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\} \end{aligned}$$

[70] If you forgot the meaning of $\mathscr{B}(X,\mathbb{R}) \times \mathscr{B}(X,\mathbb{R})$, it's time to review ch.4.1 (Cartesian products and relations) on p.69

which is the set of all points in the plane with a distance to $\vec{x}_0$ of strictly less than a number ε (the open disc around $\vec{x}_0$ with radius ε from which the points on the boundary (those with distance equal to ε) are excluded).

***C.** Given a point $\vec{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ (a point in the 3–dimensional space), we can look at*

$$N_\varepsilon(\vec{x}_0) = \{\vec{x} \in \mathbb{R}^3 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} = \{(x,y,z) \in \mathbb{R}^3 : (\vec{x} - \vec{x}_0)^2 + (\vec{y} - \vec{y}_0)^2 + (\vec{z} - \vec{z}_0)^2 < \varepsilon^2\} \tag{10.9}$$

which is the set of all points in space with a distance to $\vec{x}_0$ of strictly less than a number ε (the open ball around $\vec{x}_0$ with radius ε from which the points on the boundary (those with distance equal to ε) are excluded).

***D.** Given a normed vector space $(L, \|\cdot\|)$ and a vector $x_0 \in L$, we can look at*

$$N_\varepsilon(x_0) = \{x \in L : \|x - x_0\| < \varepsilon\} \tag{10.10}$$

which is the set of all vectors in L with a distance to x_0 of strictly less than a number ε (the open set around x_0 with "radius" ε from which the points on the boundary (those with distance equal to ε) are excluded).

There is one more item more general than neighborhoods of elements belonging to normed vector spaces, and that would be neighborhoods in metric spaces. We have arrived at the final definition:

Definition 10.4 (ε-Neighborhood)**.** Given a metric space (X, d) and an element $x_0 \in X$, let

$$N_\varepsilon(x_0) = \{x \in X : d(x, x_0) < \varepsilon\} \tag{10.11}$$

be the set of all elements of X with a distance to x_0 of strictly less than the number ε (the open set around x_0 with "radius" ε from which the points on the boundary (those with distance equal to ε) are excluded). We call $N_\varepsilon(x_0)$ the ε**–neighborhood** of x_0.

The following should be intuitively clear: Look at any point $a \in N_\varepsilon(x_0)$. You can find $\delta > 0$ such that the entire δ–neighborhood $N_\delta(a)$ of a is contained inside $N_\varepsilon(x_0)$. Just in case you do not trust your intuition, this is shown in prop. 10.3 just a little bit further down.

It then follows that any $a \in N_\varepsilon(x_0)$ is an interior point of $N_\varepsilon(x_0)$ in the following sense:

Definition 10.5 (Interior point)**.** Given is a metric space (X, d).
An element $a \in A \subseteq X$ is called an **interior point** of A if we can find some $\varepsilon > 0$, however small it may be, so that $N_\varepsilon(a) \subseteq A$.

Definition 10.6 (open set)**.** Given is a metric space (X, d).
A set all of whose members are interior points is called an **open set**.

Proposition 10.3. *$N_\varepsilon(x_0)$ is an open set*

It is worth while to examine the following proof closely because you can see how the triangle inequality is put to work.

$a \in N_\varepsilon(x_0)$ means that $\varepsilon - d(a, x_0) > 0$, say,

$$\varepsilon - d(a, x_0) = 2\delta \tag{10.12}$$

where $\delta > 0$. Let $b \in N_\delta(a)$. The claim is that any such b is an element of $N_\varepsilon(x_0)$. How so?

$$d(b, x_0) \leqq d(b, a) + d(a, x_0) \leqq \delta + d(a, x_0) < 2\delta + d(a, x_0) = \varepsilon$$

In the above chain, the first inequality is a consequence of the triangle inequality. The second one reflects the fact that $b \in N_\delta(a)$. The strict inequality is trivial because we added the strictly positive number δ. The final equality is a consequence of (10.12).

We have proved that for any $b \in N_\delta(a)$ it is true that $b \in N_\varepsilon(x_0)$ hence $N_\delta(a) \subseteq N_\varepsilon(x_0)$.

we showed earlier on that any $a \in N_\varepsilon(x_0)$ is an interior point of $N_\varepsilon(x_0)$. ■

Definition 10.7 (Neighborhoods in Metric Spaces)**.** Let (X, d) be a metric space, $x_0 \in X$. Any open set that contains x_0 is called an **open neighborhood** of x_0. Any superset of an open neighborhood of x_0 is called a **neighborhood** of x_0.

Remark 10.2 (Open neighborhoods are the important ones)**.** You will see that the important neighborhoods are the small ones, not the big ones. The definition above says that for any neighborhood A_x of a point $x \in X$ you can find an **open** neighborhood U_x of x such that $U_x \subseteq A_x$.

Because of this there are many propositions and theorems where you may assume that a neighborhood you deal with is open.

Theorem 10.2 (Metric spaces are topological spaces)**.** *The following is true about open sets of a metric space (X, d):*

(10.13a) *An arbitrary union* $\bigcup_{i \in I} U_i$ *of open sets U_i is open.*

(10.13b) *A finite intersection $U_1 \cap U_2 \cap \ldots \cap U_n$ $(n \in \mathbb{N})$ of open sets is open.*

(10.13c) *The entire set X is open and the empty set $\emptyset$ is open.*

Proof of a: Let $U := \bigcup_{i \in I} U_i$ and assume $x \in U$. We must show that x is an interior point of U. An element belongs to a union if and only if it belongs to at least one of the participating sets of the union. So there exists an index $i_0 \in I$ such that $x \in U_{i_0}$.

Because U_{i_0} is open, x is an interior point and we can find a suitable $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq U_{i_0}$. But $U_{i_0} \subseteq U$, hence $N_\varepsilon(x) \subseteq U$. It follows that x is interior point of U. But x was an arbitrary point of $U = \bigcup_{i \in I} U_i$ which therefore is shown to be an open set.

Proof of b: Let $x \in U := U_1 \cap U_2 \cap \ldots \cap U_n$. Then $x \in U_j$ for all $1 \leqq j \leqq n$ according to the definition of an intersection and it is inner point of all of them because they all are open sets. Hence, for each j there is a suitable $\varepsilon_j > 0$ such that $N_{\varepsilon_j}(x) \subseteq U_j$ Now define

$$\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n\}$$

Then $\varepsilon > 0$ and [71]

$$N_\varepsilon(x) \subseteq N_{\varepsilon_j}(x) \subseteq U_j \ (1 \leqq j \leqq n), \quad \textit{hence} \quad N_\varepsilon(x) \subseteq \bigcap_{j=1}^{n} U_j.$$

We have shown that an arbitrary $x \in U$ is interior point of U and this proves part b.

Proof of c: First we deal with the set X. Choose any $x \in X$. No matter how small or big an $\varepsilon > 0$ you choose, $N_\varepsilon(x)$ is a subset of X. But then x is an inner point of X, so all members of x are inner points and this proves that X is open.
Now to the empty set $\emptyset$. You may have a hard time to accept the logic of this statement: All elements of $\emptyset$ are interior points. But remember, the premise "let $x \in X$" is always false and you may conclude from it whatever you please (see ch.3 (Logic). ■

This last theorem provides the underpinnings for the definition of abstract topological spaces which will be touched upon in ch.10.1.4 on p.160.

10.1.3 Convergence

This chapter logically should be at the beginning of ch.10.1.8 (Contact points and closed sets) but we discuss convergence earlier because you are familiar with this concept from calculus and knowing about it will help you to better understand the structure of metric spaces.

Definition 10.8 (convergence of sequences)**.** Given is a metric space (X, d).
We say that a sequence (x_n) of elements of X **converges** to $a \in X$ for $n \to \infty$ if almost all of the x_n will come arbitrarily close to a in the following sense:

Let δ be an arbitrarily small positive real number. Then there is a (possibly extremely large) integer n_0 such that all x_j belong to $N_\delta(a)$ just as long as $j \geqq n_0$. To say this another way: Given any number $\delta > 0$, however small, you can find an integer n_0 such that

(10.14) $$d(a, x_j) < \delta \ \text{ for all } \ j \geqq n_0$$

We write either of

(10.15) $$a = \lim_{n\to\infty} x_n \qquad \text{or} \qquad x_n \to a$$

and we call a the **limit** of the sequence (x_n)

There is an equivalent way of expressing convergence towards a: No matter how small a neighborhood of a you choose: at most finitely many of the x_n will be located outside that neighborhood.

Exercise 10.1. Given is a metric space (X, d).
Prove the following: A sequence (x_n) of elements of X converges to $a \in X$ as $n \to \infty$ iff for any neighborhood U of a there exists some $n_0 \in \mathbb{N}$ such that the n_0–tail set $T_{n_0} = \{x_j : j \geqq n_0\}$ is contained in U (see def.8.12 (Tail sets of a sequence) on p.119.)

[71] by the way, this is the exact spot where the proof breaks down if you deal with an infinite intersection of open sets: the minimum would have to be replaced by an infimum and there is no guarantee that it would be strictly larger than zero.

Proof: Homework!

■

Theorem 10.3 (Limits in metric spaces are uniquely determined). *Let (X, d) be a metric space . Let $(x_n)_n$ be a convergent sequence in X Then its limit is uniquely determined.*

Proof: Otherwise there would be two different points $L_1, L_2 \in X$ such that both $\lim\limits_{n\to\infty} x_n = L_1$ and $\lim\limits_{n\to\infty} x_n = L_2$ Let $\varepsilon := d(L_1, L_2)/2$. There will be $N_1, N_2 \in \mathbb{N}$ such that

$$d(x_n, L_1) < \varepsilon \; \forall n \geqq N_1 \;\; andd(x_n, L_2) < \varepsilon \; \forall n \geqq N_2.$$

It follows that, for $n \geqq N_1 + N_2$,

$$d(L_1, L_2) \;\leqq\; d(L_1, x_n) + d(x_n, L_2) \;<\; 2\varepsilon \;=\; d(L_1, L_2)$$

and we have reached a contradiction. ■

Convergence is an extremely important concept in Mathematics, but it excludes the case of sequences such as $x_n := n$ and $y_n := -n$ ($n \in \mathbb{N}$). Intuition tells us that x_n converges to ∞ and y_n converges to $-\infty$ because we think of very big numbers as being very close to $+\infty$ and very small numbers (i.e., very big ones with a minus sign) as being very close to $-\infty$.

Definition 10.9 (Limit infinity). For this definition we do not deal with an arbitrary metric space but specifically with $X = \mathbb{R}$ and $d(x, y) = |b - a|$. Given a real number $K > 0$, we define

(10.16a) $$B_K(\infty) := \{x \in \mathbb{R} : x > K\}$$

(10.16b) $$B_K(-\infty) := \{x \in \mathbb{R} : x < -K\}$$

We call $B_K(\infty)$ the K-neighborhood of ∞ and $B_K(-\infty)$ the K-neighborhood of $-\infty$.
We say that a sequence (x_n) has limit ∞ and we write either of

(10.17) $$x_n \to \infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = \infty$$

if the following is true for any (big) K: There is a (possibly extremely large) integer n_0 such that all x_j belong to $B_K(\infty)$ just as long as $j \geqq n_0$.
We say that the sequence (x_n) has limit $-\infty$ and we write either of

(10.18) $$x_n \to -\infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = -\infty$$

if the following is true for any (big) K: There is a (possibly extremely large) integer n_0 such that all x_j belong to $B_K(-\infty)$ just as long as $j \geqq n_0$.

Note 10.1 (Notation for limits of monotone sequences). Let (x_n) be a non-decreasing sequence of real numbers and let y_n be a non-increasing sequence. If $\xi = \lim_{k\to\infty} x_k$ (that limit might be $+\infty$) then we write suggestively

$$x_n \nearrow \xi \quad (n \to \infty)$$

If $\eta = \lim_{j\to\infty} y_j$ (that limit might be $-\infty$) then we write suggestively

$$y_j \searrow \eta \quad (j \to \infty)$$

Remark 10.3 (No convergence or divergence to infinity)**.**

> The majority of mathematicians agrees that there is no "convergence to ∞" or "divergence to ∞". Rather, they will state that a sequence has the limit ∞.

10.1.4 Abstract topological spaces

Theorem 10.2 on p.157 gives us a way of defining neighborhoods for sets which do not have a metric.

Definition 10.10 (Abstract topological spaces)**.** Let X be an arbitrary non-empty set and let $\mathfrak{U}$ be a set of subsets [72] of X whose members satisfy the properties a, b and c of (10.13) on p.157:

(10.19a) An arbitrary union $\bigcup\limits_{i \in I} U_i$ of sets $U_i \in \mathfrak{U}$ belongs to $\mathfrak{U}$,

(10.19b) $U_1, U_2, \dots, U_n \in \mathfrak{U}\ (n \in \mathbb{N}) \quad \Rightarrow \quad U_1 \cap U_2 \cap \ldots \cap U_n \in \mathfrak{U}$,

(10.19c) $X \in \mathfrak{U} \quad \text{and} \quad \emptyset \in \mathfrak{U}$.

Then $(X, \mathfrak{U})$ is called a **topological space** The members of $\mathfrak{U}$ are called "open sets" of $(X, \mathfrak{U})$ and the collection $\mathfrak{U}$ of open sets is called the **topology** of X.

Definition 10.11 (Topology induced by a metric)**.** Let (X, d) be a metric space and let $\mathfrak{U}_d$ be the set of open subsets of (X, d), i.e., all sets $U \in X$ which consist of interior points only: for each $x \in U$ there exist $\varepsilon > 0$ such that

$$N_\varepsilon(x) \;=\; \{y \in X : d(x,y) < \varepsilon\} \;\subseteq\; U$$

(see (10.5) on p.156). We have seen in theorem (10.2) that those open sets satisfy the conditions of the previous definition. In other words, $(X, \mathfrak{U}_d)$ defines a topological space. We say that its topology is **induced by the metric** $d(\cdot,\cdot)$ or that it is **generated by the metric** $d(\cdot,\cdot)$. If there is no confusion about which metric we are talking about, we also simply speak about the **metric topology**.

Let X be a vector space with a norm $\|\cdot\|$. Remember that any norm defines a metric $d_{\|\cdot\|}(\cdot,\cdot)$ via $d_{\|\cdot\|}(x,y) = \|x-y\|$ (see (10.1) on p.151). Obviously, this norm defines open sets

$$\mathfrak{U}_{\|\cdot\|} \;:=\; \mathfrak{U}_{d_{\|\cdot\|}}$$

on X by means of this metric. We say that this topology is **induced by the norm** $\|\cdot\|$ or that it is **generated by the norm** $\|\cdot\|$. If there is no confusion about which norm we are talking about, we also simply speak about the **norm topology**.

Example 10.5 (Discrete topology)**.** Let X be non–empty. We had defined in (10.4) on p.152 the discrete metric as

$$d(x,y) \;=\; \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y. \end{cases}$$

[72] We encountered subsets of 2^X with special properties previously when looking at rings of sets in def.**??** (Rings and algebras of sets) on p.**??**.

The associated topology is

$$\mathfrak{U}_d \;=\; \{A : A \subseteq X\}.$$

Note that the discrete metric defines the biggest possible topology on X, i.e., the biggest possible collection of subsets of X whose members satisfy properties a, b, c of definition 10.10 on p.160. We call this topology the **discrete topology** of X.

Example 10.6 (Indiscrete topology)**.** Here is an example of a topology which is not generated by a metric. Let X be an arbitrary non–empty set and define $\mathfrak{U} := \{\emptyset, X\}$. Then $(X, \mathfrak{U})$ is a topological space. This is trivial because any intersection of members of $\mathfrak{U}$ is either $\emptyset$ (if at least one member is $\emptyset$) or X (if all members are X). Conversely, any union of members of $\mathfrak{U}$ is either $\emptyset$ (if all members are $\emptyset$) or X (if at least one member is X).

The topology $\{\emptyset, X\}$ is called the **indiscrete topology** of X. It is the smallest possible topology on X.

Definition 10.12 (Base of the topology)**.** Let $(X, \mathfrak{U})$ be a topological space.

A subset $\mathfrak{B} \subseteq \mathfrak{U}$ of open sets is called a **base of the topology** if any nonempty open set U can be written as a union of elements of $\mathfrak{B}$:

$$U \;=\; \bigcup_{i \in I} B_i \quad (B_i \in \mathfrak{B} \text{ for all } i \in I) \tag{10.20}$$

where I is a suitable index set which of course will in general depend on U.

We note that, because X is open, the above implies that $X \;=\; \bigcup \;[B : B \in \mathfrak{B}]$

definition (neighborhoods and interior points):

Let $x \in X$ and $A \subseteq X$. It is not assumed that A be open. A is called a **neighborhood** of x and x is called an **interior point** of A if you can find an open set U such that

$$x \;\in\; U \;\subseteq A. \tag{10.21}$$

Definition 10.13 (Second axiom of countability)**.** Let $(X, \mathfrak{U})$ be a topological space.
We say that X satisfies the **second axiom of countability** or X is **second countable** if we can find a countable base for $\mathfrak{U}$.

Theorem 10.4 (Euclidean space $\mathbb{R}^N$ is second countable)**.** *Let*

$$\mathfrak{B} \;:=\; \{\, N_{1/n}(q) : q \in \mathbb{Q}^N,\; n \in \mathbb{N} \,\}. \tag{10.22}$$

Here $\mathbb{Q}^N \;=\; \{q = (q_1, \dots, q_N) : q_j \in \mathbb{Q},\; 1 \leqq j \leqq N\}$ is the set of all points in $\mathbb{R}^N$ with rational coordinates. Then $\mathfrak{B}$ is a countable base.

Proof (outline): We recall from cor.7.4 on p. 113 that $\mathbb{Q}^N$is countable. Let $U \in \mathfrak{U}$ be an arbitrary open set in X. Any $x \in U$ is inner point of U, hence we can find some (large) integer n_x such that the entire $3/n_x$–neighborhood $N_{3/n_x}(x)$ is contained within U. As any vector can be approximated by vectors with rational coordinates, there exists $q = q_x \in \mathbb{Q}^N$ such that $d(x, q_x) < 1/n_x$. Draw a picture and you see that both $x \in N_{1/n_x}(q_x)$ and $N_{1/n_x}(q_x) \subseteq N_{3/n_x}(x)$. In other words, we have

$$x \;\in\; N_{1/n_x}(q_x) \;\subseteq\; U$$

for all $x \in U$. But then

$$U \subseteq \bigcup \left[\ N_{1/n_x}(q_x) : x \in U\ \right] \ \subseteq U$$

and it follows that U is the (countable union of the sets $N_{1/n_x}(q_x)$. ■

We'll conclude this chapter with a summary of what we have learned about the classification of sets with a concept of closeness of points.

Remark 10.4 (Classification of topological spaces)**.** We have seen the following:

A. All finite dimensional vector spaces $\mathbb{R}^N$ are inner product spaces (see (9.8) on p.146).

B. All inner product spaces are normed spaces (see (9.2) on p.148).

C. All normed spaces are metric spaces (see (10.1) on p.151).

D. All metric spaces are topological spaces. (see (10.10) on p.160 and (10.11) on p.160).

10.1.5 Neighborhood bases (⋆)

Note that this chapter is starred, hence optional.

Definition 10.14 (Neighborhood base)**.** Let $(X, \mathfrak{U})$ be a topological space.

The set of subsets of X

(10.23) $$\mathfrak{N}(x) \ := \ \{A \subseteq X : A \text{ is a neighborhood of } x\}$$

is called the **neighborhood system of** x

Given a point $x \in X$, any subset $\mathfrak{B} := \mathfrak{B}(x) \subseteq \mathfrak{N}(x)$ of the neighborhood system of x is called a **neighborhood base of** x if it satisfies the following condition: For any $A \in \mathfrak{N}(x)$ you can find a $B \in \mathfrak{B}(x)$ such that $B \subseteq A$.

In many propositions where proving closeness to some element is the issue, It often suffices to show that something is true for all sets that belong to a neighborhood base of x rather than having to show it for all neighborhoods of x. The reason is that often only the small neighborhoods matter and a neighborhood basis has "enough" of those.

Definition 10.15 (First axiom of countability)**.** Let $(X, \mathfrak{U})$ be a topological space.
We say that X satisfies the **first axiom of countability** or X is **first countable** if we can find for each $x \in X$ a countable neighborhood base.

Proposition 10.4 (ε-neighborhoods are a base of the topology)**.** *Let (X, d) be a metric space. Then the set $\mathscr{B}_1 \ := \ \{N_\varepsilon(x) : x \in X, \varepsilon > 0\}$ is a base for the topology of (X, d) (see 10.12 on p.161) and the same is true for the "thinner" set $\mathscr{B}_2 \ := \ \{N_{1/n}(x) : x \in X, n \in \mathbb{N}\}$.*

Proof: To show that $\mathscr{B}_1$ (resp., $\mathscr{B}_2$) is a base we must prove that any open subset of X can be written as a union of (open) sets all of which belong to $\mathscr{B}_1$ (resp., $\mathscr{B}_2$). We prove this for $\mathscr{B}_2$.

Let $U \subseteq X$ be open. As any $x \in U$ is an interior point of U we can find some $\varepsilon = \varepsilon(x) > 0$ such that $N_{\varepsilon(x)}(x) \subseteq U$. We note that for any such $\varepsilon(x)$ there is $n(x) \in \mathbb{N}$ such that $1/n(x) \leqq \varepsilon(x)$.

We observe that $U \subseteq \bigcup \left[N_{1/n(x)}(x) : x \in U \right] \subseteq U$.

The first inclusion follows from the fact that $\{x\} \subseteq N_{1/n(x)}(x)$ *for all* $x \in U$ *and the second inclusion follows from* $N_{1/n(x)}(x) \subseteq U$ *and the inclusion lemma (lemma 5.1 on p.92).*

It follows that $U = \bigcup \left[N_{1/n(x)}(x) : x \in U \right]$ *and we have managed to represent our open* U *as a union of elements of* $\mathscr{B}_2$*. This proves that* $\mathscr{B}_2$ *is a base for the topology of* (X, d)*.*

As $\mathscr{B}_2 \subseteq \mathscr{B}_1$ *it follows that* $\mathscr{B}_1$ *also is such a base.* ■

Theorem 10.5 (Metric spaces are first countable)**.** *Let* (X, d) *be a metric space. Then* X *is first countable.*

Proof (outline): For any $x \in X$ *let*

$$\mathfrak{B}(x) := \{ N_{1/n}(x) : n \in \mathbb{N} \}. \tag{10.24}$$

Then $\mathfrak{B}(x)$ *is a neighborhood base of* x*.* ■

10.1.6 Metric Subspaces

Definition 10.16 (Metric subspaces)**.** Given is a metric space (X, d) and a non–empty $A \subseteq (X, d)$. Let $d\big|_{A \times A} : A \times A \to \mathbb{R}_{\geq 0}$ be the restriction $d\big|_{A \times A}(x, y) := d(x, y)(x, y \in A)$ of the metric d to $A \times A$ (see def.4.10 on p.84). It is trivial to verify that $(A, d\big|_{A \times A})$ is a metric space in the sense of def.10.1 on p.150. We call $(A, d\big|_{A \times A})$ a **metric subspace** of (X, d) and we call $d\big|_{A \times A}$ the **metric induced by** d or the **metric inherited from** (X, d) .

Remark 10.5.

> Metric subspaces comes with their own collections of open and closed sets, neighborhoods, ε-neighborhoods, convergent sequences, ...
> You must watch out when looking at statements and their proofs whether those concepts refer to the entire space (X, d) or to the subspace $(A, d\big|_{AxA})$.

Notations 10.1.

a) Because the only difference between d and $d_{A \times A}$ is the domain, it is customary to write d instead of $d_{A \times A}$ to make formulas look simpler if doing so does not give rise to confusions.

b) We often shorten "open in $(A, d\big|_{A \times A})$" to "open in A", "closed in $(A, d\big|_{A \times A})$" to "closed in A", "convergent in $(A, d\big|_{A \times A})$" to "convergent in A", ...

10.1. DEFINITION (Traces of sets in a metric subspace)

Let (X, d) be a metric space and $A \subseteq X$ a nonempty subset of X, viewed as a metric subspace $(A, d\big|_{A \times A})$ of (X, d) (see def.10.16 on p.163). Let $Q \subseteq X$. We call $Q \cap A$ the **trace** of Q in A.

For $\varepsilon > 0$ and $a \in A$ let $N_\varepsilon(a)$ be the ε-neighborhood of a (in (X,d)). We write

$$N_\varepsilon^A(a) \;:=\; N_\varepsilon(a) \cap A,$$

i.e., $N_\varepsilon^A(a)$ is defined as the trace of $N_\varepsilon(a)$ in A.

Proposition 10.5 (Open sets in A as traces of open sets in X)**.** *Let (X,d) be a metric space and $A \subseteq X$ a nonempty subset of X.*

a. *Let $\varepsilon > 0$ and $a \in A$. Then*

$$N_\varepsilon^A(a) \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \varepsilon\}. \tag{10.25}$$

Because

$$N_\varepsilon^A(a) \;=\; N_\varepsilon(a) \cap A. \tag{10.26}$$

It follows that each ε-neighborhood in the subspace A is the trace of an ε-neighborhood in X.

b. *More generally, a set $U \subseteq A$ is open in A if and only if there is an open $V \subseteq$ in (X,d) such that*

$$U \;=\; V \cap A, \tag{10.27}$$

i.e., U is the trace of a set V which is open in X. [73]

*Proof of **a**: First we prove (10.26). As $d\big|_{A\times A}$ is the restriction of d to $A \times A$ it follows that*

$$\begin{aligned} N_\varepsilon^A(a) \;&= N_\varepsilon(a) \cap A \;=\; \{x \in X : d(x,a) < \varepsilon\} \cap A \\ &= \{x \in A : d\big|_{A\times A} < \varepsilon\} \cap A \;=\; \{x \in A : d\big|_{A\times A} < \varepsilon\} \end{aligned}$$

*This finishes the proof of **a.***

*Proof of **b**: First we show that if V is open in X then $U := V \cap A$ is open in the subspace A.*

Let $x \in U$. We must prove that x is an interior point of U with respect to $(A, d\big|_{A\times A})$ of (X,d).

Because $x \in V$ and V is open in X, there is $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq V$. It follows that

$N_\varepsilon^A(x) = N_\varepsilon(x) \cap A \subseteq V \cap A = U$ and $N_\varepsilon^A(x)$ is open in A, hence x is interior point of U with respect to the subspace $(A, d\big|_{A\times A})$.

Finally we prove that if $U \in A$ is open in A there is $V \subseteq X$ open in X such that $U = V \cap A$:

We can write $U \;=\; \bigcup \big[N_{\varepsilon(x)}^A(x) : x \in U\big]$ for suitable $\varepsilon(x) > 0$ (see the proof of prop.10.4 on p.162).

Let $V \;:=\; \bigcup \big[N_{\varepsilon(x)}(x) : x \in U\big]$ we have

$$\begin{aligned} V \cap A \;&= A \cap \bigcup \big[N_{\varepsilon(x)}(x) : x \in U\big] \;=\; \bigcup \big[N_{\varepsilon(x)}(x) \cap A : x \in U\big] \\ &= \bigcup \big[N_{\varepsilon(x)}^A(x) : x \in U\big] \;= U \end{aligned}$$

(the second equalitity follows from prop.5.1 on p.93). This finishes the proof. ■

[73] This proposition justifies to define subspaces of abstract topological spaces as follows: Let $(X, \mathfrak{U}$ be a topological space and $A \subseteq X$. We say that $V \subseteq A$ is **open in A** if V is the trace of an open set in X, i.e., if there is some $U \in \mathfrak{U}$ such that $V = U \cap A$. We denote the collection of all open sets in A as $\mathfrak{U}_A$.

Remark 10.6 (Convergence does not extend to subspaces)**.** Let $A \subseteq (X, d)$ and $x_n \in A$. Note that convergence of the sequence x_n in the space (X, d) (i.e., there exists $x \in X$ such that $x = \lim\limits_{n\to\infty} x_n$), does NOT imply convergence of the sequence in the space $(A, d\big|_{A\times A})$: Such is only the case if $x \in A$.

10.1.7 Bounded sets and bounded functions

Definition 10.17 (bounded sets)**.** Given is a subset A of a metric space (X, d). The **diameter** of A is defined as

$$\text{diam}(\emptyset) := 0, \qquad \text{diam}(A) := \sup\{d(x,y) : x, y \in A\} \text{ if } A \neq \emptyset. \tag{10.28}$$

We call A a **bounded set** if $\text{diam}(A) < \infty$.

Proposition 10.6. *Given is a metric space (X, d) and a nonempty subset A. The following are equivalent:*

(10.29) ***A.*** *$diam(A) < \infty$ i.e., A is bounded.*

(10.30) ***B.*** *There is a $\gamma > 0$ and $x_0 \in X$ such that $A \subseteq N_\gamma(x_0)$.*

(10.31) ***C.*** *For all $x \in X$ there is a $\gamma > 0$ such that $A \subseteq N_\gamma(x)$.*

*Proof of "**B** ⇒ **A**": For any $x, y \in A$ we have*

$$d(x,y) \leqq d(x,x_0) + d(x_0,y) \leqq 2\gamma$$

and it follows that $diam(A) \leqq 2\gamma$.

*Proof of "**A** ⇒ **B**": Pick an arbitrary $x_0 \in A$ and let $\gamma := diam(A)$. Then*

$$y \in A \quad \to \quad d(x_0,y) \leqq \sup_{x\in A} d(x,y) \leqq \sup_{x,z\in A} d(x,z) = diam(A) = \gamma.$$

It follows that $A \subseteq N_\gamma(x_0)$. ■

*Proof of "**C** ⇒ **A**": We pick an arbitrary $x_0 \in A$ which is possible as A is not empty. Then there is $\gamma = \gamma(x_0)$ such that $A \subseteq N_\gamma(x_0)$. For any $y, z \in A$ we then have*

$$d(y,z) \leqq d(y,x_0) + d(x_0,z) \leqq 2\gamma$$

and it follows that $diam(A) \leqq 2\gamma < \infty$.

*Proof of "**A** ⇒ **C**": Given $x \in X$, pick an arbitrary $x_0 \in A$ and let $\gamma := d(x,x_0) + diam(A)$. Then*

$$\begin{aligned} y \in A \quad \to \quad d(x,y) &\leqq d(x,x_0) + d(x_0,y) \leqq d(x,x_0) + \sup_{u\in A} d(u,y) \\ &\leqq d(x,x_0) + \sup_{u,z\in A} d(u,z) = d(x,x_0) + diam(A) = \gamma. \end{aligned}$$

It follows that $A \subseteq N_\gamma(x)$.

Definition 10.18 (bounded functions)**.** Given is a metric space (X, d).
A real-valued function $f(\cdot)$ on X is called **bounded from above** if there exists a (possibly very large) number $\gamma_1 > 0$ such that

(10.32) $$f(x) \quad < \quad \gamma_1 \qquad \text{for all arguments } x.$$

It is called **bounded from below** if there exists a (possibly very large) number $\gamma_2 > 0$ such that

(10.33) $$f(x) \quad > \quad -\gamma_2 \qquad \text{for all arguments } x.$$

It is called a **bounded function** if it is both bounded from above and below. It is obvious that if you set $\gamma := \max(\gamma_1, \gamma_2)$ then bounded functions are exactly those that satisfy the inequality

(10.34) $$|f(x)| \quad < \quad \gamma \qquad \text{for all arguments } x.$$

We note that f is bounded if and only if its range $f(X)$ is a bounded subset of $\mathbb{R}$ (compare this to definition 8.9 on p.118 on supremum and infimum of functions)

10.1.8 Contact points and closed sets

*If you look at any **closed interval** $[a, b] = \{y \in \mathbb{R} : a \leqq y \leqq b\}$, of real numbers, then all of its points are interior points, except for the end points a and b. On the other hand, a and b are contact points according to the following definition which makes sense for any topological space (X, d).*

Definition 10.19 (Contact points)**.** Given is a metric space (X, d).
Let $A \subseteq X$ and $x \in X$ (x may or may not to belong to A). x is called a **contact point** [74] of A if

(10.35) $$A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x.$$

Theorem 10.6 (Criterion for contact points)**.** *Given is a metric space (X, d).
Let $A \subseteq X$ and $x \in X$. Then x is a contact point of A if and only if there exists a sequence $x_1, x_2, x_3, \dots$ of members of A which converges to x.*

Proof of "$\Rightarrow$": Let $x \in X$ be such that $N \cap A \neq \emptyset$ for any neighborhood N of x. Let $x_n \in N_{1/n}(x) \cap A$. Such x_n exists because the neighborhood $N_{1/n}(x)$ has nonempty intersection with A.

Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be chosen such that $1/\varepsilon < N$. This is possible because $\mathbb{N}$ is not bounded (above) in $\mathbb{R}$.

For any $j \geqq N$ we obtain $d(x_j, x) < 1/j \leqq 1/N < \varepsilon$. This proves $x_n \to x$.

Proof of "$\Leftarrow$" Let $x \in X$ and assume there is $(x_n)_n$ such that $x_n \in A$ and $x_n \to x$.

We must show that if U_x is a (open) neighborhood of x then $U_x \cap A \neq \emptyset$. Let $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq U_x$.

It follows from $x_n \to x$ that there is $N = N(\varepsilon)$ such that $x_n \in N_\varepsilon(x)$ for all $n \geqq N$, especially, $x_N \in N_\varepsilon(x)$. By assumption, $x_N \in A$, hence $x_N \in N_\varepsilon(x) \cap A \subseteq U_x \cap A$, hence $U_x \cap A \neq \emptyset$. ■

[74] German: Berührungspunkt - see [12] Von Querenburg, p.21

Note 10.2. Note that any $a \in A$ is a contact point of A but not necessarily the other way around:

a. Let $a \in A$. Then any neighborhood N_a of a contains a, hence $U_A \cap A$ is not empty, hence a is a contact point of A. This proves that any $a \in A$ is a contact point of A.

b. Here is a counterexample which shows that the converse need not be true.
Let $(X, d) := \mathbb{R}$ with the standard Euclidean metric and let A be the subset $]0, 1[$. We show now that 0 is a contact point of A.
Any (open) neighborhood N_0 of 0 contains for some small enough $\delta > 0$ the entire interval $] - \delta, \delta[$. We may assume that $\delta < 2$, i.e., $\delta/2 \in A$.
Clearly, $\delta/2 \in] - \delta, \delta[\subseteq N_0$.
It follows that $\delta/2 \in A \cap N_0$. As N_0 was an arbitrary neighborhood of 0, we have proved that 0 is a contact point of A, even though $0 \notin A$.

Note 10.3 (Contact points vs Limit points)**.** Besides contact points there also is the concept of a limit point. Here is the definition (see [9] Munkres, a standard book on topology):

Given is a metric space (X, d). Let $A \subseteq X$ and $a \in X$. a is called a **limit point** or **cluster point** or **point of accumulation** of A if any neighborhood U of a intersects A in at least one point <u>other than a</u>. This definition excludes "isolated points" [75] of A from being limit points of A.

Definition 10.20 (Closed sets)**.** Given is a metric space (X, d)
and a subset $A \subseteq X$. We call

$$\bar{A} := \{x \in X : x \text{ is a contact point of } A\} \tag{10.36}$$

the **closure** of A. A set that contains all its contact points is called a **closed set**.

Remark 10.7. It follows from note 10.2.**a** that $A \subseteq \bar{A}$.

Proposition 10.7. *The complement of an open set is closed.*

Proof of 10.7: Let A be an open set. Each point $a \in A$ is an interior point which can be surrounded by a δ-neighborhood $N_\delta(a)$ which, for small enough δ, will be entirely contained within A.

Let $B = A^{\complement} = X \setminus A$ and assume $x \in X$ is a contact point of B. We want to prove that B is a closed set, so we must show that $x \in B$.

We assume the opposite and show that this will lead to a contradiction. So let us assume that $x \notin B$.

That means, of course, that x belongs to B's complement which is A. But A is open, so x must necessarily be an interior point of A. This means that there is a neighborhood $N_\delta(x)$ surrounding x which is entirely contained in A and hence has no points in common with the complement B of A.

On the other hand we assumed that x is a contact point of B. That again means that there must be points in $B \cap N_\delta(x)$.

We have proved on the one hand that $N_\delta(x) \cap B = \emptyset$ and on the other hand that there must be points in B that also are contained in $N_\delta(x)$.

We have reached a contradiction. ■

[75] $a \in A$ is called an **isolated point** of A if there is a neighborhood U of a such that $U \cap A = \{a\}$.

Proposition 10.8. *The complement of a closed set is open.*

We will give two complete proofs of the above. The first one is based on the definition of contact points and is based on the concept of neighborhoods and interior points. The second proof works with the criterion for contact points (theorem 10.6) and works with sequences.

a. *First proof of prop.10.8:*
Let A be closed set. Let $B = A^{\complement} = X \setminus A$. If B is not open then there must some be $b \in B$ which is not an interior point of B.

We show now that this assumption leads to a contradiction.

Because b is not an interior point of B, there is no δ-neighborhood, for whatever small δ, that entirely belongs to B. So, for each $j \in \mathbb{N}$, there is an $x_j \in N_{1/j}(b)$ which does not belong to B.

We have constructed a sequence x_j which is entirely contained in A and which also converges to b. The latter is true because, for any j, all but finitely many members are contained in $N_{1/j}(b)$.

The closed set A contains all its contact points and it follows from the criterion for contact points that $b \in A$.

But we had assumed at the outset that $b \in B$ which is the complement of A and we have a contradiction. ■

b. *Alternate proof of prop.10.8 which is entirely based on the concept of neighborhoods and interior points:*

Let A be closed set. Let $B = A^{\complement} = X \setminus A$. Let $b \in B$.

The closed set A contains all its contact points, so $b \notin A$ implies that b is not a contact point of A: according to def.10.19 there exists some neighborhood V of b such that $V \cap A = \emptyset$, i.e., $V \subseteq A^{\complement} = B$.

We have proved that an arbitrary $b \in B$ is an interior point of B, i.e., the complement B of the closed set A is open. ■

Theorem 10.7 (Open iff complement is closed)**.** *Let (X, d) be a metric space and $A \subseteq X$. Then A is open if and only if $A^{\complement}$ is closed.*

Proof: Immediate from prop.10.7 and prop.10.8 ■

Remark 10.8. a. We have seen that def.10.19 for contact points and hence def.10.20 for closed sets are entirely based on the concept of neighborhood which itself is entirely based on that of open sets. It follows that those two definitions make perfect sense not only in metric spaces but, more generally, in abstract topological spaces $(X, \mathfrak{U})$ which are characterized by the set $\mathfrak{U}$ of all open subsets of X (see def.10.10 on p.160).

b. Moreover the proof for prop.10.7 (complements of open sets are closed) and the alternate proof for prop.10.8 (complements of closed sets are open) are based on those definitions and do not employ specific properties of metric spaces either; theorem 10.7 also works for abstract topological spaces.

c. Matter of fact, many books <u>define</u> closed sets as the complements of open sets and only afterwards define contact points as we did. No surprise then that our definition of closed sets becomes their theorem: It can be proved that closed sets are exactly those that contain all their contact points.

DEFINITION *(contact points and closed sets in topological spaces)* [76]

Given is an abstract topological space $(X, \mathfrak{U})$.

Let $A \subseteq X$ *and* $x \in X$ *(*x *may or may not to belong to* A*).* x *is called a* ***contact point*** *of* A *if*

$$A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x.$$

We call

$$\bar{A} \quad := \{x \in X : x \text{ is a contact point of } A\}$$

the ***closure*** *of* A*. A set that contains all its contact points is called a* ***closed set****.*

REMARK*:*

We note that $A \subseteq \bar{A}$*: Let* $a \in A$ *and let* V_a *be a neighborhood of* a*. Because* $a \in V_a$*, we obtain* $a \in V_a \cap A$*, hence* $V_a \cap A \neq \emptyset$*, hence* $a \in \bar{A}$*.*

It follows that A *is closed if and only if* $A = \bar{A}$ *(which justifies the name "closure of* A*" for* $\bar{A}$*.)*

Proposition 10.9. *Let* $(X, \mathfrak{U})$ *be a topological space.*

The closed sets of X *satisfy the following property:*

(10.37)
- ***a.*** *An arbitrary intersection of closed sets is closed.*
- ***b.*** *A finite union of closed sets is closed.*
- ***c.*** *The entire set* X *is closed and* $\emptyset$ *is closed.*

Remark 10.9. Take another look at definition 10.8 on p.158. A sequence converges to a point a if, for any $\varepsilon > 0$ there is an integer n_0 such that all x_j belong to $N_\varepsilon(a)$ just as long as $j \geqq n_0$. You can also redefine this for abstract topological spaces: For any (open) neighborhood U_a there is an integer n_0 such that all x_j belong to U_a for all $j \geqq n_0$.

You see that convergence of sequences is another concept that makes sense in any kind of topological spaces.

Proof of ***a****: The proof is an easy consequence of De Morgan's law (the duality principle for sets) (see (5.1) on p.93). Observe that* X *is a universal set because all members* U *of* $\mathfrak{U}$ *and their complements* $U^\complement$ *are subsets of* X*.*

Let (C_α) *be an arbitrary familiy of closed sets. Then* $U_\alpha := C_\alpha^\complement$ *is an open set for each* α*. Observe that* $C_\alpha^\complement = U_\alpha$ *because the complement of the complement of any set gives you back that set. Let* $C := \bigcap\limits_\alpha C_\alpha$*. Then*

$$C^\complement \;=\; \Big(\bigcap_\alpha C_\alpha\Big)^\complement = \bigcup_\alpha C_\alpha^\complement = \bigcup_\alpha U_\alpha$$

In other words $C^\complement$ *is an arbitrary union of open sets which is open by the very definition of open sets of a topological space. We have proved* ***a****.*

[76] see def.10.19 and def.10.20 for the following definitions in metric spaces.

*Proof of **b**: Let $C_1, C_2, \dots C_n$ be closed sets. Then $U_j := C_j^{\complement}$ is an open set for each j. Let $C := \bigcup_{1 \leqq j \leqq n} C_j$. Then*

$$C^{\complement} = \big(\bigcup_j C_j\big)^{\complement} = \bigcap_j C_j^{\complement} = \bigcap_j U_j$$

*Hence, $C^{\complement}$ is the intersection of finitely many open sets. This shows that $C^{\complement}$ is open, i.e., C is closed. We have proved **b**.*

*Proof of **c**: Trivial because*

$$X^{\complement} = \emptyset \qquad \text{and} \qquad \emptyset^{\complement} = X.$$

■

10.1.9 Completeness in metric spaces

Often you are faced with a situation where you need to find a contact point a and all you have is a sequence which behaves like one converging to a contact point in the sense of inequality 10.14 (page 158)

Definition 10.21 (Cauchy sequences)**.** Given is a metric space (X, d).
A sequence (x_n) in X is called a **Cauchy sequence** [77] or, in short, it is Cauchy if it has the following property: Given any whatever small number $\varepsilon > 0$, you can find a (possibly very large) number n_0 such that

(10.38) $$d(x_i, x_j) < \varepsilon \qquad \text{for all } i, j \geqq n_0$$

This is called the **Cauchy criterion for convergence** of a sequence.

Example 10.7 (Cauchy criterion for real numbers)**.** In $\mathbb{R}$ we have $d(x, y) = |x - y|$ and the Cauchy criterion requires for any given $\varepsilon > 0$ the existence of $n_0 \in \mathbb{N}$ such that

(10.39) $$|x_i - x_j| < \varepsilon \qquad \text{for all } i, j \geqq n_0$$

The following theorem of the completeness of the set of all real numbers [78] states that any Cauchy sequence converges to a real number. This is a big deal: To show that a sequence has a finite limit you need not provide the actual value of that limit. All you must show is that this sequence satisfies the Cauchy criterion. One can say that this preoccupation with proving existence rather than computing the actual value is one of the major points which distinguish Mathematics from applied Physics and the engineering disciplines.

Here is the formal definition of a complete set in a metric space.

Definition 10.22 (Completeness in metric spaces)**.** Given is a metric space (X, d). A subset $A \subseteq X$ is called **complete** if any Cauchy sequence (x_n) with elements in A converges to an element of A.

[77] so named after the great french mathematician Augustin–Louis Cauchy (1789–1857) who contributed massively to the most fundamental ideas of Calculus.

[78] Remember the completeness axiom for $\mathbb{R}$ (axiom 8.1 on p.116) which states that any subset A of $\mathbb{R}$ which possesses upper bounds has a least upper bound (the supremum $\sup(A)$). This axiom was needed to establish the validity of thm.8.2 (Characterization of limits via limsup and liminf) on p.126 which will be used now to prove the completeness of $\mathbb{R}$ as a metric space.

Remark:

a. In particular, X itself is complete if any Cauchy sequence in X converges.

b. A is complete as a subset of (X,d) iff $((A, d\big|_{A\times A})$ is complete "in itself".

Theorem 10.8 (Completeness of the real numbers)**.** *The following is true for the real numbers with the metric $d(a,b) = |b-a|$ but will in general be false for arbitrary metric spaces: Let (x_n) be a Cauchy sequence in $\mathbb{R}$. then there exists a real number L such that $L = \lim\limits_{n\to\infty} x_n$.*

Proof: Part 1: We first show that x_n is bounded. There is $N = N(1)$ such that $|x_i - x_j| < 1/2$ for all $i, j \geqq N$. In particular, $|x_i - x_N| < 1/2$.

Hence $|x_i| = |x_i - x_N + x_N| \leqq |x_i - x_N| + |x_N| \leqq |x_N| + 1$ for all $i \geqq N$.

Let $M := \max\{|x_j| : j \leqq N\}$. Then $|x_j| \leqq M + 1$ and we have proved that the sequence is bounded. It follows that $(x_n)_n$ possesses finite liminf and limsup. [79]

Part 2: We now show that $\liminf\limits_{n\to\infty} x_n \;=\; \limsup\limits_{n\to\infty} x_n$.

Let $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $|x_i - x_j| \leqq \varepsilon$ for all $i, j \geqq N$.

Let $T_n := \{x_j : j \geqq n\}$ be the tail set of the sequence $(x_n)_n$. Let $\alpha_N := \inf T_N, \beta_N := \sup T_N$.

There is some $i \geqq N$ such that $|x_i - \alpha_N| = x_i - \alpha_N \leqq \varepsilon$ and there is some $j \geqq N$ such that $|\beta_N - x_j| = \beta_N - x_j \leqq \varepsilon$. It follows that

$$0 \;\leqq\; \beta_N - \alpha_N \;= |\beta_N - \alpha_N| \;\leqq\; |(\beta_N - x_j) \;+\; (x_j - x_i) \;+\; (x_i - \alpha_N)| \leqq 3\varepsilon.$$

Further, if $k > N$ then $T_k \subseteq T_N$, hence $\alpha_k \geqq \alpha_N$ and $\beta_k \leqq \beta_N$. It follows that

$$\beta_k - \alpha_k \leqq \beta_N - \alpha_N \leqq 3\varepsilon.$$

But then

$$0 \;\leqq\; \limsup_{k\to\infty} x_k - \liminf_{k\to\infty} x_k \;=\; \lim_{k\to\infty} \beta_k - \lim_{k\to\infty} \alpha_k \;\leqq\; \beta_N - \alpha_N \;\leqq\; 3\varepsilon.$$

$\varepsilon > 0$ was arbitrary, hence $\limsup\limits_{k\to\infty} x_k \;=\; \liminf\limits_{k\to\infty} x_k$.

Part 3: It follows from theorem 8.2 on p.126 that the sequence $(x_n)_n$ converges to $L := \limsup\limits_{k\to\infty} x_k$ and the proof is finished. ■

Now that the completeness of $\mathbb{R}$ has been established, it is not very difficult to see that N–dimensional space $\mathbb{R}^N$ also is complete.

Theorem 10.9 (Completeness of $\mathbb{R}^N$)**.** *Let $(\vec{x}_n)$ be a Cauchy sequence in $\mathbb{R}^N$.*
Then there exists a vector $\vec{a} \in \mathbb{R}^N$ such that $\vec{a} = \lim\limits_{n\to\infty} \vec{x}_n$.

[79] See ch.8.2 (Maxima, suprema, limsup ...).

Proof (outline): Let $\vec{x}_j = (x_{j,1}, x_{j,2}, \dots, x_{j,N})$ be Cauchy in $\mathbb{R}^N$. For fixed k, each coordinate sequence $(x_{j,k})_j$ is Cauchy because, if $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that if $i, j \geqq K$ then $\|\vec{x}_i - \vec{x}_j\|_2 < \varepsilon$. Hence

$$|x_{i,k} - x_{j,k}| \;=\; \sqrt{|x_{i,k} - x_{j,k}|^2} \;\leqq\; \sqrt{\sum_{k=1}^{N} |x_{i,k} - x_{j,k}|^2} \;=\; \|\vec{x}_i - \vec{x}_j\|_2 \;<\; \varepsilon.$$

It follows from the completeness of $\mathbb{R}$ as a metric space that there exist real numbers

$$a_1, a_2, a_3, \dots, a_N \qquad \text{such that } a_k \;=\; \lim_{n\to\infty} x_{n,k} \; (1 \leqq k \leqq N).$$

For a given number ε we can find natural numbers $n_{0,1}, n_{0,2}, \dots, n_{0,N}$ such that

$$|x_{n,k} - a_k| \;<\; \frac{\varepsilon}{N} \quad \text{for all } n \geqq n_{0,j} \text{ and for all } 1 \leqq k \leqq N.$$

Let $n^\star := \max(n_{0,1}, n_{0,2}, \dots, n_{0,N})$. It follows that

$$d(\vec{x}_n - \vec{a}) \;=\; \sqrt{\sum_{k=1}^{N} |x_{n,k} - a_k|^2} \;\leqq\; \sqrt{N \cdot \left(\frac{\varepsilon}{N}\right)^2} \;=\; \frac{\varepsilon}{\sqrt{N}} \;\leqq\; \varepsilon \qquad \text{for all } n \geqq n^\star. \;\blacksquare$$

Example 10.8 (Approximation of decimals)**.** The following illustrates Cauchy sequences and completeness in $\mathbb{R}$. Take any real number $x \geqq 0$ and write it as a decimal:

$$x \;=\; m \;+\; \sum_{j=1}^{\infty} d_j \cdot 10^{-j} \quad (m \in \mathbb{Z},\; d_j \in \{0, 1, 2, \dots, 9\})$$

As I explained in (2.9) on (p.15), anything that can be written as a decimal number is a real number. Let's say, x starts out on the left as

$$x \;=\; 258.1408926584207531\dots$$

If we define as x_k the leftmost part of x, truncated k digits after the decimal points:

$$x_1 \;=\; 258.1, \quad x_2 \;=\; 258.14, \quad x_3 \;=\; 258.140, \quad x_4 \;=\; 258.1408, \quad x_5 \;=\; 258.14089, \quad \dots$$

and as y_k the leftmost part of x, truncated k digits after the decimal points, but the rightmost digit incremented by 1 (where you then might obtain a carry-over to the left when you add 1 to 9)

$$y_1 \;=\; 258.2, \quad y_2 \;=\; 258.15, \quad y_3 \;=\; 258.141, \quad y_4 \;=\; 258.1409, \quad y_5 \;=\; 258.14090, \quad \dots$$

then the sequence (x_n) is non-decreasing: $x_{n+1} \geqq x_n$ for all n and the sequence (y_n) is non-increasing: $y_{n+1} \leqq y_n$ for all n. We have the sandwich property: $x_n \leqq x \leqq y_n$ for all n. Both sequences are Cauchy because both

$$d(x_{n+i}, x_{n+j}) \;= |x_{n+i} - x_{n+j}| \;\leqq\; 10^{-n} \;\to\; 0 \quad (n \to \infty)$$

$$d(y_{n+i}, y_{n+j}) \;= |y_{n+i} - y_{n+j}| \;\leqq\; 10^{-n} \;\to\; 0 \quad (n \to \infty)$$

It is obvious that $x \;=\; \lim_{n\to\infty} x_n \;=\; \lim_{m\to\infty} y_m$.

What just has been illustrated is that there a natural way to construct for a given $x \in \mathbb{R}$ Cauchy sequences that converge towards x. The completeness principle states that the reverse is true: For any Cauchy sequence there is an element x towards which the sequence converges.

We won't really talk about completeness in general until the chapter on compact spaces. Just to mention one of the simplest facts about completeness:

Theorem 10.10 (Complete sets are closed)**.** *Any complete subset of a metric space is closed.*

Proof: Let (X, d) be a metric space and $A \subseteq X$. Let a be a contact point of A. The theorem is proved if we can show that $a \in A$.

a) *We employ prop.10.19 on p.166: A point $x \in X$ is a contact point of A if and only if $A \cap V \neq \emptyset$ for any neighborhood V of x.*

Let $m \in \mathbb{N}$. Then $N_{1/m}(a)$ is a neighborhood of the contact point a, hence hence $A \bigcap N_{1/m}(a) \neq \emptyset$ and we can pick a point from this intersection which we name x_m.

b) *We prove next that $(x_m)_m$ is Cauchy. Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be such that $N > 1/\varepsilon$. if $j \in \mathbb{N}$ and $k \in \mathbb{N}$ both exceed N then*

$$d(x_j, x_k) \leqq d(x_j, a) + d(a, x_k) \leqq \frac{1}{j} + \frac{1}{k} \leqq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It follows that the sequence (x_j) is Cauchy.

c) *Because A is complete, this sequence must converge to some $b \in A$. But b cannot be different from a Otherwise we could "separate" a and b by two disjoint neighborhoods: choose the open ρ–balls $N_\rho(a)$ and $N_\rho(b)$ where ρ is one half the distance between a and b (see the proof of thm.10.3 on p.159).*

Only finitely many of the x_n are allowed to be outside $N_\rho(a)$ and the same is true for $N_\rho(b)$. That is a contradiction and it follows that $b = a$, i.e., $a \in A$.

d) *We summarize: if a is a contact point of A then $a \in A$. It follows that A is closed.* ■

The following is the reverse of thm.10.10.

Theorem 10.11 (Closed subsets of a complete space are complete)**.** *Let (X, d) be a complete metric space and let $A \subseteq X$ be closed. Then A is complete.*

Proof: Let $(x_n)_n$ be a Cauchy sequence in A. We must show that there is $a \in A$ such that $x_n \to a$. (x_n) also is Cauchy in X because the Cauchy criterion is entirely specified in terms of members of the sequence (x_n).

Because X is complete there exists $x \in X$ such that $x_n \to x$. If we can show that x is a contact point of A then we are done:

As the set A is assumed to be closed, it contains all its contact points. It follows that $x \in A$, i.e., the arbitrary Cauchy sequence (x_n) in A converges to an element of A. We conclude that A is complete. ■

Theorem 10.12 (Convergent sequences are Cauchy)**.** *Let $(x_{n_j})_n$ be a convergent sequence in a subset A of a metric space (X, d). Then $(x_{n_j})_n$ is a Cauchy sequence (in A).*

Proof: Let $x_n \to L$ ($L \in A$). Let $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that

(10.40) $$k \geqq N \Rightarrow d(x_k, L) < \varepsilon/2. \qquad (\star)$$

It follows from $(\star)$ *that, for any* $i, j \geqq N$,

(10.41) $$i, j \geqq N \;\Rightarrow\; d(x_i, x_j) \;\leqq\; d(x_i, L) + d(L, x_j) \;<\; \varepsilon/2 + \varepsilon/2 \;=\; \varepsilon.$$

It follows that the sequence satisfies (10.38) *of the definition* *10.21* *on p.170 of a Cauchy sequence.* ■

10.2 Continuity (Study this!)

10.2.1 Definition and characterization of continuous functions

Informally speaking a continuous function

$$f(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R} \qquad x \longmapsto y = f(x)$$

is one whose graph in the xy plane is a continuous line without any disconnections or gaps. This can be stated slightly more formal by saying that if the x-values are closely together then the $f(x)$-values must be closely together too. The latter makes sense for any sets X, Y where closeness can be measured, i.e., for metric spaces (X, d_1) and (Y, d_2). Here is the formal definition:

Definition 10.23 (Sequence continuity)**.** Given are two metric spaces (X, d_1) and (Y, d_2). Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. We say that $f(\cdot)$ is **sequence continuous at $\boldsymbol{x_0}$** and we write

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{10.42}$$

if the following is true for any sequence (x_n) with values in A:

$$\text{if } x_n \to x_0 \text{ then } f(x_n) \to f(x_0). \tag{10.43}$$

We say that $f(\cdot)$ is **sequence continuous** if $f(\cdot)$ is sequence continuous in a for all $a \in A$.

In other words, the following must be true for any sequence (x_n) in A

$$\lim_{n \to \infty} x_n = x_0 \quad \Rightarrow \quad \lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n) = f(x_0) \tag{10.44}$$

Remark 10.10. Important points to notice:

- **a)** It is not enough for the above to be true for some sequences that converge to x_0. Rather, it must be true for all such sequences!
- **b)** We restrict our universe to the domain A of f: x_0 and the entire sequence $(x_n)_{n \in \mathbb{N}}$ must belong to A because there must be function values for all x-values. In other words, f is continuous at $x_0 \in A$ if and only if f is continuous at x_0 in the metric subspace $(A, d\big|_{A \times A})$.

Definition 10.24 (ε-δ continuity)**.** Given are two metric spaces (X, d_1) and (Y, d_2). Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. We say that $f(\cdot)$ is **ε-δ continuous at $\boldsymbol{x_0}$**

if the following is true:

For any (whatever small) $\varepsilon > 0$ there exists a (most likely very small) $\delta > 0$ such that

$$f\big(N_\delta(x_0) \cap A\big) \ \subseteq \ N_\varepsilon(f(x_0)), \tag{10.45}$$

which is another way of saying that, for all $x \in A$,

$$d_1(x, x_0) < \delta \ \Rightarrow \ d_2(f(x), f(x_0)) < \varepsilon. \tag{10.46}$$

We say that $f(\cdot)$ is **ε-δ continuous** if $f(\cdot)$ is ε-δ continuous at a for all $a \in A$.

Remark 10.11. We recall from thm.10.25 on p.164 that

$$N_\varepsilon \cap A \;=\; N_\varepsilon^A(a) \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \varepsilon\}$$

and it follows that $f(\cdot)$ is ε-δ continuous at x_0 if and only if $f\big(N_\delta^A(x_0)\big) \;\subseteq\; N_\varepsilon(f(x_0))$.

Theorem 10.13 (ε-δ characterization of continuity)**.** *Let (X,d_1) and (Y,d_2) be two metric spaces. Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. Then f is sequence continuous at x_0 if and only if f is ε-δ continuous at x_0.*

In particular f is sequence continuous if and only if f is ε-δ continuous.

a) **⇒***: Proof that sequence continuity implies ε-δ-continuity:*
We assume to the contrary that there exists some function f which is sequence continuous but not ε-δ-continuous at x_0, i.e., there exists some $\varepsilon > 0$ such that neither (10.45) nor the equivalent (10.46) is true for any $\delta > 0$.

a.1. *In other words, No matter how small a δ we choose, there is at least one $x = x(\delta) \in A$ such that $d_1(x,x_0) < \delta$ but $d_2(f(x), f(x_0)) \geqq \varepsilon$. In particular we obtain for $\delta := 1/m (m \in \mathbb{N})$ that*

(10.47) *there exists some $x_m \in N_{1/m}(x_0) \cap A$; such that; $d_2(f(x_m), f(x_0)) \geqq \varepsilon$.*

a.2. *We now show that the sequence $(x_m)_{m\in\mathbb{N}}$ converges to x_0: Let $\gamma > 0$ and pick $N := N(\gamma) \in \mathbb{N}$ so big that $N > 1/\gamma$, i.e., $1/N < \gamma$. As $x_m \in N_{1/m}(x_0)$, we obtain for all $m \geqq N$ that*

$$d_1(x_m, x_0) < 1/m \leqq 1/N < \gamma$$

and this proves that $x_m \to x_0$.

a.3. *It is clear that the sequence $\big(f(x_m)\big)_{m\in\mathbb{N}}$ does not converge to $f(x_0)$ as that requires $d_2(f(x_m), f(x_0)) < \varepsilon$ for all sufficiently big m, contrary to (10.47) which implies that there is not even one such m. In other words, the function f is not sequence continuous, contrary to our assumption. We have our contradiction.*

b) **⇐***: Proof that ε-δ-continuity implies sequence continuity:*
Let $x_n \to x_0$. Let $y_n := f(x_n)$ and $y := f(x_0)$. We must prove that $y_n \to y$ as $n \to \infty$.

b.1. *Let $\varepsilon > 0$. We can find $\delta > 0$ such that (10.45) and hence (10.46) is satisfied. We assumed that $x_n \to x_0$. Hence there exists $N := N(\delta) \in \mathbb{N}$ such that $d_1(x_n, x_0) < \delta$ for all $n \geqq N$.*

b.2. *It follows from (10.46) that $d_2(y_n, y) = d_2(f(x_n), f(x_0)) < \varepsilon$ for all $n \geqq N$. In other words, $y_n \to y$ as $n \to \infty$ and the proof of "⇐" is finished.* ■

[1] B/G: Art of Proof defines in appendix A, p.136, continuity of a function f as follows: "f^{-1}(open) $=$ open". The following proposition proves that their definition coincides with the one given here: the validity of (10.45) for all $x_0 \in X$.

a) *In the interest of simplicity f now is defined on all of X and not just on some subset A of X. Note that the general case of $f : A \to Y$ is covered by replacing (X, d_1) with $(A, d_1\big|_{A\times A})$, i.e., we deal with $f : (A, d_1\big|_{A\times A}) \to (Y, d_2)$.*

b) *Also note that this next proposition addresses continuity of f for **all** $x \in X$ and **not** at a specific x_0.*

Proposition 10.10 ("$f^{-1}(\text{open}) = \text{open}$" continuity). *Let (X, d_1) and (Y, d_2) be two metric spaces and let $f(\cdot) : X \to Y$ be a mapping from X to Y. Then $f(\cdot)$ is continuous if and only if the following is true: Let V be an open subset of Y. Then the inverse image $f^{-1}(V)$ is open in X.*

Proof of "$\Rightarrow$": Let V be an open set in Y. Let $U := f^{-1}(V)$, $a \in U$ and $b := f(a)$. Then $b \in V$ by the definition of inverse images. b is inner point of the open set V and there is $\varepsilon > 0$ such that $N_\varepsilon(b) \subseteq V$.

It follows from def.10.24 (ε-δ continuity) that there is $\delta > 0$ such that $f(N_\delta(a)) \subseteq N_\varepsilon(b)$. It follows from the monotonicity of direct and inverse images and prop.6.1 on p.106 that

$$N_\delta(a) \subseteq f^{-1}\big(f(N_\delta(a))\big) \subseteq f^{-1}(N_\varepsilon(b)) \subseteq f^{-1}(V) = U.$$

It follows that the arbitrarily chosen $a \in U$ is an interior point of U and this proves that U is open.

Proof of "$\Leftarrow$": We now assume that all inverse images of open sets in Y are open in X. Let $a \in X, b = f(a)$ and $\varepsilon > 0$. We must find $\delta > 0$ such that $f(N_\delta(a)) \subseteq N_\varepsilon(b)$. Let $U := f^{-1}\big(N_\varepsilon(b)\big)$. Then U is open as the inverse image of the open neighborhood $N_\varepsilon(b)$ and there will be $\delta > 0$ such that $N_\delta(a) \subseteq U$. It follows from the monotonicity of direct and inverse images and prop.6.6 on p.106 that

$$f\big(N_\delta(a)\big) \subseteq f(U) = f\big(f^{-1}(N_\varepsilon(b))\big) = N_\varepsilon(b) \cap f(X) \subseteq N_\varepsilon(b).$$

■

Remark 10.12 (continuity for real functions of real numbers). Let $(X, d_1) = (Y, d_2) = \mathbb{R}$. In this case equation (10.46) on p.175 looks like this:

$$|x - x_0| < \delta \;\Rightarrow\; |f(x) - f(x_0)| < \varepsilon$$

Proposition 10.11 (continuity of the identity mapping). *Let $X, d)$ be a metric space and*

$$id(\cdot) : E \to E \qquad x \mapsto x$$

be the identity function on E. Then $id(\cdot)$ is continuous.

Proof: Given any $\varepsilon > 0$, let $\delta := \varepsilon$. Let $x, y \in X$. Assume that $d(x, y) < \delta$. Then

$$d(id(x), id(y)) = d(x, y) < \delta = \varepsilon$$

and we have satisfied condition (10.46) of the $\varepsilon - \delta$ characterization of continuity. This proves that the identity mapping is continuous. ■

10.2.2 Continuity of constants and sums and products

For all the following, unless stated differently, let (X, d) be a metric space and $A \subseteq X$. Let

$$f : A \to \mathbb{R}$$
$$g : A \to \mathbb{R}$$

be two real functions which both are continuous in a point $x_0 \in A$. Moreover, let a, b be two (constant) real numbers. You can think of any constant number a as a function on $\mathbb{R}$ as follows:

$$a(\cdot): \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto a$$

In other words, the function $a(\cdot)$ assigns to each $x \in X$ one and the same value a. We called such a function a constant function (see (8.3) on p.114).

Proposition 10.12. *Given is a metric space (X,d). Let the functions*

$$f(\cdot), g(\cdot), f_1(\cdot), f_2(\cdot), f_3(\cdot), \dots, f_n(\cdot) : A \longrightarrow \mathbb{R}$$

all be continuous at $x_0 \in A \subseteq X$. Then

a: Constant functions are continuous everywhere on A.

b: The product $fg(\cdot) : x \mapsto f(x)g(x)$ is continuous at x_0. Especially $af(\cdot) x \mapsto a \cdot f(x)$ is continuous at x_0 and , using -1 as a constant, $-f(\cdot) : x \mapsto -f(x)$ is continuous at x_0

c: The sum $f+g(\cdot) : x \mapsto f(x)+g(x)$ is continuous at x_0

d: Any linear combination [80] $\sum\limits_{j=0}^{n} a_j f_j(\cdot) : x \mapsto \sum\limits_{j=0}^{n} a_j f_j(x)$ is continuous in x_0.

Proof of a: Let $\varepsilon > 0$. We do not even have to look for a suitable δ to restrict the distance between two arguments x and x_0 because it is always true that

$$|a(x) - a(x_0)| = |a - a| = 0 < \varepsilon$$

and we are done.

Proof of b: In the following chain of calculations each inequality results from applying the triangle inequality (2.11) which states, just to remind you, that $|a+b| \leqq |a| + |b|$ for any two real numbers a and b:

$$\begin{aligned}
&|f(x_0)g(x_0) - f(x)g(x)| \\
&= |f(x_0)g(x_0) - f(x)g(x_0) + f(x)g(x_0) - f(x)g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad |f(x)| \cdot |g(x_0) - g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad |f(x) - f(x_0) + f(x_0)| \cdot |g(x_0) - g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad (|f(x) - f(x_0)| + |f(x_0)|) \cdot |g(x_0) - g(x)|
\end{aligned}$$

Now write x_n rather than x and assume that (x_n) is a sequence which converges to x_0. We have shown above that

$$|f(x_0)g(x_0) - f(x_n)g(x_n)| \leqq K_1 + K_2 \tag{10.48}$$

where

$$\begin{aligned}
K_1 &= |g(x_0)| \cdot |f(x_0) - f(x_n)| \\
K_2 &= (|f(x_n) - f(x_0)| + |f(x_0)|) \cdot |g(x_0) - g(x_n)|
\end{aligned}$$

The continuity of $f(\cdot)$ and $g(\cdot)$ in x_0 and the convergence $x_n \to x_0$ for $n \to \infty$ implies that $f(x_n) \to f(x_0)$ and $g(x_n) \to g(x_0)$ (see (10.43) on p.175).

[80] See def.9.6 (linear combinations) on p.141

So both $|f(x_0) - f(x_n)|$ and $|g(x_0) - g(x_n)|$ will converge to zero as $n \to \infty$ and the same will be true if those expressions are multiplied by the constant value $|g(x_0)|$, no matter how big it may be, or by $|f(x_n) - f(x_0)| + |f(x_0)|$ (for big n, $f(x_n)$ is very close to $f(x_0)$ so that $|f(x_n) - f(x_0)| + |f(x_0)|$ will be bounded by the constant value $1 + |f(x_0)|$) for big enough n.

This means that both K_1 and K_2 will converge to zero and (10.48) shows that $fg(x_n) = f(x_n)g(x_n)$ converges to $fg(x_0)$ as $n \to \infty$. But we made no special assumption about (x_n) besides its converging against x_0 and we have proved the continuity of $(fg)(\cdot)$ in x_0. This concludes the proof of b.

Proof of c: Let $\varepsilon > 0$ and let $\tilde{\varepsilon} = \frac{\varepsilon}{2}$. Because $f(\cdot)$ and $g(\cdot)$ are both continuous in x_0, there is $\delta > 0$ such that $|f(x_0) - f(x_n)| < \tilde{\varepsilon}$ and $|g(x_0) - g(x_n)| < \tilde{\varepsilon}$ Again, we make heavy use of the triangle inequality:

$$\begin{aligned} |f(x_0) + g(x_0) \; - \; (f(x_n) + g(x_n))| \quad &= \quad |(f(x_0) - f(x)) \; + \; (g(x_0) - g(x))| \\ &\leqq \quad |f(x_0) - f(x)| \; + \; |g(x_0) - g(x))| \\ &\leqq \quad \tilde{\varepsilon} + \tilde{\varepsilon} \quad = \quad \varepsilon \end{aligned}$$

and we are done with the proof of c.

proof of d: For linear combinations of two functions f_1 and f_2, the proof is obvious from parts a, b and c. The proof for sums of more than two terms needs a simple (strong) induction argument: Write

$$\sum_{j=0}^{n+1} a_j f_j(x) \; = \Big(\sum_{j=0}^{n} a_j f_j(x) \Big) \; + \; a_{n+1} f_{n+1}(x) \; = \; I \; + \; II.$$

The left term "I" is continuous by the induction assumption and the entire sum $I \; + \; II$ then is continous as the sum of two continuous functions. ■

Remark 10.13 (Opposite of continuity)**.** Given a metric space (X, d), what is the opposite of $\lim\limits_{k \to \infty} x_k \; = \; L$?
Beware! It is NOT the statement that $\lim\limits_{k \to \infty} x_k \; \neq \; L$ because such a statement would mislead you to believe that such a limit exists, it just happens not to coincide with L

The correct answer: There exists some $\varepsilon \; > \; 0$ such that for **all** $N \; \in \; \mathbb{N}$ there exists some natural number $j = j(N)$ such that $j \geq N$ and $d(x_j, L) \geq \varepsilon$.

From the above we conclude the following:

Proposition 10.13. *A sequence $(x_k)_k$ with values in (X, d) does not have $L \in X$ as its limit if and only if there exists some $\varepsilon > 0$ and $n_1 < n_2 < n_3 < \cdots \in \mathbb{N}$ such that $d(x_{n_j}, L) \geq \varepsilon$ for **all** j. In other words, we can find a subsequence $(x_{n_j})_j$ which completely stays out of some ε–neighborhood of L.*

Proof: Obvious from remark 10.13 ■

10.2.3 Function spaces (Understand this!)

Definition 10.25 (linear combinations (imprecise))**.** textbfREMOVED

Definition 10.26 (linear mappings (imprecise))**.** textbfREMOVED

Example 10.9. textbfREMOVED

Example 10.10 (Vector space of continuous real functions)**.** The set

$$\mathscr{C}(X,\mathbb{R}) \;:=\; \{f(\cdot) : f(\cdot) \text{ is a continuous real function on } X\}$$

of all real continuous functions on an arbitrary non–empty set X is a vector space if you define addition and scalar multiplication as in (8.2) on p.114. The reason is that you can verify the properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ and for each one fo those you just deal with ordinary real numbers. The "sup–norm"

$$\|f(\cdot)\| \;=\; \sup\{|f(x)| : x \in X\}$$

(see (10.3) on p.154) is **not a real function** on all of $\mathscr{C}(X,\mathbb{R})$ because $\|f(\cdot)\| = +\infty$ for any unbounded $f(\cdot) \in \mathscr{C}(X,\mathbb{R})$.

The subset

$$\mathscr{C}_{\mathscr{B}}(X,\mathbb{R}) \;:=\; \{h(\cdot) : h(\cdot) \text{ is a bounded continuous real function on } X\}$$

(see prop.10.1 on p. 154) is a subspace of the normed vector space of all bounded real functions on X. On this subspace the sup–norm truly is a real function in the sense that $\|f(\cdot)\| < \infty$.

10.2.4 Continuity of Polynomials (Understand this!)

Definition 10.27 (polynomials)**.** Anything that has to do with polynomials takes place in $\mathbb{R}$ and not on a metric space.

Let A be subset of the real numbers and let $p(\cdot) : A \to \mathbb{R}$ be a real function on A. $p(\cdot)$ is called a **polynomial**. if there is an integer $n \geqq 0$ and real numbers $a_1, a_2, \dots, a_n$ which are constant (they do not depend on x) so that $p(\cdot)$ can be written as a sum

$$p(x) \quad = \quad a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \tag{10.49}$$

Remember that $x^0 = 1$ and $x^1 = x$ and we have

$$p(x) \quad = \quad a_0x^0 + a_1x^1 + a_2x^2 + \ldots + a_nx^n \quad = \quad \sum_{j=0}^{n} a_jx^j \tag{10.50}$$

In other words, polynomials are linear combinations of the **monomials** $x \to x^k \quad (k \in (N)_0$.

Proposition 10.14 (All polynomials are continuous)**.**

Proof: It suffices to show that the monomials $m_j(x) := x^j$ are continuous for all $j = 0, 1, 2, \ldots$ because of proposition (10.12), part d and because all polynomials are linear combinations of monomials. $m_0(\cdot)$ is continuous because it is the constant function $x \to 1$. $m_1(\cdot) : x \to x$ is continuous according to def.10.24 [81] *(p.175) because for any given $\varepsilon > 0$ we choose $\delta := \varepsilon$ and this will ensure that $|m_1(x) - m_1(y)| < \varepsilon$ whenever $|x - y| < \delta$. But if $m_1(\cdot)$ is continuous then so is the product $m_2(\cdot) = m_1(\cdot)m_1(\cdot)$. But then so is the product $m_3(\cdot) = m_2(\cdot)m_1(\cdot)$. But then so is the product $m_j(\cdot) = m_{j-1}(\cdot)m_1(\cdot)$ for any choice of $j > 0$. We have shown that all monomials are continuous and so are polynomials as their linear combinations.* ■

[81] besides, $m_1(\cdot)$ is the identity mapping on $\mathbb{R}$ and we know from proposition (10.11) on p.177 that identity mappings are always continuous.

Proposition 10.15 (Vector space property of polynomials)**.** *Sums and scalar products of polynomials are polynomials.*

Proof of a. Additivity:
Let

$$p_1(x) \;=\; a_0 + a_1x + a_2x^2 + \ldots + a_nx_1^n \;=\; \sum_{j=0}^{n_1} a_jx^j$$

and

$$p_2(x) \;=\; b_0 + b_1x + b_2x^2 + \ldots + b_nx_2^n \;=\; \sum_{j=0}^{n_2} b_jx^j$$

be two polynomials. Might as well assume that $n_1 \leqq n_2$*. Let* $a_{n_1+1} = a_{n_1+2} = \ldots = a_{n_2} = 0$*. This does not change anything and we get*

$$\begin{aligned} p_1(x) + p_1(x) \;&=\; \sum_{j=0}^{n_2} a_jx^j \;+\; \sum_{j=0}^{n_2} b_jx^j \\ &=\; \sum_{j=0}^{n_2}(a_j + b_j)x^j \\ &=\; \sum_{j=0}^{n_2} c_jx^j \qquad (c_j := a_j + b_j) \end{aligned}$$

This proves that the function $p_1(\cdot) + p_2(\cdot)$ *is of the form* (10.50) *and we have shown that it is a polynomial. The proof for the sum of more than two polynomials now follows by the principle of proof by mathematical induction (see* (2.11) *on p.17).*

Proof of b. Scalar product:
Let

$$p(x) \;=\; a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \;=\; \sum_{j=0}^{n} a_jx^j$$

be a polynomial. Let λ *be a real number. Then*

$$\begin{aligned} (\lambda p)(x) \;&=\lambda p(x) \;=\; \lambda\sum_{j=0}^{n} a_jx^j \\ &=\; \sum_{j=0}^{n} \lambda a_jx^j \;=\; \sum_{j=0}^{n} c_jx^j \qquad (c_j := \lambda a_j) \end{aligned}$$

This proves that the function $\lambda p(\cdot)$ *is of the form* (10.50) *and we are done.* ■

Polnomials may not always be given in their ***normalized form*** *(10.50) on p.180. Here is an example:*

$$\begin{aligned} p(x) \;&=\; a_0x^0(1-x)^n + a_1x^1(1-x)^{n-1} + a_2x^2(1-x)^{n-2} + \ldots + a_{n-1}x^{n-1}(1-x)^1 + a_nx^n \\ &=\; \sum_{k=0}^{n} a_kx^k(1-x)^{n-k} \end{aligned}$$

is a linear combination of monomials and hence a polynomial. All you need to do is "multiply out" the $x^k(1-x)^{n-k}$ *terms and then regroup the resulting mess. The so called* ***Bernstein polynomials***

$$p(x) \quad = \quad \sum_{k=0}^{n} \binom{n}{k} f(\frac{k}{n}) x^k (1-x)^{n-k} \qquad \textit{see note}^{82}$$

are of that form.

Example 10.11 (Vector space of polynomials)**.** Let $A \subseteq \mathbb{R}$. I follows from (10.15) and (10.14) that the set

$$\{p(\cdot) : p(\cdot) \text{ is a polynomial on } A\}$$

of all polynomials on an arbitrary non–empty subset A of the real numbers is a subspace of the vector space $\mathscr{C}(A, \mathbb{R})$. (see example (10.10) on p.180. The "sup–norm"

$$\|f(\cdot)\| \; = \; \sup\{|f(x)| : x \in A\}$$

is **not a real function** on the set of all polynomials on A as its value may be ∞.. Matter of fact, it can be shown that, if the set A itself is not bounded, then the only polynomials for which $\|p(\cdot)\| < \infty$ are the constant functions on A(!)

10.2.5 Uniform continuity

It will be proved in theorem 11.8 (Uniform continuity on sequence compact spaces) on p.211 [83] *that continuous real functions on the compact set* $[0,1]$ *are uniformly continuous in the sense of the following definition which you should compare, for the special case of* $(X, d) = (\mathbb{R}, d_{|\cdot|})$ *where* $d_{|\cdot|}(x, y) = |y - x|$*, to [1] Beck/Geoghegan, Appendix A.3, "Uniform continuity".*

Definition 10.28 (Uniform continuity of functions)**.** Let (X, d_1), (Y, d_2) be metric spaces and let A be a subset of X. A function

$$f(\cdot) : A \to Y \qquad \text{is called } \textbf{uniformly continuous}$$

if for any $\varepsilon > 0$ there exists a (possibly very small) $\delta > 0$ such that

(10.51) $$d_2(f(x) - f(y)) \; < \varepsilon \quad \text{for any } x, y \in A \text{ such that } d_1(x, y) \; < \; \delta.$$

Remark 10.14 (Uniform continuity vs. continuity)**.** Note the following:

A. Condition (10.51) for uniform continuity looks very close to the ε–δ characterization of ordinary continuity (10.46) on p.175. Can you spot the difference? Uniform continuity is more demanding than plain continuity because when dealing with the latter you can ask for specific values of both ε and x_0 according to which you had to find a suitable δ. In other words, for plain continuity

$$\delta \; = \; \delta(\varepsilon, x_0).$$

[82] Here $f(\cdot)$ is a function, not necessarily continuous, on the unit interval $[0,1]$. The binomial coefficient $\binom{n}{k}$ is defined as $\frac{n!}{k!(n-k)!}$ where $0! = 1$ and $n! = 1 \cdot 2 \cdot 3 \cdots n$ for $n \in !\mathbb{N}$ (see ch.4 of [1] B/G Art of Proof)

[83] see chapter 11.5 (Continuous functions and compact spaces) on p.210

But in the case of uniform continuity all you get is ε and you must come up with a suitable δ regardless of what arguments are thrown at you. To write that one in functional notation,

$$\delta \;=\; \delta(\varepsilon).$$

B. In case you missed the point, uniform continuity implies continuity but the opposite need not be true.

Example 10.12 (Uniform continuity of the identity mapping)**.** Have another look at proposition(10.11) where we proved the continuity of the identity mapping on a metric space. We chose $\delta = \varepsilon$ no matter what value of x we were dealing with and it follows that the identity mapping is always uniformly continuous.

Remark 10.15. Now that you have learned the definitions for both continuity and uniform continuity, have another look at example 3.28, p.54 in ch.3.6.3 (Quantifiers for statement functions of more than two variables) where it was explained how you could obtain one definition from the other just by switching around a $\forall$ quantifier and a $\exists$ quantifier.

10.2.6 Continuity of linear functions (Understand this!)

Lemma 10.1. *Let $f : (V, \|\cdot\|) \to (W, |\cdot|)$ be a linear function between two normed vector spaces. Let*

$$\begin{aligned} a \;&:=\; \sup\{\, \big| f(x) \big| \; : x \in V, \|x\| = 1\}, \\ b \;&:=\; \sup\{\, \big| f(x) \big| \; : x \in V, \|x\| \leqq 1\}, \\ c \;&:=\; \sup\{\, \frac{\big| f(x) \big|}{\|x\|} : x \in V, x \neq 0\}. \end{aligned}$$

Then $a = b = c$.

Proof: We introduce the following three sets for this proof:

$$\begin{aligned} A :=&\{\, \big| f(x) \big| \; : x \in V, \|x\| = 1\}, \\ B :=&\{\, \big| f(x) \big| \; : x \in V, \|x\| \leqq 1\}, \\ C :=&\Big\{ \, \frac{\big| f(x) \big|}{\|x\|} : x \in V, x \neq 0 \Big\}. \end{aligned}$$

Proof that $a = b$:
It follows from $A \subseteq B$ that $a \leqq b$. On the other hand let $x \in B$ such that $x \neq 0$ (if $x = 0$ then $f(x) = 0$ certainly could not exceed a). Let $y := \|x\|^{-1}x$. Then $y \in A$ and $\|x\|^{-1} \geqq 1$, hence

$$\big| f(y) \big| \;=\; \big| f\big(x/\|x\|\big) \big| \;=\; (1/\|x\|) \, \big| f(x) \big| \;\geqq\; \big| f(x) \big|$$

and it follows that the sup over the bigger set B does not exceed the sup over A, hence $a = b$.

Proof that $a = c$:
Let $x \in C$ and $y := \|x\|^{-1}x$. Then $y \in A$ and

$$\big| f(x) \big| \,/\|x\| \;=\; \big| f(x)/\|x\| \big| \;=\; \big| f\big(x/\|x\|\big) \big| \;=\; \big| f(y) \big|$$

and it follows that the sup over the bigger set C does not exceed the sup over A, hence $c = b$. ■

Definition 10.29 (norm of linear functions)**.** Let $f : (V, \|\cdot\|) \to (W, |\cdot|)$ be a linear function between two normed vector spaces. We denote the quantity $a = b = c$ from lemma 10.1 by $\|f\|$, i.e.,

$$\begin{aligned}\|f\| &= \sup\{ |f(x)| : x \in V, \|x\| = 1\} \\ &= \sup\{ |f(x)| : x \in V, \|x\| \leqq 1\} \\ &= \sup\{ \frac{|f(x)|}{\|x\|} : x \in V, x \neq 0\}.\end{aligned} \tag{10.52}$$

$\|f\|$ is called the **norm of** f. [84]

We note that $\|f\|$ need not be finite.

Theorem 10.14 (Continuity criterion for linear functions)**.** *Let $f : (V, \|\cdot\|) \to (W, |\cdot|)$ be a linear function between two normed vector spaces. Then the following are equivalent.*

A. f *is continuous at* $x = 0$,
B. f *is continuous in all points of* V,
C. f *is uniformly continuous on* V,
D. $\|f\| < \infty$.

Moreover we then have

$$|f(x)| \leqq \|f\| \cdot \|x\| \quad \text{for all } x \in V. \tag{10.53}$$

Proof: Clearly we have C ⇒ B ⇒ A. We now show A ⇒ D.

It follows from the continuity of f at 0 that there exists $\delta > 0$ such that

$$\text{if } z \in V \text{ and } \|z\| < \delta \text{ then } |f(z)| = |f(z) - f(0)| < 1. \tag{10.54}$$

Let $x \in V$ such that $\|x\| \leqq 1$. Then $\|\delta/2 \cdot x\| \leqq \delta/2 < \delta$, hence, according to (10.53),

$$|f(\delta/2 \cdot x)| = \delta/2 \cdot |f(x)| < 1.$$

But then $|f(x)| < 2/\delta$ for all $x \in V$ with norm bounded by 1, hence

$$\|f\| = \sup\{ |f(x)| : x \in V, \|x\| \leqq 1\} < 2/\delta < \infty$$

and we have shown that A ⇒ D.

We now show D ⇒ (10.53). The inequality trivially holds for $x = 0$ because linearity of f implies $f(0) = 0$. If $x \neq 0$ then $\|x\| > 0$ (norms are positive definite) and the inequality follows from the last characterization of $\|f\|$ in (10.52).

We finally show D ⇒ C. Let $\varepsilon > 0$ and $\delta := \varepsilon/\|f\|$. Let $x, y \in V$ such that $\|x - y\| < \delta$. If we can show that this implies $|f(x) - f(y)| < \varepsilon$ then f is indeed uniformly continuous and the proof is done.

It follows from (10.53) that $|f(x-y)| \leqq \|f\| \cdot \|x - y\|$ and we obtain

$$|f(x) - f(y)| = |f(x-y)| \leqq \|f\| \cdot \|x - y\| \leqq \|f\| \cdot \delta = \|f\| \cdot \varepsilon/\|f\| = \varepsilon. \blacksquare$$

[84] Note that we use the same notation $\|\cdot\|$ for both the norm on V and the norm of the linear function f. **Do not confuse the two!**

10.3 Function sequences and infinite series

10.3.1 Convergence of function sequences (Study this!)

Vectors are more complicated than numbers because an n–dimensional vector $v \in \mathbb{R}^n$ represents a grouping of a finite number n of real numbers. Matter of fact, any such vector $(x_1, x_2, x_3, \cdots, x_n)$ can be interpreted as a real function (remember: a real function is one which maps it arguments into $\mathbb{R}$)

(10.55) $$f(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R} \qquad j \mapsto x_j$$

(see (9.4) on p.133).

Next come sequences $(x_j)_{j \in \mathbb{N}}$ which can be interpreted as real functions

(10.56) $$g(\cdot) : \mathbb{N} \to \mathbb{R} \qquad j \mapsto x_j$$

Finally we deal with any kind of real function

(10.57) $$h(\cdot) : X \to \mathbb{R} \qquad x \mapsto h(x)$$

as the most general case

Now we add more complexity by not just dealing with one or two or three real functions but with an entire sequence

(10.58) $$f_n(\cdot) : X \to \mathbb{R} \qquad x \mapsto f_n(x)$$

For any fixed argument x_0 we have a sequence $f_1(x_0), f_2(x_0), f_3(x_0), \cdots$ which we can examine for convergence. This sequence may converge for some or all arguments $x_0 \in X$ to a real number. Time for some definitions.

Definition 10.30 (Pointwise convergence of function sequences)**.** Let X be a non-empty set, (Y, d) a metric space and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $A \subseteq X$ be a subset of X. We say that $f_n(\cdot)$ **converges pointwise** or, simply, **converges** to $f(\cdot)$ on A and we write $f_n(\cdot) \to f(\cdot)$ if

(10.59) $$f_n(x) \to f(x) \quad \text{for all } x \in A$$

Definition 10.31 (Uniform convergence of function sequences)**.** Let X be a non-empty set, (Y, d) a metric space and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $A \subseteq X$ be a subset of X. We say that $f_n(\cdot)$ **converges uniformly** to $f(\cdot)$ on A and we write [85]

(10.60) $$f_n(\cdot) \overset{uc}{\to} f(\cdot)$$

if the following is true: For each $\varepsilon > 0$ (no matter how small) there exists a (possibly huge) number n_0 which can be chosen once and for all, independently of the specific argument x, such that

(10.61) $$d(f_n(x), f(x)) < \varepsilon \quad \text{for all } x \in A \quad \text{and } n \geqq n_0$$

[85] I must confess that " $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ " is a notation that I coined myself because it is not as tedious as writing " $f_n(\cdot) \to f(\cdot)$ uniformly"

Remark 10.16 (Uniform convergence implies pointwise convergence)**.** Look at definition (10.8) on p.158 of convergence of sequences and you should immediately see that (10.61) implies, for any given $x \in A$, ordinary convergence $f(x) = \lim_{n\to\infty} f_n(x)$ because the number $n_0 = n_0(\varepsilon)$ chosen in (10.61) will also satisfy (10.14) (p.158) for $x_n = f_n(x)$ and $a = f(x)$.

In other words, unform convergence implies pointwise convergence. But what is the difference between pointwise and uniform convergence? The difference is that, for poinwise convergence, the number n_0 will depend on both ε and x: $n_0 = n_0(\varepsilon, x)$. In the case of uniform convergence, the number n_0 will still depend on ε but can be chosen independently of the argument $x \in A$.

Example 10.13 (a. Constant sequence of functions)**.** Let X be a set and let $f : X \to \mathbb{R}$ be a real function on X which may or may not be continuous anywhere. Define a sequence of functions

$$f_n : X \to \mathbb{R}\ (n \in \mathbb{N}) \quad \text{as} \quad f_1 = f_2 = \cdots = f$$

i.e.,

$$f_1(x) = f_2(x) = \cdots = f(x)\ \forall n \in \mathbb{N},\ \forall x \in X.$$

In other words, we are looking at a constant sequence of functions (not to be confused with a sequence of constant functions – seriously!).

Then $f_n(\cdot) \overset{uc}{\to} f(\cdot)$

Proof of the example: This is trivial. No matter how small an ε and n_0 we choose and no matter what argument $x \in X$ we are looking at, we have

$$|f_n(x) - f(x)| = 0 < \varepsilon \quad \textit{for all } x \in A \quad \textit{and } n > n_0$$

■

Example 10.14 (b. Pointwise but not uniformly convergent sequence of functions)**.** Let $X = [0, 1]$, i.e., X is the closed unit interval $\{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. Let the functions $f_n(\cdot)$ be defined as follows on X:

$$f_n(x) = \begin{cases} n^2 x & \text{for } 0 \leqq x \leqq \frac{1}{n} \\ \frac{1}{x} & \text{for } \frac{1}{n} \leqq x \leqq 1 \end{cases}$$

Note that both pieces fit together in the point $a = 1/n$ because the "$\frac{1}{x}$ definition" gives $f_n(a) = \frac{1}{1/n} = n$ and the "$n^2 x$ definition" gives the same value $n = n^2 \frac{1}{n}$. We do not give a formal proof that each $f_n(\cdot)$ is continuous in every point of $[0, 1]$. Just accept it from the fact that the two graphs flow into each other at the "splicing point" $1/n$.

Now we define the function $f(\cdot) : [0, 1] \to \mathbb{R}$ as

$$f(x) = \begin{cases} \frac{1}{x} & \text{for } 0 < x \leqq 1 \\ 0 & \text{for } x = 0 \end{cases}$$

Then the functions $f_n(\cdot)$ converge pointwise but not uniformly to $f(\cdot)$ on the entire unit interval.

Proof of pointwise convergence:
first we look separately at the point $a = 0$. We have $f(0) = 0 = n^2 0 = f^n(0)$ and the constant sequence

of zeroes certainly converges against zero. Now assume $a > 0$. If $n > 1/a$ then $f_n(a) = \frac{1}{a}$ for all such n. Again, we have a constant sequence $(1/a)$ except for finitely many n and it converges against $1/a = f(a)$. We have thus proved pointwise convergence.

Proof that there is no uniform convergence:
To prove that (10.61) is not satisfied, we must find $\varepsilon > 0$ and points x_N so that for no matter how big a natural number N we choose, there will be at least one $n > N$ such that $|f_n(x) - f(x)| \geqq \varepsilon$. Let $N \in \mathbb{N}$ be any natural number. Then

$$f_N(\frac{1}{N^2}) \;=\; \frac{N^2}{N^2} = 1$$

and

$$f_{2N}(\frac{1}{N^2}) \;=\frac{(2N)^2}{N^2} \;= 4$$

So

$$|f_{2N}(\frac{1}{N^2}) - f_N(\frac{1}{N^2})| \;= 3$$

To recap: We found $\varepsilon > 0$ so that for each $N \in \mathbb{N}$ we were able to find an $n \geqq N$ and $x_N \in [0,1]$ such that $|f_n(X_N) - f_N(x_N)| > \varepsilon$: we chose

$$\varepsilon = 2, \quad n = 2N, \quad x_N = \frac{1}{N^2}$$

We have thus prove that the pointwise convergence is not uniform. ■

Proposition 10.16 (Uniform convergence is $\|\cdot\|$ convergence). *Let X be a nonempty set and $\mathscr{B}(X,\mathbb{R})$ the set of all bounded real functions on X. We remember that this set is a vector space with the norm $\|f\|_\infty = \sup\{|g(x) - f(x)| : x \in X\}$ and a metric space with the corresponding metric*

$$d_{\|\cdot\|_\infty}(f,g) = \sup\{|g(x) - f(x)| : x \in X\}$$

(see example 10.2 on p.152).

The following is true: $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ if and only if $f_n(\cdot) \overset{\|\cdot\|_\infty}{\to} f(\cdot)$, i.e., the sequence f_n converges to f in the metric space $\big(\mathscr{B}(X,\mathbb{R}), d_{\|\cdot\|_\infty}(\cdot,\cdot)\big)$.

This justifies to call $\|f\|_\infty$ the norm of uniform convergence and $d_{\|\cdot\|_\infty}(\cdot,\cdot)$ the metric of uniform convergence.

Proof of "⇒": Assume that $f_n(\cdot) \overset{uc}{\to} f(\cdot)$. Let $\varepsilon > 0$. According to def.10.31 (Uniform convergence of function sequences) on p.185 there exists an index $n_0 = n_0(\varepsilon)$ (which does not depend on the function argument $x \in X$) such that

$$d(f_n(x), f(x)) = |f_n(x) - f(x)| < \varepsilon/2 \quad \textit{for all } x \in X \quad \textit{and } n \geqq n_0$$

(note that the metric space Y in def.10.31 is $\mathbb{R}$ here, so $d(f_n(x), f(x))$ becomes $|f_n(x) - f(x)|$). But then

$$\sup\{|f_n(x) - f(x)| : x \in X\} \;\leqq\; \varepsilon/2 \textit{ for all } n \geqq n_0,$$

i.e., $d_{\|\cdot\|_\infty}(f_n, f) < \varepsilon$ for all $n \geqq n_0$. It follows that $f_n(\cdot) \overset{\|\cdot\|_\infty}{\to} f(\cdot)$.

Proof of "⇐": Assume that $f_n(\cdot) \overset{\|\cdot\|_\infty}{\to} f(\cdot)$, i.e., $\lim\limits_{n\to\infty} f_n \;=\; f$ in the metric space $\big(\mathscr{B}(X,\mathbb{R}), d_{\|\cdot\|_\infty}(\cdot,\cdot)\big)$.

Let $\varepsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that

$$d_{\|\cdot\|_\infty}(f_n, f) \;=\; \|f_n - f\|_\infty \;=\; \sup\{|f_n(x) - f(x)| : x \in X\} \;<\; \varepsilon \quad \text{for all } n \geqq n_0$$

But then

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } x \in X \quad \text{and } n \geqq n_0$$

and it follows that $f_n(\cdot) \stackrel{uc}{\to} f(\cdot)$. ■

Theorem 10.15 (Uniform limits of continuous functions are continuous)**.** *Let (X, d_1) and (Y, d_2) be metric spaces and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $x_0 \in X$ and let $V \subseteq X$ be a neighborhood of x_0. Assume **a)** that the functions $f_n(\cdot)$ are continuous at x_0 for all n and **b)** that $f_n(\cdot) \stackrel{uc}{\to} f(\cdot)$ on V. Then f is continuous at x_0*

Proof: Let $\varepsilon > 0$.

A. Uniform convergence $f_n(\cdot) \stackrel{uc}{\to} f(\cdot)$ on V guarantees the existence of some $N = N(\varepsilon)$ such that

$$d_2\big(f_n(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V \text{ and } n \geqq \mathbb{N}.$$

In particular, for $n = N$,

$$d_2\big(f_N(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V\,. \tag{10.62}$$

B. All functions f_n and in particular f_N are continuous in V. There is $\tilde{\delta} > 0$ such that

$$d_2(f_N(x), f_N(x_0)) < \frac{\varepsilon}{3} \text{ for all } x \in N_{\tilde{\delta}}(x_0). \tag{10.63}$$

C. As x_0 is an interior point of V, there exists $\hat{\delta} > 0$ such that $N_{\hat{\delta}}(x_0) \subseteq V$. Let δ be the smaller of $\hat{\delta}$ and $\tilde{\delta}$.

Then (10.62) and (10.63) both hold for $x \in N_\delta(x_0)$. We note that $x_0 \in N_\delta(x_0)$ and obtain

$$d(f(x), f(x_0)) \;\leqq\; d(f(x), f_N(x)) \;+\; d(f_N(x), f_N(x_0)) \;+\; d(f_N(x_0), f(x_0)) \;<\; \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \;=\; \varepsilon.$$

The proof is finished. ■

10.3.2 Infinite Series (Understand this!)

*We start by repeating the definition of a sequence given in section 4.2 on p.71: A **sequence** (x_j) is nothing but a family of things x_j which are indexed by integers, usually the natural numbers or the non-negative integers. We make throughout this entire document the following*

Assumption 10.1 (indices of sequences)**.**

Unless explicitly stated otherwise, sequences are always indexed $1, 2, 3, \ldots$, i.e., the first index is 1 and, given any index, you obtain the next one by adding 1 to it.

The simplest things that a mathematician deals with are numbers. One nice thing that is always possible with numbers, is that you can add them. Here is a very simple definition:

Definition 10.32 (Numeric Sequences and Series)**.** A sequence (a_j) of real numbers is called a **numeric sequence** if each a_j is a real number. For any such sequence, we can build another sequence (s_n) as follows:

$$s_1 := a_1; \quad s_2 := a_1 + a_2; \quad s_3 := a_1 + a_2 + a_3; \cdots \quad s_n := \sum_{k=1}^{n} a_k \tag{10.64}$$

We write this more compactly as

$$a_1 + a_2 + a_3 + \cdots = \sum a_k \tag{10.65}$$

and we call any such object, which represents a sequence of partial sums, a **series**. Loosely speaking, a series is an infinite sum. We call (s_n) the sequence of **partial sums** associated with the series $\sum a_k$.

We say that the series converges to a real number s and we write

$$s = \sum_{k=1}^{\infty} a_k \tag{10.66}$$

if this is true for the associated sequence of partial sums (10.64). We say that the series has limit ∞ (has limit $-\infty$) if this true for the associated partial sums and we write

$$\sum_{k=1}^{\infty} a_k = \infty \quad \left(\sum_{k=1}^{\infty} a_k = -\infty\right) \tag{10.67}$$

Proposition 10.17 (Convergence criteria for series)**.** *A series $s := \sum a_k$ of real numbers converges if and only if for all $\varepsilon > 0$ there exists $n_o \in \mathbb{N}$ such that one of the following is true:*

$$\left|\sum_{k=n}^{\infty} a_k\right| < \varepsilon \qquad \textit{for all } n \geqq n_0 \tag{10.68a}$$

$$\left|\sum_{k=n}^{m} a_k\right| < \varepsilon \qquad \textit{for all } m, n \geqq n_0 \tag{10.68b}$$

Proof: Write

$$s = \sum_{k=1}^{\infty} a_k = \sum_{k=1}^{n} a_k + \sum_{k=n+1}^{\infty} a_k = s_n + \sum_{k=n+1}^{\infty} a_k \tag{10.69}$$

Remember the convergence criteria for numeric sequences. Convergence of a sequence (s_n) *to a real number* s *means that for any* $\varepsilon > 0$, *all but finitely many members* s_n *will be inside the* ε*–neighborhood* $N_\varepsilon(s)$ *of* s. *Written in terms of the distance to* s *this means there exists a suitable* $n_0 \in \mathbb{N}$ *such that*

$$|s - s_n| < \varepsilon \qquad \text{for all } n \geqq n_0$$

(see (10.8) *on p.158). According to* (10.69) *we can write that as*

$$\Big|\sum_{k=n+1}^{\infty} a_k\Big| < \varepsilon \qquad \text{for all } n \geqq n_0$$

which is the same as (10.68.a) because it does not matter whether we look at the sum of all terms bigger than n *or* $n+1$.

Alternatively, there was the Cauchy criterion

$$|s_i - s_j| < \delta \text{ for all } i, j \geqq n_0$$

(see (10.21) *on p.170) which ensures convergence to some number* s *without specifying what it might actually be. Again we use* (10.69) *and obtain, assuming without loss of generality that* $i < j$,

$$\Big|\sum_{k=i+1}^{j} a_j\Big| < \delta \text{ for all } j > i \geqq n_0 \quad \blacksquare$$

Corollary 10.1. *If a series* $\sum a_j$ *converges then* $\lim\limits_{n\to\infty} a_n = 0$.

Proof: Let $\varepsilon > 0$. *It follows from 10.68b that there is some* $n_0 \in \mathbb{N}$ *such that* $|a_m - 0| = \big|\sum_{k=m}^{m} a_k\big| < \varepsilon$ *for all* $m \geqq n_0$. *But this means that the sequence* a_n *converges to zero.* $\blacksquare$

Here is a second corollary.

Corollary 10.2 (Dominance criterion)**.** *Let* $N \in \mathbb{N}$ *and let* $\sum a_j$ *and* $\sum b_j$ *be two series such that* $|b_k| \leqq a_k$ *for all* $k \geqq N$. *It follows that if* $\sum a_k$ *converges then* $\sum b_k$ *converges.*

In particular, if $|b_k| \leqq a_k$ *for all* $k \in \mathbb{N}$ *then* $\Big|\sum_{k=1}^{\infty} b_j\Big| \leqq \sum_{k=1}^{\infty} a_j$

Proof: Let $\varepsilon > 0$. *It follows from 10.68b that there is some* $n_0 \in \mathbb{N}$ *such that* $\big|\sum_{k=m}^{n} a_k\big| < \varepsilon$ *for all* $m, n \geqq n_0$. *Let* $M := \max(n_0, N)$. *We obtain*

$$\Big|\sum_{k=i+1}^{j} b_j\Big| \leqq \sum_{k=i+1}^{j} |b_j| \leqq \sum_{k=i+1}^{j} a_j < \varepsilon \quad \text{for all } j > i \geqq M.$$

We conclude from (10.68b) that $\sum b_k$ *converges.*

Let $s_n := \sum_{k=1}^{n} |a_k|, \quad s := \lim_{n\to\infty} s_n, \quad t_n := \sum_{k=1}^{n} a_k, \quad t := \lim_n t_n.$

It follows from the triangle inequality that $|t_n| \leqq s_n$ *for all* $n \in \mathbb{N}$*, hence* $|t| = \lim_n |t_n| \leqq \lim_n s_n = s$*. This completes the proof.* ■

It is very important to remember that a series either converges to a finite number or it diverges. If it diverges it may be the case that $\sum_{k=1}^{\infty} a_k = \infty$ *or* $\sum_{k=1}^{\infty} a_k = -\infty$ *or there is no limit at all. As an example for a series which has no limit, look at the oscillating sequence*

$$a_0 = 1; \quad a_1 = -1; \quad a_2 = 1; \quad a_3 = -1; \cdots \quad s_n = \sum_{k=0}^{n} (-1)^n \tag{10.70}$$

The above is an example of a series that starts with an index other than 1 *(zero).* s_n *obviously does not have limit* $+\infty$ *or* $-\infty$ *because* s_n *is* 1 *for all even* n *and* 0 *for all odd* n*. Do not make the mistake of thinking that the limit of the series is zero because you fail to notice the odd indices and only see that* $s_0 = s_2 = s_4 = \cdots = s_{2j} = 0$*.*

Note that for any $j \in \mathbb{N}$ *we have* $|s_j - s_{j-1}| = 1$ *because at each step we either add or subtract* 1*. This means that no matter what real number* a *and how big a number* $n_0 \in \mathbb{N}$ *we choose, it will never be true that* $|a - s_j| < 1$ *for all* $j \in \mathbb{N}$ *and* a *cannot be a limit of the series.*

Just so you understand the difference between limits and contact points (see *(def.10.19) on p.166**): Even though neither* $(a_j)_j$ *nor* $(s_j)_j$ *has a limit, both have two contact points each.* $(a_j)_j$ *has the contact points* $\{1, -1\}$ *and* $(s_j)_j$ *has the contact points* $\{0, 1\}$*.*

We now turn our attention to convergence properties of series.

Definition 10.33 (Finite permutations)**.** Let $N \in \mathbb{N}$ and let $[N] := \{1, 2, 3, \dots, N\}$ denote the set of the first N integers. [86] A **permutation** of $[N]$ is a mapping

$$\pi(\cdot) : [N] \to [N]; \qquad j \mapsto \pi(j)$$

which is both surjective: each element k of $[N]$ is the image $\pi(j)$ for a suitable $j \in [N]$ and injective: different arguments $i \neq j \in [N]$ will always map to different images $\pi(i) \neq \pi(j) \in [N]$ (see (4.9) on p.81). Remember that

$$\text{surjective} + \text{injective} = \text{bijective}$$

and that under our assumptions the inverse mapping

$$\pi^{-1}(\cdot) : [N] \to [N]; \qquad \pi(j) \mapsto \pi^{-1}\pi(j) = j,$$

which associates with each image $\pi(j)$ the unique argument j which maps into $\pi(j)$, exists (see def. 4.9 on p.81 for properties of the inverse mapping).

It is customary to write

$$i_1 \text{ instead of } \pi(1), \quad i_2 \text{ instead of } \pi(2), \quad \dots, \quad i_j \text{ instead of } \pi(j), \quad \dots$$

[86] This notation was copied from chapter 13 (Cardinality) of [1] B/G (Beck/Geoghegan). It has nothing to do with equivalence classes!

Definition 10.34 (Permutations of $\mathbb{N}$)**.** A **permutation** of $\mathbb{N}$ is a mapping

$$\pi(\cdot) : \mathbb{N} \to \mathbb{N}; \qquad j \mapsto \pi(j)$$

which is both surjective: each element k of $\mathbb{N}$ is the image $\pi(j)$ for a suitable $j \in \mathbb{N}$ and injective: different arguments $i \neq j \in \mathbb{N}$ will always map to different images $\pi(i) \neq \pi(j) \in \mathbb{N}$.

Permutations are the means of describing a A ***rearrangement*** *or* ***reordering*** *of the members of a finite or infinite sequence or series. Look at any sequence* (a_j)*. Given a permutation* $\pi(\cdot)$ *of the natural numbers, we can form the sequence* $(b_k) := (a_{\pi(k)})$*, i.e.,*

$$b_1 = a_{\pi(1)}, \quad b_2 = a_{\pi(2)}, \quad \dots, \quad b_k = a_{\pi(k)}, \quad \dots$$

We can use the inverse permutation, $\pi^{-1}(\cdot)$*, to regain the* a_j *from the* b_j *because*

$$b_{\pi^{-1}(k)} \;=\; a_{\pi^{-1}(\pi(k))} \;=\; a_k$$

Proposition 10.18. *Let* (a_n) *be a sequence of non–negative members:* $a_n \geqq 0$ *for all* $n \in \mathbb{N}$*. Then exactly one of the following is true:*

A*: the series* $\sum a_n$ *converges (to a finite number). In that case*

$$\sum_{n=1}^{\infty} a_n \;=\; \sum_{n=1}^{\infty} a_{\pi(n)} \text{ for any permutation } \pi(\cdot) \text{ of } \mathbb{N}.$$

B*: the series* $\sum\limits_{n=1}^{\infty} a_n$ *has limit* ∞*. In that case it is true for any permutation* $\pi(\cdot)$ *of* $\mathbb{N}$ *that the reordered series* $\sum\limits_{n=1}^{\infty} a_{\pi(n)}$ *also has limit* ∞*.*

Proof of ***A****: Let* $b_j := a_{\pi(j)}$ *and, hence,* $a_k = b_{\pi-1(j)}$*. Let* $N \in \mathbb{N}$ *. Let*

$$\alpha \;:=\; \max\{\pi(j) : j \leqq N\} \qquad \text{and} \qquad \beta \;:=\; \max\{\pi^{-1}(k) : k \leqq N\}. \tag{10.71}$$

Note that $\alpha \geqq N$ *and* $\beta \geqq N$*. Because all terms* a_j, b_k *are non–negative it follows that*

$$\sum_{j=1}^{N} b_j \;=\; \sum_{j=1}^{N} a_{\pi(j)} \;\leqq\; \sum_{k=1}^{\alpha} a_k \;\leqq\; \sum_{k=1}^{\alpha} a_k \;+\; \sum_{k=\alpha+1}^{\infty} a_k \;=\; \sum_{k=1}^{\infty} a_k,$$
$$\sum_{k=1}^{N} a_k \;=\; \sum_{k=1}^{N} b_{\pi^{-1}(k)} \;\leqq\; \sum_{j=1}^{\beta} b_j \;\leqq\; \sum_{j=1}^{\beta} b_j \;+\; \sum_{j=\beta+1}^{\infty} b_j \;=\; \sum_{j=1}^{\infty} b_j.$$

We take limits as $N \to \infty$ *and it follows that*

$$\sum_{j=1}^{\infty} b_j \;\leq\; \sum_{k=1}^{\infty} a_k \quad \text{and} \quad \sum_{k=1}^{\infty} a_k \;\leqq\; \sum_{j=1}^{\infty} b_j.$$

This proves part ***A*** *of the proposition.*

Proof of ***B***: *Assume that* $\sum a_j$ *diverges. Because all terms* a_j *are non–negative, the sequence* s_n *of the partial sums is non–decreasing and hence has a limit* s. $s \notin \mathbb{R}$ *because we assumed that* $\sum a_j$ *is not convergent and we can rule out* $s = -\infty$ *because* $s \geqq a_0 \geqq 0$. *It follows that* $s = \infty$.

Assume to the contrary that there is a rearrangement $\sum b_j := \sum a_{\pi(j)}$ *of* $\sum a_j$ *which converges to a limit* $t \in \mathbb{R}$. *According to the already proved part* ***A*** *the rearrangement* $\sum a_j = \sum b_{\pi^{-1}(j)}$ *converges to the same (finite) limit* t. *We have reached a contradiction.* ■

Definition 10.35 (absolutely convergent series)**.** A series $\sum a_j$ is **absolutely convergent** if the corresponding series $\sum |a_j|$ of its absolute values converges.

Proposition 10.19. *Let* $\sum a_k$ *be an absolutely convergent series. Then* $\sum a_k$ *converges and*

$$|\sum_{k=1}^{\infty} a_k| \leqq \sum_{k=1}^{\infty} |a_k|. \tag{10.72}$$

Proof: This follows from the dominance criterion (cor.10.2) ■

It follows from prop.10.18 on p.192 that if a series of non-negative terms converges then its limit is invariant under rearrangements of that series. The next theorem states that any absolutely convergent series has that property.

Theorem 10.16. [87] *Let* $\sum a_k$ *be an absolutely convergent series. Let* $\pi : \mathbb{N} \to \mathbb{N}$ *be a permutation of* $\mathbb{N}$, *i.e., the series* $\sum b_k$ *with* $b_k := a_{\pi(k)}$ *is a rearrangement of the series* $\sum a_k$. *Then* $\sum b_k$ *converges and has the same limit as* $\sum a_k$. [88]

Proof: Let $\varepsilon > 0$. *Since* $\sum |a_k|$ *converges, there exists* $n_0 \in \mathbb{N}$ *such that*

$$\sum_{k=n_0+1}^{n_0+m} |a_k| \leqq \sum_{k=n_0+1}^{\infty} |a_k| < \varepsilon \quad \textit{for all } m \in \mathbb{N}. \tag{10.73}$$

For $n \in \mathbb{N}$ *let* $s_n := \sum_{k=1}^{n} a_k$ *and* $t_n := \sum_{k=1}^{n} b_k$.

Let $A := \{\pi(j) : 1 \leqq j \leqq n_0\}$ *and* $p_0 := \max(A)$. *This maximum exists because the set* A *is finite.*

Then $p_0 \geqq n_0$. *Each of* $a_1, a_2, \dots, a_{n_0}$ *is a term of* s_{n_0}, *hence of* s_{p_0}.
Moreover each of $b_1 = a_{\pi(1)}, b_2 = a_{\pi(2)}, \dots, b_{p_0} = a_{\pi(p_0)}$ *is a term of* t_{p_0}.

Let $n, p \geqq p_0$. *Then each of* $a_1, a_2, \dots, a_{n_0}$ *is a term of* s_n
and each of $b_1 = a_{\pi(1)}, b_2 = a_{\pi(2)}, \dots, b_{p_0} = a_{\pi(p_0)}$ *is a term of* t_p.

We recall that p_0 *was chosen so big that each of* $a_1, \dots, a_{n_0}$ *is one of* $b_1, \dots, b_{p_0}$.

It follows from all this that each of $a_1, \dots, a_{n_0}$ *is a term both of* s_n *and* t_p, *hence none of those terms appears in the difference* $s_n - t_p$. *We obtain for big enough* $m \in \mathbb{N}$ *(the bigger of* $\max(\{\pi(j) : 1 \leqq j \leqq n\})$ *and* p*)*

$$|s_n - t_p| \leqq \sum_{k=n_0+1}^{n_0+m} |a_k| < \varepsilon.$$

[87] This was proved by the German mathematician Peter Gustav Lejeune Dirichlet (1805-1859).
[88] $\sum a_k$ converges according to prop.10.19.

This implies

$$|s - t_p| \leqq |s - s_n| + |s_n - t_p| \ \leqq \ |s - s_n| \ + \sum_{k=n_0+1}^{n_0+m} |a_k| \ < \ |s - s_n| + \varepsilon.$$

We had chosen $n \geqq n_0$ *and it follows from* (10.73) *that* $|s - s_n| < \varepsilon$, *hence* $|s - t_p| < 2\varepsilon$.

We remember how p *was defined and see that it could be any integer* $\geqq p_0$, *a number which depends (via* n_0*) only on* ε.

To summarize: for all $\varepsilon > 0$ *there exists* p_0 *such that* $p \geqq p_0$ *implies* $|s - t_p|$. *But this means* $\lim\limits_{p\to\infty} t_p \ = \ s$. *On the other hand,* $\lim\limits_{p\to\infty} t_p \ = \ t \ = \ \sum\limits_{p\to\infty} b_k$.

This concludes the proof that $\sum\limits_{p\to\infty} a_k \ = \ \sum\limits_{p\to\infty} b_k$. ■

There are series which are convergent but not absolutely convergent. Such series are given a special name:

Definition 10.36 (conditionally convergent series)**.** A series $\sum a_j$ is called **conditionally convergent** if it is convergent but not absolutely convergent.

We introduce alternating series to give a simple example of a conditionally convergent series.

Definition 10.37 (alternating series)**.** A series $\sum a_j$ is called an **alternating series** if it is of the form $\sum (-1)^j a_j$ with either all terms a_j being strictly positive or all of them being strictly negative.

Proposition 10.20 (Leibniz test for alternating series)**.** *Let* (a_k) *be an alternating series such that the sequence* $|a_k|$ *is non–increasing:* $|a_1| \geqq |a_2| \geqq |a_3| \geqq \dots$. *Then* (a_k) *converges.*

Proof: Left as an exercise. ■

Example 10.15 (Alternating series)**.** The series $\sum (-1)^n$ and the alternating harmonic series $\sum (-1)^n/n$ are examples of alternating series.

It is known from calculus that the harmonic series $\sum 1/n$ is divergent: $\sum\limits_{j=1}^{\infty} \dfrac{1}{n} = \infty$. On the other hand, according to the Leibniz test, $\sum (-1)^n/n$ converges. It follows that the alternating harmonic series is convergent but not absolutely convergent, i.e., it is conditionally convergent.

We are going to prove Riemann's Reordering Theorem, from which it can be easily deduced that if $\sum a_j$ *is conditionally convergent and* $x \in \mathbb{R}$, *a rearrangement* $\sum a_{\pi_j}$ *can be found which converges to* x. *In preparation we must prove a lemma.*

Lemma 10.2. *Let* $\sum a_k$ *be a series. We split it into two series* $\sum p_k$ *and* $\sum q_k$ *as follows.*

p_j *is the* j*th strictly positive member of the sequence* $(a_k)_k$ *and* q_j *is the* j*th strictly negative member of that sequence.*

The following is true:

a. *If $\sum a_k$ is absolutely convergent then both $\sum p_k$ and $\sum q_k$ are (absolutely) convergent.*
b. *If $\sum a_k$ is convergent but not absolutely convergent then $\sum p_k$ has limit ∞ and $\sum q_k$ has limit $-\infty$.*

*Proof of **a**: Let $\alpha := \sum_{i=1}^{\infty} |a_i|$. Let $j \in \mathbb{N}$. p_j is the jth strictly positive member of the sequence $(a_k)_k$.*

Let m be the index such that a_m is the jth strictly positive member of the sequence $(a_k)_k$. Then

$$\sum_{i=1}^{j} p_i \;\leqq\; \sum_{i=1}^{m} |a_i| \;\leqq\; \sum_{i=1}^{\infty} |a_i| \;<\; \infty.$$

The above is true for all $j \in \mathbb{N}$ and it follows that $\sum_{i=1}^{\infty} p_i < \infty$. The proof that $\sum q_k$ has a finite limit is similar.

*Proof of **b**: For any $n \in \mathbb{N}$ we have*

$$\sum_{k=1}^{n} |a_k| \;\leqq\; \sum_{k=1}^{n} p_k \;+\; \sum_{k=1}^{n} (-q_k).$$

This is true because each one of $a_1, \dots, a_n$ is one of the first n strictly positive numbers $p_1, \dots, p_n$ or one of the strictly negative numbers $a_1, \dots, a_n$ or it is zero, in which case it contributes nothing to the series. Both series $\sum a_k$ and $\sum(-q_k)$ are non–decreasing, hence for each fixed n,

$$\sum_{k=1}^{n} |a_k| \;\leqq\; \sum_{k=1}^{\infty} p_k \;-\; \sum_{k=1}^{\infty} q_k.$$

It follows that if both $\sum p_k$ and $\sum q_k$ are convergent then so is $\sum |a_k|$, i.e., this series is absolutely convergent. We have a contradiction. ■

Theorem 10.17 (Riemann's Rearrangement Theorem). [89]

Let $\alpha, \beta \in \mathbb{R}$ such that $\alpha \leqq \beta$. and let the series $\sum a_k$ be conditionally convergent. Then a rearrangement $\sum b_k$ of $\sum a_k$ exists such that

$$\liminf_{n\to\infty} \sum_{k=1}^{n} b_k \;=\; \alpha \quad \textit{and} \quad \limsup_{n\to\infty} \sum_{k=1}^{n} b_k \;=\; \beta.$$

Proof: We may assume that $a_j \neq 0$ for all $j \in \mathbb{N}$ because those terms do not contribute anything to the partial sums, hence leave the limit of the series and any rearrangement unchanged.

We split $\sum a_j$ into the series $\sum p_j$ of its positive members and $\sum q_j$ of its negative members in the same way as was done in lemma 10.2:

p_j is the jth strictly positive member of the sequence $(a_k)_k$;
q_j is the jth strictly negative member of $(a_k)_k$.

[89] This was proved by the German mathematician Bernhard Riemann (1826-1866).

It was proved in lemma 10.2 that $\sum_{k=1}^{\infty} p_k = \infty$ *and* $\sum_{k=1}^{\infty} q_k = -\infty$.

case 1: $\beta \geqq 0$.

Let $U_1 := \{k \in \mathbb{N} : p_1 + p_2 + \cdots + p_k > \beta\}$. U_1 *is not empty because* $\sum p_j$ *has limit* ∞*, hence* $u_1 := \min(U_1)$ *exists. We call the list* $p_1, p_2, \ldots, p_{u_1}$ *the* ***first upcrossing*** *of the (unfinished) series* $\sum b_k$.

We now construct the first piece of the desired rearrangement $\sum b_k$. *Let*

$$n_1 := u_1; \quad b_1 := p_1,\ b_2 := p_2,\ \ldots,\ b_{n_1} := p_{u_1}; \quad \sigma_1 := \sum_{j=1}^{n_1} b_j.$$

Note that n_1 *is the first (and so far, only) index* n *of the series* $\sum b_k$ *for which* $\sum_{k=1}^{n} b_k$ *exceeds* β.

Let $L_1 := \{k \in \mathbb{N} : \sigma_1 + \sum_{j=1}^{k} q_j < \alpha\}$. L_1 *is not empty because* $\sum q_j$ *has limit* $-\infty$*, hence* $l_1 := \min(L_1)$ *exists. We call the list* $q_1, q_2, \ldots, q_{l_1}$ *the* ***first downcrossing*** *of* $\sum b_k$.

We add more terms to $b_1, b_2, \ldots, b_{n_1}$.

$$n_2 := n_1 + l_1; \quad b_{n_1+1} := q_1, b_{n_1+2} := q_2, \ldots, b_{n_2} := q_{l_1}; \quad \sigma_2 := \sum_{j=1}^{n_2} b_j.$$

Note that n_2 *is the first index* n *of* $\sum b_k$ *for which* $\sum_{k=1}^{n} b_k$ *drops below* α.

Let $U_2 := \left\{k \in \mathbb{N} : k > u_1 \text{ and } \sigma_2 + \sum_{j=u_1+1}^{u_1+k} p_j > \beta\right\}$. U_2 *is not empty because* $\sum_{j=u_1+1}^{\infty} p_j$ *has limit* ∞*, hence* $u_2 := \min(U_2)$ *exists. We call* $p_{u_1+1},\ p_{u_1+2}, \cdots,\ p_{u_2}$ *the* ***second upcrossing*** *of* $\sum b_k$.

We add more terms to $b_1, b_2, \ldots, b_{n_2}$.

$$n_3 := n_2 + u_2; \quad b_{n_2+1} := p_{u_1+1},\ b_{n_2+2} := p_{u_1+2}, \ldots,\ b_{n_3} := p_{u_1+u_2}; \quad \sigma_3 := \sum_{j=1}^{n_3} b_j.$$

Note that n_3 *is the second index* n *of the series* $\sum b_k$ *for which* $\sum_{k=1}^{n} b_k$ *exceeds* β.

Let $L_2 := \left\{k \in \mathbb{N} : k > l_1 \text{ and } \sigma_3 + \sum_{j=l_1+1}^{l_1+k} q_j < \alpha\right\}$. L_2 *is not empty because* $\sum_{j=l_1+1}^{\infty} q_j$ *has limit* $-\infty$*, hence* $l_2 := \min(L_2)$ *exists. We call* $q_{l_1+1},\ q_{l_1+2}, \cdots,\ q_{l_2}$ *the* ***second downcrossing*** *of* $\sum b_k$.

We add more terms to $b_1, b_2, \ldots, b_{n_3}$.

$$n_4 := n_3 + l_2; \quad b_{n_3+1} := q_{l_1+1},\ b_{n_3+2} := q_{l_1+2}, \ldots,\ b_{n_4} := q_{l_1+l_2}; \quad \sigma_4 := \sum_{j=1}^{n_4} b_j.$$

Note that n_4 *is the second index* n *of the series* $\sum b_k$ *for which* $\sum_{k=1}^{n} b_k$ *drops below* α.

It should be clear how we proceed. Let us assume that we have constructed the Nth upcrossing $p_{u_{N-1}+1},\ p_{u_{N-1}+2}, \cdots,\ p_{u_N}$ and from it

$$\begin{aligned} n_{(2N-1)} &:= n_{(2N-2)} + u_N; \\ b_{(n_{(2N-2)}+1)} &:= p_{(u_{(N-1)}+1)},\ b_{(n_{(2N-2)}+2)} := p_{(u_{(N-1)}+2)},\ \ldots,\ b_{(n_{(2N-1)})} := p_{u_N}, \\ \sigma_{(2N-1)} &:= \sum_{j=1}^{n_{(2N-1)}} b_j. \end{aligned}$$

Let us further assume that we have constructed the Nth downcrossing $q_{l_{N-1}+1},\ q_{l_{N-1}+2}, \cdots,\ q_{l_N}$ and from it

$$\begin{aligned} n_{(2N)} &:= n_{(2N-1)} + l_N; \\ b_{(n_{(2N-1)}+1)} &:= q_{(l_{(N-1)}+1)},\ b_{(n_{(2N-1)}+2)} := q_{(l_{(N-1)}+2)},\ \ldots,\ b_{n_{(2N)}} := q_{l_N}, \\ \sigma_{2N} &:= \sum_{j=1}^{n_{(2N)}} b_j. \end{aligned}$$

We proceed to construct the $(N+1)$th upcrossing and the $(N+1)$th downcrossing as follows.

Let $U_{N+1} := \left\{k \in \mathbb{N} : k > u_N \text{ and } \sigma_{2N} + \sum_{j=u_N+1}^{u_N+k} p_j > \beta\right\}$. U_{N+1} is not empty because $\sum_{j=u_N+1}^{\infty} p_j$ has limit ∞, hence $u_{N+1} := \min(U_{N+1})$ exists. We call $p_{(u_N+1)},\ p_{(u_N+2)}, \cdots,\ p_{u_{(N+1)}}$ the $(N+1)$th upcrossing of $\sum b_k$.

We add more terms to $b_1, b_2, \ldots, b_{n_{2N}}$.

$$\begin{aligned} n_{(2N+1)} &:= n_{(2N)} + u_{(N+1)}; \\ b_{(n_{(2N)}+1)} &:= p_{(u_N+1)},\ b_{(n_{(2N)}+2)} := p_{(u_N+2)},\ \ldots,\ b_{(n_{(2N+1)})} := p_{u_{(N+1)}}, \\ \sigma_{(2N+1)} &:= \sum_{j=1}^{n_{(2N+1)}} b_j. \end{aligned}$$

Let $L_{N+1} := \left\{k \in \mathbb{N} : k > l_N \text{ and } \sigma_{2N+1} + \sum_{j=l_N+1}^{l_N+k} q_j < \alpha\right\}$. L_{N+1} is not empty because $\sum_{j=l_N+1}^{\infty} q_j$ has limit ∞, hence $l_{N+1} := \min(L_{N+1})$ exists. We call $q_{(l_N+1)},\ q_{(l_N+2)}, \cdots,\ q_{l_{(N+1)}}$ the $(N+1)$th downcrossing of $\sum b_k$.

We add more terms to $b_1, b_2, \ldots, b_{n_{(2N+1)}}$.

$$\begin{aligned} n_{(2(N+1))} &:= n_{(2N+1)} + l_{(N+1)}; \\ b_{(n_{(2N+1)}+1)} &:= q_{(l_N+1)},\ b_{(n_{(2N+1)}+2)} := q_{(l_N+2)},\ \ldots,\ b_{(n_{2(N+1)})} := q_{l_{(N+1)}}, \\ \sigma_{2(N+1)} &:= \sum_{j=1}^{n_{2(N+1)}} b_j. \end{aligned}$$

We have defined by recursion $\sum_{k=1}^{n_N} b_k$ *for all* $N \in \mathbb{N}$ *Note that the increasing sequence* $(n_N)_{N\in\mathbb{N}}$ *is not bounded above because* $n_(2N)$ *is the number of terms that belong to the first* N *upcrossings plus the first* N *downcrossings. But each upcrossing and each downcrossing must have at least one term because at least one term* p_j *is needed to move a partial sum from below* α *to above* β *and at least one term* q_j *is needed to move a partial sum from above* β *to below* α*, hence* $n_{2N} \geqq 2N$*. It follows that* $\sum b_k$ *indeed has infinitely many terms.*

We note that all positive terms p_j *and all negative terms* q_j *are being used in sequence, starting with the first one. It follows that each one of the terms of* $\sum a_k$ *has become part of* $\sum b_k$*. It follow that* $\sum b_k$ *is indeed a rearrangement of* $\sum a_k$*.*

Let $s_n := \sum_{j=1}^{n} b_j$*.* $n_1, n_3, n_5, \dots$ *are (precisely the) integers* n *for which* $s_n > \beta$ *and* $n_2, n_4, n_6, \dots$ *are (precisely the) integers* n *for which* $s_n < \alpha$*. There are infinitely many of each and it follows from thm.8.1 (Characterization of limsup and liminf) on p.124 that*

(10.74)
$$\liminf_{n\to\inf} s_n \leqq \alpha \quad \textit{and} \quad \limsup_{n\to\inf} s_n \geqq \beta.$$

We now prove that for any $\varepsilon > 0$

(10.75)
$$\liminf_{n\to\inf} s_n \geqq \alpha - \varepsilon \quad \textit{and} \quad \limsup_{n\to\inf} s_n \leqq \beta + \varepsilon.$$

Let $\varepsilon > 0$ *The terms* $(a_n)_n$ *of the original series* $\sum a_k$ *converge to zero because* $\sum a_k$ *converges (see cor.10.1 on p.190). It follows that there exists* $n_0 \in \mathbb{N}$ *such that* $|a_j| < \varepsilon$ *for all* $j \geqq n_0$*. We show next that*

(10.76)
$$|p_j| \ = \ p_j < \varepsilon \textit{ and } |q_j| \ = \ -q_j < \varepsilon \textit{ for all } j \geqq n_0.$$

$|p_j| = p_j < \varepsilon$ *is true whenever* $j \geqq n_0$ *because* p_j *is the* j*th positive member of* $(a_n)_n$*, hence* $p_j = a_i$ *for some* $i \geqq j \geqq n_0$*. Likewise,* $|q_j| = -q_j < \varepsilon$ *whenever* $j \geqq n_0$ *because* q_j *is the* j*th negative member of* $(a_n)_n$*, hence* $q_j = a_i$ *for some* $i \geqq j \geqq n_0$*. We have proved (10.76).*

We recall that $n_1, n_3, n_5, \dots$ *are precisely the integers* n *for which* $s_n > \beta$*, so*

$$s_{(n_1-1)} \leqq \beta,\ s_{(n_3-1)} \leqq \beta,\ \dots,\ s_{(n_{(2j-1)}-1)} \leqq \beta,\ \dots$$

. But then $s_{(n_{(2j-1)})} \leqq \beta + \varepsilon$ *because less than* ε *was added to the previous term (which is no bigger than* β*) for any* j *so big that the last item in the* j*th upcrossing is less than* ε

It follows from (10.76) that this is certainly true if $j \geqq n_0$ *because each upcrossing has size of at least 1. It follows that there are at most finitely many indices* n *such that* $s_n > \beta + \varepsilon$ *and we conclude that* $\limsup_n s_n \leqq \beta + \varepsilon$*. A similar reasoning allows us to conclude that* $\liminf_n s_n \geqq \alpha - \varepsilon$*.*

We have proved that (10.75) is in fact true and this yields, together with (10.74), that

(10.77)
$$\liminf_{n\to\inf} s_n = \alpha \quad \textit{and} \quad \limsup_{n\to\inf} s_n = \beta.$$

We have proved the theorem for case 1: $\beta \geqq 0$

case 2: $\beta < 0$. *We proceed exactly as in case 1. The only difference is that we start with a downcrossing that gets us below* α *rather than an upcrossing to obtain a rearrangement* $\sum c_k$ *for which a partial sum exceed* α *exactly when its last terms are all terms of an upcrossing and it drops below* β *exactly when its last terms are all terms of a downcrossing. Because* a_j *converges to zero there will again only be finitely many upcrossings and downcrossings with terms that exceed* ε. *For all others the partial sums cannot exceed* β *or drop below* α *by more than* ε *and we conclude as before that*

$$\liminf_{n\to\inf} \sum_{k=1}^{n} c_k = \alpha \quad \textit{and} \quad \limsup_{n\to\inf} \sum_{k=1}^{n} c_k = \beta. \blacksquare \tag{10.78}$$

Corollary 10.3. *Let the series* $\sum a_k$ *be conditionally convergent and let* $\alpha \in \mathbb{R}$. *Then a rearrangement* $\sum b_k$ *of* $\sum a_k$ *exists such that*

$$\lim_{n\to\infty} \sum_{k=1}^{n} b_k = \alpha.$$

Proof: Let $\beta := \alpha$.

We apply Riemann's Reordering Theorem to the special case $\beta = \alpha$: *There is a rearrangement* $\sum b_j$ *of* $\sum a_j$ *such that*

$$\liminf_{n\to\infty} \sum_{k=1}^{n} b_k = \alpha \quad \textit{and} \quad \limsup_{n\to\infty} \sum_{k=1}^{n} b_k = \alpha$$

*It follows now from thm.*8.2 *on p.*126 *that* $\sum b_j$ *converges to* α. ■

We have seen that if a series is absolutely convergent then it is convergent and each rearrangement converges to the same limit. Here is the reverse.

Corollary 10.4. *Let series* $\sum a_k$ *be a convergent series with limit* $\alpha \in \mathbb{R}$ *such that each rearrangement* $\sum b_k$ *also converges to* α.

Then $\sum a_k$ *is absolutely convergent.*

Proof: We assume to the contrary that $\sum a_k$ *is not absolutely convergent. This series is convergent but not absolutely convergent, hence conditionally convergent. We apply Riemann's Reordering Theorem and find that there is a rearrangement of* $\sum a_j$ *which converges to a different real number, contrary to our assumption.* ■

Corollary 10.5 (Dichotomy for convergent series)**.** *Let series* $\sum a_k$ *be a convergent series. Then either* ***a*** *or* ***b*** *is true:*

- ***a.*** *All rearrangements of* $\sum a_k$ *converge to the same limit.*
- ***b.*** *For any* $\alpha \in \mathbb{R}$ *there is a rearrangement of* $\sum a_k$ *which converges to* α.

Proof: Either $\sum a_k$ *is absolutely convergent and* ***a*** *is true according to Riemann's Reordering Theorem or the series it is conditionally convergent and* ***b*** *is true according to cor.*10.3. ■

10.4 Addenda to chapter 10 (Metric spaces)

10.4.1 Misc. Addenda to Metric Spaces

Proposition 10.21. *Let (X, d) be a metric space. Let $n \in \mathbb{N}$ and $x_1, x_2, \dots, x_n \in X$. Then*

$$d(x_1, x_n) \leqq \sum_{j=1}^{n-1} d(x_j, x_{j+1}) = d(x_1, x_2) + d(x_2, x_3) + d(x_{n-1}, x_n). \tag{10.79}$$

Proof: The proof is required as part of an upcoming homework.

10.4.2 Exercises

Exercise 10.2. It was stated in example 10.5 on p.160 that the discrete topology which is induced by the discrete metric $d(x, y) = 1$ if $x \neq y$ and 0 if $x = y$ is the entire powerset 2^X of X. Prove it.

Exercise 10.3. If (X, d) is a metric space and $x_1, \dots x_n \in X$ then $d(x_1, x_n) \leqq \sum_{j=1}^{n-1} d(x_j, x_{j+1})$ (Prop.10.21)

Exercise 10.4. Let (X, d) be a metric space and $A \subseteq X$, $A \neq \emptyset$. Let

$$\gamma := \gamma(A) := \inf\{d(x, y) : x, y \in A \text{ and } x \neq y\}. \tag{10.80}$$

a. Prove that if $\gamma > 0$ then A is complete.

b. The reverse is not true. Find a counterexample.

Exercise 10.5. Let (X, d) be a metric space and let $A \subseteq X$ be a finite subset. Prove that A is complete.

Exercise 10.6. Given is $\mathbb{R}$ with the Euclidean metric $d(x, y) = |x - y|$. We look at $\mathbb{N}$ and $\mathbb{Q}$ as metric subspaces of $\mathbb{R}$. We know that $\mathbb{Q}$ is not complete.

a. Is $\mathbb{N}$ complete as a subspace of $\mathbb{Q}$?
b. Is $\mathbb{N}$ complete as a subspace of $\mathbb{R}$?
Prove your answer.

Exercise 10.7. Let X be a nonempty set with the discrete metric $d(x, y) = 1 - 1_{\{x\}}(y)$, i.e., $d(x, y) = 0$ if $x = y$ and 1 else. Prove that (X, d) is complete.

11 Compactness (Study this!)

11.1 Introduction: Closed and bounded sets in Euclidean space (Understand this!)

One of the results that are true for N-dimensional space is the "sequence compactness" of closed and bounded subsets: Any sequence that lives in such a set has a convergent subsequence. We will discuss that next.

Theorem 11.1 (Convergent subsequences in closed and bounded sets of $\mathbb{R}$)**.** *Let A be a bounded and closed set of real numbers and let (z_n) be an arbitrary sequence in A. Then there exists $z \in A$ and a subset*

$$n_1 < n_2 < \ldots < n_j < \ldots \text{ of indices such that } z = \lim_{j\to\infty} z_{n_j}$$

i.e., the subsequence (z_{n_j}) converges to z.

Proof: Let m be the midpoint between $a := \inf(A)$ and $b := \sup(A)$. Because A is bounded, a and b must exist as finite numbers. Let

(11.1) $$A_{\star 1} := A \cap [a, m]; \qquad A^{\star}{}_1 := A \cap [m, b].$$

Then at least one of $A_{\star 1}$, $A^{\star}{}_1$ must contain infinitely many of the z_n because $A_{\star 1}$ and $A^{\star}{}_1$ form a "covering" of A (the formal definition will be given later in def.11.4 on p.212), i.e., $A_{\star 1} \cup A^{\star}{}_1 \supseteq A$. We pick one with infinitely many elements and call it A_1. In case both sets contain infinitely many of the z_n, it does not matter which one we pick. Do you see that $diam(A_1) \leqq diam(A)/2$?

Let m_1 be the midpoint between $a_1 := \inf(A_1)$ and $b_1 := \sup(A_1)$. Let

(11.2) $$A_{\star 2} := A_1 \cap [a_1, m_1]; \qquad A^{\star}{}_2 := A \cap [m_1, b_1].$$

Then at least one of $A_{\star 2}$, $A^{\star}{}_2$ must contain infinitely many of the z_n. We pick one with infinitely many elements and call it A_2. In case both sets contain infinitely many of the z_n, it does not matter which one we pick. Note that

$$diam(A_2) \leqq diam(A_1)/2 \leqq diam(A)/2^2$$

We keep picking the midpoints m_j of the sets A_j each of which has at most half the diameter of the previous one. (Why?) In other words, we have constructed a sequence

(11.3) $$\begin{aligned} &A \supseteq A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots \qquad \text{such that} \\ &diam(A) \geqq 2diam(A_1) \geqq 2^2 diam(A_2)/ \ldots \geqq 2^n diam(A_n)/ \ldots \end{aligned}$$

which means that $diam(A_n) \leqq diam(A)/2^n \to 0$ as $n \to \infty$.

We pick a subsequence $(x_j) = (z_{n_j})$ of the original sequence (z_n) such that $z_{n_j} \in A_j$ for all $j \in \mathbb{N}$. This is not too hard because the sets A_j were picked in such a way that each one of them contains infinitely many of the z_k.

The following inequality is true because the sequence of sets (A_j) is "nested": each A_j is contained in its predecessor A_{j-1}. It follows that A_m contains all A_k for any $k > m$ and this implies that A_m contains all members $x_k = z_{n_k}$, for all $k > m$. Thus

$$|x_m - x_k| \leqq diam(A_m) \leqq \frac{diam(A)}{2^m} \qquad \text{for all } m \text{ and } k \text{ such that } \quad k > m.$$

This means that (x_n) is a Cauchy sequence (p.170). According to theorem 10.8 about the completeness of $\mathbb{R}$ (p.171) there is a contact point x such that $x_n \to x$ for $n \to \infty$.
Because A is a closed set it contains all its contact points. It follows that $x \in A$ and we have found a subsequence of the original sequence (z_n) which converges to an element of A. ■

Theorem 11.2 (Convergent subsequences in closed and bounded sets of $\mathbb{R}^N$)**.** *Let A be a bounded and closed set of $\mathbb{R}^N$ and let $(\vec{z}_n)$ be an arbitrary sequence of N–dimensional vectors in A. Then there exists $\vec{z} \in A$ and a subset*

$$n_1 < n_2 < \ldots < n_j < \ldots \text{ of indices such that } \vec{z} = \lim_{j\to\infty} \vec{z}_{n_j}$$

i.e., the subsequence $(\vec{z}_{n_j})$ converges to $\vec{z}$.

Proof (outline): We review the above proof for $\mathbb{R}$:

The base idea was to chop A in half during each step to obtain a sequence of sets A_n which become smaller and smaller in diameter but yet contain infinitely many points. of the original sequence z_n.

In higher dimensions we would still find the center point $\vec{m}_n$ which is determined by the fact that it is the center of a γ–neighborhood (N–dimensional ball) that contains A_n and does so with the smallest radius possible. We then take the minimal square (in $\mathbb{R}^2$) or the minimal N–dimensional cube (in $\mathbb{R}^N$) that is parallel to the coordinate axes and still contains that sphere or ball.

We then divide that N–dimensional cube (a square in 2 dimensions, a cube in 3 dimensions) into 2^N sectors (4 quadrants in $\mathbb{R}^2$, 8 sectors in $\mathbb{R}^3$) and partition A_n into at most 2^N pieces by intersecting it with those 2^N sectors).

The set A_{n+1} is then chosen from one of those pieces which contain infinitely many of the z_n. Again, we get a nested sequence A_n whose diameters contract towards 0. You'll find more detail about the messy calculations required in the proof of prop.11.2 on p.204.

Each A_n contains infinitely many of the $(\vec{z}_k)$. Now pick $\vec{x}_j := \vec{z}_{n_j}$ where $\vec{z}_{n_j}$ is one of the infinitely many members of the original sequence $(\vec{z}_n)$ which are contained in A_j.

Because $A_j \subseteq A_K$ for $j \geqq K$ and $\lim\limits_{K\to\infty} diam(A_K) = 0$, we do the following for a given $\varepsilon > 0$: choose K so big that $diam(A_K) \leqq \varepsilon$. Note that

$$\text{if } i,j \geqq K \text{ then } d(\vec{x}_i, \vec{x}_j) = d(\vec{z}_{n_i}, \vec{x}_{n_j}) \leqq diam(A_K) \leqq \varepsilon$$

because $n_i \geqq i$ (and $n_j \geqq j$), hence $\vec{x}_i, \vec{x}_j \in A_K$. It follows that the sequence $\vec{x}_j$ is Cauchy. We have seen in thm.10.9 on p.171 that $\mathbb{R}^N$ is complete, and it follows that $\vec{L} := \lim_{j\to\infty} \vec{x}_j$ exists in $\mathbb{R}^N$.

The proof is complete if it can be shown that $\vec{L} \in A$.

But we know that all $\vec{x}_i = \vec{z}_{n_i}$ belong to A. $\vec{L}$ is a contact point of A because any neighborhood $N_\varepsilon(\vec{L})$ contains an entire tail set of the sequence $(\vec{x}_i)_i$. As the closed set A owns all its contact points, it follows that $L \in A$ and the theorem is proved. ■

Theorem 11.3. *Let A be a bounded and closed set of real numbers and let $f(\cdot) : A \to \mathbb{R}$ be a continuous function on A. Then $f(\cdot)$ is a bounded function.*

Proof: Let us assume that $f(\cdot)$ is not bounded and conclude something that is impossible.
An unbounded function is not bounded from above, from below, or both. We might as well assume that $f(\cdot)$ is not bounded from above because otherwise it is not bounded from below and we can work with $-f(\cdot)$ which then is not bounded from above. We conclude that there is a sequence $(z_n) \in A$ such that

$$f(z_n) > n \qquad \text{for all } n \in \mathbb{N}. \tag{11.4}$$

According to thm.11.1 ("Convergent subsequences in closed and bounded sets") there exists a subsequence $(x_j) = (z_{n_j})$ and $x_0 \in A$ such that $x_0 = \lim_n \to \infty x_n$.

In particular, $f(x_0)$ exists as a finite value and $f(x_n) \to f(x_0)$ because $f(\cdot)$ is continuous in x_0. But the x_n were constructed as a subsequence of the z_j which have the property that $f(z_j) > j$ for all j.

The subsequence $(f(x_n))$ cannot converge to $f(x_0)$ because $f(x_j) = f(z_{n_j}) > n_j$, i.e., $\lim_{j\to\infty} f(z_{n_j}) = \infty$.
We have reached a contradiction and it follows that $f(\cdot)$ is bounded. ■

Corollary 11.1. *Let $a < b$ be two real numbers and let $f(\cdot) : [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then $f(\cdot)$ is a bounded function.*

Proof: The interval interval $[a, b]$ is closed and bounded (diam$([a, b]) = b - a$). and the proof follows from theorem 11.3. ■

11.2 Four definitions of compactness

We now look at ways to extend those results to general metric spaces by looking at the concept of compactness.

Compact sets are a wonderful thing to deal with because they allow you in some sense to go from dealing with "arbitrarily many" to dealing with "countably many" and even "finitely many". This chapter will show that A, B and C below are equivalent statements for any subspace $(K, d|_{K\times K})$ of a metric space (X, d):

A.	*any sequence in K has a convergent subsequence*
B.	*K is complete and contains only finitely many point of a grid of length ε*
C.	*any open covering of K has a finite subcovering*
D.	*K is bounded and closed - ONLY works in $\mathbb{R}^N$!*

Such metric spaces K will be called "compact" (see def.11.3 (Sequence compactness) on p.206 and def.11.5 (Compact sets) on p.213).

We can now state theorems 11.1 and 11.2 of the introduction as follows: In $\mathbb{R}^N$ statement D implies A.

When you take a course on real analysis you will probably be given the definition of compactness as that in C: any open covering of K has a finite subcovering. In this document this definition is pushed into the background as it is the most difficult to understand. But full proofs will be given of the equivalence of sequence compactness (def.A) on the one hand and completeness plus "total boundedness" (def.B) on the other hand.

One of the important results of this chapter on compactness is that, if you look at $\mathbb{R}^N$ with the Euclidean norm and its associated metric

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2} \qquad (\,\vec{x} = (x_1, x_2, \dots),\ \vec{y} = (y_1, y_2, \dots) \in \mathbb{R}^N\,)$$

(see (9.3) on p.135) then all four statements A, B, C, D coincide.

11.3 ε-nets and total boundedness

We now briefly discuss ε–nets and decreasing sequences of closed sets which contract to a single point.

Definition 11.1 (ε–nets)**.** Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \subseteq X$. let $G \subseteq A$ be a subset of A with the following property:

For each $x \in A$ there exists $g \in G$ such that $x \in N_\varepsilon(g)$.

In other words, the points of G form a "grid" or "net" fine enough so that no matter what point x of A you choose, you can always find a "grid point" g with distance less than ε to x, because that is precisely the meaning of $x \in N_\varepsilon(g)$.

We call G an ε**–net** or ε**–grid** for A and we call $g \in G$ a **grid point** of the net.

Proposition 11.1 (ε–nets and coverings)**.** *Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \subseteq X$. Let $G \subseteq A$ be an ε–grid for A. Then $\{N_\varepsilon(g)\}_{g \in G}$ is an open covering of A in the sense of def.11.4 on p.212: It is a collection of open sets the union of which "covers", i.e., contains, A.*

Proof: *Let $x \in A$. We can choose a point $g = g(x) \in G$ such that $x \in N_\varepsilon\big(g(x)\big)$. It follows from $\{x\} \subseteq N_\varepsilon\big(g(x)\big)$ and $g(x) \in G$ for all $x \in A$ that*

$$A = \bigcup_{x \in A} \{x\} \subseteq \bigcup_{x \in A} N_\varepsilon\big(g(x)\big) \subseteq \bigcup_{g \in G} N_\varepsilon(g). \blacksquare$$

Proposition 11.2 (ε–nets in $\mathbb{R}^N$)**.** *Let (X, d) be $\mathbb{R}^N$ with the Euclidean metric.*

A. *Let*

$$\mathbb{Z}^N = \{\vec{z} = (z_1, z_2, \dots z_N) : z \in \mathbb{Z}\}$$

In other words, $\mathbb{Z}^N$ is the set of all points in $\mathbb{R}^N$ with integer coordinates. [90]
Then $\mathbb{Z}^N$ is a $\sqrt{N}$–net of $\mathbb{R}^N$.

B. *Let $\varepsilon > 0$ and $G_\varepsilon^{\mathbb{R}^N} := \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^N\}$. Then $G_\varepsilon^{\mathbb{R}^N}$ is an $\varepsilon\sqrt{N}$–net of $\mathbb{R}^N$.*

C. *Let A be a bounded set in $\mathbb{R}^N$ and $\varepsilon > 0$. Then there is $n \in \mathbb{N}$ and $g_1, \dots g_n \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$ such that*

$$A \subseteq N_\varepsilon(g_1) \cup N_\varepsilon(g_2) \cup \dots, \cup N_\varepsilon(g_n),$$

i.e., A is covered by finitely many ε–neighborhoods of points in the $(\varepsilon/\sqrt{N})$–grid $G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$.

(Skip this proof!) *(all three parts A, B, C)*

*Proof of **A.***
Let $\vec{x} = (x_1, x_2, \dots x_N) \in \mathbb{R}^N$. For each x_j let $x_j^\star$ be the integer closest to x_j.

[90] That is as intuitive a grid as you can think of, especially if you look at the 2–dimensional plane or 3–dimensional space.

Before we continue, let's have an example, if $N = 5$ and $\vec{x} = (12.85, -12.35, 2/3, 9, -\pi)$ then its associated grid point is $\vec{x^\star} = (13, -12, 1, 9, -3)$. Let's compute the distance:

$$d(\vec{x}, \vec{x^\star}) = \sqrt{.15^2 + .35^2 + 1/3^2 + 0 + (\pi - 3)^2} \leqq \sqrt{(1/2 + 1/2 + 1/2 + 0 + 1/2)} \leqq \sqrt{N}$$

*and we see that part **A** of the lemma is true for this specific example.*

Now to the real proof. It is not really more complicated if you notice that $|x_j - x_j^\star| < 1$ for all $1 \leqq j \leqq N$. We get

$$d(\vec{x}, \vec{x^\star}) = \sqrt{\sum_{j=1}^{N} (x_j - x_j^\star)^2} < \sqrt{N \cdot 1} = \sqrt{N}$$

So, for each point you can find a grid point with integer coordinates at a distance of less than $\sqrt{N}$. That proves that $\mathbb{Z}^N$ is a $\sqrt{N}$–net of $\mathbb{R}^N$.

*Proof of **B.***
*Let $\vec{y} \in \mathbb{R}^N$. Let $\vec{x} := (\sqrt{N}/\varepsilon)\vec{y}$ and let $\vec{x^\star}$ be the vector where we discard the decimal parts of $\vec{x}$. According to part **A.** we know that $\|\vec{x} - \vec{x}^\star\|_2 = d(\vec{x^\star}, \vec{x}) < \sqrt{N}$. Thus*

$$d\big(\vec{y}, \frac{\varepsilon}{\sqrt{N}}\,\vec{x}^\star\big) = d\big(\frac{\varepsilon}{\sqrt{N}}\,\vec{x}, \frac{\varepsilon}{\sqrt{N}}\vec{x}^\star\big) = \| \frac{\varepsilon}{\sqrt{N}}\,\vec{x} - \frac{\varepsilon}{\sqrt{N}}\vec{x}^\star\|_2 = \frac{\varepsilon}{\sqrt{N}}\,\|\vec{x} - \vec{x}^\star\|_2 < \frac{\varepsilon}{\sqrt{N}}\,\sqrt{N} = \varepsilon$$

In other words, for any $\vec{y} \in \mathbb{R}^N$ there is a vector $\vec{x}^\star \in \mathbb{Z}^N$ such that $d(\vec{y}, (\varepsilon\,\sqrt{N})\vec{x}^\star) < \varepsilon$.

Rephrase that: For any $\vec{y} \in \mathbb{R}^N$ there is a vector $\vec{g} \in G^{\mathbb{R}^N}_{(\varepsilon/\sqrt{N})} = \{(\varepsilon/\sqrt{N})\vec{z} : \vec{z} \in \mathbb{Z}^N\}$

such that $d(\vec{y}, \vec{g}) < \varepsilon$ (choose $\vec{g} = \vec{x}^\star$).

So, for each point you can find a grid point in $G^{\mathbb{R}^N}_\varepsilon$ at a distance of less than $\varepsilon\sqrt{N}$. It follows that $G^{\mathbb{R}^N}_\varepsilon$ is an $\varepsilon\sqrt{N}$–net of $\mathbb{R}^N$.

*Proof of **C.***
Intuitively clear but very messy. Here is an outline.

***a.** You can choose a radius R_1 so big that $A \subseteq N_{R_1}(\vec{0})$ (see prop.10.6 on p.165).*

***b.** We enlarge the radius by ε: Let $R := R_1 + \varepsilon$. The enlarged "N-dimensional ball" of radius R $N_R(\vec{0})$ is contained in the "N-dimensional cube"*

$$Q_R := \{\vec{x} = (x_1, x_2, \ldots x_n) : -R \leqq x_j \leqq R \text{ for all } 1 \leqq j \leqq N\}.$$

This is true because if $\vec{x} = (x_1, x_2, \ldots x_n) \in N_R(\vec{0})$ then $|x_j| = \sqrt{x_j^2} \leqq \sqrt{\sum_i x_i^2} = \|\vec{x}\|_2 = R$.

***c.** Let $\vec{z} = (z_1, z_2, \ldots z_n)$ be a grid point, i.e., $z_j = m_j\varepsilon$ for the j-th coordinate ($m_j \in \mathbb{Z}$). There are only finitely many integers m, say K, for which $-R \leqq \varepsilon \cdot m \leqq R$.*

d. *Hence there are only K possible values for the first coordinate $z_1 = m_1\varepsilon$. For each one of those there are only K possible values for z_2, so there are at most K^2 possible combinations (z_1, z_2) for which $\vec{z} \in A$. We keep going and find that there are at most K^N possible grid points $\vec{z} \in Q_R$.*

e. *Any point in $\mathbb{R}^N$ with distance less than ε from some point in A must belong to $B_R(\vec{0})$ (now you know why we chose to augment R_1 by ε). In particular, all grid points $g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}}$ whose neighborhoods $N_\varepsilon(g)$ intersect A belong to $B_R(\vec{0})$ and hence to Q_R. We conclude that $A \cap N_\varepsilon(g) = \emptyset$ for all grid points outside Q_R, hence*

$$\bigcup[\, A \cap N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}}\,] = \bigcup[\, A \cap N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}} \cap Q_R\,].$$

*We know from part **B** which was already proved that $\mathbb{R}^N = \bigcup[\, N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}}\,]$. Hence,*

$$\begin{aligned} A = A \cap \mathbb{R} = A \cap \bigcup[\, N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}}\,] \\ = \bigcup[\, A \cap N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}}\,] = \bigcup[\, A \cap N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}} \cap Q_R\,]. \end{aligned}$$

*It follows that $A \subseteq \bigcup[N_\varepsilon(g) : g \in G^{\mathbb{R}^N}_{\varepsilon/\sqrt{N}} \cap Q_R\,]$. We have proved **C** as there are only finitely many grid points in Q_R.* ■

Remark 11.1. The observant reader will have noted that, in part **C.** of the previous proposition, it was not stated that the gridpoints belong to the subset A of $\mathbb{R}^N$. Here is a trivial counterexample. Look at the "standard" ε–grid $G^{\mathbb{R}^N}_\varepsilon = \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^N\}$ defined in prop.11.2, part **B**. Take any $A \subseteq \mathbb{R}^N$ you like and look at $B := A \setminus G^{\mathbb{R}^N}_\varepsilon$, i.e., we have removed all grid points. It is clear that B cannot be covered by $\sqrt{N}\varepsilon$ balls belonging to grid points in B. Matter of fact, B cannot be covered by balls of any radius if their centers are grid points in B.

Definition 11.2 (Total boundedness)**.** Let (X, d) be a metric space and let A be a subset of X. We say that A is **totally bounded** if for each $\varepsilon > 0$ there is a finite collection $\mathscr{G}_\varepsilon := \{g_1, \dots g_n\}$ of points in A whose ε–balls $N_\varepsilon(g_j)$ cover A: For any $a \in A$ there is $j = j(a)$ such that $d(a, g_j) < \varepsilon$.

We will use this definition in connection with sequence compactness which is defined in the next section.

11.4 Sequence compactness

We saw in the introductory section that, for the space $\mathbb{R}^N$ with the Euclidean metric, closed and bounded sets have the property that any sequence contains a convergent subsequence. We named this property in section 11.2, p.203 on Four definitions of compactness "sequence compactness" and we will examine that property in this chapter.

Definition 11.3 (Sequence compactness)**.** Let (X, d) be a metric space and let A be a subset of X. We say that A is **sequence compact** or **sequentially compact** if it has the following property: Given any sequence (x_n) of elements of A, there exists $L \in A$ and a subset

$$n_1 < n_2 < \ldots < n_j < \ldots \quad \text{of indices such that} \quad L = \lim_{n\to\infty} x_{n_j},$$

i.e., there exists a subsequence (x_{n_j}) which converges to L.

Proposition 11.3 (Sequence compactness implies total boundedness). *Let (X, d) be a metric space and let A be a sequentially compact subset of X. Then A is totally bounded.*

Proof: Nothing needs to be shown if A is empty, so we may assume that $A \neq \emptyset$. This is proved by contradiction.
a. *Assume that A is not totally bounded. Then there is $\varepsilon > 0$ such that for any finite collection of points $z_1, z_2, \dots z_n \in A$ the union $\bigcup_{1 \leq j \leq n} N_\varepsilon(z_j)$ does not cover A: There exists $z \in A$ outside any one of those ε–neighborhoods, i.e., $z \in A \setminus \bigcup \left[N_\varepsilon(z_j) : j \leqq n \right]$.*

This allows us to create an infinite sequence $(x_j)_{j \in \mathbb{N}}$ such that $d(x_j, x_n) \geqq \varepsilon$ for all $j, n \in \mathbb{N}$ such that $j \neq n$, say, $j < n$, as follows: We pick

$$x_1 \in A; \quad x_2 \in A \setminus N_\varepsilon(x_1); \quad x_3 \in A \setminus \big(N_\varepsilon(x_1) \cup N_\varepsilon(x_2) \big); \ \dots \ x_n \in A \setminus \bigcup_{j<n} N_\varepsilon(x_j); \ \dots$$

b. *The proof is done if we can show that $(x_j)_{j \in \mathbb{N}}$ does not possess a convergent subsequence. Assume to the contrary that there is $L \in A$ and $n_1 < n_2 < \dots$ such that $lim_{j \to \infty} x_{n_j} = L$.*

*We pick the number $\varepsilon > 0$ that was used in part **a** of the proof. There exists $N = N(\varepsilon)$ such that $d(x_{n_m}, L) < \varepsilon/2$ for all $m \geqq N$. Because $i \leqq n_i$ and $j \leqq n_j$, it follows for all $i, j \geqq N$ that*

$$d(x_{n_i}, x_{n_j}) \leqq d(x_{n_i}, L) + d(L, x_{n_j}) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

*But the x_n were constructed in such a fashion that $d(x_m, x_k) \geqq \varepsilon$ for **all** $m \neq k$, in particular for $m := n_i \neq k := n_j$. We have arrived at a contradiction because $n_i \neq n_j$ whenever $i \neq j$.* ■

Proposition 11.4 (Sequence compact implies completeness). *Let (X, d) be a metric space and let A be a sequence compact subset of X. Then A is complete, i.e., any Cauchy sequence (x_{n_j}) in A converges to a limit $L \in A$.*

Proof: Let (x_n) be a Cauchy sequence in A and let $\varepsilon > 0$. There exists $N_1 \in \mathbb{N}$ such that

$$k, l \geqq N_1 \ \Rightarrow \ d(x_k, x_l) \ < \ \varepsilon/2. \qquad (\star) \tag{11.5}$$

Because A is sequence compact, we can extract a subsequence $z_j := x_{n_j}$ and find $L \in A$ such that $z_j \to L$ as $j \to \infty$. It follows that for ε chosen above there exists $N_2 \in \mathbb{N}$ such that

$$j \geqq N_2 \ \Rightarrow \ d(x_{n_j}, L) \ < \ \varepsilon/2. \qquad (\star\star) \tag{11.6}$$

Let $N := max(N_1, N_2)$ and $j \geqq N$. We observe that $n_j \geqq j$ for all j, hence $n_j \geqq N$ if $j \geqq N$. Hence $j \geqq N_1$ and $n_j \geqq j \geqq N \geqq N_2$

It follows from $(\star)$ that $d(x_j, x_{n_j}) < \varepsilon/2$ and from $(\star\star)$ that $d(x_{n_j}, L) < \varepsilon/2$, hence $d(x_j, L) < \varepsilon$ for all $j \geqq N$. We have proved that the arbitrarily chosen Cauchy sequence (x_n) converges. ■

The last two propositions have proved that any sequence compact set in a metric space is both totally bounded and complete. The reverse is also true:

Theorem 11.4 (Sequence compact iff totally bounded and complete). *Let A be a subset of a metric space (X, d). Then A is sequence compact if and only if A is totally bounded and complete.*

Proof: We have already seen in prop.11.3 on p.207 and prop.11.4 on p.207 that if A is sequentially compact then A is totally bounded and complete. We now show the other direction.

Let A be totally bounded and complete and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in A. All we need to show is the existence of a subsequence $z_j = x_{n_j}$ which is Cauchy: As A is complete, such a Cauchy sequence must converge to a limit $L \in A$, i.e., $x_{n_j} \to L$ as $n \to \infty$. We now are going to extract a convergent subsequence $(x_{n_j})_j$ from $(x_n)_n$.

***a.** Because A is totally bounded, there will be a finite net for $\varepsilon = 1/2$: there exists*

$$\mathscr{G}_1 = \{g_{1,1}, g_{1,2}, \dots, g_{1,k_1}\} \subseteq A \text{ such that } A \subseteq U_1 := \bigcup \left[\, N_{1/2}(g_{1,j}) : j \leqq k_1 \,\right].$$

It follows that $x_k \in U_1$ for each k. There are infinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_1$. Hence there exists some $g_1 \in \mathscr{G}_1$ such that $B_1 := N_{1/2}(g_1)$ contains $x_{1,j} := x_{n_j}$ for an entire (infinite) subsequence n_j. [91]

***b.** Because A is totally bounded, there will be a finite net for $\varepsilon = 1/3$: there exists*

$$\mathscr{G}_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,k_2}\} \subseteq A \text{ such that } A \subseteq U_2 := \bigcup \left[\, N_{1/3}(g_{2,j}) : j \leqq k_2 \,\right].$$

It follows that $x_{1,k} \in U_2$ for each k. There are infinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_2$. Hence there exists some $g_2 \in \mathscr{G}_2$, such that $N_{1/3}(g_2)$ contains $x_{2,j} := x_{1,n_j}$ for an entire subsequence n_j. As the entire sequence $(x_{1,k})$ belongs to B_1, it follows that our new subsequence $(x_{2,j})$ of $(x_{1,k})$ belongs to $B_2 := B_1 \cap N_{1/3}(g_2)$.

***c.** Having constructed a subsequence $(x_{n-1,j})$ of the original sequence (x_k) which lives in a set N_{n-1} contained in $N_{1/n}(g_{n-1})$ for a suitable $g_{n-1} \in A$, total boundedness of A, guarantees the existence of a finite net for $\varepsilon = 1/(n+1)$: there exists*

$$\mathscr{G}_n = \{g_{n,1}, g_{n,2}, \dots, g_{n,k_n}\} \subseteq A \text{ such that } A \subseteq U_n := \bigcup \left[\, N_{1/(n+1)}(g_{n,j}) : j \leqq k_n \,\right].$$

It follows that $x_{n,k} \in U_n$ for each k. There are infinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_n$. Hence there must be at least one of those which we name g_n, such that $N_{1/(n+1)}(g_n)$ contains $x_{n,j} := x_{n-1,n_j}$ for an entire subsequence n_j. As the entire sequence $(x_{n-1,k})$ belongs to N_{n-1}, it follows that our new subsequence $(x_{n,j})$ of $(x_{n-1,k})$ belongs to $B_n := N_{n-1} \cap N_{1/(n+1)}(g_n)$.

We note that the maximal distance $d(x_{n,i}, x_{n,j})$ between any two members of that new subsequence is bounded by $2/(n+1)$, the diameter of $N_{1/(n+1)}$. It follows that

$$\text{diam}(B_n) \;\leqq\; 2/(n+1). \qquad (\star)$$

***d.** Diagonalization procedure: The following trick is employed quite frequently in real analysis. We create the "diagonal sequence" $z_1 := x_{1,1}, z_2 := x_{2,2}, \dots$ which is a subsequence of the original sequence (x_n). If we can show that it is Cauchy then the proof is complete.*

By construction, if $j \geqq n$ then

$$z_j \in B_j \subseteq B_n \subseteq N_{n-1} \subseteq \cdots \subseteq B_2 \subseteq B_1 \quad \text{and } \text{diam}(B_j) \leq \frac{2}{j+1} \quad \text{(see } (\star)) .$$

[91] Note that it is not claimed that there would be infinitely many different points x_{n_j}, only infinitely many indices n_j. Indeed, what would you do if the original Cauchy sequence was chosen to be $x_1 = x_2 = \cdots = a$ for some $a \in A$?

Let $\varepsilon > 0$. We can find $N \in \mathbb{N}$ such that $\frac{1}{N+1} < \frac{\varepsilon}{2}$. We remember from part c of this proof that

$$B_j = N_{j-1} \cap N_{1/(j+1)}(g_j) \subseteq N_{1/(j+1)}(g_j) \text{ for some suitable } g_j \in A,$$

hence all its points have distance from g_j bounded by $(j+1)^{-1}$. We obtain for any $i, j \geqq N$ that

$$d(z_i, z_j) \ \leq \ d(z_i, g_N) + d(g_N, z_j) \ \leqq \ \frac{1}{N+1} \ + \ \frac{1}{N+1} < \varepsilon.$$

It follows that $(z_n)_n$ is indeed Cauchy and the proof is completed. ■

Corollary 11.2 (Sequence compact sets are complete). *Let (X, d) be a metric space and let K be a sequence compact subset of X. Then K is complete.*

Proof: Immediate from the last theorem.

Theorem 11.5 (Sequence compact sets are closed and bounded). *Let A be sequence compact subset of a metric space (X, d). Then A is a bounded and closed set.*

***a**. Proof of boundedness:*

We may assume that A is not empty because otherwise there is nothing to prove. We assume that A is not bounded, i.e., diam$(A) = \infty$. It will be proved by induction that there exists a sequence $x_n \in A$ such that $d(x_i, x_j) \geqq 1$ for any $i \neq j$.

Base case: Let $x_0 \in A$. There exists $x_1 \in A$ such that $r_1 := d(x_0, x_1) \geqq 1$.

Induction step: We assume that n elements $x_1, \dots x_n$ such that $d(x_i, x_j) \geqq 1$ for any $1 \leqq i < j \leqq n$ have aready been chosen. Let $k := \max\{d(x_0, x_j) : j \leqq n\}$ and $r := k + 1$. As A is not bounded, we can pick $x_{n+1} \in A \setminus B_r(x_0)$. We obtain

$$k + 1 \ \leq d(x_{n+1}, x_0) \leq d(x_{n+1}, x_j) + d(x_j, x_0) \ \leq \ d(x_{n+1}, x_j) + k, \quad \text{i.e.,} \ \ 1 \ \leq \ d(x_{n+1}, x_j).$$

We have constructed a sequence x_n for which any two items have distance no less than 1. It follows that there is no Cauchy subsequence, hence no convergent subsequence and we have a contradiction to the sequence compactness of A.

***b**. Proof of closedness: If A is not closed then A has a contact point $x \in A^{\complement}$.*

As $N_{1/m}(x) \cap A \neq \emptyset$ we can pick a sequence $x_m \in A$ such that $d(x_m, x) < 1/m$ for all $m \in \mathbb{N}$.

Clearly x_m converges to $x \in A^{\complement}$. Sequence compactness of A allows us to extract a subsequence $z_j = x_{n_j}$ which converges to $z \in A$.

Both z and x are limit of z_j. According to thm.10.3 on p.159, $x = z$. It follows that both $x \in A^{\complement}$ and $x \in A$, a contradiction. This proves that sequence compact sets are closed. ■

Corollary 11.3 (Sequence compact sets are bounded). *Let (X, d) be a metric space and let K be a sequence compact subset of X. Then K is bounded, i.e., diam$(K) = \sup\{d(x, y) : x, y \in K\} < \infty$.*

Proof: Obvious from thm.11.5. ■

Remark 11.2. It follows from the results of this chapter and the introductory chapter on Closed and bounded sets in Euclidean space (11.1 on p.201) that, in $\mathbb{R}^N$, three of the definitions of compactness given in section 11.2 on Four definitions of compactness (p.203 are equivalent:

A subset of $\mathbb{R}^N$ is sequentially compact iff it is totally bounded and complete iff it is bounded and closed.

We will see later that any metric space is sequentially compact if and only if it is compact, i.e., covering compact (thm.11.11 on p.thm-x:compact-iff-seq-compact).

In other words, in $\mathbb{R}^{\mathbb{N}}$ all four of the definition given in section 11.2 on p.203 coincide.

11.5 Continuous functions and compact spaces

Theorem 11.6 (Closed subsets of compact spaces are compact)**.** *Let A be a closed subset of a compact metric space (X, d). Then $(A, d\big|_{A\times A})$ is a compact subspace.*

Proof:

Let $(U_j)_{j\in J}$ be a family of sets open in A whose union is A. According to prop.10.5 on p.164 there are open sets V_j in X such that $U_j = V_j \cap A$. It follows that $\bigcup\limits_{j\in J} V_j \supseteq A$, hence the family $(V_j)_{j\in J}$, augmented by the (open!) set $A^{\complement}$ is an open cover of (X, d). As X is compact we can extract finitely many members from that extended family such that they still cover X. If one of them happens to be $A^{\complement}$ then we remove it and we still obtain that the remaining ones, say, $V_{i_1}, V_{i_2}, \dots, V_{i_n}$, cover A. But then the traces in A (def.10.1 on p.163)

$$U_{i_1} = V_{i_1} \cap A,\ U_{i_2} = V_{i_2} \cap A,\ \dots,\ U_{i_n} = V_{i_n} \cap A \ = \ A$$

of those open sets form an open covering of the subspace A (see prop.10.5 on p.164). We have proved that the given open covering of A has a finite subcover of A. ■

Theorem 11.7 (Continuous images of compact spaces are compact)**.** *Let (X, d_1) and (Y, d_2) be two metric spaces. and let $f : X \to Y$ be continuous on X. If X is compact then the direct image $f(X)$ is compact, i.e., the metric subspace $f(X)$ of Y is compact.*

Proof: Let $(V_j)_{j\in J}$ be a family of sets open in Y whose union contains $B := f(X)$. Let the sets $W_j := V_j \cap f(X)$ be the traces of V_j in $f(X)$. Then the W_j are open in the subspace $(f(X), d_2)$ of Y and they form an open cover of $f(X)$. We note that any open cover of $f(X)$ is obtained in this manner from open sets in Y.

Let $U_j := f^{-1}(V_j)$.

$$\bigcup_{j\in J} U_j \ = \ \bigcup_{j\in J} f^{-1}(V_j) \ = \ f^{-1}\Big(\bigcup_{j\in J} V_j\Big) \ \supseteq \ f^{-1}(B) \ = \ f^{-1}\big(f((X)\big) \ = \ X. \tag{11.7}$$

In the above the second equation follows from prop. 6.2 (f^{-1} is compatible with all basic set ops) on p.102 and the last one follows from the fact that $f^{-1}\big(f((A)\big) \supseteq A$ for any subset of the domain of f (see cor. 6.1 on p. 106).

According to prop.10.10 ("f^{-1}(open) = open" continuity) on p.177, each U_j is open as the inverse image of the open set V_j under the continuous function f. It follows from (11.7) that $(U_j)_{j\in J}$ is an open covering of the compact space X. We can extract a finite subcover $U_{i_1}, U_{i_2}, \dots, U_{i_n}$.

It follows from the interchangeability of unions with direct images (see (6.18) on p.104) that

$$\begin{aligned} f(X) &= f(U_{j_1} \cup \dots \cup U_{j_n}) = f(U_{j_1}) \cup \dots \cup f(U_{j_n}) \\ &= f\big(f^{-1}(V_{j_1})\big) \cup \dots \cup f\big(f^{-1}(V_{j_n})\big) \subseteq V_{j_1} \cup \dots \cup V_{j_n}. \end{aligned}$$

The inclusion relation above follows from the fact that $f(f^{-1}(B)) = B \cap f(X)$ for any subset B of the codomain of f (see prop.6.6 on p. 106).

We have proved that the arbitrary open cover $(V_j)_{j\in J}$ of $f(X)$ contains a finite subcover $V_{j_1} \cup \dots \cup V_{j_n}$ and it follows that $f(X)$ is indeed a compact metric subspace of Y. ■

Read the following remark for an easier way to prove the above theorem.

Remark 11.3. We could have proved the last two theorems more easily using sequence compactness instead of covering compactness but the proofs that were given generalize to abstract topological spaces.

The missing ingredient: We did not define topological subspaces $(A, \mathfrak{U}_A)$ of an abstract topological space $(X, \mathfrak{U})$ This is done as follows: Let $\mathfrak{U}_A := \{U \cap A : U \in \mathfrak{U}\}$. It is easy to prove that $\mathfrak{U}_A$ satisfies the axioms for the open subsets of a topological space and the proofs above will go through with almost no alterations.

here is an alternate proof of theorem 11.6 which uses sequence compactness.

Given is a sequence $x_n \in A$. X is compact, hence sequence compact and it follows that there is $x \in X$ and a subsequence x_{n_j} such that x_{n_j} converges to x. It follows from theorem 10.6 on p.166 that $x \in \bar{A} = A$ and this proves that A is (sequence) compact. ■

and here is the outline of an alternate proof of theorem 11.7 which uses sequence compactness.

Given a sequence $y_n \in f(X)$ we construct a convergent subsequence y_{n_j} as follows: For each n there is some $x_n \in X$ such that $y_n = f(x_n)$ X is compact, hence sequence compact and it follows that there is $x \in X$ and a subsequence x_{n_j} such that x_{n_j} converges to x. We now use (sequence) continuity of f at x to conclude that $y_{n_j} = f(x_{n_j})$ converges to $f(x) \in f(X)$. ■

The following theorem relates compactness and uniform continuity. [92]

Theorem 11.8 (Uniform continuity on sequence compact spaces)**.** *Let (X, d_1), (Y, d_2) be metric spaces and let A be a sequence compact subset of X. Then any continuous real function on A is uniformly continuous on A.*

Proof: Let us assume that $f(\cdot)$ is continuous but not uniformly continuous and find a contradiction. Because $f(\cdot)$ is not uniformly continuous, there exists $\varepsilon > 0$ such that no $\delta > 0$, however small, will satisfy (10.51) for all pairs x, y such that $d_1(x,y) < \delta$. Looking specifically at $\delta := 1/j$ for all $j \in \mathbb{N}$, we can find

[92] See def.10.28 on p.182.

$x_j, x'_j \in A$ *such that*

(11.8) $$d_1(x_j, x'_j) < \frac{1}{j} \quad but \quad d_2(f(x), f(x')) \geqq \varepsilon.$$

Because A is sequence compact there is a subsequence (x_{j_k}) of the x_j which converges to an element $x \in A$ We have

(11.9) $$d_1(x'_{j_k}, x) \leqq d_1(x'_{j_k}, x_{j_k}) + d_1(x_{j_k}, x) \leqq \frac{1}{j_k} + d_1(x_{j_k}, x).$$

Both right hand terms converge to zero as $k \to \infty$. This is obvious for $1/j_k$ because $j_k \geqq k$ for all k and it is true for $d_1(x_{j_k}, x)$ because x_{j_k} converges to x.

It follows from (11.9) that (x'_{j_k}) also converges to x. It follows from the ordinary continuity of $f(\cdot)$ that

$$f(x) = \lim_{k\to\infty} f(x'_{j_k}) = \lim_{k\to\infty} f(x_{j_k}).$$

It follows from the "ordinary" (non-uniform) convergence of sequences that there exist $N, N' \in \mathbb{N}$ such that

$$d_2(f(x), f(x_{j_k})) < \frac{\varepsilon}{2} \text{ for } k > N; \quad d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} \text{ for } k > N'.$$

Both inequalities are true whenever $k > \max(N, N')$. It follows for all such k that

(11.10) $$d_2(f(x_{j_k}), f(x'_{j_k})) < d_2(f(x_{j_k}), f(x)) + d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and we have a contradiction to (11.8). ■

Corollary 11.4 (Uniform continuity on closed intervals)**.** *Let a, b be two real numbers such that $a \leqq b$. Any continuous real function on the closed interval $[a, b]$ is uniformly continuous on $[a, b]$:*

For any $\varepsilon > 0$, there exists $\delta > 0$ such that

(11.11) $$d(f(x) - f(y)) < \varepsilon \quad \text{for all } x, y \in [a, b] \text{ such that } d(f(x) - f(y)) < \delta$$

Proof: This follows from the previous theorem (11.8) because closed intervals $[a, b]$ are closed and bounded sets and, in $\mathbb{R}$, any closed and bounded set is sequence compact . ■

11.6 Open coverings and the Heine–Borel theorem

We now discuss families of open sets called "open coverings". You should review the concept of an indexed family and how it differs from that of a set (see (4.12) on p.85).

Definition 11.4 (Coverings and open coverings)**.** Let X be an arbitrary non–empty set and $A \subseteq X$. Let $(U_i)_{i\in I}$ be an indexed family of subsets of X such that $A \subseteq \bigcup_{i\in I} U_i$. Then we call $(U_i)_{i\in I}$ a **covering** of A.

A **finite subcovering** of a covering $(U_i)_{i\in I}$ of the set A is a finite collection

(11.12) $$U_{i_1}, U_{i_2}, U_{i_3}, \dots, U_{i_n} \quad (i_j \in I \quad \text{for } 1 \leqq j \leqq n) \qquad \text{such that} \quad A \subseteq U_{i_1} \cup U_{i_2} \cup \ldots \cup U_{i_n}.$$

Assume in addition that X is a topological space — this includes the case of a metric space (X, d). If all members U_i of the family $(U_i)_{i\in I}$ are open then we call this family an **open covering** of A.

Definition 11.5 (Compact sets)**.** Let $(X, \mathfrak{U})$ be a topological space and $K \subseteq X$. We say that K is **compact** if it has the "extract finite subcovering" property: Given any **open** covering $(U_i)_{i \in I}$ of K, one can extract a finite subcovering. In other words, there is a (possibly very large) $n \in \mathbb{N}$ and indices

$$i_1, i_2, \dots, i_n \in I \qquad \text{such that} \quad A \subseteq \bigcup_{j=1}^{n} U_{i_j}.$$

Remark 11.4. a. An open covering for the entire space X is a collection of open sets $(U_i)_{i \in I}$ such that $X = \bigcup [U_i : i \in I]$

b. Let (X, d) be a metric space. Then $K \subseteq (X, d)$ is compact if and only if the metric subspace $(K, d\big|_{K \times K})$ is compact, i.e., for any collection of subsets $(U_i)_{i \in I}$ of K which are open in K there exist finitely many indices $i_1, \dots, i_n \in I$ such that $K = U_{i_1} \cup \dots \cup U_{i_n}$. This is true because the open subsets of (K, d) are the traces in K of sets which are open in (X, d) (see def.10.1 on p.163).

Example 11.1. Here are some simple examples.

- **a.** Any finite topological space is compact.
- **b.** Any topological space that only contains finitely many open sets is compact. In particular a set with the indiscrete topology (example 10.6 on p.161) is compact
- **c.** A space with the discrete metric (example 10.4 on p.152) is compact if and only if it is finite.

And here is a counterexample.

> The open interval $]0, 1[$ with the Euclidean metric is not compact because it is not possible to extract an finite covering from the open covering $\left(]\frac{1}{n}, 1[\right)_{n \in \mathbb{N}}$.

Example 11.2. Here are the corresponding results for sequence compact metric spaces.

- **a.** Any finite metric space is sequence compact.
- **b.** Any metric space that only contains finitely many open sets is compact. [93]
- **c.** A space with the discrete metric is compact if and only if it is finite.

The counterexample also fits in:

> The open interval $]0, 1[$ with the Euclidean metric is not sequence compact because it is not possible to extract a convergent subsequence from the sequence $x_n := 1/n$ (the limit zero does not belong to $]0, 1[$).

We will now see that the correspondence in the above two examples is not a coincidence. We will see that (subspaces of) metric spaces are compact if and only if they are sequentially compact. We will prove each direction separately.

Theorem 11.9 (Compact metric spaces are sequence compact)**.** *Let (X, d) be a compact metric space. Then X is sequence compact.*

[93] We had to remove the example of the indiscrete topology because this topology does not come from a metric.

Let X be compact and let $(x_n)_n$ be a sequence in X from which one cannot extract a convergent subsequence. We will prove that this leads to a contradiction.

Let $F := \{x \in X : x = x_j$ for some $j \in \mathbb{N}\}$ [94] *be the set of distinct members of $(x_n)_n$. Let $z \in X$. We first prove that there exists an open neighborhood U_z of z such that $U_z \cap F$ is finite.*

This is true because otherwise for each for each $m \in \mathbb{N}$ there exists some index j_m such that $x_{j_m} \neq z$ and $x_{j_m} \in N_{1/m}(z)$. We have constructed a sequence $(x_{j_m})_m$ which converges to z, i.e., $(x_n)_n$ possesses a convergent subsequence, contrary to our assumption.

It follows from $\{z\} \subseteq U_z$ that $(U_z)_{z \in X}$ is an open covering of X. X is compact and we can extract a finite subcovering $U_1, U_2, \dots, U_k$. Each U_j contains at most finitely many distinct member of the sequence $(x_n)_n$ and we conclude that this entire sequence consists of only finitely many distinct members.

But then at least of those members, say $x_{k^\star}$, will appear infinitely often in that sequence: there is $k_1 < k_2 < \dots$ such that $x_{k_1} = x_{k_2} = \dots = x_{k^\star}$. This subsequence converges (to $x_{k^\star}$). We have reached a contradiction. ■

Lemma 11.1. *Let (X, d) be a metric space. Let $x, y \in X$ and $\varepsilon > 0$ such that $y \in N_\varepsilon(x)$.*

If $\delta > 0$ Then $N_\delta(y) \subseteq N_{\delta+\varepsilon}(x)$.

Proof: Let $z \in N_\delta(y)$. Then

$$d(z, x) \leqq d(z, y) + d(y, x) < \delta + \varepsilon.$$

In other words, each element z of $N_\delta(y)$ is $\delta + \varepsilon$–close to x. But then $N_\delta(y) \subseteq N_{\delta+\varepsilon}(x)$. ■

Proposition 11.5. *Let (X, d) be a sequence compact metric space. Let $(U_i)_{i \in I}$ be an open covering of X. Then there exists $\rho > 0$ as follows: For each $x \in X$ there exists $i \in I$ such that $N_\rho(x) \subseteq U_i$.* [95]

Proof: Assume to the contrary that no such $\rho > 0$ exists. We then can find for any $n \in \mathbb{N}$ some $x_n \in X$ such that $N_{1/n}(x_n)$ is not contained in any of the U_i. X is sequence compact — there exists $x \in X$ and a subsequence $(x_{n_j})_j$ which converges to x. $(U_i)_{i \in I}$ covers X, so there exists $i_0 \in I$ such that $x \in U_{i_0}$.

$(\star)$ Because U_{i_0} is an open neighborhood of x there exists $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq U_{i_0}$.

$(\star\star)$ Because $(x_{n_j})_j$ converges to x there are infinitely many $j \in \mathbb{N}$ such that $d(x_{n_j}, x) < \varepsilon/2$, hence we may assume that $j > 2/\varepsilon$. It follows from $n_j \geqq j$ that $n_j > 2/\varepsilon$, i.e. $1/n_j < \varepsilon/2$.

$(\star\star\star)$ It follows from $d(x_{n_j}, x) < \varepsilon/2$ and lemma 11.1 on p.214 that $N_{\varepsilon/2}(x_{n_j}) \subseteq N_\varepsilon(x)$.

We apply first $\star\star$, then $\star\star\star$, then $\star$ to obtain

$$N_{1/n_j}(x_{n_j}) \subseteq N_{\varepsilon/2}(x_{n_j}) \subseteq N_\varepsilon(x) \subseteq U_{i_0}.$$

[94] We could have written more concisely $F := \{x_j : j \in \mathbb{N}\}$ but the above definition was chosen to remind you that F does not contain any duplicates.

[95] The number $\lambda = 2\rho$ is called a **Lebesgue number** of $(U_i)_{i \in I}$. In other words, the Lebesgue number is the diameter of the ρ–neighborhoods. Note that if λ is a Lebesgue number of an open covering then any λ' which satisfies $0 < \lambda' < \lambda$ also is a Lebesgue number.

But this contradicts our assumption that each x_{n_j} was chosen in such a fashion that $N_{1/n}(x_n)$ is not contained in any of the U_i. ■

We now can prove the converse of thm.11.9.

Theorem 11.10. *Sequence compact metric spaces are compact.*

Proof: Let (X, d) be a sequence compact metric space and let $(U_i)_{i \in I}$ be an open covering of X. According to prop.11.5 there exists $\rho > 0$ as follows: For each $x \in X$ there exists $i(x) \in I$ such that $N_\rho(x) \subseteq U_{i(x)}$.

Since X is totally bounded (see thm.11.4 on p.207) there exist finitely many $x_1, \dots, x_k \in X$ such that $\{N_\rho(x_j) : j = 1, ..k\}$ forms an open covering of X. It then follows from $N_\rho(x_j) \subseteq U_{i(x_j)}$ that $U_{i(x_1)}, U_{i(x_2)}, \dots, U_{i(x_k)}$ also forms an open covering of X. We have extracted a finite subcovering from $(U_i)_{i \in I}$. ■

Theorem 11.11 (Sequence compact is same as compact in metric spaces)**.** *Let (X, d) be a metric space and let A be a subset of X. Then A is sequence compact if and only if A is compact, i.e., every open covering of A has a finite subcovering.*

Proof: Theorems 11.9 and 11.10. ■

An easy consequence is the Heine–Borel theorem.

Theorem 11.12 (Heine–Borel)**.** *Let (X, d) be $\mathbb{R}^N$ with the Euclidean norm and its associated metric. A subset $K \subseteq \mathbb{R}^N$ is compact if and only if it is closed and bounded.*

For a general metric space it is still true that any compact subset is closed and bounded.

Proof: We have seen in thm.11.2 on p.202 that closed and bounded subsets of $\mathbb{R}^N$ are sequence compact. We have further seen that sequence compact sets are closed and bounded (thm.11.5 on p.209). Our assertion now follows from the equivalence of compactness and sequence compactness in metric spaces (thm.11.11 on p.215).
■

12 Applications of Zorn's Lemma

12.1 More on Partially Ordered Sets (Study this!)

Some of this was copied almost literally from [6] Dudley.

Definition 12.1 (Linear orderings)**.** Given is a non-empty set X and a partial ordering $\preceq$ on X (see def.4.4 on p.70). $\preceq$ is a **linear ordering** , also called a **total ordering** of X if and only if, for all x and $y \in X$ such that $x \neq y$, either $x \preceq y$ or $y \preceq x$. Then $(X, \preceq)$ is called a **linearly ordered set** or a **totally ordered set** set. The classic example of a linearly ordered set is the real line $\mathbb{R}$, with its usual ordering. Actually, $(\mathbb{R}, \leq)$ and $(\mathbb{R}, \geq)$ are all linearly ordered sets.

Definition 12.2. A **chain** is a linearly ordered subset of a partially ordered set (with the same ordering). In a partially ordered set $(X, \preceq)$, an element m of X is called **maximal** iff there is no $x \neq m$ with $m \preceq x$. A **maximum** of X is an $m \in X$ such that $x \preceq m$ for all $x \in X$.

Note 12.1 (Notes on maximal elements and maxima)**.**

- **A.** If X is not linearly ordered, it may have many maximal elements. For example, for the trivial partial ordering $x \preceq y$ if and only if $x = y$, every element is maximal. A maximum is a maximal element, but the converse is often not true.
- **B.** If an ordering is not specified, then we always mean set inclusion.
- **C.** If $m \in X$ is a maximum of X then this implies that m must be related to all other elements of X.

Example 12.1 (Maximal elements and maxima)**.** Let X be the collection of all intervals $[a, b] \in \mathbb{R}$ of length $b - a \leq 2$ such that $a \leq b$. These intervals are partially ordered by inclusion. Any interval of length equal to 2 is a maximal element. There is no maximum.

Axiom 12.1 (Zorn's Lemma)**.** A hundred years ago the following was seen as extremely controversial by mathematicians who specialize in the foundations of mathematics.

Zorn's Lemma: Let $(X, \preceq)$ be a partially ordered set with the **ZL property**:

$$\text{Every chain } C \subseteq X\text{, possesses an upper bound } u \in X\text{, i.e., } x \preceq u \text{ for all } x \in C. \qquad \textbf{(ZL)}$$

Then X has a maximal element.

Zorn's lemma is an axiom rather than a theorem or a proposition in the following sense: It is impossible to verify its truth or falsehood from the axioms of "a" (meaning there are more than one) "reasonable" axiomatic set theory. In that sense mathematicians are free to accept or reject this statement when building their mathematical theories. Two notes on that remark:

a. Today the mathematicians who refuse to accept proofs which make use, directly or indirectly, of Zorn's lemma, are a very small minority.

b. It can be proved that if one accepts (rejects) Zorn's lemma as a mathematical tool then this is equivalent to accepting (rejecting) the **Axiom of Choice** [96] which is so abstract in its precise formulation that it was relegated to a footnote.

[96] Remember that if Ω is any set then 2^Ω is the set of all subsets of Ω . **Axiom of Choice**: Let Ω be any non-empty set. Then there is a function $f : 2^\Omega \setminus \emptyset \to \Omega$ such that $f(A) \in A$ for each $A \in 2^\Omega \setminus \emptyset$, i.e., for each non-empty $A \subseteq \Omega$.

Remark 12.1. We will see in a later chapter how Zorn's lemma allows a surprisingly simple proof to the effect that **any** vector space has a basis.

12.2 Existence of bases in vector spaces (Study this!)

The following is thematically a continuation of the material in chapter 9 (Vectors and vector spaces).

We now focus on proving that every vector space, ***even if it does not possess a finite subset which spans the entire space****, possesses a basis.*

For the remainder of this chapter we assume that V is a vector space and that $\mathfrak{B}$ denotes the set

$$\mathfrak{B} := \{A \subseteq V : A \text{ is linearly independent } \}. \tag{12.1}$$

$\mathfrak{B}$ is a partially ordered set with respect to set inclusion. The next lemma allows us to apply Zorn's lemma.

Lemma 12.1. *Every chain* [97] *$\mathfrak{C}$ in $(\mathfrak{B}, \subseteq)$ possesses an upper bound.*

Proof:
Let $U := \bigcup \left[C : C \in \mathfrak{C}\right]$. We will show that $U \in \mathfrak{B}$. As $U \supseteq C$ for all $C \in \mathfrak{C}$ it then follows that U is an upper bound of $\mathfrak{C}$ and the proof is finished.

Let $x_1, x_2, \dots x_k \in U$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$) such that

$$\sum_{j=1}^{k} \alpha_j x_j \;=\; 0. \tag{12.2}$$

We must show that each α_j is zero. For each $0 \leqq j \leqq k$ there is some $C_j \in \mathfrak{C}$ such that $x_j \in C_j$. $\mathfrak{C}$ is totally ordered, hence $C_i \subseteq C_j$ or $C_j \subseteq C_i$ for any two indices $0 \leqq i, j \leqq k$. But then there exists an index j_0 such that $C_{j_0} \supseteq C_j$ for all j, hence $x_1, x_2, \dots x_k \in C_{j_0}$. But C_{j_0} is linearly independent because $C_{j_0} \in \mathfrak{C} \subseteq \mathfrak{A}$. It follows that $\alpha_1 = \dots = \alpha_k = 0$. ■

Theorem 12.1. *Every vector space V has a basis.*

Proof: It follows from lemma 12.1 and Zorns Lemma (axiom 12.1 on p. 216) that the set $\mathfrak{B}$ of all independent subsets of the vector space V contains a maximal element (subset of V) which we denote by B. As membership in $\mathfrak{B}$ guarantees its linear independence we only need to prove that $span(B) = V$.

Let us assume to the contrary that there exists $y \in span(B)^{\complement}$. It follows from lemma 9.2 on p.144 that the set $B' := B \cup \{y\}$ is linearly independent, hence $B' \in \mathfrak{B}$. Clearly, $B \subsetneq B'$. This contradicts the maximality of B in the partially ordered set $(\mathfrak{B}, \subseteq)$. ■

We now turn our attention to extending a linear real function f from a subspace $F \subseteq V$ to the entire vector space V.

[97] see def.12.2 on p.216.

Lemma 12.2. *Let V be a vector space and let F be a (linear) subspace of V. Let $f : F \to \mathbb{R}$ be linear.*

Let $\mathscr{G} := \{(W, f_W) : W$ is a subspace of $V, W \supseteq F,\ f_W : W \to \mathbb{R}$ is a linear extension of f to $W\}$.

Then the following defines a partial ordering on $\mathscr{G}$: $(U, f_U) \preceq (W, f_W) \Leftrightarrow V \subseteq W$ and $f_W\big|_U = f_U$.

Moreover this ordering satisfies the requirements of Zorn's Lemma: Every chain in $(\mathscr{G}, \preceq)$ possesses an upper bound (in $\mathscr{G}$).

Proof: Reflexivity and transitivity of "$\preceq$" are trivial. The latter is true because the extension of an extension is again an extension.

Antisymmetry: If both $(U, f_U) \preceq (W, f_W)$ and $(W, f_W) \preceq (U, f_U)$ then both $U \subseteq W$ and $W \subseteq U$, hence $U = W$. But then f_W is an extension of f_U to itself, i.e., $f_U = f_W$. It follows that $\preceq$ is indeed a partial order on $\mathscr{G}$.

Now let $\mathscr{C}$ be a chain in $\mathscr{G}$. We must find an upper bound for $\mathscr{C}$. Let $W := \bigcup \big[U : (U, f_U) \in \mathscr{C}\big]$.

We show that W is a subspace of E: If $x, y \in W$ and $\lambda \in \mathbb{R}$ then there are $(C_1, f_1), (C_2, f_2) \in \mathscr{C}$ such that $x \in C_1$ and $y \in C_2$. Because $\mathscr{C}$ is a chain we have $C_1 \subseteq C_2$ or $C_2 \subseteq C_1$, say, $C_1 \subseteq C_2$. It follows that $x, y \in C_2$. But C_2 is a subspace of V and we conclude that $x + \lambda y \in C_2$, hence $x + \lambda y \in W$. It follows that W is a subspace of V.

Let $f_W : W \to \mathbb{R}$ be defined as follows: If $x \in W$ then there is some $(C, f_C) \in \mathscr{C}$ such that $x \in C$. We define $f_W(x) := f_C(x)$. This definition is unambiguous even if x belongs to (possibly infinitely) many elements of $\mathscr{C}$. To see this let $(C, f_C), (D, f_D) \in \mathscr{C}$ such that $x \in C$ and $x \in D$. Then $C \subseteq D$ or $D \subseteq C$. We may assume that $C \subseteq D$. But as $f_D\big|_C = f_C$ we conclude that $f_C(x) = f_D(x)$, i.e., the definition of $f_W(\cdot)$ is unambiguous. The above specifically holds for $x \in F$ and we note that f_W is an extension of f.

Next we show linearity of f_W. Let $x, y \in W$ and $\alpha \in \mathbb{R}$. Then there are $(C, f_C), (D, f_D) \in \mathscr{C}$ such that $x \in C$ and $y \in D$. Again we may assume that $C \subseteq D$. It now follows from the linearity of f_D that

$$f_W((x + \alpha y) = f_D((x + \alpha y) = f_D((x) + \alpha f_D(y) = f_W((x) + \alpha f_W(y).$$

and we have proved that f_W is linear (on all of W.

To summarize, W is a subspace of V and f_W is a linear extension of f to W. But then $(W, f_W) \in \mathscr{G}$ and $(W, f_W) \succeq (C, f_C)$ for all $(C, f_C) \in \mathscr{C}$. It follows that (W, f_W) is an upper bound of $\mathscr{C}$. ■

Lemma 12.3. *Let V be a vector space and let F be a (linear) subspace of V such that $F \subsetneqq V$. Let $f : F \to \mathbb{R}$ be linear. Let $a \in F^{\complement}$ and $\gamma \in \mathbb{R}$.*

Let $\tilde{F} := span(F \uplus \{a\})$. Then a and γ define a linear extension $\tilde{f}$ of f to $\tilde{F}$ as follows:

$$\tilde{f}(x + \alpha a) := f(x) + \alpha\gamma$$

Proof: Do this as an exercise. ■

Theorem 12.2 (Extension theorem for linear real functions)**.** *Let V be a vector space and let F be a (linear) subspace of V. Let $f : F \to \mathbb{R}$ be a linear mapping.*

Then there is an extension of f to a linear mapping $\tilde{f} : V \to \mathbb{R}$.

Proof:

Let $\mathscr{G} := \{(W, f_W) : W$ *is a subspace of* $V, W \supseteq F,\ f_W : W \to \mathbb{R}$ *is a linear extension of* f *to* $W\}$

and let $(U, f_U) \preceq (W, f_W) \Leftrightarrow V \subseteq W$ *and* $f_W\big|_U = f_U$.

We have seen in lemma 12.2 that $\preceq$ *is a partial ordering on* $\mathscr{G}$ *such that any chain in* $(\mathscr{G}, \preceq)$ *possesses an upper bound. We apply Zorn's Lemma (axiom 12.1 on p.216) to conclude that* $\mathscr{G}$ *possesses a maximal element* (F', f').

We now show that $F' = V$.

If this was not true then we could find $a \in V \setminus F'$ *and, according to lemma 12.3 extend* f' *to a linear function* $\tilde{f}$ *on* $F' \uplus \{a\}$. *It follows that* $(F' \uplus \{a\}, \tilde{f}) \in \mathscr{G}$ *and* $(F', f') \precneqq (F' \uplus \{a\}, \tilde{f})$.

This contradicts the maximality of (F', f'). *and we have reached a contradiction.* ■

12.3 The Hahn-Banach extension theorem [98] ($\star$)

Let V *be a vector space and let* F *be a (linear) subspace of* V. *Let* $f : F \to \mathbb{R}$ *be linear mapping. This chapter examines the ability to extend* f *from its domain* F *to the entire space* V *subject to some condition which guarantees that if* V *is a normed space (hence a metric space) and if* f *is continuous on* F *then this linear extension will be continuous on all of* V.

In preparation for this we must study sublinearity, a generalization of linearity and norms.

12.3.1 Sublinear functionals

Definition 12.3 (Sublinear functionals)**.** Let V be a vector space and $p : V \to \mathbb{R}$ such that

a.	if $\lambda \in \mathbb{R}_{\geqq 0}$ and $x \in V$ then $p(\lambda x) = \lambda p(x)$	(positive homogeneity) ;
a.	if $x, y \in V$ then $p(x + y) = p(x) + p(y)$	(subadditivity) .

Then we call p a **sublinear functional** on V.

Proposition 12.1. *Let* V *be a vector space and* $p : V \to \mathbb{R}$ *sublinear. Let* $x \in V$. *Then*
a. $p(0) = 0$,
b. $-p(x) \leqq p(-x)$,

Proof of ***a****:* $p(0) = p(0 \cdot 0) = 0 \cdot p(0) = 0$.

Proof of ***b****: This follows from* $0 = p(0) = p(x + (-x)) \leqq p(x) + p(-x)$. ■

Example 12.2 (Norms are sublinear)**.** Let $(V, \|,\cdot\|)$ be a normed vector space. Then the function $p(x) := \|x\|$ is sublinear.

Indeed, norms are homogenous: We have $\|\lambda x\| = |\lambda| \cdot \|x\|$ not only for $\lambda \geqq 0$ but for all $\lambda \in \mathbb{R}$.

Further subadditivity is just the validity of the triangle inequality.

[98] **This chapter is optional**. The proof given here is a more detailed version of the one found in [5] Choquet.

Example 12.3 (Linear functions are sublinear)**.** Let V be a vector space and let $f := V \to \mathbb{R}$ be a linear function. Then f is sublinear.

Indeed, linear functions f satisfy $f(\lambda x) = \lambda \cdot f(x)$ not only for $\lambda \geqq 0$ but for all $\lambda \in \mathbb{R}$.

Further linear functions satisfy additivity: $f(x+y) = f(x) + f(y)$,
hence also subadditivity $f(x+y) \leqq f(x) + f(y)$.

More about sublinearity can be found in chapter 12.4 on p.223

12.3.2 The Hahn-Banach extension theorem and its proof

The subject of the Hahn-Banach extension theorem is extension of a linear function from a subspace to the entire vector space. The following remark is about first extending it to "one more dimension".

Remark 12.2. Let V be a vector space, let F be a linear subspace of V and let $f := F \to \mathbb{R}$ be a linear function. Let $a \in V \setminus F$. Then any linear extension $\tilde{f}$ of f to $span(F \uplus \{a\})$ is uniquely determined by its value $k := \tilde{f}(a)$.

Indeed, any $x \in span(F \uplus \{a\})$ can be written as $u + \lambda a$ for some $u \in F$ and $\lambda \in \mathbb{R}$. It follows from the linearity of $\tilde{f}$ that

$$\tilde{f}(x + \lambda a) = \tilde{f}(x) + \lambda \tilde{f}(a) = f(x) + \lambda k. \tag{12.3}$$

Theorem 12.3 (Hahn–Banach extension theorem)**.** *Let V be a vector space and $p : V \to \mathbb{R}$ a sublinear mapping. Suppose F is a (linear) subspace of V and $f : F \to \mathbb{R}$ is a linear mapping with $f \leqq p$ on F. Then there is an extension of f to a linear map $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f} \leqq p$ on V.*

Before proving this theorem, first we prove two lemmata.

Lemma 12.4. *Suppose F is a subspace of V, $f : F \to \mathbb{R}$ is a linear mapping and $a \in V \setminus F$. Let $k \in \mathbb{R}$ and*

$$\tilde{f}(x + \lambda a) := f(x) + \lambda k, \quad i.e., \quad k = \tilde{f}(a) \tag{12.4}$$

(see (12.3)). Then

$$k \leqq \inf_{u \in F}\{p(u+a) - f(u)\} \quad \Leftrightarrow \quad \tilde{f}(x + \lambda a) \leqq p(x + \lambda a) \text{ for all } \lambda > 0 \text{ and } x \in F, \tag{12.5}$$

$$k \geqq \sup_{v \in F}\{f(v) - p(v-a)\} \quad \Leftrightarrow \quad \tilde{f}(x + \lambda a) \leqq p(x + \lambda a) \text{ for all } \lambda < 0 \text{ and } x \in F. \tag{12.6}$$

***Proof of* (12.5), $\Rightarrow$):** *Let us assume that $\lambda > 0$. Then, on account of the left side of (12.5),*

$$\tilde{f}(x + \lambda a) = f(x) + \lambda k = \lambda\big(f(x/\lambda) + k\big) \leqq \lambda\Big(f(x/\lambda) + \big(p(x/\lambda + a) - f(x/\lambda)\big)\Big) = \lambda p(x/\lambda + a)$$

We use the positive homogeneity of p: $\lambda p(x/\lambda + a) = p(x + \lambda a)$ to obtain $\tilde{f}(x + \lambda a) \leqq p(x + \lambda a)$.

***Proof of* (12.6), $\Rightarrow$):** *Let us assume that $\lambda < 0$. Because of the left side of (12.6) and $\lambda < 0$ and positive homogeneity of p,*

$$\begin{aligned} k \geqq f(v) - p(v-a) &\Rightarrow \lambda k \leqq f(\lambda v) - \lambda p(v-a) \\ &\Rightarrow -f(\lambda v) + \lambda k \leqq (-\lambda)p(v-a) = p\big((-\lambda)(v-a)\big) = p((-\lambda)v + \lambda a). \end{aligned}$$

We substitute $v := x/\lambda \in F$:

$$-f(x) + \lambda k \leqq p(-x + \lambda a), \quad \textit{hence} \quad \tilde{f}(-x + \lambda a) = f(-x) + \lambda k \leqq p(-x + \lambda a)$$

We can switch from $-x$ to x as the above holds for all x in the subspace F and because $-x \in F$ if and only if $x \in F$. It follows that p indeed dominates $\tilde{f}$ for all $x \in F$ and $\lambda < 0$.

***Proof of* (12.5), $\Leftarrow$):** *we assume $\tilde{f}(x + \lambda a) \leqq p(x + \lambda a)$ for all $\lambda > 0$ and $x \in F$. We now show that $k = \tilde{f}(a) \leqq p(u + a) - f(u)$ for all $u \in F$.*

$$p(u + a) - f(u) \geqq \tilde{f}(u + a) - f(u) = \tilde{f}(u) + \tilde{f}(a) - f(u) = f(u) + \tilde{f}(a) - f(u) = \tilde{f}(a) = k.$$

***Proof of* (12.6), $\Leftarrow$):** *we assume $\tilde{f}(x + \lambda a) \leqq p(x + \lambda a)$ for all $\lambda < 0$ and $x \in F$. We now show that $k = \tilde{f}(a) \geqq f(v) - p(v - a)$ for all $v \in F$.*

$$-p(v - a) + f(v) \leqq -\tilde{f}(v - a) + f(v) = \tilde{f}(a - v) + f(v) = \tilde{f}(a) - \tilde{f}(v) + f(v) = \tilde{f}(a) = k.$$

■

Lemma 12.5. *Let $F \subset V$ be a genuine subspace of V and $a \in V \setminus F$. Let $G := span(F \uplus \{a\})$ be the subspace of all linear combinations that can be created by a and or vectors in F. Then*

a. *there exists a linear extension $\tilde{f}$ of f to G.*

b. *This extension is unique if and only if $\sup_{v \in E}\{f(v) - p(v - a)\} = \inf_{u \in E}\{p(u + a) - f(u)\}$.*

Proof of a. *For $u, v \in F$ we have*

$$f(u) + f(v) = f(u + v) \leqq p(u + v) = p\big((u + a) + (v - a)\big) \leqq p(u + a) + p(v - a)$$

and hence $f(v) - p(v - a) \leqq p(u + a) - f(u)$. Therefore

$$\sup_{v \in F}\{f(v) - p(v - a)\} \leqq \inf_{u \in F}\{p(u + a) - f(u)\}.$$

Now for a fixed $k \in \mathbb{R}$, we define $\tilde{f}(x + \lambda a) = f(x) + \lambda k$. We claim that $\tilde{f} \leqq p$ if and only if we have

$$\sup_{v \in F}\{f(v) - p(v - a)\} \leqq k \leqq \inf_{u \in F}\{p(u + a) - f(u)\} \tag{12.7}$$

*which will conclude the proof of **a** since such a k exists by the above work. Our claim holds because $f(x) + \lambda k = \tilde{f}(x + \lambda a) \leqq p(x + \lambda a)$ for all λ if and only if*

$$\begin{aligned} k &\leqq p(u + a) - f(u) \quad \textit{for all } u \in F \\ \textit{and} \quad k &\geqq f(v) - p(v - a) \quad \textit{for all } v \in F \end{aligned}$$

(the cases $\lambda > 0$ and $\lambda < 0$ respectively). This is proved above in lemma 12.4.

Proof of b. *From (12.7) we deduce that k is unique if and only if $\sup_{v \in E}\{f(v) - p(v - a)\} = \inf_{u \in E}\{p(u + a) - f(u)\}$. This exactly the case where $\tilde{f}$ which was defined by $\tilde{f}(x + \lambda a) = f(x) + \lambda k$ is unique.* ■

Proof of thm.12.3 (Hahn-Banach):

Let $\mathscr{G} = \{(V, g) : V$ *is a subspace of* $E,\ g : V \to \mathbb{R}$ *is linear and* $g \leqq p$ *on* $V\}$.

We define a partial order "$\preceq$*" on* $\mathscr{G}$ *as follows:*

$$(V_1, g_1) \preceq (V_2, g_2) \ \Leftrightarrow\ V_1 \subseteq V_2 \text{ and } g_2 \text{ is an extension of } g_1. \tag{12.8}$$

a. *We first prove that any chain* $\mathscr{C} \subseteq (\mathscr{G}, \preceq)$ *has an upper bound: Let* $W := \bigcup\big[V : V \in \mathscr{C}\big]$. *Then* W *is a subspace of* E *because if* $x, y \in W$ *and* $\lambda \in \mathbb{R}$ *then there are* $(V_1, g_1), (V_2, g_2) \in \mathscr{C}$ *such that* $x \in V_1$ *and* $y \in V_2$. *Because* $\mathscr{C}$ *is a chain we have* $V_1 \subseteq V_2$ *or* $V_2 \subseteq V_1$*, say,* $V_1 \subseteq V_2$. *It follows that* $x, y \in V_2$. *But* V_2 *is a subspace of* E *and we conclude that* $x + \lambda y \in V_2$*, hence* $x + \lambda y \in W$. *It follows that* W *is a subspace of* E.

We now find a linear $h : W \to \mathbb{R}$ *such that* $h \leqq p$ *on* W *and* $h\big|_V = g$ *for all* $(V, g) \in \mathscr{C}$*, i.e.,* h *is a linear extension of* g *for all* $(V, g) \in \mathscr{C}$. *If we find such* h *then it follows that* $(W, h) \in \mathscr{G}$ *and* (W, h) *is an upper bound of* $\mathscr{C}$. *Let* $x \in W$. *Then* $x \in V_1$ *for some* $(V_1, g_1) \in \mathscr{C}$. *We define* $h(x) := g_1(x)$. *This assignment is unambiguous because if* $x \in V_2$ *for some other* $(V_2, g_2) \in \mathscr{C}$ *then one of them, say* V_1*, is contained in the other and* g_2 *is an extension of* g_1*, i.e.,* $h(x) = g_1(x) = g_2(x)$. *As* $(V_1, g_1) \in \mathscr{G}$ *we conclude that* $h(x) = g_1(x) \leqq p(x)$*, i.e.,* $h \leqq p$ *on* W.

Next we show that h *is linear. Let* $x, y \in W$ *and* $\lambda \in \mathbb{R}$. *We repeat the argument from the proof that* W *is a subspace of* V *to conclude that both* x, y *belong to some* $(V, g) \in \mathscr{C}$. *We obtain*

$$h(x + \lambda y) = g(x + \lambda y) = g(x) + \lambda g(y) = h(x) + \lambda h(y)$$

This completes the proof that $(W, h) \in \mathscr{G}$. *Let* $(V, g) \in \mathscr{C}$. *Clearly,* $V \subseteq W = \bigcup\big[U : U \in \mathscr{C}\big]$. *Further* h *is linear, dominated by* p *and is constructed in such a manner that* $h(x) = g(x)$ *for all* $x \in V$. *It follows that* $(W, h) \succeq (V, g)$ *for all* $(V, g) \in \mathscr{C}$ *and we have proved that* $\mathscr{C}$ *has an upper bound in* $(\mathscr{G}, \preceq)$.

b. *We are now in a position to apply Zorn's Lemma (axiom 12.1 on p.216) to conclude that* $\mathscr{G}$ *possesses a maximal element* (F', f').

We now show that $F' = E$. *If this was not true then we could find* $a \in E \setminus F'$ *and, according to lemma 12.5 on p.221, extend* f' *to a linear function* $\tilde{f}$ *on* $F' \uplus \{a\}$ *in such a fashion that* $\tilde{f} \leqq p$. *It follows that* $(F' \uplus \{a\}, \tilde{f}) \in \mathscr{G}$ *and* $(F', f') \precneqq (F' \uplus \{a\}, \tilde{f})$. *This contradicts the maximality of* (F', f') *and we have reached a contradiction.* ■

Corollary 12.1 (Continuous extensions of continuous linear functions)**.** *Let* $(V, \|\cdot\|)$ *be a normed vector space. Let* F *be a (linear) subspace of* V *and let* $f : F \to \mathbb{R}$ *be a continuous linear mapping on* F. *Then there is an extension of* f *to a continuous linear map* $\tilde{f} : V \to \mathbb{R}$.

Proof: Let $p(x) := \|f\| \cdot \|x\|$ *(see def.10.29 on p.184). Because* p *is a positive multiple of a norm it also is a norm on* V *(see prop.9.10) on p.148), hence sublinear by example 12.2 on p.219. According to the Hahn-Banach extension theorem there exists a linear extension* $\tilde{f}$ *of* f *to all of* V *such that*

$$\tilde{f}(x) \leqq p(x) \text{ for all } x \in V. \tag{12.9}$$

We replace x *with* $-x$ *and obtain from the linearity of* $\tilde{f}$ *that* $-\tilde{f}(x) = \tilde{f}(-x) \leqq p(-x)$.
We note that $p(x) = p(-x)$ *because* p *is a norm. Hence*

$$-\tilde{f}(x) \leqq p(-x) = p(x), \text{ i.e., } \tilde{f}(x) \geqq -p(x).$$

Together with (12.9) this shows that

$$-p(x) \leqq \tilde{f}(x) \leqq p(x) \text{ for all } x \in V$$

and thus

$$|\tilde{f}(x)| \;\leqq\; p(x) \;=\; \|f\| \cdot \|x\| \text{ for all } x \in V \tag{12.10}$$

It follows from (12.10) that f has been extended in such a way that $\|\tilde{f}\| \leqq \|f\|$. We aply the continuity criterion for linear functions (thm.10.14 on p.184) twice in a row to finish the proof as follows: It follows from the continuity of f that $\|f\| < \infty$. But then $\|\tilde{f}\| < \infty$ and this proves the continuity of $\tilde{f}$. ■

12.4 Convexity (⋆)

Definition 12.4 (Concave-up and convex functions)**.** Let $f : \mathbb{R} \to \mathbb{R}$.

a. The **epigraph** of f is the set $epi(f) := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geqq f(x_1)\}$ of all points in the plane that lie above the graph of f.

b. f is **convex** if for any two vectors $\vec{a}, \vec{b} \in epi(f)$ the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leqq \lambda \leqq 1$ is contained in $epi(f)$.

c. Let f be differentiable at all points $x \in \mathbb{R}$. Then f is **concave-up**, if the function $f' : x \mapsto f'(x)$ is non-decreasing.

Proposition 12.2 (Convexity criterion)**.** *f is convex if and only if the following is true: For any $a \leqq x_0 \leqq b$ let $S(x_0)$ be the unique number such that the point $(x_0, S(x_0))$ is on the line segment that connects the points $(a, f(a))$ and $(b, f(b))$. Then*

$$f(x_0) \;\leqq\; S(x_0). \tag{12.11}$$

Note that any x_0 between a and b can be written as $x_0 = \lambda a + (1-\lambda)b$ for some $0 \leqq \lambda \leqq 1$ and that the corresponding y-coordinate $S(x_0) = S(\lambda a + (1-\lambda)b)$ on the line segment that connects $(a, f(a))$ and $(b, f(b))$ then is $S(\lambda a + (1-\lambda)b) = \lambda f(a) + (1-\lambda) f(b)$. Hence we can rephrase the above as follows:

f is convex if and only if for any $a < b$ and $0 \leqq \lambda \leqq 1$ it is true that

$$f(\lambda a + (1-\lambda)b) \;\leqq \lambda f(a) + (1-\lambda) f(b). \tag{12.12}$$

Proof of "⇒": Any line segment S that connects the points $(a, f(a))$ and $(b, f(b))$ in such a way that is entirely contained in the epigraph of f will satisfy $f(x_0) \leqq S(x_0)$ for all $a \leqq x_0 \leqq b$ and it follows that convexity of f implies (12.11).

Proof of "⇐": Now assume that (12.11) is valid. Let $\vec{a} = (a_1, a_2), \vec{b} = (b_1, b_2) \in epi(f)$, i.e.,

$$a_2 \geqq f(a_1) \;\text{ and }\; b_2 \geqq f(b_1). \tag{12.13}$$

We must show that the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leqq \lambda \leqq 1$ is contained in $epi(f)$.

Let $\vec{a}' := (a_1, f(a_1))$. Let $S' := \{\lambda\vec{a}' + (1-\lambda)\vec{b}\} : 0 \leqq \lambda \leqq 1$ be the line segment obtained by leaving the right endpoint $\vec{b}$ unchanged and pushing the left one downward until a_2 matches $f(a_1)$. Clearly, S' nowhere exceeds S.

Let $\vec{b}'' := (b_1, f(b_1))$. Let $S'' := \{\lambda\vec{a}' + (1-\lambda)\vec{b}''\} : 0 \leqq \lambda \leqq 1$ be the line segment obtained by leaving the left endpoint $\vec{a}'$ unchanged and pushing the right one downward until the b_2 matches $f(b_1)$. Clearly, S'' nowhere exceeds S'.

We view any line segment T between two points with abscissas a_1 and b_1 as a function $T(\cdot) : [a_1, b_1] \to \mathbb{R}$ which assigns to $x \in [a_1, b_1]$ that unique value $T(x)$ for which the point $(x, T(x)$ lies on T.

The segment S'' connects the points $(a, f(a))$ and $(b, f(b))$ and it follows from assumption ***b'*** *that for any $a \leqq x_0 \leqq b$ we have $f(x_0) \leqq S''(x_0)$. We conclude from $S(\cdot) \geqq S'(\cdot) \geqq S''(\cdot)$ that $S(x_0) \geqq f(x_0)$, i.e. $(x_0, S(x_0)) \in epi(f)$. As this is true for any $a \leqq x_0 \leqq b$ it follows that the line segment S is entirely contained in the epigraph of f.* ■

Example 12.4 (Sublinear functions are convex)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be sublinear. Then f is convex.

Let $a, b \in \mathbb{R}$ and $0 \leqq \lambda \leqq 1$. It follows from subadditivity and positive homogeneity of f that

$$f(\lambda a + (1-\lambda)b) \leqq f(\lambda a) + f((1-\lambda)b)) = \lambda f(a) + (1-\lambda)f(b).$$

According to prop.12.2 this implies convexity of f.

Proposition 12.3 (Convex vs concave-up)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be concave-up. Then f is convex.*

Proof: Assume to the contrary that f is (differentiable and) concave-up and that $f(x_0) > S(x_0)$ for some $a \leqq x_0 \leqq b$.

It follows that $a < x_0 < b$ because $f(a) = S(a)$ and $f(b) = S(b)$.

Let $S : x \mapsto S(x)$ be the line through the points $(a, f(a))$ and $(b, f(b))$. and let m be the slope of S, i.e.,

$$m = \frac{S(b) - S(a)}{b - a}.$$

It then follows that

$$m = \frac{S(b) - S(x_0)}{b - x_0} > \frac{S(b) - f(x_0)}{b - x_0} = \frac{f(b) - f(x_0)}{b - x_0} = f'(\xi) \tag{12.14}$$

for some $x_0 < \xi < b$ (according to the mean value theorem for derivatives). Further

$$m = \frac{S(x_0) - S(a)}{x_0 - a} < \frac{f(x_0) - S(a)}{x_0 - a} = \frac{f(x_0) - f(a)}{x_0 - a} = f'(\eta) \tag{12.15}$$

for some $a < \eta < x_0$ (according to the mean value theorem for derivatives).

Because f is concave up we have

$$f'(a) \leqq f'(\eta) \leqq f'(x_0) \leqq f'(\xi) \leqq f'(b).$$

From (12.14) *and* (12.15) *we obtain*

$$m \; < \; f'(\eta) \; \leqq \; f'(x_0) \; \leqq \; f'(\xi) \; < \; m$$

and we have reached a contradiction. ■

Proposition 12.4 (Sublinear functions are convex)**.** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be sublinear. Then* f *is concave-up.*

Proof: Let $0 \leqq \lambda \leqq 1$ *and* $x, y \in \mathbb{R}$*. Then*

$$p(\lambda x + (1-\lambda)y) \; \leqq p(\lambda x) \; + \; p((1-\lambda)y) \; = \; \lambda p(x) \; + \; (1-\lambda)p(y) \tag{12.16}$$

and it follows that f *is concave-up.* ■

13 Algebraic structures ($\star$)

This chapter is at its very beginnings. It has been created because it is mentioned in one of the first lectures that the axiomatically defined set $\mathbb{Z}$ of the first chapter of [1] B/G forms a group.

Note that this chapter is starred and hence optional.

13.1 Semigroups and groups ($\star$)

Definition 13.1 (Semigroups and monoids)**.** Given is a nonempty set S with a binary operation $\Box$,

i.e. an "assignment rule" $(s,t) \mapsto s\Box t$ which assigns to any two elements $s,t \in S$ a third element $u := s\Box t \in S$. [99] $(G, \Box)$ is called a **semigroup** if the operation $\Box$ satisfies

$$(13.1) \qquad \textbf{associativity:} \quad (s\Box t)\Box u = s\Box(t\Box u) \text{ for all } s,t,u \in S.$$

A semigroup for which there exists in addition a **neutral element** with respect to the operation$(s,t) \mapsto s\Box t$, i.e., some $e \in S$ such that

$$(13.2) \qquad s\Box e = e\Box s \text{ for all } s \in S$$

is called a **monoid**.

Example 13.1. $(\mathbb{Z}, +)$ (the integers with addition) and $(\mathbb{Z}, \cdot)$ (the integers with multiplication) are monoids: Both $+$ and $\cdot$ are associative and addition has zero, multiplication has 1 as neutral element.

This is also true for the other number systems: $(\mathbb{N}, +)$ and $(\mathbb{N}, \cdot)$ (natural numbers), $(\mathbb{Q}, +)$ and $(\mathbb{Q}, \cdot)$ (rational numbers), $(\mathbb{R}, +)$ and $(\mathbb{R}, \cdot)$ (real numbers), $(\mathbb{C}, +)$ and $(\mathbb{C}, \cdot)$ (complex numbers) all are monoids.

Example 13.2. You need to know from linear algebra or ch. on p. about vector spaces to understand this example:

If V is a vector space with addition "$+$" and scalar multiplication "$\cdot$" then $(V, +)$ is a monoid but $(V, \cdot)$ is not. (Why not?)

Example 13.3. You need to know function composition to understand this example.

Let A be a nonempty set and let

$S := \{f : f \text{ is a function } A \to A\}$. [100]

Then the operation $(f,g) \mapsto g \circ f$ which assigns to any two functions f and g the function [101] $x \mapsto g \circ f(x) := g\big(f(x)\big)$ is associative, i.e., S is a semigroup. Moreover S is a monoid because the identity function $id_A : x \mapsto x$ for all $x \in A$ which does nothing with its arguments satisfies (13.2).

[99] In other words, we have a function $\Box : S \times S \to S$, $(s,t) \mapsto \Box(s,t) := s\Box t$ in the sense of def.4.6 (Mappings (functions)) on p.74.

[100] If this is too abstract for you, choose $A := \mathbb{R}$, the set of real numbers. Then the elements of S will be functions such as $f(x) = 3x^2$ and $g(x) = 7x + 5e^x$. for those two specific functions you get $g \circ f(x) = g\big(f(x)\big) = g(3x^2) = 21x^2 + 5e^{3x^2}$.

[101] See def.4.7 (Function composition) on p.76.

Proof of associativity: For any three functions $f, g, h \in S$ and any $x \in A$ it is true that

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

. In other words, both the left side $x \mapsto \big((h \circ g) \circ f\big)(x)$ and the right side $h \circ (g \circ f)(x)$ are, for each argument $x \in A$, equal to $h(g(f(x))$. This then means that those two functions coincide and we have proved associativity.

Proof that id_A is a neutral element: We have

$$\big(id_A \circ f\big)(x) = id_A\big(f(x)\big) = f(x) = f\big(id_A(x)\big) = \;\big(f \circ id_A\big)(x)$$

for all $x \in A$. It follows that the three assignments $x \mapsto \big(id_A \circ f\big)(x)$, $x\big(f \circ id_A\big)(x)$ and $x \mapsto f(x)$ coincide for all x, i.e., they all represent the same function $x \mapsto f(x)$. This proves (13.2) and we have proved the existence of a neutral element. ■

Definition 13.2 (Groups and Abelian groups)**.** Let $(G, \square)$ be a monoid with neutral element e which satisfies the following: For each $g \in G$ there exists some $g' \in G$ such that

(13.3) $$g \square g' \;=\; g' \square g \;=\; e \;\text{ for all } g \in G.$$

Then $(G, \square)$ is called a **group**.

Assume moreover that the operation $\square$ satisfies

(13.4) **commutativity:** $g \square h \;= h \square g$ for all $g, h \in G$.

Then G is called a **commutative group** or **abelian group**. [102]

Theorem 13.1 (Uniqueness of the inverse in groups)**.** *Let $(G, \square)$ be a group and let $g \in G$. Assume that there exists besides g' another $g'' \in G$ which satisfies (13.3). Then $g'' = g'$.*

Proof: We have

(13.5) $$g'' \overset{(13.2)}{=} e \square g'' \overset{(13.3)}{=} (g' \square g) \square g'' \overset{assoc}{=} g' \square (g \square g'') \overset{(13.3)}{=} g' \square e \overset{(13.2)}{=} g'$$

and this proves uniqueness. ■

Definition 13.3 (inverse element g^{-1})**.** From now on we are allowed to write g^{-1} for the unique element of G that is associated with the given $g \in G$ by means of the formula (13.3). We call g^{-1} the **inverse** element of g.

Example 13.4. a. $(\mathbb{Z}, +)$ (the integers with addition) is an abelian group: We have already seen that $(\mathbb{Z}, +)$ is a monoid.

Inverse element to $k \in \mathbb{Z}$ is $-k$ because $k + (-k) = (-k) + k = 0$ for all k.

and this group is abelian because $m + k = k + m$ for all $k, m \in \mathbb{Z}$.

[102] named so after the Norwegian mathematician Niels Henrik Abel who lived in the first half of the 19th century and died at age 26.

and $(\mathbb{Z}, \cdot)$ (the integers with multiplication) are monoids: Both $+$ and $\cdot$ are associative and addition has zero, multiplication has 1 as neutral element.

b. $(\mathbb{Z}, \cdot)$ (the integers with multiplication) is **not** a group: Let $k = 5$. Then $k \in \mathbb{Z}$ but $1/5$, the only number m such that $5 \cdot m = m \cdot 5 = 1$ (1 is the neutral element with respect to "$\cdot$") is a fraction and does not belong to $\mathbb{Z}$.

Example 13.5. Being a group is a lot more specific than just being a semigroup or monoid. Not all types of numbers form groups for addition and/or multiplication:

Natural numbers:	Neither $(\mathbb{N}, +)$ nor $(\mathbb{N}, \cdot)$ are groups: $(\mathbb{N}, +)$ does not even have a neutral element, $(\mathbb{N}, \cdot)$ has 1 as a neutral element but there is no multiplicative inverse for, say, 5 because $1/5 \notin \mathbb{N}$.
Integers:	We have seen in example 13.4 that $(\mathbb{Z}, +)$ is an abelian group but $(\mathbb{Z}, +)$ is not a group.
Rational numbers:	$(\mathbb{Q}, +)$ is an abelian group but $(\mathbb{Q}, \cdot)$ is **not** a group because the number 0 does not have a multiplicative inverse: There is no number x such that $0 \cdot x = 1$. But note that the set $\mathbb{Q}^\star$ of all non-zero rational numbers is an abelian group.
Real numbers:	$(\mathbb{R}, +)$ is an abelian group but $(\mathbb{R}, \cdot)$ is **not** a group for the same reason as $(\mathbb{Q}, \cdot)$. Again, the set $\mathbb{R}^\star$ of all non-zero real numbers is an abelian group.
Complex numbers:	$(\mathbb{C}, +)$ is an abelian group but $(\mathbb{C}, \cdot)$ is **not** a group for the same reason as $(\mathbb{Q}, \cdot)$. Again, the set $\mathbb{C}^\star$ of all non-zero complex numbers is an abelian group.

Example 13.6. You need to know from linear algebra or ch.ssec-general-vector-spaces on p.137 about general vector spaces to understand this example:

If V is a vector space with addition "$+$" and scalar multiplication "$\cdot$" then $(V, +)$ is a monoid but $(V, \cdot)$ is not. (Why not?)

Example 13.7. As in example 13.3 (function composition) let A be a nonempty set and let $S := \{f : f$ is a function $A \to A\}$ with the operation $(f, g) \mapsto g \circ f$ defined as $g \circ f(x) = g(f(x))$. We have seen that S is a monoid but S is not a group because not every $f \in S$ has an inverse function f^{-1} which satisfies $f \circ f^{-1}(x) = f^{-1}(f(x)) = x$ for all $x \in A$.

Counterexample: Let $A := \mathbb{R}$. Some functions will have an inverse. For example $f(x) = x - 7$ has inverse $f^{-1}(x) = x + 7$.

But $f(x) = x^2$ does not have an inverse: $g(x) := \sqrt{x}$ will not work because $g(f(-2)) = g(4) = 2$, not -2 as required, and $h(x) := -\sqrt{x}$ will not work because $h(f(2)) = h(4) = -2$, not 2 as required.

Example 13.8. Let $(G, \Box)$ and $(H, \bullet)$ be defined as follows:

$$G := \{g \in \mathbb{R} : g = e^x \text{ for some } x \in \mathbb{R}\}, \quad e^x \Box e^y := e^x \cdot e^y = e^x + y, \tag{13.6}$$

$$H := \{h \in \mathbb{R} : h = \ln(x) \text{ for some } u \in]0, \infty\}, \quad \ln u \bullet \ln v := \ln u + \ln v = \ln(xy). \tag{13.7}$$

a. Both $(G, \Box)$ and $(H, \bullet)$ are abelian groups G has neutral element 1 and H has neutral element 0. (Exercise: Prove it. What are the inverses?)

b. Let the functions φ and ψ be defined as follows:

$$\varphi : (G, \square) \to (H, \bullet), \quad \varphi(g) := \ln g, \tag{13.8}$$

$$\psi : (H, \bullet) \to (G, \square), \quad \psi(h) := e^h. \tag{13.9}$$

Then φ and ψ satisfy the following:

$$\varphi(g_1 \square g_2) = \varphi(g_1) \bullet \varphi(g_2), \quad \varphi(1) = 0, \quad \varphi(g^{-1}) = \varphi(g)^{-1}, \tag{13.10}$$

$$\psi(h_1 \bullet h_2) = \psi(h_1) \square \psi(h_2), \quad \psi(0) = 1, \quad \psi(h^{-1}) = \psi(h)^{-1}. \tag{13.11}$$

Further, the functions φ and ψ are inverse to each other, i.e.,

$$\psi\big(\varphi(g)\big) = g \quad \text{and} \quad \varphi\big(\psi(h)\big) = h \tag{13.12}$$

for all $g \in G$ and $h \in H$.

If you talk about φ and ψ as "the functions" and $\square$ and $\bullet$ as "the operations" you might state the results (13.10) and (13.10) as follows:

*The functions are **structure compatible** with the operations on their domains and codomains:*

It does not matter whether you first apply the operation to two items in the domain and then apply the function to the result or whether you first map those two items into the codomain and then apply the operation to the two function values. Further, the inverse of the function value is the function value of the inverse and the function maps the neutral element to the neutral element.

Such functions between two groups have a special name:

Definition 13.4 (Homomorphisms and isomorphisms)**.** Let $(G, \square)$ and $(H, \bullet)$ be two groups with neutral elements e_G and e_H and let us write g^{-1} and h^{-1} for the inverses (in the sense of def. 13.3.

Let $\varphi : (G, \square) \to (H, \bullet)$ be a function which satisfies the following:

$$\varphi(g_1 \square g_2) = \varphi(g_1) \bullet \varphi(g_2), \quad \varphi(e_G) = e_H, \quad \varphi(g^{-1}) = \varphi(g)^{-1}. \tag{13.13}$$

Then we call φ a **homomorphism**, more specifically, a **group homomorphism**, from the group $(G, \square)$ to the group $(H, \bullet)$.

Let $\psi : (H, \bullet) \to (G, \square)$ be a group homomorphism from $(H, \bullet)$ to $(G, \square)$ such that φ and ψ are inverse to each other. [103]

We call such a bijective homomorphism an **isomorphism**.

[103] They then of course also are bijective functions (see def. 4.9).

14 Construction of the number systems

14.1 The Peano axioms (Skim this!)

Definition 14.1 (Set of non-negative integers)**.** We define the set $\mathbb{N}_0$ (the non-negative integers) axiomatically as follows:

Ax.1 There is an element "0" contained in $\mathbb{N}_0$.
Ax.2 There is a function $\sigma : \mathbb{N}_0 \to \mathbb{N}_0$ such that
- **Ax.2.1** σ is injective,
- **Ax.2.2** $0 \notin \sigma(\mathbb{N}_0)$ (range of f),
- **Ax.2.3** Induction axiom: Let $U \subseteq \sigma(\mathbb{N}_0)$ such that **a.** $0 \in U$, **b.** If $n \in U$ then $\sigma(n) \in U$. Then $U = \mathbb{N}_0$.

We define $\mathbb{N} := \mathbb{N}_0 \setminus \{0\}$.

Definition 14.2 (Iterative function composition)**.** Let $X \neq \emptyset$ and $f : X \to X$. We now use the induction axiom above to define f^n for an arbitrary function $f : X \to X$.

a. $f^0 := \mathrm{id}_X : x \mapsto x$, **b.** $f^1 := f$, **c.** $f^2 := f \circ f$ (function composition), **c.** $f^{\sigma(n)} := f \circ f^n$.

Proposition 14.1. *f^n is defined for all $n \in \mathbb{N}_0$.*

Proof: Let $U := \{k \in \mathbb{N}_0 : f^k$ is defined $\}$. Then $0 \in U$ as $f^0 = id_A$ and if $k \in U$, i.e., f^k is defined then $f^{\sigma(k)} = f \circ f^k$ also is defined, i.e., $\sigma(k) \in U$. It follows from Ax.2.3 that $U = \mathbb{N}_0$. ■

Remark 14.1 ($\sigma(\cdot)$ as successor function)**.** Of course the meaning of $\sigma(n)$ will be that of $n+1$:

$$0 \overset{\sigma}{\mapsto} 1 \overset{\sigma}{\mapsto} 2 \overset{\sigma}{\mapsto} 3 \overset{\sigma}{\mapsto} \dots$$

Definition 14.3 (Addition and multiplication on $\mathbb{N}_0$)**.** Let $m, n \in \mathbb{N}_0$. Let

$$m + n := \sigma^n(m), \tag{14.1}$$
$$m \cdot n := (\sigma^m)^n(0). \tag{14.2}$$

Note that we know the meaning of $(\sigma^m)^n$: $f := \sigma^m$ is a function $A \to A$ and we have established in prop.14.1 the meaning of f^n, i.e., $(\sigma^m)^n$.

Proposition 14.2. *Addition and multiplication satisfy all rules of arithmetic we learned in high school such as*

$$m + n = n + m \quad \textit{commutativity of addition} \tag{14.3}$$
$$k + (m + n) = (k + m) + n \quad \textit{associativity of addition} \tag{14.4}$$
$$m \cdot n = n \cdot m \quad \textit{commutativity of multiplication} \tag{14.5}$$
$$k \cdot (m \cdot n) = (k \cdot m) \cdot n \quad \textit{associativity of multiplication} \tag{14.6}$$
$$k \cdot (m + n) = k \cdot m + k \cdot n \quad \textit{distributivity of addition} \tag{14.7}$$
$$n \cdot 1 = 1 \cdot n = n \quad \textit{neutral element for multiplication} \tag{14.8}$$
(14.9)

Here 1 *is defined as* $1 = \sigma(0)$.

Proof: Drudge work. ■

Definition 14.4 (Order relation $m < n$ on $\mathbb{N}_0$)**.** Let $m, n \in \mathbb{N}_0$.

a. We say m is less than n and we write $m < n$ if there exists $x \in \mathbb{N}$ such that $n = m + x$.

b. We say m is less or equal than n and we write $m \leqq n$ if $m < n$ or $m = n$.

c. We say m is greater than n and we write $m > n$ if $n < m$. We say m is greater or equal than n and we write $m \geqq n$ if $n \leqq m$.

Proposition 14.3. *"$<$" and "$\leqq$" satisfy all the usual rules we learned in high school such as*

Trichotomy of the order relation: Let $m, n \in \mathbb{N}_0$. *Then exactly one of the following is true:*

$$m < n, \quad m = n, \quad m > n.$$

Proof: Drudge work. ■

14.2 Constructing the integers from $\mathbb{N}_0$

For the following look at B/G project 6.9 in ch.6.1 and B/G prop.6.25 in ch.6.3.

Definition 14.5 (Integers as equivalence classes)**.** We define the following equivalence relation $(m_1, n_1) \sim (m_2, n_2)$ on the cartesian product $\mathbb{N}_0 \times \mathbb{N}_0$:

$$(m_1, n_1) \sim (m_2, n_2) \;\Leftrightarrow\; m_1 + n_2 = n_1 + m_2 \tag{14.10}$$

We write $\mathbb{Z} := \{[(m, n)] : m, n \in \mathbb{N}_0\}$. In other words, $\mathbb{Z}$ is the set of all equivalence classes with respect to the equivalence relation (14.10).

We "embed" $\mathbb{N}_0$ into $\mathbb{Z}$ with the following injective function $e : \mathbb{N}_0 \to \mathbb{Z}$: e(m) := [(m,0)].

From this point forward we do not distinguish between $\mathbb{N}_0$ and its image $e(\mathbb{N}_0) \subseteq \mathbb{Z}$ and we do not distinguish between $\mathbb{N}$ and its image $e(\mathbb{N}) \subseteq \mathbb{Z}$. In particular we do not distinguish between the two zeroes 0 and $[(0, 0)]$ and between the two ones 1 and $[(1, 0)]$.

Finally we write $-n$ for the integer $[(0, n)]$.

With those abbreviation we then obtain

Proposition 14.4 (Trichotomy of the integers)**.** *Let* $z \in \mathbb{Z}$. *Then exactly one of the following is true:*

Either ***a.*** $z \in \mathbb{N}$, *i.e.,* $z = [(m, 0)]$ *for some* $m \in \mathbb{N}$ *or* ***b.*** $-z \in \mathbb{N}$, *i.e.,* $z = [(0, n)]$ *for some* $n \in \mathbb{N}$ *or* ***c.*** $z = 0$.

Proof: Drudge work. ■

Remark 14.2. a. The intuition that guided the above definition is that the pairs $(4, 0), (7, 3), (130, 126)$ all define the same integer 4 and the pairs $(0, 4), (3, 7), (126, 130)$ all define the same integer -4.

b. If it had been possible to define subtraction $m - n$ for all $m, n \in \mathbb{N}_0$ then (14.10) could be rewritten as

$$(m_1, n_1) \sim (m_2, n_2) \;\Leftrightarrow\; m_1 - n_1 = m_2 - n_2.$$

Looking at the equivalent pairs $(4, 0), (7, 3), (130, 126)$ we get $4 - 0 = 7 - 3 = 130 - 126 = 4$ and for $(0, 4), (3, 7), (126, 130)$ we get $0 - 4 = 3 - 7 = 126 - 130 = -4$.

Definition 14.6 (Addition, multiplication and subtraction on $\mathbb{Z}$)**.** Let $[(m_1, n_1)]$ and $[(m_2, n_2)] \in \mathbb{Z}$. We define

$$-[(m_1, n_1)] := [n_1, m_1], \tag{14.11}$$

$$[(m_1, n_1)] + [(m_2, n_2)] := [(m_1 + m_2, n_1 + n_2)] \tag{14.12}$$

$$[(m_1, n_1)] \cdot [(m_2, n_2)] := [(m_1 m_2 + n_1 n_2, m_1 n_2 + n_1 m_2)] \tag{14.13}$$

We write $[(m_1, n_1)] - [(m_2, n_2)]$ ("$[(m_1, n_1)]$ minus $[(m_2, n_2)]$")
as an abbreviation for $[(m_1, n_1)] + \big(-[(m_2, n_2)]\big)$.

We write $[(m_1, n_1)] < [(m_2, n_2)]$ if $[(m_2, n_2)] - [(m_1, n_1)] \in \mathbb{N}$, i.e., if there is $k \in \mathbb{N}$ such that $[(m_2, n_2)] - [(m_1, n_1)] = [(k, 0)]$. We then say that $[(m_1, n_1)]$ is less than $[(m_2, n_2)]$.

We write $[(m_1, n_1)] \leqq [(m_2, n_2)]$ if $[(m_1, n_1)] < [(m_2, n_2)]$ or if $[(m_1, n_1)] = [(m_2, n_2)]$ and we then say that $[(m_1, n_1)]$ is less than or equal to $[(m_2, n_2)]$.

We write $[(m_1, n_1)] > [(m_2, n_2)]$ if $[(m_2, n_2)] < [(m_1, n_1)]$ and we then say that $[(m_1, n_1)]$ is greater than $[(m_2, n_2)]$.

We write $[(m_1, n_1)] \geqq [(m_2, n_2)]$ if $[(m_2, n_2)] \leqq [(m_1, n_1)]$ and we then say that $[(m_1, n_1)]$ is greater than or equal to $[(m_2, n_2)]$.

We write $\mathbb{Z}_{\geqq 0}$ for the set of all integers z such that $z \geqq 0$ and $\mathbb{Z}_{\neq 0}$ for the set of all integers z such that $z \neq 0$. You should convince yourself that $\mathbb{Z}_{\geqq 0} = \mathbb{N}_0$.

It turns out that all three operations are "well defined" in the sense that the resulting equivalence classes on the right of each of the three equations above do not depend on the choice of representatives in the classes on the left. Further we have

Proposition 14.5. *Let $m, n \in \mathbb{N}_0$. Then*

$$[(m, n)] + [(0, 0)] = [(0, 0)] + [(m, n)] = [(m, n)], \tag{14.14}$$

$$\big(-[(m, n)]\big) + [(m, n)] = [(m, n)] + \big(-[(m, n)]\big) = [0, 0] \tag{14.15}$$

$$[(m, n)] \cdot [(1, 0)] = [(1, 0)] \cdot [(m, n)] = [(m, n)], \tag{14.16}$$

i.e., $[(0, 0)]$ becomes the neutral element with respect to addition, $[(1, 0)]$ becomes the neutral element with respect to multiplication and $-[(m, n)]$ becomes the additive inverse of $[(m, n)]$.

Proof: Drudge work. ■

Remark 14.3. Again, if it had been possible to define subtraction $m - n$ for all $m, n \in \mathbb{N}_0$ then it would be easier to see why addition and multiplication have been defined as you see it in def.14.6:

Addition is defined such that $(m_1 - n_1) + (m_2 - n_2) = (m_1 + m_2) - (n_1 + n_2)$
and multiplication: $(m_1 - n_1) \cdot (m_2 - n_2) = \big(m_1 m_2 + (-n_1)(-n_2)\big) \; - \; \big(m_1 n_2 + n_1 m_2\big)$.

14.3 Constructing the rational numbers from $\mathbb{Z}$

For the following look again at B/G project 6.9 in ch.6.1 and B/G prop.6.25 in ch.6.3.

Definition 14.7 (Fractions as equivalence classes)**.** We define the following equivalence relation $(p, q) \sim (r, s)$ on the cartesian product $\mathbb{Z} \times \mathbb{Z}_{\neq 0}$:

$$(p, q) \sim (r, s) \;\Leftrightarrow\; p \cdot s \;=\; q \cdot r \tag{14.17}$$

We write $\mathbb{Q} := \{[(p, q)] : p, q \in \mathbb{Z} \text{ and } q \neq 0\}$. In other words, $\mathbb{Q}$ is the set of all equivalence classes with respect to the equivalence relation (14.17).

We "embed" $\mathbb{Z}$ into $\mathbb{Q}$ with the injective function $e : \mathbb{Z} \to \mathbb{Q}$ defined as $e(z) := [(z, 1)]$.

Remark 14.4. a. The intuition that guided the above definition is that the pairs $(12, 4), (-21, -7), (105, 35)$ all define the same fraction $3/1$ and the pairs $(4, -12), (-7, 21), (-35, 105$ all define the same fraction $-1/3$.

b. If it had been possible to define division p/q for all $p, q \in \mathbb{Z}$ for which $q \neq 0$ then (14.17) could be rewritten as

$$(p, q) \sim (r, s) \;\Leftrightarrow\; p/q \;=\; r/s$$

Looking at the equivalent pairs $(12, 4), (-21, -7), (105, 35)$ we get $12/4 \;=\; (-21)/(-7) \;=\; 105/35 \;=$ 3 and for $(4, -12), (-7, 21), (-35, 105)$ we get $4/(-12) \;=\; (-7)/21 \;=\; (-35)/105 \;=\; -1/3$.

c. It is easy to see that $(p, q) \sim (r, s)$ if and only if there is (rational) $\alpha \neq 0$ such that $r = \alpha p$ and $s = \alpha q$. A formal proof is just drudgework.

Definition 14.8 (Addition, multiplication, subtraction and division in $\mathbb{Q}$)**.** Let $[(p_1, q_1)]$ and $[(p_2, q_2)] \in \mathbb{Q}$. We define

$$-[(p_1, q_1)] \;:= [(-p_1, q_1)], \tag{14.18}$$
$$[(p_1, q_1)] + [(p_2, q_2)] \;:= [(p_1 q_2 + q_1 p_2, n_1 n_2)] \tag{14.19}$$
$$[(p_1, q_1)] - [(p_2, q_2)] \;:= [(p_1, q_1)] + \big(-[(p_2, q_2)]\big) \tag{14.20}$$
$$[(p_1, q_1)] \cdot [(p_2, q_2)] \;:= [(p_1 p_2, q_1 q_2)] \tag{14.21}$$
$$[(p_1, q_1)]^{-1} \;:= [(1, 1)]/[(p_1, q_1)] \;:=\; [(q_1, p_1)] \text{ (if } p_1 \neq 0), \tag{14.22}$$
$$[(p_1, q_1)]/[(p_2, q_2)] \;:= [(p_1 q_2, q_1 p_2)] \;=\; [(p_1, q_1)] \cdot [(p_2, q_2)]^{-1} \text{ (if } p_2 \neq 0) \tag{14.23}$$

It turns out that operations above are "well defined" in the sense that the resulting equivalence classes on the right of each of the three equations above do not depend on the choice of representatives in the classes on the left. [104]

Further we have

Proposition 14.6 (Trichotomy of the rationals)**.** *Let $x \in \mathbb{Q}$. Then exactly one of the following is true:*

*Either **a.** $x > 0$, i.e., $x = [(p, q)]$ for some $p, q \in \mathbb{N}$ or **b.** $-x > 0$, i.e., $x = [(-p, q)]$ for some $p, q \in \mathbb{N}$ or **c.** $x = 0$.*

Proof: Drudge work. ■

14.4 Constructing the real numbers via Dedekind Cuts

The material presented here, including the notation, follows [10] Rudin, Walter: Principles of Mathematical Analysis.

*Note that in this section small greek letters denote **sets** of rational numbers!*

The idea behind real numbers as intervals of rational numbers with no lower bounds, called Dedekind cuts, is as follows:

Given a real number x you can associate with it the set $\{q \in \mathbb{Q} : q < x\}$ which we call the cut or Dedekind cut associated with x The mapping

$$\Phi : x \mapsto \Phi(x) := \{q \in \mathbb{Q} : q < x\} \tag{14.24}$$

*is injective because if $x, y \in \mathbb{R}$ such that $x \neq y$, say, $x < y$, then we have $\{q \in \mathbb{Q} : q < x\} \subsetneq \{q \in \mathbb{Q} : q < y\}$ because there are (infinitely many) rational numbers in the open interval $]x, y[$ and we get surjectivity of Φ for free if we take as codomain the set of all cuts. Because Φ is bijective we can "identify" any real number with its cut. We now go in reverse: we start with a definition of cuts which does not reference the real number x, i.e., we define them just in terms of rational numbers and define addition, multiplication and the other usual operations on those cuts and show that those cuts have all properties of the real numbers as they were axiomatically defined in B/G ch.8, including the **completeness axiom** which states that each subset A of $\mathbb{R}$ with upper bounds has a least upper bound $\sup(A)$, i.e., a minimum in the set of all its upper bounds.*

Definition 14.9 (Dedekind cuts)**.** (Rudin def.1.4)

We call a subset $\alpha \subseteq \mathbb{Q}$ a **cut** or **Dedekind cut** if it satisfies the following:

a. $\alpha \neq \emptyset$ and $\alpha^{\complement} \neq \emptyset$
b. Let $p, q \in \mathbb{Q}$ such that $p \in \alpha$ and $q < p$. Then $q \in \alpha$.
c. α does not have a max: $\forall p \in \alpha\ \exists q \in \alpha$ such that $p < q$.

Given a cut α, let $p \in \alpha$ and $q \in \alpha^{\complement}$. We call p a **lower number** of the cut α and we call q an **upper number** of α.

[104] This was shown for multiplication $[(p_1, q_1)] \cdot [(p_2, q_2)] = [(p_1 p_2, q_1 q_2)]$ in exercise 4.5 on p.89.

Theorem 14.1. *(Rudin thm.1.5)*

Let $\alpha \subseteq \mathbb{Q}$ be a cut. Let $p \in \alpha, q \in \alpha^{\complement}$. Then $p < q$.

Assume to the contrary that $q \leqq p$. Then we either have $p = q$ which means that either both p, q belong to α or both belong to its complement, a contradiction to our assumption. Or we have $q < p$. It then follows from $p \in \alpha$ and def.14.9.b that $q \in \alpha$, contrary to our assumption. ■

Theorem 14.2. *(Rudin thm.1.6)*

Let $r \in \mathbb{Q}$. Let $r^\star := \{p \in \mathbb{Q} : p < r\}$. Then $r^\star$ is a cut and $r = \min\left((r^\star)^{\complement}\right)$.

Proof: In the following let $p, q, r \in \mathbb{Q}$.

Proof of def.14.9.a: $r - 1 < r \;\Rightarrow\; r - 1 \in r^\star \;\Rightarrow\; r^\star \neq \emptyset$. Further, $r \in (r^\star)^{\complement} \;\Rightarrow\; (r^\star)^{\complement} \neq \emptyset$.

Proof of def.14.9.b: Let $q < p$ and $p \in r^\star$. Then also $q \in r^\star = \{p' \in \mathbb{Q} : p' < r\}$.

Proof of def.14.9.c: Let $p \in r^\star$. Then $p < (p+r)/2 < r$, hence $(p+r)/2 \in r^\star$ and r cannot be the max of $r^\star$.
■

Definition 14.10 (Rational cuts)**.** Let $r \in \mathbb{Q}$. The cut $r^\star = \{p \in \mathbb{Q} : p < r\}$ from the previous theorem is called the **rational cut** associated with r.

Remark 14.5. If we define intervals in $\mathbb{Q}$ in the usual way for $p, q \in \mathbb{Q}$:

$$]p, q[:= \{r \in \mathbb{Q} : p < r < q\}, \quad [p, q] := \{r \in \mathbb{Q} : p \leqq r \leqq q\}, \quad \text{etc.}$$

then rational cuts $r^\star (r \in \mathbb{Q})$ are those for which $r^\star =]-\infty, r[$ and $(r^\star)^{\complement} = [r, \infty[$ whereas for non-rational cuts α we cannot specify the "thingy" that should take the role of r. It would be the $\sup(\alpha)$ if we already had defined the set of all real numbers and we could understand α as a subset of those real numbers.

Definition 14.11 (Ordering Dedekind cuts)**.** (Rudin def.1.9) Let α, β be two cuts.

We say $\alpha < \beta$ if $\alpha \subsetneq \beta$ (strict subset) and we say $\alpha \leqq \beta$ if $\alpha < \beta$ or $\alpha = \beta$, i.e., $\alpha \subseteq \beta$.

Proposition 14.7 (Trichotomy of the cuts)**.** *(Rudin thm.1.10)*

Let α, β be two cuts. Then either $\alpha < \beta$ or $\alpha > \beta$ or $\alpha = \beta$.

Proof: We only need to show that if $\alpha \nsubseteq \beta$ then $\beta \subsetneq \alpha$.
So let $\alpha \nsubseteq \beta$. Then $\alpha \setminus \beta$ is not empty and there exists $q \in \alpha \setminus \beta$.
But then $q > b$ for all $b \in \beta$. Also, if $a \in \mathbb{Q}$ and $a \leqq q$ then $a \in \alpha$ (we applied def.14.9.b twice.)
As $b < q$ for all $b \in \beta$ it follows that $\beta \subseteq \alpha$. We saw earlier that $\alpha \setminus \beta \neq \emptyset$ and this proves that $\beta \neq \alpha$, i.e., $\beta \subsetneq \alpha$. ■

Theorem 14.3 (Addition of two cuts)**.** *(Rudin thm.1.12) Let α, β be two cuts and let*

$$\alpha + \beta := \{a + b : a \in \alpha, b \in \beta\}.$$

Then the set of all cuts is an abelian group with this operation. In other words, $+$ is commutative and associative with a neutral element (which turns out to be $0^\star$, the rational cut corresponding to $0 \in \mathbb{Q}$) and a suitably defined cut $-\alpha$ for a given cut α which satisfies $\alpha + (-\alpha) = (-\alpha) + \alpha = 0^\star$

Having defined negatives $-\alpha$ for all cuts we then also can define their absolute values

$$|\alpha| := \begin{cases} \alpha & \text{if } \alpha \geqq 0^\star, \\ -\alpha & \text{if } \alpha < 0^\star. \end{cases}$$

Proof: Not given here. ■

Theorem 14.4 (Multiplication of two cuts)**.** *Let $\alpha \geqq 0^\star, \beta \geqq 0^\star$ be two non-negative cuts. Let*

$$\alpha \cdot \beta := \begin{cases} \{q \in \mathbb{Q} : q < 0\} \cup \{ab : a \in \alpha, b \in \beta\} & \text{if } \alpha \geqq 0^\star, \beta \geqq 0^\star, \\ -|\alpha| \cdot |\beta| & \text{if } \alpha < 0^\star, \beta \geqq 0^\star \text{ or } \alpha \geqq 0^\star, \beta < 0^\star, \\ |\alpha| \cdot |\beta| & \text{if } \alpha < 0^\star, \beta < 0^\star. \end{cases}$$

Then the set $\alpha \cdot \beta$ is a cut, called the product of α and β.

It can be proved that for each cut $\alpha \neq 0^\star$ there is a cut α^{-1} uniquely defined by the equation $\alpha \cdot \alpha^{-1} = 1^\star$.

Theorem 14.5 (The set of all cuts forms a field)**.** *Let Let $\mathbb{R}$ be the set of all cuts. Then $\mathbb{R}$ satisfies axioms 8.1 - 8.5 of B/G:*

Addition and multiplication are both commutative and associative and the law of distributivity $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$ holds.

The cut $0^\star$ is the neutral element for addtition and the cut $1^\star$ is the neutral element for multiplication.

$-\alpha$ is the additive inverse of any cut α and α^{-1} is the multiplicative inverse of $\alpha \neq 0^\star$.

Further the set $\mathbb{R}_{>0} := \{\alpha \in \mathbb{R} : \alpha > 0^\star\}$ satisfies B/G axiom 8.26.

Proof: It follows from prop.14.7 on p.235 that $\mathbb{R}_{>0}$ satisfies B/G axiom 8.26. Proofs of the other properties of $\mathbb{R}$ are not given here. ■

In the remainder of this section we will see that the completeness axiom B/G ax.8.52 (every subset of $\mathbb{R}$ with upper bounds has a supremum) is a consequence from the properties of cuts and there is no need to state it as an axiom.

Theorem 14.6. *(Rudin thm.1.29) Let $\alpha, \beta \in \mathbb{R}$ and let $\alpha < \beta$. Then there exists $q \in \mathbb{Q}$ such that $\alpha < q^\star < \beta$*

Proof: Any $q \in \beta \setminus \alpha$ will do. ■

Theorem 14.7. *(Rudin thm.1.30) Let $\alpha \in \mathbb{R}, p \in \mathbb{Q}$. Then $p \in \alpha \Leftrightarrow p^\star < \alpha$, i.e., $p^\star \subsetneq \alpha$*

Proof of $\Leftarrow$): Let $p \in \alpha$. it follows for any $q \in p^\star$ that $q < p \in \alpha$, hence $q \in \alpha$, hence $p^\star \subseteq \alpha$. As $p \notin p^\star = \{p' \in \mathbb{Q} : p' < p\}$ but $p \in \alpha$ we have strict inclusion $p^\star \subsetneq \alpha$.

Proof of $\Rightarrow$): As $p^\star \subsetneq \alpha$ there exists $q \in \alpha \setminus p^\star$. As $q \geqq p$ and $q \in \alpha$ we obtain $p \in \alpha$ from def.14.9.b. ■

Theorem 14.8 (Dedekind's Theorem)**.** *(Rudin thm.1.32) Let $\mathbb{R} = A \biguplus B$ a partitioning of $\mathbb{R}$ such that*

$$\textbf{a. } A \neq \emptyset \text{ and } B \neq \emptyset$$
$$\textbf{b. } \alpha \in A, \beta \in B \ \Rightarrow \ \alpha < \beta \ \ (\text{i.e., } \alpha \subsetneq \beta).$$

Then there exists a unique cut $\gamma \in \mathbb{R}$ such that if $\alpha \in A$ then $\alpha \leqq \gamma$ and if $\beta \in B$ then $\gamma \leqq \beta$.

Proof: We first prove uniqueness and afterwards the existence of γ.

Proof of uniqueness: Assume there is $\gamma'' \in \mathbb{R}$ which satisfies $\alpha \leqq \gamma''$ for all $\alpha \in A$ and $\gamma'' \leqq \beta$ for all $\beta \in B$.

We may assume that $\gamma < \gamma''$. It follows from thm.14.6 on p.236 that there is $\gamma' \in \mathbb{R}$ (matter of fact, a rational cut) such that $\gamma < \gamma' < \gamma''$. But $\gamma < \gamma'$ implies that $\gamma' \in B$ and $\gamma' < \gamma''$ implies that $\gamma' \in A = B^{\complement}$. We have reached a contradiction and conclude that γ must be unique.

Proof of existence of γ: Let $\gamma := \bigcup [\alpha : \alpha \in A]$.

Step 1: We now show that γ is a cut.

We first show that def.14.9.a is satisfied. As $B \neq \emptyset$ there is some $\beta \in B$. As $\beta^{\complement} \neq \emptyset$ there is some $q \in \beta^{\complement}$. It follows from $\alpha \subseteq \beta$ for all $\alpha \subseteq \gamma = \bigcup [\alpha : \alpha \in A]$ that $\gamma \subseteq \beta$, hence $\gamma^{\complement} \supseteq \beta^{\complement}$. It follows from $q \in \beta^{\complement}$ that $q \in \gamma^{\complement}$, hence $\gamma^{\complement} \neq \emptyset$. Further, it follows from $A \neq \emptyset$ that $\gamma \neq \emptyset$. We conclude that def.14.9.a is satisfied.

Next we show the validity of def.14.9.b. Let $p \in \gamma$, i.e., $p \in \alpha_0$ for some $\alpha_0 \in A$. Let $q < p$. Then $q \in \alpha_0 \subseteq \bigcup [\alpha : \alpha \in A]$, i.e., $q \in \gamma$. We conclude that def.14.9.b is satisfied.

Now we show the validity of def.14.9.c. Let $p \in \gamma$, i.e., $p \in \alpha_0$ for some $\alpha_0 \in A$. As the cut α_0 does not have a maximum there exists some $q \in \alpha_0$ such that $q > p$. As $\alpha_0 \subseteq \gamma$, hence $q \in \gamma$ We have seen that any $p \in \gamma$ is strictly dominated by some $q \in \gamma$. It follows that γ does not have a max and this shows that def.14.9.c is satisfied. We conclude that γ is a cut and step 1 of the proof for existence is completed.

Step 2: It remains to show that $\alpha \leqq \gamma \leqq \beta$ for all $\alpha \in A$ and $\beta \in B$. It is trivial that $\alpha \leqq \gamma$ for all $\alpha \in A$ because $\gamma := \bigcup [\alpha : \alpha \in A]$.

To show that $\gamma \leqq \beta$ for all $\beta \in B$ we prove that the opposite statement that

$$\gamma > \beta, \ \ \text{i.e., } \gamma \setminus \beta \neq \emptyset \ \text{ for some cut } \beta \in B \tag{14.25}$$

will lead to a contradiction. As $q \in \gamma$ there is some $\alpha_0 \in A$ such that $q \in \alpha_0$. Actually, $q \in \alpha_0 \setminus \beta$ because $q \notin \beta$. But then $\alpha_0 \not< \beta$ even though $\alpha_0 \in A$ and $\beta \in B$, contrary to the assumptions about the partitioning $A \biguplus B$ of $\mathbb{R}$. ■

Corollary 14.1. *Let $\mathbb{R} = A \uplus B$ be a partitioning of $\mathbb{R}$ such that*

$$\textbf{a. } A \neq \emptyset \text{ and } B \neq \emptyset$$
$$\textbf{b. } \alpha \in A, \beta \in B \ \Rightarrow \ \alpha < \beta \ \ (\text{i.e., } \alpha \subsetneq \beta).$$

Then either $\max(A) (= l.u.b.(A))$ exists or $\min(B) (= g.l.b.(B))$ exists.

Proof: *According to thm.14.8 there exists* $\gamma \in \mathbb{R}$ *such that if* $\alpha \in A$ *then* $\alpha \leqq \gamma$ *and if* $\beta \in B$ *then* $\gamma \leqq \beta$. *Clearly* γ *is an upper bound of* α *and a lower bound of* β. *It follows that if* $\gamma \in A$ *then* $\max(A) = \gamma$ *and if* $\gamma \notin A$, *i.e.,* $\gamma \in B$, *then* $\min(B) = \gamma$.

■

Theorem 14.9 (Completeness theorem for $\mathbb{R}$). *(Rudin thm.1.36)*

Let $\emptyset \neq E \subset \mathbb{R}$ *and assume that* E *is bounded above. Then* E *has a least upper bound which we denote by* $\sup(E)$ *or l.u.b.(E).*

Proof: *Let* B *be the set of all upper bounds for* E, *i.e.,* $b \in B$ *if and only if* $b \geqq x$ *for all* $x \in E$. *Then* B *is not empty by assumption. Let* $A := B^{\complement} = \{\alpha \in \mathbb{R} : \alpha < x$ *i.e.,* $\alpha \subsetneqq x$ *for some* $x \in E\}$. *In other words,* $\alpha \in A$ *if and only if* α *is not an upper bound of* E.

A *is not empty either: As* $E \neq \emptyset$ *there is some* $x \in E$. *Let* $\alpha := x - 1$. *Cleary* $x \leqq \alpha$ *is not true for all* $x \subseteq E$. *It follows that* α *is not an upper bound of* E, *hence* $\alpha \in A$, *hence* A *is not empty.*

Moreover we have $\alpha < \beta$ *for all* $\alpha \in A$ *and* $\beta \in B$. *Because for any* $\alpha \in A$ *there is some* $x \in E$ *such that* $\alpha < x$ *and we have* $x \leqq \beta$ *for all upper bounds* β, *i.e., for all* $\beta \in B$.

It follows that the sets A *and* B *form a partition which satisfies the requirements of Dedekind's Theorem (thm.14.8). Hence there exists* $\gamma \in \mathbb{R}$ *such that* $\alpha \leqq \gamma \leqq \beta$ *for all* $\alpha \in A$ *and* $\beta \in B$.

We now show that that the assumption $\gamma \in A$ *leads to a contradiction. As* γ *is not an upper bound of* A *there exists* $x \in E$ *such that* $\gamma < x$. *According to thm.14.6 on p. 236 there exists* $\gamma' \in \mathbb{R}$ *such that* $\gamma < \gamma' < x$. *It follows that* $\gamma' \notin B$, *i.e.,* $\gamma' \in A$, *in contradiction to the fact that* $\gamma \geqq a$ *for all* $a \in A$.

It follows that $\gamma \notin A$, *i.e.,* $\gamma \in B$ *and we conclude from cor.14.1 that* $\gamma = \min(B)$, *i.e.,* $\gamma = \sup(E)$. ■

14.5 Constructing the real numbers via Cauchy Sequences

This chapter was created after discussions with Nguyen-Phan Tam about teaching the Math 330 course: she plans to construct the real numbers from the rationals by means of equivalence classes of Cauchy sequences in $\mathbb{Q}$.

In the following we always assume that $i, j, k, m, n \in \mathbb{N}$, $\varepsilon, p, q, r, s, p_n, p_{i,j}, \dots \in \mathbb{Q}$, $x, y, z, x_n, x_{i,j}, \dots \in \mathbb{R}$.

- **a.** *def. convergence in* $\mathbb{Q}$: $\lim\limits_{n\to\infty} q_n = q \Leftrightarrow \forall$ *pos.* $\varepsilon \in \mathbb{Q}\ \exists\ N \in \mathbb{Q}$ *such that if* $n \geqq N$ *then* $|q_n - q| < \varepsilon$.
- **b.** *def. Cauchy seqs. in* $\mathbb{Q}$: $(q_n)_n$ *is Cauchy* $\Leftrightarrow \forall$ *pos.* $\varepsilon \in \mathbb{Q}\ \exists\ N \in \mathbb{Q}$ *such that if* $i, j \geqq N$ *then* $|q_i - q_j| < \varepsilon$.
- **c.** *Let* $\mathscr{C} := \{$ *all Cauchy sequences in* $\mathbb{Q}\}$. *For* $(q_n)_n, (r_n)_n$ *we define* $(q_n)_n \sim (r_n)_n$ *iff* $\lim\limits_{n\to\infty}(r_n - q_n) = 0$.

d. *Let $q \in \mathbb{Q}$ and $q_n := q \,\forall\, n$. Write q for $[(q_n)_n]$.*

e. *Let $\mathbb{R} := \mathscr{C}_{/\sim}$. Show that for $[(p_n)], [(q_n)] \in \mathscr{C}$ the operations $([(p_n)_n], [(q_n)_n]) \mapsto [(p_n + q_n)_n]$ and $([(p_n)_n], [(q_n)_n]) \mapsto [(p_n \cdot q_n)_n]$ are well defined (do not depend on the particular members chosen from the equivalence classes).*

f. *Let $[(p_n)_n] \neq 0$ (i.e., $\lim_n p_n \neq 0$), i.e., we may assume $p_n \neq 0$ for all n. Show $-[(q_n)_n] := [(-q_n)_n]$ and $[(p_n)_n]^{-1} := [(1/p_n)_n]$ are additive and multiplicative inverses*

g1. *Define $[(p_n)_n] < [(q_n)_n]$ iff $\exists\, \varepsilon > 0$ and $N \in \mathbb{N}$ such that $q_n - p_n \geqq \varepsilon \,\forall\, n \geqq N$.*

g2. *Define $[(p_n)_n] \leqq [(q_n)_n]$ iff $\forall\, \varepsilon > 0$ $exists$ $N \in \mathbb{N}$ such that $q_n - p_n \geqq -\varepsilon \,\forall\, n \geqq N$.*

g3. *show that $[(p_n)_n] < [(q_n)_n]$ iff $[(p_n)_n] \leqq [(q_n)_n]$ and $[(p_n)_n] \neq [(q_n)_n]$.*

h. *Show that $(\mathbb{R}, +, \cdot, <)$ satisfies the axioms of B/G ch.8 with the exception of the completeness axiom.*
*Easy to see this specific item: If $[(p_n)_n] > 0$ then there is $[(q_n)_n] > 0$ such that $[(q_n)_n] < [(p_n)_n]$: choose $\varepsilon > 0$ as in **g1** (remember: $\varepsilon \in \mathbb{Q}$) and set $q_n := \varepsilon/2$.*

i. *Embed $\mathbb{Q}$ into $\mathbb{R}$: $q \mapsto \bar{q} := [(q, q, q, \dots)]$.*

j. *Define limits and Cauchy sequences in $\mathbb{R}$ just as in **a** and **b**.*

k. *Let $(q_n)_n$ be Cauchy in $\mathbb{Q}$. Prove that $\bar{q_n} \to [(q_j)_j]$*

l. *Let $x_n \in \mathbb{R}$ such that $(x_n)_n$ is Cauchy in $\mathbb{R}$. With a density argument we find $q_n \in \mathbb{Q}$ such that $x_n \leqq \bar{q}_n \leqq x_n + 1/n$. Now show that **1.** $(q_n)_n$ is Cauchy and then **2.** $\lim_n x_n = [(q_n)_n]$.*

m. *Prove completeness according to B/G: If nonempty $A \subseteq \mathbb{R}$ is bounded above then its set of upper bounds U has a min: Let $Q_n := \{i/j \,:\, i, j \in \mathbb{Z}$ and $j \leqq n\}$. Let $U_n := U \cap Q_n$. Let $u_n := \min(U_n)$ (exists because $n \cdot U_n \subset Z$ is bounded below and has a min. Easy to see that u_n is Cauchy (in $\mathbb{Q}$ and, because distance(u_n, A) $\leqq 1/n$, $[(u_n)_n]$ is the least upper bound of A.*

*Proofs for **k** and **l** in particular and an entire section on constructing $\mathbb{R}$ from $\mathbb{Q}$ by means of equivalence classes of Cauchy sequences can be found in [7] Haaser/Sullivan: Real Analysis.*

15 Measure Theory ($\star$)

Introduction:

The following are the best known examples of measures ($a_j, b_j \in \mathbb{R}$):

$$\begin{aligned}
&\textit{Length}: \quad \lambda^1([a_1,b_1]) \ := \ b_1 - a_1,\\
&\textit{Area}: \quad \lambda^2([a_1,b_1]\times[a_2,b_2]) \ := \ (b_1-a_1)(b_2-a_2),\\
&\textit{Volume}: \quad \lambda^3([a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]) \ := \ (b_1-a_1)(b_2-a_2)(b_3-a_3).
\end{aligned}$$

Then there also are probability measures: $P\{$ a die shows a 1 or a $6\} = 1/3$.

We will explore in this chapter some of the basic properties of measures.

15.1 Basic Definitions

Definition 15.1 (Extended real functions)**.**

$$\overline{\mathbb{R}}_+ \ := \ \mathbb{R}_+ \cup \{+\infty\} \ = \ \{x \in \mathbb{R} : x \geqq 0\} \cup \{+\infty\}$$

is the set of all non–negative real numbers augmented by the element ∞.

A mapping F whose codomain is a subset of

$$\overline{\mathbb{R}} \ := \ \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$$

is called an **extended real function**.

There are many issues with functions that allow some arguments to have infinite value (hint: if $F(x) = \infty$ and $F(y) = \infty$, what is $F(x) - F(y)$?)

We only list the following rule which might come unexpected to you:

$$0 \cdot \pm\infty \ = \ \pm\infty \cdot 0 \ = \ 0 \tag{15.1}$$

Definition 15.2 (Abstract measures)**.** Let Ω be a non–empty set and let $\mathfrak{F}$ be a set that contains some, but not necessarily all, subsets of Ω. $\mathfrak{F}$ is called a σ**–algebra** or σ**–field** for Ω if it satisfies the following:

$$\emptyset \in \mathfrak{F} \quad \text{and} \quad \Omega \in \mathfrak{F} \tag{15.2a}$$

$$A \in \mathfrak{F} \implies \complement A \in \mathfrak{F} \tag{15.2b}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F} \quad \text{and} \quad \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F} \tag{15.2c}$$

The pair $(\Omega, \mathfrak{F})$ is called a **measurable space**. Note that $\mathfrak{F}$ is a set whose elements themselves are sets! The elements of $\mathfrak{F}$ are called **measurable sets**.

A **measure** on $\mathfrak{F}$ is an extended real function

$$\mu(\cdot) : \mathfrak{F} \to \overline{\mathbb{R}}_+ \qquad A \mapsto \mu(A)$$

with the following properties:

(15.3a) $\mu(\emptyset) \;=\; 0$

(15.3b) $A, B \in \mathfrak{F}$ and $A \subseteq B \implies \mu(A) \leqq \mu(B)$ (monotony)

(15.3c) $(A_n)_{n\in\mathbb{N}} \in \mathfrak{F}$ mutually disjoint $\implies \mu\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) \;=\; \sum_{n\in\mathbb{N}} \mu(A_n)$ (σ–additivity)

where mutual disjointness means that $A_i \cap A_j = 0$ for any $i, j \in \mathbb{N}$ such that $i \neq j$ (see def.2.4 on p.12). The triplet $(\Omega, \mathfrak{F}, \mu)$ is called a **measure space**

We call μ a **finite measure** on $\mathfrak{F}$ if $\mu(\Omega) < \infty$.

A measure space can support many different measures.

If $\mu(\Omega) = 1$ then $\mu(\cdot)$ is called a **probability measure**. Traditionally, mathematicians write $P(A)$ rather than $\mu(A)$ for probability measures and the elements of $\mathfrak{F}$ (the measurable subsets) are thought of as **events** for which $P(A)$ is interpreted as the probability with which the event A might happen.

Example 15.1 (Lebesgue measure)**.** The most important measures we encounter in real life are those that measure the length of sets in one dimension, the area of sets in two dimensions and the volume of sets in three dimensions. Given intervals $[a, b] \in \mathbb{R}$, rectangles $[a_1, b_1] \times [a_2, b_2] \in \mathbb{R}^2$, boxes or quads $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \in \mathbb{R}^3$ and n**-dimensional parallelepipeds** $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \in \mathbb{R}^n$, we define

$$
\begin{aligned}
\lambda^1([a, b]) \;&:=\; b - a,\\
\lambda^2([a_1, b_1] \times [a_2, b_2]) \;&:=\; (b_1 - a_1)(b_2 - a_2),\\
\lambda^3([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) \;&:=\; (b_1 - a_1)(b_2 - a_2)(b_3 - a_3),\\
\lambda^n([a_1, b_1] \times \cdots \times [a_n, b_n]) \;&:=\; (b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)
\end{aligned} \tag{15.4}
$$

It can be shown that any measure that is defined on all parallelepipeds in $\mathbb{R}^n$ can be uniquely extended to a measure on the σ-algebra $\mathcal{B}^n$ generated by those parallelepipeds [105] λ^n is called n-**dimensional Lebesgue measure**

Note that Lebesgue measure is not finite.

Example 15.2. You can easily verify that the following set function defines a measure on an arbitrary non-empty set Ω with an arbitrary σ-field $\mathfrak{F}$.

$$\mu(\emptyset) \;:= 0; \qquad \mu(A) \;:=\; \infty \text{ if } A \neq \emptyset$$

Keep this example in mind if you contemplate infinity of measures.

Remark 15.1 (Finite disjoint unions)**.** The σ–additivity of measures is what makes working with them such a pleasure in many ways. You can now express it as follows: Given any mutually disjoint sequence of measurable sets, the measure of the disjoint union is the sum of the measures. The last

[105] This is not entirely correct: we must demand that the measure is σ-finite, i.e., there are measurable sets with finite measure whose union is the entire space. Such is the case for Lebesgue measure: Let $A_k \;:= [-k, k]^n$. The union of those sets is $\mathbb{R}^k$ and $\lambda^n(A_K) = (2k)^n < \infty$.

property (15.2c) for σ–algebras is required for exactly that reason: you cannot take advantage of the σ–additivity of a measure μ if its domain does not contain countable unions and intersections of all its constituents.

Note that if we have only finitely many sets then "σ–additivity" which stands for "additivity of countably many" becomes simple additivity. We obtain the following by setting $A_{N+1} = A_{N+2} = \ldots = 0$:

$$\begin{aligned} &A_1, A_2, \ldots, A_N \in \mathfrak{F} \text{ mutually disjoint} \\ &\quad \Rightarrow \mu(A_1 \uplus A_2 \uplus \ldots \uplus A_N) = \mu(A_1) + \mu(A_2) + \ldots + \mu(A_N) \quad \text{(additivity).} \end{aligned} \tag{15.5}$$

In the case of only two disjoint measurable sets A and B the above simply becomes

$$\mu(A \uplus B) = \mu(A) + \mu(B).$$

In many circumstances you have a set function on a σ–algebra which behaves like a measure but you can only prove that it is additive instead of σ–additive. You should not be surprised that there is a special name for those "generalized measures":

Definition 15.3 (Contents as additive measures)**.** Let Ω be a non–empty set and let $\mathfrak{F}$ be a σ–algebra for Ω.

A **content** on $\mathfrak{F}$ is a real function $m(\cdot) : \mathfrak{F} \to \mathbb{R}$, $A \mapsto m(A)$ which satisfies

(15.6a) $m(\emptyset) = 0$ (positivity)

(15.6b) $A, B \in \mathfrak{F}$ and $A \subseteq B \Rightarrow m(A) \leqq m(B)$ (monotony)

(15.6c) $A_1, A_2, \ldots, A_N \in \mathfrak{F}$ mutually disjoint $\Rightarrow m\Big(\biguplus_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} m(A_n)$ (additivity).

Note that $\mu(\Omega) < \infty$ for a content μ. After this digression on contents let us go back to measures.

Proposition 15.1 (Simple properties of measures)**.** ***A*** : *Let $A, B, \in \mathfrak{F}$ and let μ be a measure on $\mathfrak{F}$. Then*

(15.7a) $\mu(A) \geqq 0$ *for all* $A \in \mathfrak{F}$

(15.7b) $A \subseteq B \Rightarrow \mu(B) = \mu(A) + \mu(B \setminus A)$

(15.7c) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$

If μ is finite then we can write

(15.8a) $A \subseteq B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$

(15.8b) $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$

*Proof of **A** : The first property follows from the fact that $\mu(\emptyset) = 0$, $\emptyset \subseteq A$ for all $A \in \mathfrak{F}$ and (15.3b). To prove the second property, observe that $B = A \uplus (B \setminus A)$.*

Proving the third property is more complicated because neither A nor B may be a subset of the other. We first note that because $A \setminus B \subseteq A$, $B \setminus A \subseteq A$ and $A \cap B \subseteq A$, $\mu(A \cup B) = \infty$ can only be true if $\mu(A) = \infty$ or $\mu(B) = \infty$. In this case (15.7c) is obviously true. Hence we may assume that $\mu(A \cup B) < \infty$. We have

$$A \cup B = (A \cap B) \uplus (B \setminus A) \uplus (A \setminus B) \tag{15.9a}$$

$$A \cup B = A \uplus (B \setminus A) = B \uplus (A \setminus B) \tag{15.9b}$$

It follows from (15.9a) that

$$\mu(A \cup B) = \mu(A \cap B) + \mu(B \setminus A) + \mu(A \setminus B) \tag{15.10}$$

It follows from (15.9b) that

$$2 \cdot \mu(A \cup B) = \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) \tag{15.11}$$

We subtract the left and right sides of (15.10) *from those of* (15.11) *and obtain*

$$\begin{aligned} \mu(A \cup B) &= \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) - \mu(A \cap B) - \mu(B \setminus A) - \mu(A \setminus B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B) \end{aligned}$$

and the third property is proved. ■

Proposition 15.2 (Minimal sigma–algebras)**.** *Let Ω be a non–empty set.*

A: *The intersection of arbitrarily many σ–algebras is a σ–algebra.*

B: *Let $\mathfrak{K}$ be a set which contains subsets of Ω. It is not assumed that $\mathfrak{K}$ is a σ–algebra. Then there exists a σ–algebra which contains $\mathfrak{K}$ and is minimal in the sense that it is contained in any other σ–algebra that also contains $\mathfrak{K}$. We name this σ–algebra $\sigma(\mathfrak{K})$ because it clearly depends on $\mathfrak{K}$. It is constructed as follows:*

$$\sigma(\mathfrak{K}) = \bigcap \{\mathfrak{A} : \mathfrak{A} \supseteq \mathfrak{K} \text{ and } \mathfrak{A} \text{ is a } \sigma\text{–algebra for } \Omega\}. \tag{15.12}$$

*Proof of **A**:*
We must prove (15.2a), (15.2b) and (15.2c). Let $(\mathfrak{A}_\alpha)_\alpha$ be an arbitrary family of σ–algebras for Ω. Let

$$\mathfrak{A} := \bigcap_\alpha \mathfrak{A}_\alpha.$$

$\emptyset$ and Ω belong to each σ–algebra according to (15.2a). It follows that they both belong to the intersection $\bigcap_\alpha \mathfrak{A}_\alpha$, i.e., $\mathfrak{A}$ satisfies (15.2a). Let $A \in \mathfrak{A}$. Then $A \in \mathfrak{A}_\alpha$ for each α. $\complement A$ belongs to each σ–algebra according to (15.2b). It follows that $\complement A \in \bigcap_\alpha \mathfrak{A}_\alpha$, i.e., $\mathfrak{A}$ satisfies (15.2b). Finally, let $A_n \in \mathfrak{A}$ for all $n \in \mathbb{N}$. Then $A_n \in \mathfrak{A}_\alpha$ for all $n \in \mathbb{N}$ and for each α, $\cup_{n \in \mathbb{N}} A_n$ and $\bigcap_{n \in \mathbb{N}} A_n$ both belong to each σ–algebra according to (15.2c). It follows that they both belong to the intersection $\bigcap_\alpha \mathfrak{A}_\alpha$, i.e., $\mathfrak{A}$ satisfies (15.2c). It follows that $\mathfrak{A}$ is a σ–algebra.

*Proof of **B**:*
*First of all, we know that $\sigma(\mathfrak{K})$ is an intersection of σ–algebras and, according to part **A** of this proposition, really is a σ–algebra. We now prove that $\sigma(\mathfrak{K})$ contains $\mathfrak{K}$ and is the minimal σ–algebra with that property. First let us prove that $\sigma(\mathfrak{K}) \supseteq \mathfrak{K}$. But that is obvious because it is the intersection of sets all of which contain $\mathfrak{K}$. On the other hand, $\sigma(\mathfrak{K})$ is the intersection of **all** σ–algebras that contain $\mathfrak{K}$, so it is impossible for any other σ–algebra to both be a strict subset of $\sigma(\mathfrak{K})$ and also contain $\mathfrak{K}$.* ■

15.2 Sequences of sets - limsup and liminf

Assumption 15.1 (Existence of a universal set)**.** We assume the existence of a set X which contains all sets A_n, B_n, C_n that are used here in sequences.

Definition 15.4 (Monotone set sequences)**.** A sequence A_k of arbitrary subsets of X is called

(15.13a) **non–decreasing** if $A_1 \subseteq A_2 \subseteq \dots$

(15.13b) **non–increasing** if $A_1 \supseteq A_2 \supseteq \dots$

(15.13c) **strictly increasing** if $A_1 \subsetneq A_2 \subsetneq \dots$

(15.13d) **strictly decreasing** if $A_1 \supsetneq A_2 \supsetneq \dots$

Each one of those sequences is called a **monotone set sequence**.

Might as well define limits of monotone sequences of sets. It's certainly intuitive enough:

Definition 15.5 (Limits of monotone set sequences)**.** Given are sets $A_n, B_n \subseteq X$ $(n \in \mathbb{N})$. Assume that

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \qquad \text{and let } A := \bigcup_{k \in \mathbb{N}} A_k$$

$$B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots \qquad \text{and let } B := \bigcap_{k \in \mathbb{N}} B_k$$

We say that A is the limit of the sequence $(A_j)_{j \in \mathbb{N}}$ and B is the limit of the sequence $(B_j)_{j \in \mathbb{N}}$ and we write

(15.14a) $$A = \lim_{n \to \infty} A_n \quad \text{or} \quad A_n \nearrow A \text{ for } n \to \infty$$

(15.14b) $$B = \lim_{n \to \infty} B_n \quad \text{or} \quad B_n \searrow B \text{ for } n \to \infty$$

The above are not terribly useful definitions. What does it matter whether we write $A = \lim\limits_{n \to \infty} A_n$ *or* $A = \cup_{k \in \mathbb{N}} A_k$*? Things would be very different if we went further and defined limits of sequences of sets. Doing so is at the very beginning of a branch of Mathematics called Measure Theory and its (slightly) more applied version, Abstract Probability Theory.*

Definition 15.6 (lim inf and lim sup of set sequences)**.** Given are sets $A_n, B_n \subseteq X$ $(n \in \mathbb{N})$. Let

(15.15) $$\liminf_{n \to \infty} A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{k \geqq n} A_k \quad \textbf{(limit inferior)}$$

(15.16) $$\limsup_{n \to \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \geqq n} A_k \quad \textbf{(limit superior)}$$

In general those two will not coincide. But if they do then we define

(15.17) $$\lim_{n \to \infty} A_n := \liminf_{n \to \infty} A_n = \limsup_{n \to \infty} A_n$$

We call $\lim\limits_{n\to\infty} A_n$ the **limit** of the sequence (A_n) and we write

$$A_n \to A \quad \text{for } n \to \infty$$

The following comments should make matters easier to understand if you abbreviate

Lemma 15.1 (lim inf and lim sup as monotone limits)**.** *Given are sets* $A_n, B_n \subseteq X$ $(n \in \mathbb{N})$. *Let*

(15.18) $$A_{\star n} := \bigcap_{k \geqq n} A_k \qquad \textit{Then} \quad A_{\star n} \nearrow \liminf_{n\to\infty} A_n$$

(15.19) $$A^{\star}{}_n := \bigcup_{k \geqq n} A_k \qquad \textit{Then} \quad A^{\star}{}_n \searrow \limsup_{n\to\infty} A_n$$

Proof: Let $m, n \in \mathbb{N}$ *such that* $m < n$. *Then*

$$\begin{aligned} A_{\star m} &= \bigcap_{k=m}^{n-1} A_k \cap \bigcap_{k \geqq n} A_k &= \bigcap_{k=m}^{n-1} A_k \cap A_{\star n} &\subseteq A_{\star n} \\ A^{\star}{}_m &= \bigcup_{k=m}^{n-1} A_k \cup \bigcup_{k \geqq n} A_k &= \bigcup_{k=m}^{n-1} A_k \cup A^{\star}{}_n &\supseteq A^{\star}{}_n \end{aligned}$$

This proves that $A_{\star n}$ *is non–decreasing and* $A * \star_n$ *is non–increasing. By the very definition of the limit of a monotone sequence of sets it is true that*

$$\begin{aligned} \lim_{n\to\infty} A_{\star n} &= \bigcup_{n \in \mathbb{N}} A_{\star n} = \liminf_{n\to\infty} A_n \\ \lim_{n\to\infty} A^{\star}{}_n &= \bigcap_{n \in \mathbb{N}} A^{\star}{}_n = \limsup_{n\to\infty} A_n \end{aligned}$$

■

15.3 Conditional expectations as generalized averages

16 Appendices

16.1 Greek Letters

The following section lists all greek letters that are commonly used in mathematical texts. You do not see the entire alphabet here because there are some letters (especially upper case) which look just like our latin alphabet letters. For example: $A =$ Alpha $B =$ Beta. On the other hand there are some lower case letters, namely epsilon, theta, sigma and phi which come in two separate forms. This is not a mistake in the following tables!

α	*alpha*	θ	*theta*	ξ	*xi*	ϕ	*phi*
β	*beta*	ϑ	*theta*	π	*pi*	φ	*phi*
γ	*gamma*	ι	*iota*	ρ	*rho*	χ	*chi*
δ	*delta*	κ	*kappa*	ϱ	*rho*	ψ	*psi*
ϵ	*epsilon*	$\varkappa$	*kappa*	σ	*sigma*	ω	*omega*
ε	*epsilon*	λ	*lambda*	ς	*sigma*		
ζ	*zeta*	μ	*mu*	τ	*tau*		
η	*eta*	ν	*nu*	υ	*upsilon*		

Γ	*Gamma*	Λ	*Lambda*	Σ	*Sigma*	Ψ	*Psi*
Δ	*Delta*	Ξ	*Xi*	Υ	*Upsilon*	Ω	*Omega*
Θ	*Theta*	Π	*Pi*	Φ	*Phi*		

16.2 Notation

This appendix on notation has been provided because future additions to this document may use notation which has not been covered in class. It only covers a small portion but provides brief explanations for what is covered.

For a complete list check the list of symbols and the index at the end of this document.

Notations 16.1. a) If two subsets A and B of a space Ω are disjoint, i.e., $A \cap B = \emptyset$, then we often write $A \uplus B$ rather than $A \cup B$ or $A + B$. Both $A^{\complement}$ and, occasionally, $\complement A$ denote the complement $\Omega \setminus A$ of A.

b) $\mathbb{R}_{>0}$ or $\mathbb{R}^{+}$ denotes the interval $]0, +\infty[$, $\mathbb{R}_{\geqq 0}$ or $\mathbb{R}_{+}$ denotes the interval $[0, +\infty[$,

c) The set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of all natural numbers excludes the number zero. We write $\mathbb{N}_0$ or $\mathbb{Z}_{+}$ or $\mathbb{Z}_{\geqq 0}$ for $\mathbb{N} \uplus \{0\}$. $\mathbb{Z}_{\geqq 0}$ is the B/G notation. It is very unusual but also very intuitive.

Definition 16.1. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. We call that sequence **non-decreasing** or **increasing** if $x_n \leqq x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **strictly increasing** if $x_n < x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **non-increasing** or **decreasing** if $x_n \geqq x_{n+1}$ for all n.

We call it **strictly decreasing** if $x_n > x_{n+1}$ for all $n \in \mathbb{N}$.

References

[1] *Matthias Beck and Ross Geoghegan.* The Art of Proof. *Springer, 1st edition, 2010.*

[2] *Regina Brewster and Ross Geoghegan.* Business Calculus, a text for Math 220, Spring 2014. *Binghamton University, 11th edition, 2014.*

[3] *Matthew Brin and Gerald Marchesi.* Linear Algebra, a text for Math 304, Spring 2016. *Binghamton University, 13th edition, 2016.*

[4] *John Bryant and Penelope Kirby.* Course Notes for MAD 2104 Discrete Mathematics I. *Florida State University.*

[5] *Gustave Choquet.* Lectures on Analysis, Vol. 2: Representation Theory. *Benjamin, New York, 1st edition, 1969.*

[6] *Richard M. Dudley.* Real Analysis and Probability. *Cambridge University Press, Cambridge, New York, 2nd edition, 2002.*

[7] *Norman B. Haaser and Joseph A. Sullivan.* Real Analysis. *publisher = Van Nostrand, year = 1971, 1st edition.*

[8] *A.N. Kolmogorov and S.V. Fomin.* Introductory Real Analysis. *Dover, Mineola, 1st edition, 1975.*

[9] *James R. Munkres.* Topology. *Prentice-Hal, 1st edition, 2000.*

[10] *Walter Rudin.* Principles of Mathematical Analysis. *McGraw-Hill, New York, San Francisco, Toronto, London, 2nd edition, 1964.*

[11] *James Stewart.* Single Variable Calculus. *Thomson Brooks Cole, 7th edition, 2012.*

[12] *Boto Von Querenburg.* Mengentheoretische Topologie. *Springer, Berlin, Heidelberg, New York, 1st edition, 1973.*

[13] *Richard E. Williamson, Richard H. Crowell, and Hale F. Trotter.* Calculus of Vector Functions. *Prentice Hall, Englewood Cliffs, 3rd edition, 1972.*

List of Symbols

Index