Math 330 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

Last update: November 25, 2024

Contents

1 Before You Start

Errors detected by Math 330 students, Spring 2017:

Date	Topic
2024-11-25	Error in Introduction 2.3**(d)**. **Incorrect version:** function $f(x) : D \to C$ has domain C and codomain D. **Correct version:** function $f(x) : D \to C$ has domain D and codomain C. Detected by **Ryan Wang**.
2017-01-26	Error in Definition 5.2. **Incorrect version:** A relation is symmetric if $x_1 R x_2$ implies $x_1 R x_2$ for all $x_1, x_2 \in X$. **Correct version:** A relation is symmetric if $x_1 R x_2$ implies $x_2 R x_1$ for all $x_1, x_2 \in X$. Detected by **Brad Whistance**.

1.1 About This Document

Remark 1.1 (The purpose of this document)**.** The original version of this document was written in 2005 under the title "Introduction to Abstract Math – A Journey to Approximation Theory". Since then parts of it were discarded and others have been added. It now serves as lecture notes for the course "Math 330: Number systems" which is held at the Department of Mathematical Sciences at Binghamton University. Parts of the remainder of this chapter are specifically addressed to the students of this course.

These notes serve at least two purposes:

(a) They contain material on topics that cannot found in sufficient detail or generality in the textbook [2] Beck/Geoghegan: The Art of Proof. That book serves as the primary reference for the first two thirds of the Math 330 course. It is often simply referred to as "B/G" in these notes.

(b) This document covers material which is beyond the scope of [2] 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.

Prof. Geoghegan has graciously given permission to let me copy definitions, proofs and theorems verbatim from this text. I have indicated for such items how they are referenced there. An example is, e.g., proposition 3.20 on p.66 of this document which is stated here for integral domains and shows its origin by the references B/G prop.1.13 and B/G prop.8.14. No proof is given in the B/G student edition for this proposition, and in all such circumstances I too do not furnish a proof to the student unless I have one that is quite different from the one to be found in the instructor's edition. □

Remark 1.2 (Acknowledgements)**.** Chapters 2 and 4 of this document draw on [5] Bryant, Kirby Course Notes for MAD 2104. Moreover such a document cannot be written with the intent to supplement the [2] B/G book without strongly borrowing from it. □

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

a. “Understand this” directive: When you read “Understand this”, you should know the definitions, propositions and theorems without worrying about proofs. It is quite likely that this kind of material will be referenced in more important sections of this document. As of May 2022 only the chapter on logic contains this directive.

b. “ ★ ” directive: Chapters marked “ ★ ” are optional. The student need not worry about learning the material, although it may be referenced in the non optional chapters if doing so benefits the interested reader. This symbol is also used to indicate that a certain statement or maybe just its proof can be skipped. Again, it may be referenced in the non optional chapters to provide some background for the interested reader. Moreover certain definitions are tagged with this symbol, but the reader should NOT skip those definitions: My students are not expected to give precise equivalents of such definitions in quizzes and exams, but they will need to know where to find them to do their homework and to make sense of the propositions and theorems and their proofs.

Notation Alert: All directives discussed above apply to the entire subtree, and a lower level directive overrides the “parent directives”. 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.

Homework: 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, also those tagged with a “ ★ ” symbol, so that I can move along at a fast pace except for some topics that I will focus on in detail.

We use colored boxes like the two above according to the following. Generally speaking,

These boxes contain important definitions or parts thereof.

These boxes contain important theorems and proposiitions or parts thereof.

These boxes contain other kinds of important items that are worth while to know. □

There are definitions and theorems that contain two or even three small boxes rather than a big one. There is a technical reason: such boxes do not span pages and will needlessly inflate the page count of the docum

Remark 1.4 (How to navigate this document II)**.** I believe that, particularly in Math, more words take a lot less time to understand than a skeletal write-up like one often finds in the [2] B/G text. Accordingly, almost all of the material in this document comes with quite detailed proofs. Those proofs are there for you to study.

Some of those proofs, notably those in prop. 8.4 on p.240, make use of “⇔” 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 “⇔” between statements in your proofs. You very likely lack the experience to use this technique without errors.

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 a younger audience of 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 brief chapter to understand some of the dos and don'ts when submitting your homework.

To prove the validity of an equation such as $A = Z$, do one of the following:

Method a.

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

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

Transitivity of equality means that if A = B and B = C then A = C.

Method 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:

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

Right side:

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

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

⚠

You are not allowed to structure your proof that A = Z as follows.

Method c.

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

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 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 that this is so. If this equation is wrong then doing the same thing to both of its sides will potentially lead to a true equation.

Here is a simple example that demonstrates why method c **is not allowed** We will use this method in two different ways to prove that $-2 = 2$.

First proof that $-2 = 2$:

$$\begin{aligned} -2 &= 2 \qquad \text{(want to prove)} \\ -2 \cdot 0 &= 2 \cdot 0 \quad \text{(multiply both sides from the right w. } 0\text{)} \\ 0 &= 0 \qquad \text{(anything times zero = zero)} \end{aligned}$$

We are done. ■

Second proof that $-2 = 2$:

$$\begin{aligned} -2 &= 2 \quad \text{(want to prove)} \\ (-2)^2 &= 2^2 \quad \text{(square both sides)} \\ 4 &= 4 \quad \text{(minus times minus = plus)} \end{aligned}$$

We are done. ■

Now you know why you must never use "method c" for a proof. [1]

[1] You will learn later in this document 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 about Sets, Numbers and Functions

Introduction 2.1. This document strives for mathematically exact definition of proofs, but not in this chapter. Here we want to provide the reader with a refresher of some material that a student with an interest in mathematics should have previously encountered in beginner's calculus, or even in high school. Examples are union, intersection, and inclusion of sets, integers, rational and real numbers, functions $y = f(x)$ of a real–valued variable x with a real–valued function value y, and some facts about differentiation and integration.

Much of this material will be given again in later chapters, with an accuracy that is satisfactory to a mathematician, so why waste some effort here? For example, you will find in this chapter a preliminary definition of the real numbers. See page 25. You will have to wait for the real thing (no pun intended) until p.255 of chapter 9!

We will be careful not to use those preliminary definitions before we reach the exact ones when developing the general theory, unless the instances where we do so will be covered a second time with precision. Examples for this are the preliminary definitions of the integers, the rational numbers, and the real numbers. However we will use the concepts defined here in examples, in clarifying remarks, and in exercises to help the student achieve a better understanding of the material.

The most important concept which we use before it is properly defined is that of "finitely many". We all know what it means, but do we know it in such a way that it can be used for the study of abstract math? This author does not think so, and the proper definition of finiteness is deferred until definition 7.1 on p.214, at the beginning of ch.7.1. Thus you will see examples of infinite sets and we will use phrases like "finitely many" and "infinitely many" in examples and preliminary definitions, but we will avoid using the concepts of finiteness and infiniteness when developing the general theory. □

The students should read this chapter carefully, with the expectation that it contains material that they are not familiar with, as much of it will be used in lecture without comment. Very likely candidates are power sets, a function $f : X \to Y$ where domain X and codomain Y are part of the definition.
We do not expect that the student has a background in proofs by induction and definitions by recursion. Those concepts are introduced here as tools, and the student is expected to familiarize herself/himself with those techniques before the mathematical underpinnings are provided in chapter 6.1 (The Integers, the Induction Axiom, and the Induction Principles) on p.167.

2.1 Sets and Basic Set Operations

Introduction 2.2. This first subchapter of ch.2 is different from the following ones in that the treatment of sets given here is sufficiently exact for a PhD in math unless s/he works in the areas of logic or axiomatic set theory. The only exception is the end of the chapter where the preliminary definition of the size of a set (Definition 2.12 on p.23) needs to refer to finiteness.

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:

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

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 we use instead **setbuilder notation**, i.e., instructions that specify what belongs to the set.

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

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

[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. Therefore, examples as the one above have spawned 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. □

We sometimes refer in the examples to the sets of numbers $\mathbb{N}$ (natural numbers), $\mathbb{Z}$ (integers), $\mathbb{R}$ (real numbers). If you are not familiar with those set please review briefly Definitions 2.13 and 2.14 at the start of section 2.3 (Numbers). This will come in handy for understanding the following example which demonstrates that some or all elements of a set can be sets themselves.

Example 2.2.

(a) $\mathscr{A} := \{\,]a,b[: a,b \in \mathbb{R}, 0 < b-a < 2\}$ is the set of all open intervals of length less than 2

(b) $\mathscr{B} := \{K$:K is a set of integers $\}$ We will later refer to $\mathscr{B}$ as the power set of $\mathbb{Z}$. [3] □

Definition 2.2 (empty set)**.** $\emptyset$ or $\{\,\}$ denotes the **empty set**. [4] 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 all $x, y, z \in \emptyset$ 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 a set. If $x \in \emptyset$ then $x \in A$.

[3] See Definition 2.10 on p.22.

[4] We discourage the use of { } since it makes expressions with nested braces hard to read.

As you will learn later, **(c1)+(c2)+(c3)** means that "$\sim$" is an equivalence relation (see Definition 5.3 on p.128) and **(d)** means that the empty set is a subset (see the next definition) of all sets, including itself. □

Definition 2.3 (subsets, supersets and equality of sets)**.**

(a) We say that a set A is a **subset** of the set B and we write $A \subseteq B$ if each 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 all sets A.

(b) If $A \subseteq B$ but $A \neq B$, i.e., there is at least one $x \in B$ such that $x \notin A$, then we say that A is a **strict subset** or a **proper subset** of B. We write "$A \subsetneq B$" or "$A \subset B$". [5] Alternatively we say that B is a **strict superset** or a **proper superset** of A and we write "$B \supsetneq A$") or "$B \supset A$".

(c) We say that two sets A and B are **equal** if both $A \subseteq B$ and $B \subseteq A$ □

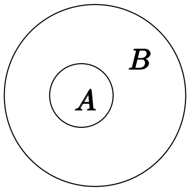


Figure 2.1: Set inclusion: $A \subseteq B, \ B \supseteq A$

Note that $A = B$ iff each element of A also belongs to B and each element of B also belongs to A iff both A and B contain the same elements. Here "**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. [6]

> To show that two sets A and B are equal you show that
> **(a)** if $x \in A$ then $x \in B$,
> **(b)** if $x \in B$ then $x \in A$.

Definition 2.4 (Unions and intersections of two sets)**.** Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or that either one is not empty!

(a) 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. [7]

[5] We try to avoid the notation "$A \subset B$"

[6] A formal definition of "if and only if" will be given in Definition 4.10 on p.93 where we will also introduce the symbolic notation $P \Leftrightarrow Q$. Informally speaking, a statement is something that is either true or false.

[7] We could have shortened the phrase "all elements which belong to A or B or both" to "all elements which belong to A or B", and we will almost always do so because it is understood among mathematicians that "or" always means at least one of the choices. If they mean instead exactly one of the choices $\#1, \#2, \dots \#n$ then they will use the phrase "either #1 or #2 or … or $\#n$. See rem3.3 on p.63. We will also see in a moment that there is a special symbol $A \triangle B$ which denotes the items which belong to either A or B (but not both).

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

It is obvious how to define unions and intersections of more than two sets: If $A_1, A_2, \dots, A_n$ is a collection of n sets then we define

Definition 2.5 (Unions and intersections of n sets)**.** Let $A_1, A_2, \dots, A_n$ be arbitrary sets.

(a) The **union** $\bigcup\limits_{j=1}^{n} A_j \;:=\; A_1 \cup A_2 \cup \dots \cup A_n$ is defined as the set of all those items which belong to at least one of the sets, i.e.,

$$x \in \bigcup_{j=1}^{n} A_j \;\Leftrightarrow\; x \in A_j \text{ for at least one index } j. \tag{2.2}$$

(b) The **intersection** $\bigcap\limits_{j=1}^{n} A_j \;:=\; A_1 \cap A_2 \cap \dots \cap A_n$ is defined as the set of all those items which belong to each and everyone of the sets, i.e.,

$$x \in \bigcap_{j=1}^{n} A_j \;\Leftrightarrow\; x \in A_j \text{ for each index } j. \;\square \tag{2.3}$$

Definition 2.6 (Disjoint unions)**.** We call two sets A and B **disjoint** , also **mutually disjoint**, , if $A \cap B = \emptyset$. More generally, we say that a collection of sets $A_1, A_2, \dots, A_n$ is (mutually) disjoint if each pair A_i, A_j for different indices i and j is disjoint. We often write "$\uplus$" (pronounced "disjoint union") rather than "$\cup$" to remind the reader that we are dealing with unions of disjoint sets, i.e., we write

$$A \uplus B \qquad A_1 \uplus A_2 \uplus \dots \uplus A_n\,, \qquad \biguplus_{j=1}^{n} A_j\,,$$

rather than $A \cup B,\; A_1 \cup A_2 \cup \dots \cup A_n,\; \bigcup\limits_{j=1}^{n} A_j$. □

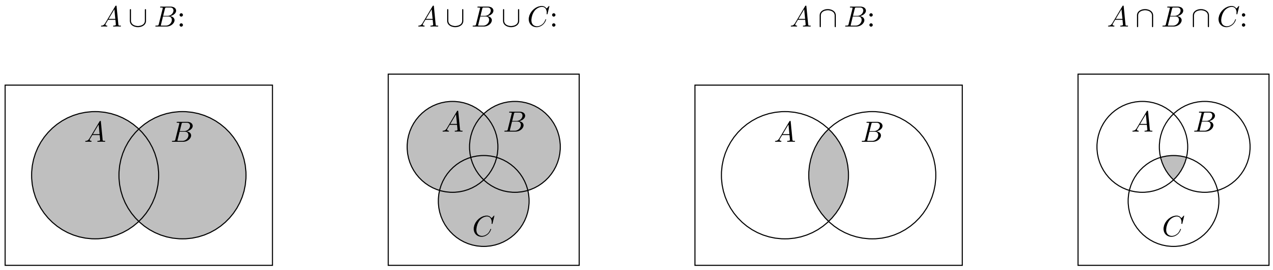


Figure 2.2: Union and intersection of sets

Remark 2.4. It is obvious from the definition of unions and intersections and the meaning of the phrases " all elements which belong to A or B or both", "all elements which belong to both A and B" and "$A \subseteq B$ if each element of A also belongs to B" that the following is true for arbitrary sets A, B and C.

$$A \cap B \subseteq A \subseteq A \cup B, \tag{2.4}$$

$$A \subseteq B \Rightarrow A \cap B = A \text{ and } A \cup B = B, \tag{2.5}$$

$$A \subseteq B \Rightarrow A \cap C \subseteq B \cap C \text{ and } A \cup C \subseteq B \cup C. \tag{2.6}$$

The symbol $\Rightarrow$ stands for "allows us to conclude that". So $A \subseteq B \Rightarrow A \cap B = A$ means "From the truth of $A \subseteq B$ we can conclude that $A \cap B = A$ is true". Shorter: "From $A \subseteq B$ we can conclude that $A \cap B = A$". Shorter: "If $A \subseteq B$ then it follows that $A \cap B = A$". Shorter: "If $A \subseteq B$ then $A \cap B = A$". More technical: "$A \subseteq B$ implies $A \cap B = A$".

You will learn more about implication in ch.4 of this document and in ch.3 (Some Points of Logic) of [2] Beck/Geoghegan: The Art of Proof. □

Definition 2.7 (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 is not empty!

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:

$$A \setminus B := \{x \in A : x \notin B\} \tag{2.7}$$

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:

$$A \triangle B := (A \cup B) \setminus (A \cap B) \quad \square \tag{2.8}$$

Definition 2.8 (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 $\mathbb{R}$. [8] If there is a universal set, it makes perfect sense to talk about the complement of a set:

Definition 2.9 (Complement of a set)**.** Let Ω be a universal set. The **complement** of a set $A \subseteq \Omega$ consists of all elements of Ω which do not belong to A. We write $A^{\complement}$. or $\complement A$ In other words:

$$A^{\complement} := \complement A := \Omega \setminus A = \{\omega \in \Omega : x \notin A\} \quad \square \tag{2.9}$$

[8] $\mathbb{R}$ 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 ch.2.3 ("Classification of numbers") on p.24).

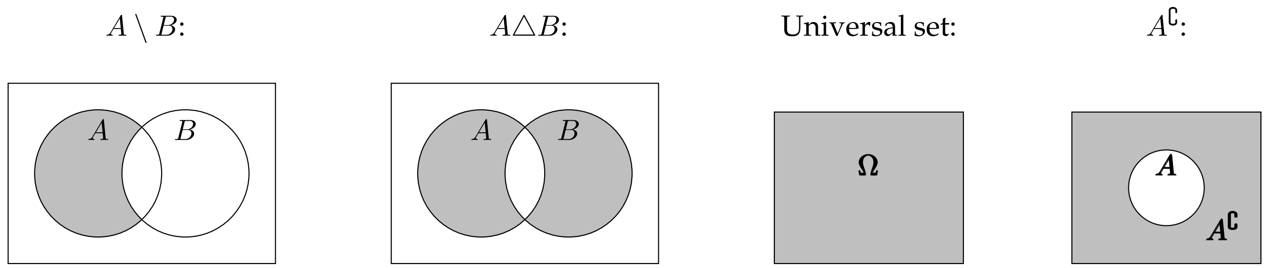

$A \setminus B$: $A \triangle B$: Universal set: $A^{\complement}$:

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

Remark 2.5. Note the following: If Ω is a universal set then

$$\Omega^{\complement} = \emptyset, \qquad \emptyset^{\complement} = \Omega. \ \square \tag{2.10}$$

Example 2.3 (Complement of a set relative to the unit interval)**.** Assume we are exclusively dealing with the unit interval, i.e., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$. Let $a \in [0,1]$ and $\delta > 0$ and

$$A = \{x \in [0,1] : a - \delta < x < a + \delta\} \tag{2.11}$$

the δ–neighborhood [9] 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

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

Draw some Venn diagrams to visualize the following formulas.

Proposition 2.1. *Let A, B, X be subsets of a universal set Ω and assume $A \subseteq X$. Then*

$$A \cup \emptyset = A; \qquad A \cap \emptyset = \emptyset \tag{2.12a}$$
$$A \cup \Omega = \Omega; \qquad A \cap \Omega = A \tag{2.12b}$$
$$A \cup A^{\complement} = \Omega; \qquad A \cap A^{\complement} = \emptyset \tag{2.12c}$$
$$A \triangle B = (A \setminus B) \uplus (B \setminus A) \tag{2.12d}$$
$$A \setminus A = \emptyset \tag{2.12e}$$
$$A \triangle \emptyset = A; \qquad A \triangle A = \emptyset \tag{2.12f}$$
$$X \triangle A = X \setminus A \tag{2.12g}$$
$$A \cup B = (A \triangle B) \uplus (A \cap B) \tag{2.12h}$$
$$A \cap B = (A \cup B) \setminus (A \triangle B) \tag{2.12i}$$
$$A \triangle B = \emptyset \ \textit{ if and only if } \ B = A \tag{2.12j}$$

[9] Neighborhoods of a point will be discussed in the chapter on the topology of $\mathbb{R}^n$ (see (12.6) on p.366). In short, the δ–neighborhood of a is the set of all points with distance less than δ from a.

PROOF: The proof is left as exercise 2.2. See p.47. ■

Next we give a very detailed and rigorous proof of a simple formula for sets. The reader should make an effort to understand it line by line.

Proposition 2.2 (Distributivity of unions and intersections for two sets)**.** *Let A, B, C be sets. Then*

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \tag{2.13}$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{2.14}$$

PROOF: We only prove (2.13). The proof of (2.14) is left as exercise 2.1.

PROOF of "$\subseteq$": Let $x \in (A \cup B) \cap C$. It follows from (2.4) on p.18 that $x \in (A \cup B)$, i.e., $x \in A$ or $x \in B$ (or both). It also follows from (2.4) that $x \in C$. We must show that $x \in (A \cap C) \cup (B \cap C)$ regardless of whether $x \in A$ or $x \in B$.

Case 1: $x \in A$. Since also $x \in C$, we obtain $x \in A \cap C$, hence, again by (2.4), $x \in (A \cap C) \cup (B \cap C)$, which is what we wanted to prove.

Case 2: $x \in B$. We switch the roles of A and B. This allows us to apply the result of case 1, and we again obtain $x \in (A \cap C) \cup (B \cap C)$.

PROOF of "$\supseteq$": Let $x \in (A \cap C) \cup (B \cap C)$, i.e., $x \in A \cap C$ or $x \in B \cap C$ (or both). We must show that $x \in (A \cup B) \cap C$ regardless of whether $x \in A \cap C$ or $x \in B \cap C$.

Case 1: $x \in A \cap C$. It follows from $A \subseteq A \cup B$ and (2.6) on p.18 that $x \in (A \cup B) \cap C$, and we are done in this case.

Case 2: $x \in B \cap C$. This time it follows from $A \subseteq A \cup B$ that $x \in (A \cup B) \cap C$. This finishes the proof of (2.13).

Epilogue: The proofs both of "$\subseteq$" and of "$\supseteq$" were **proofs by cases**, i.e., we divided the proof into several cases (to be exact, two for each of "$\subseteq$" and "$\supseteq$"), and we proved each case separately. For example we proved that $x \in (A \cup B) \cap C$ implies $x \in (A \cap C) \cup (B \cap C)$ separately for the cases $x \in A$ and $x \in B$. Since those two cases cover all possibilities for x the assertion "if $x \in (A \cup B) \cap C$ then $x \in (A \cap C) \cup (B \cap C)$" is proven. ■

Proposition 2.3 (De Morgan's Law for two sets)**.** *Let $A, B \subseteq \Omega$. Then the complement of the union is the intersection of the complements, and the complement of the intersection is the union of the complements:*

$$\textbf{\textit{(a)}}\ (A \cup B)^{\complement} = A^{\complement} \cap B^{\complement} \qquad \textbf{\textit{(b)}}\ (A \cap B)^{\complement} = A^{\complement} \cup B^{\complement} \tag{2.15}$$

PROOF of **(a)**:

(1) First we prove that $(A \cup B)^{\complement} \subseteq A^{\complement} \cap B^{\complement}$:

Assume that $x \in (A \cup B)^{\complement}$. Then $x \notin A \cup B$, which is the same as saying that x does not belong to either of A and B. That in turn means that x belongs to both $A^{\complement}$ and $B^{\complement}$ and hence also to the intersection $A^{\complement} \cap B^{\complement}$.

(2) Now we prove that $(A \cup B)^{\complement} \supseteq A^{\complement} \cap B^{\complement}$:

Let $x \in A^{\complement} \cap B^{\complement}$. Then x belongs to both $A^{\complement}, B^{\complement}$, hence neither to A nor to B, hence $x \notin A \cup B$. Therefore x belong to the complement of $A \cup B$. This completes the proof of formula **(a)**.

PROOF of **(b)**:

The proof is very similar to that of formula **(a)** and left as an exercise. ■

Formulas **(a)** through **(g)** of the next proposition are very useful. You are advised to learn them by heart and draw pictures to visualize them. You also should examine closely the proof of the next proposition. It shows how a proof which involves 3 or 4 sets can be split into easily dealt with cases.

Proposition 2.4. *Let A, B, C, Ω be sets such that $A, B, C \subseteq \Omega$. Then*

(a) $(A \triangle B) \triangle C \ = \ A \triangle (B \triangle C)$

(b) $A \triangle \emptyset \ = \ \emptyset \triangle A \ = \ A$

(c) $A \triangle A \ = \ \emptyset$

(d) $A \triangle B \ = \ B \triangle A$

Further we have the following for the intersection operation:

(e) $(A \cap B) \cap C \ = \ A \cap (B \cap C)$

(f) $A \cap \Omega \ = \ \Omega \cap A \ = \ A$

(g) $A \cap B \ = \ B \cap A$

And we have the following interrelationship between $\triangle$ and $\cap$:

(h) $A \cap (B \triangle C) \ = \ (A \cap B) \triangle (A \cap C)$

PROOF:

The proof of **(a)** is very tedious and there is a much more elegant proof, but that one requires knowledge of indicator functions [10] and of base 2 modular arithmetic (see, e.g., [2] B/G (Beck/Geoghegan) ch.6.2).

By definition $x \in U \triangle V$ if and only if either $x \in U$ or $x \in V$, i.e.,
(either) $\big[\ x \in U$ and $x \notin V\ \big]$ or $\big[\ x \in V$ and $x \notin U\ \big]$. Hence,

- $x \in (A \triangle B) \triangle C \ \Leftrightarrow$ either $x \in (A \triangle B)$ or $x \in C$
  $\Leftrightarrow$ either $\big[\ x \in A, x \notin B$ or $x \in B, x \notin A\ \big]$ or $x \in C$.
- $x \in A \triangle (B \triangle C) \ \Leftrightarrow$ either $x \in A$ or $x \in (B \triangle C)$
  $\Leftrightarrow$ either $\big[\ x \in B, x \notin C$ or $x \in C, x \notin B\ \big]$ or $x \in A$.

Thus, we obtain the following "truth table" for the eight possible combinations:

$x \in \dots$	A	B	C	$A \triangle B$	$B \triangle C$	$(A \triangle B \triangle C)$	$A \triangle (B \triangle C)$
	F	**F**	**F**	**F**	**F**	**F**	**F**
	F	**F**	**T**	**F**	**T**	**T**	**T**
	F	**T**	**F**	**T**	**T**	**T**	**T**
	F	**T**	**T**	**T**	**F**	**F**	**F**
	T	**F**	**F**	**T**	**F**	**T**	**T**
	T	**F**	**T**	**T**	**T**	**F**	**F**
	T	**T**	**F**	**F**	**T**	**F**	**F**
	T	**T**	**T**	**F**	**F**	**T**	**T**

We have a perfect match of set membership in the two rightmost columns. Thus,

$$x \in (A \triangle B \triangle C) \ \Leftrightarrow \ A \triangle (B \triangle C)\,.$$

This proves **(a)**.

The proofs of **(b)**, **(c)**, **(d)** are easy if you work with

$$U \triangle V \ = \ (U \setminus V) \uplus (V \setminus U) \ = \ (U \cap V^{\complement}) \uplus (V \cap U^{\complement})\,.$$

[10] Indicator functions will be discussed in ch.8.5 on p.246 and in ch.9.9 on p.296.

For example the proof of **(c)** is as follows.

$$A \triangle A \;=\; (A \cap A^{\complement}) \uplus (A \cap A^{\complement}) \;=\; \emptyset \uplus \emptyset \;=\; \emptyset\,.$$

The proofs of **(e)**, **(f)**, **(g)** are immediate. The proof of **(h)** can be done by cases, similarly to the proof of **(a)**. ■

Definition 2.10 (Power set)**.**

The **power set**

$$2^{\Omega} \;:=\; \{A : A \subseteq \Omega\}$$

of a set Ω is the set of all its subsets. Note that many older texts also use the notation $\mathfrak{P}(\Omega)$ for the power set. □

Example 2.4.

(a) $]3.2, 4.8[\, \in 2^{[3.2,4.8]}$ because $]3.2, 4.8[$ is a subset of $[3.2, 4.8]$,
but $[3.2, 4.8] \notin 2^{]3.2,4.8[}$ because $]3.2, 4.8[$ is not a subset of $[3.2, 4.8]$.

(b) Let $Z := \{\, 5.4, \{19\}, \pi \,\}$. Then

$$2^{Z} \;=\; \{\, \emptyset, \{5.4\}, \{\{19\}\}, \{\pi\}, \{5.4, \{19\}\}, \{5.4, \pi\}, \{\{19\}, \pi\}, \{5.4, \{19\}, \pi\} \,\}\,. \;\square$$

Remark 2.6. Note that $\emptyset \in 2^{\Omega}$ for all sets Ω, even if $\Omega = \emptyset$, since $2^{\emptyset} = \{\emptyset\}$. In particular, the power set of the empty set is not empty. □

Definition 2.11 (Partition)**.** Let Ω be a set and $\mathfrak{A} \subseteq 2^{\Omega}$, i.e., the elements of $\mathfrak{A}$ are subsets of Ω.

We call $\mathfrak{A}$ a **partition** or a **partitioning** of Ω if

(a) If $A, B \in \mathfrak{A}$ such that $A \neq B$ then $A \cap B = \emptyset$. In other words, $\mathfrak{A}$ consists of mutually disjoint subsets of Ω (see Definition 2.6),

(b) Each $x \in \Omega$ is an element of some $A \in \mathfrak{A}$. □

Remark 2.7. Let Ω be a set and $\mathfrak{A} \subseteq 2^{\Omega}$. Then $\mathfrak{A}$ is a partition of Ω if and only if

$$\text{For each } x \in \Omega, \text{ there exists a UNIQUE } A \in \mathfrak{A} \text{ such that } x \in A\,. \;\square$$

Example 2.5.

a. For $n \in \mathbb{Z}$ let $A_n := \{n\}$. Then $\mathfrak{A} := \{A_n : n \in \mathbb{Z}\}$ is a partition of $\mathbb{Z}$. $\mathfrak{A}$ is not a partition of $\mathbb{N}$ because not all its members are subsets of $\mathbb{N}$ and it is not a partition of $\mathbb{Q}$ or $\mathbb{R}$. The reason: $\frac{1}{2} \in \mathbb{Q}$ and hence $\frac{1}{2} \in \mathbb{R}$, but $\frac{1}{2} \notin A_n$ for any $n \in \mathbb{Z}$, hence condition **b** of def.2.11 is not satisfied.

b. For $n \in \mathbb{N}$ let $B_n := [\, n^2, (n+1)^2 [\, = \{x \in \mathbb{R} : n^2 \leq x < (n+1)^2\}$. Then $\mathfrak{B} := \{B_n : n \in \mathbb{N}\}$ is a partition of $[1, \infty[$. □

Definition 2.12 (Size of a set (preliminary))**.**

(a) Let X be a finite set, i.e., a set which only contains finitely many elements. We write $\left|X\right|$ for the number of its elements, and we call $\left|X\right|$ the **size** of the set X.

(b) For infinite, i.e., not finite sets Y, we define $|Y| := \infty$. □

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

2.2 The Proper Use of Language in Mathematics: Any vs All, etc

Mathematics must be very precise in its formulations. Such precision is achieved not only by means of symbols and formulas, but also by its use of the English language. We will list some important points to consider early on in this document.

2.2.0.1 All vs. ANY

Assume for the following that X is a set of numbers. Do the following two statements mean the same?

(1) It is true for ALL $x \in X$ that x is an integer.

(2) It is true for ANY $x \in X$ that x is an integer.

You will hopefully agree that there is no difference and that one could rewrite them as follows:

(3) ALL $x \in X$ are integers.

(4) ANY $x \in X$ is an integer.

(5) EVERY $x \in X$ is an integer.

(6) EACH $x \in X$ is an integer.

(7) IF $x \in X$ THEN x is an integer.

Is it then always true that ALL and ANY means the same? Consider

(8a) It is NOT true for ALL $x \in X$ that x is an integer.

(8b) It is NOT true for ANY $x \in X$ that x is an integer.

Completely different things have been said: Statement **(8)** asserts that as few as one item and as many as all items in X are not integers, whereas **(9)** states that no items, i.e., exactly zero items in X, are integers.

My suggestion: Express formulations like **(8b)** differently. You could have written instead

(8c) There is no $x \in X$ such that x is an integer.

2.2.0.2 AND vs. IF ... THEN

Some people abuse the connective AND to also mean IF ... THEN. However, mathematicians use the phrase "p AND q" exclusively to mean that something applies to both p and q. Contrast the use of AND in the following statements:

(9) "Jane is a student AND Joe likes baseball". This phrase means that both are true: Jane is indeed a student and Joe indeed likes baseball.

(10) "You hit me again AND you'll be sorry". **Never, ever use the word AND in this context!** A mathematician would express the above as "IF you hit me again THEN you'll be sorry".

2.2.0.3 OR vs. EITHER ... OR

The last topic we address is the proper use of "OR". In mathematics the phrase

(11) "p is true OR q is true"

is always to be understood as

(12) "p is true OR q is true OR BOTH are true", i.e., at least one of p, q is true.

This is in contrast to everyday language where "p is true OR q is true" often means that exactly one of p and q is true, but not not both.

When referring to a collection of items then the use of "OR" also is inclusive If the items $a, b, c, \dots$ belong to a collection $\mathscr{C}$, e.g., if those items are elements of a set, then

(13) "a OR b OR c OR ..." means that we refer to at least one of $a, b, c, \dots$.

Note that "OR" in mathematics always is an **inclusive or**, i.e., "A OR B" means "A OR B OR BOTH". More generally, "A OR B OR ..." means "at least one of A, B, ...".
To rule out that more than one of the choices is true you must use a phrase like "EXACTLY ONE OF A, B, C, ..." or "EITHER A OR B OR C OR ...". We refer to this as an **exclusive or**.

2.3 Numbers

We start with an informal classification of numbers. It is not meant to be mathematically exact. We will give exact definitions of the integers, rational numbers and real numbers in chapter 9 (The Real Numbers).

Definition 2.13 (Integers and decimal numerals)**.** A **digit** or **decimal digit** Is one of the numbers $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$.

We call numbers that can be expressed as a finite string of digits, possibly preceded by a minus sign, **integers**. In particular we demand that an integer can be written without a decimal point. Examples of integers are

$$3,\ -29,\ 0,\ 3\cdot 10^6,\ -1,\ 2.\bar{9},\ 12345678901234567890,\ -2018. \tag{2.16}$$

Note that $3\cdot 10^6 = 3000000$ is a finite string of digits and that $2.\bar{9}$ equals 3 (see below about the period of a decimal numeral). 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**.

An integer n is an **even** integer if it is a multiple of 2, i.e., there exists $j \in \mathbb{Z}$ such that $n = 2j$, and it is an **odd** integer otherwise. One can give a strict proof that n is odd if and only if there exists $j \in \mathbb{Z}$ such that $n = 2j + 1$. See prop6.26 on p.196.

A **decimal** or **decimal numeral** is a finite or infinite list of digits, possibly preceded by a minus sign, which is separated into two parts by a point, the **decimal point**. The list to the left of the decimal point must be finite or empty, but there may be an infinite number of digits to its right. Examples are

$$3.0,\ -29.0,\ 0.0,\ -0.75,\ .\bar{3},\ 2.74\bar{9},\ \pi = 3.141592.....,\ -34.56. \tag{2.17}$$

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, i.e., $.\bar{3} = 0.33333333333\dots$ and $1.234\overline{567} = 1.234567567567\dots$.

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.

If the list to the right of the decimal point is of the form

$$d_1d_2d_3\dots d_k00\dots,$$

i.e., all digits $d_{k+1}, d_{k+2}, d_{k+3}, \dots$ are zero, then we may remove them from the list. For example, the following all denote the same decimal:

$$-12.34 \;=\; -12.340 \;=\; -12.340000 \;=\; -12.34\bar{0}.$$

The above example shows that any decimal numeral which can be represented by finitely many digits, can also be represented as a repeating decimal (with period $\bar{0}$).

Any integer can be transformed into a decimal numeral of same value by appending the pattern ".0" to its right. Hence the first three integers of (2.16) are equal in value to the first three decimals of (2.17). The mathematician says that we **identify** the integer $\pm d_1d_2d_3\dots d_k$ and the decimal numeral $\pm d_1d_2d_3\dots d_k.0$, i.e., we do not distinguish those expressions and we consider them as equal, just as we would "six" and "half a dozen". □

We are ready to give an informal definition of the most important kind of numbers. The formal, axiomatic, definition will be given in axiom 9.1 on p.255.

Definition 2.14 (Real numbers)**.** We call any kind of number which can be represented as a decimal numeral, a **real number**. We write $\mathbb{R}$ for the set of all real numbers. It follows from what was remarked at the end of Definition 2.13 that integers, in particular natural numbers, are real numbers. Thus we have the following set relations:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{R}. \quad \Box \tag{2.18}$$

We next define rational numbers. The formal definition will be given in Definition 9.4 on p.255,

Definition 2.15 (Rational numbers)**.** A number that is an integer or can be written as a fraction of integers, i.e., as $\frac{m}{n}$ where $m, n \in \mathbb{Z}$ and $n \neq 0$, is called a **rational number**. We write $\mathbb{Q}$ for the set of all rational numbers. Examples of rational numbers are

$$\tfrac{3}{4},\ -0.75,\ -\tfrac{1}{3},\ .\bar{3},\ \tfrac{7}{1},\ 16,\ \tfrac{13}{4},\ -5,\ 2.99\bar{9},\ -37\tfrac{2}{7}.$$

Note that a mathematician does not care whether a rational number is written as a fraction

$$\frac{\text{numerator}}{\text{denominator}}$$

or as a decimal numeral. The following all are representations of one third:

$$0.\bar{3} \;=\; .\bar{3} \;=\; 0.33333333333\dots \;=\; \tfrac{1}{3} \;=\; \tfrac{-1}{-3} \;=\; \tfrac{2}{6}, \tag{2.19}$$

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

$$-\,4 \;=\; -4.000 \;=\; -3.\bar{9} \;=\; -\tfrac{12}{3} \;=\; \tfrac{4}{-1} \;=\; \tfrac{-4}{1} \;=\; \tfrac{12}{-3} \;=\; -\tfrac{400}{100}. \tag{2.20}$$

If $q \in \mathbb{Q}$ then there are unique integers n and d such that $q = \frac{n}{d}$ and

- **(a)** $d \in \mathbb{N}$,
- **(b)** d is minimal: there are no numbers $n' \in \mathbb{Z}$ and $d' \in \mathbb{N}$ such that $q = \frac{n'}{d'}$ and $d' < d$.

We say that this choice of n and d is a representation of q in **lowest terms** or that q is written in lowest terms. For example, the representation of $.\bar{3}$ in lowest terms is $\frac{1}{3}$ and the representation of -4 in lowest terms is $\frac{-4}{1}$.

Note that if $q \in \mathbb{Q}$ is strictly positive and if $\frac{d}{n}$ represents q in lowest terms then $d \in \mathbb{N}$. □

There are real numbers which cannot be expressed as integers or fractions of integers.

Definition 2.16 (Irrational numbers)**.** We call real numbers that are not rational **irrational numbers**. They hence fill the gaps that exist 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 ch.9.6 (Decimal Expansions of Real and Rational Numbers). Irrational numbers must then be those with infinitely many decimal digits without a continually repeating pattern. □

Example 2.6. 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 is certainly a fraction. □

Remark 2.8. Examples of irrational numbers are $\sqrt{2}$ *and* $\boldsymbol{\pi}$. A proof that $\sqrt{2}$ is irrational (actually that $\sqrt[n]{2}$ is irrational for any integer $n \geq 2$) is given in prop.9.30 on p.276. □

Remark 2.9. We will see in ch.9.7 (Countable and Uncountable Subsets of the Real Numbers) on p.285 that, in a sense, 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 2.17 (Types of numbers)**.** We summarize what was said sofar about the classification of numbers:

$\mathbb{N} := \{1, 2, 3, \dots\}$ denotes the set of **natural numbers**.
$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \dots\}$ 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 of some often referenced sets of numbers:

$\mathbb{N}_0 := \mathbb{Z}_+ := \mathbb{Z}_{\geq 0} := \{0, 1, 2, 3, \dots\}$ denotes the set of nonnegative integers,
$\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$ denotes the set of all nonnegative 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 2.18 (Translation and dilation of sets of numbers)**.** For a set of numbers A and numbers λ and b, we define [11]

(2.21) $$\lambda A + b := \{\lambda a + b : a \in A\}.$$

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

(2.22) $$A + b = \{a + b : a \in A\},$$
(2.23) $$-A = \{-a : a \in A\}. \ \square$$

Definition 2.19 (Intervals of Numbers [12])**.** For $a, b \in \mathbb{R}$ we have the following intervals.

- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ is the **closed interval** with endpoints a and b.
- $]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ is the **open interval** with endpoints a and b.
- $[a, b[:= \{x \in \mathbb{R} : a \leq x < b\}$ and $]a, b] := \{x \in \mathbb{R} : a < x \leq b\}$ are **half-open intervals** with endpoints a and b.

The symbol "∞" stands for an object which itself is not a number but is larger than any (real) number, and the symbol "$-\infty$" stands for an object which itself is not a number but is smaller than any number. We thus have $-\infty < x < \infty$ for any number x. This allows us to define the following intervals of "infinite length":

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

You should always work with $a < b$. In case you don't, you get

- $[a, a] = \{a\}$; $[a, a[=]a, a[=]a, a] = \emptyset$
- $[a, b] = [a, b[=]a, b[=]a, b] = \emptyset$ for $a \geq b$ □

[11] See Definition 3.8 in ch.3.2

[12] The following will be generalized in Definition 3.12 on p.70 to so called ordered integral domains.

Notation 2.1 (Notation Alert for intervals of integers or rational numbers)**.**

It is at times convenient to also use the notation $[\dots]$, $]\dots[$, $[\dots[$, $]\dots]$, for intervals of integers or rational numbers. We will subscript them with $\mathbb{Z}$ or $\mathbb{Q}$. For example,

$$\begin{aligned} [\,3, n\,]_{\mathbb{Z}} &= [\,3,\ n] \cap \mathbb{Z} \ = \ \{k \in \mathbb{Z} : 3 \leq k \leq n\}, \\ ]-\infty, 7\,]_{\mathbb{Z}} &=]-\infty, 7\,] \cap \mathbb{Z} \ = \ \{k \in \mathbb{Z} : k \leq 7\} \ = \ \mathbb{Z}_{\leq 7}, \\ ]a, b[_{\mathbb{Q}} &=]a, b[\cap \mathbb{Q} \ = \ \{q \in \mathbb{Q} : a < q < b\}. \end{aligned}$$

An interval which is not subscripted always means an interval of real numbers, but we will occasionally write, e.g., $[a, b]_{\mathbb{R}}$ rather than $[a, b]$, if the focus is on integers or rational numbers and an explicit subscript helps to avoid confusion. □

Definition 2.20 (Absolute value)**.** For a real number x we define its **absolute value** as

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

Example 2.7. $|3| = 3$; $|-3| = 3$; $|-5.38| = 5.38$. □

Remark 2.10. For any real number x we have

$$\sqrt{x^2} \ = \ |x|. \ \square \tag{2.25}$$

Assumption 2.1 (Square roots are always assumed nonnegative)**.** 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 $b \geq 0$

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

We will always assume that "$\sqrt{b}$" is the **positive** value unless the opposite is explicitly stated.

Example: $\sqrt{9} = +3$, not -3. □

Proposition 2.5 (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:*

$$\textit{Triangle Inequality}: \ |\boldsymbol{a} + \boldsymbol{b}| \leq |\boldsymbol{a}| + |\boldsymbol{b}| \tag{2.26}$$

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

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

The next definition should be familiar to anyone who has worked with matrix algebra.

Definition 2.21 (Kronecker symbol). $\boxed{\star}$
For $i, j \in \mathbb{N}$, the **Kronecker symbol** δ_{ij}, also called the **Kronecker delta**, is defined as follows.

$$\delta_{ij} := \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \ \square \end{cases}$$

2.4 A First Look at Functions, Sequences and Families

The material on functions presented in this section will be discussed again and in greater detail in chapter 5 (Functions and Relations) on p.126.

Introduction 2.3. You are familiar with functions from calculus. Examples are $f_1(x) = \sqrt{x}$ 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 $\big((x, y), \ln(x - y)\big)$ in the plane. At other times this expression 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.5.2 on p.131 which precedes the mathematically precise definition of a function (Definition 5.7 on p.134). You are encouraged to look at it once you have read the remainder of this short section as ch.5.2 contains everything you see here.

To obtain 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 $\{\ (x, y) : x, y \in \mathbb{R} \textit{ and } x - y > 0\}$ 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.

For 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. Rather domain and codomain can be a very different kind of animal. In chapter 4 on logic you will learn about statement functions $A(x)$ which assign arguments x from some 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 one can think of the graph of a function $f(x)$ with domain D and codomain C (see earlier in this note) as the set

$$\Gamma_f := \{\big(x, f(x)\big) : x \in D\}.$$

Alternatively one can characterize this function by the assignment rule which specifies how $f(x)$ 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 one does 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

$$f : D \longrightarrow C$$

as a short way of saying that the function $f(x)$ has domain D and codomain C.
In case of the function $f_1(x) = \sqrt{x}$ for which we might 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

$$\text{either } \; f_1 : [\,2.5, 7\,] \to]1, 3[; \quad x \mapsto \sqrt{x} \qquad \text{or } \; f_1 : [\,2.5, 7\,] \to]1, 3[; \quad f(x) = \sqrt{x}.$$

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

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

We incorporate what we noted above into this preliminary definition of a function.

Definition 2.22 (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

$$f : X \to Y, \qquad x \mapsto f(x). \tag{2.27}$$

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**. The **graph** of such a function is the collection of pairs

$$\Gamma_f := \{\big(x, f(x)\big) : x \in X\}. \ \square \tag{2.28}$$

Remark 2.11. The name given to the argument variable is irrelevant. Let f_1, f_2, X, Y, U, V be as defined in **(d)** of the introduction to ch.2.4 (A First Look at Functions, Sequences and Families). The function

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

is identical to the function f_1. The function

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

is identical to the function f_2 and so is the function

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

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

There are times when we write $f(\cdot)$ rather than f for a function f when this avoids confusion. For example physicists and engineers often write $x = x(t)$ to denote the height x of a particle as a function of time t. In such a case we would write

$$x(\cdot) : [0, \infty[\longrightarrow \mathbb{R}\,; \qquad t \mapsto x(t)$$

rather than

$$x : [0, \infty[\longrightarrow \mathbb{R}\,; \qquad t \mapsto x(t)$$

and refer to $x(\cdot)$ rather than x. $\square$

Now some remarks about inverse functions.

Remark 2.12. We all know what it means that $f(x) = \sqrt{x}$ has the function $g(x) = x^2$ as its inverse function: f and f^{-1} cancel each other, i.e.,

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

That certainly is the most important aspect, but there is more. There is an issue with how free one is in the choice of domains and codomains of both functions. Let us replace g with the more familiar f^{-1}, let us write Dom_f and Cod_f for domain and codomain of f and $Dom_{f^{-1}}$ and $Cod_{f^{-1}}$ for domain and codomain of f^{-1}. Thus we have

$$f : Dom_f \longrightarrow Cod_f \qquad \text{and} \qquad f^{-1} : Dom_{f^{-1}} \longrightarrow Cod_{f^{-1}}\,.$$

We want that f^{-1} cancels the effect of f for **all** arguments of f, and we want that f cancels the effect of f^{-1} for **all** arguments of f^{-1}. In other words we want

$$f^{-1}\big(f(x)\big) \;=\; x \;\text{ for } \textbf{all } x \in Dom_f, \tag{2.29}$$

$$f\big(f^{-1}(y)\big) \;=\; y \;\text{ for } \textbf{all } y \in Dom_{f^{-1}}. \tag{2.30}$$

(a) We choose $Dom_f := [0, \infty[$ since that's the biggest set of real numbers for which the square root exists, and let us choose $Cod_f := \mathbb{R}$. Since everything can be squared we choose $Cod_{f^{-1}} := \mathbb{R}$ and $Cod_{f^{-1}} := \mathbb{R}$.
We have a problem. Let $x := -2$. Then $f\big(f^{-1}(-2)\big) = f(4) = 2$, thus (2.30) does not hold. We have to exclude negative numbers from $Dom_{f^{-1}}$, so we try again with $Dom_{f^{-1}} := [0, \infty[$, leaving everything else unchanged. Now (2.30) is satisfied.

(b) Some abstract considerations for the inverse: Let $f : X \to Y,\; x \mapsto f(x)$. Since the inverse should satisfy $f^{-1}\big(f(x)\big) = x$ it must accept items of the form $f(x)$ as arguments, thus its domain must be part of or maybe even all of Cod_f. Likewise, since the f itself should satisfy $f\big(f^{-1}(y)\big) = y$, it must accept items of the form $f^{-1}(y)$ as arguments, thus its domain must be part of or maybe even all of $Cod_{f^{-1}}$.
There are mathematical reasons to demand equality in the above: We want

$$Dom_{f^{-1}} \;=\; Cod_f \;=\; Y; \qquad \text{and} \qquad Cod_{f^{-1}} \;=\; Dom_f \;=\; X.$$

Thus, if $f : X \to Y,\; x \mapsto f(x)$ has an inverse then it must be of the form

$$f^{-1} : Y \to X, \quad y \mapsto f^{-1}(y). \;\square$$

We are now ready to give the preliminary definition of an inverse function.

Definition 2.23 (Preliminary definition of the inverse function)**.**

Given are two nonempty sets X and Y and a function $f : X \to Y$ with domain X and codomain Y. We say that f has an **inverse function** if it satisfies all of the following conditions which uniquely determine this inverse function, so that we are justified to give it the symbol f^{-1}:

(a) $f^{-1} : Y \to X$, i.e., f^{-1} has domain Y and codomain X.

(b) $f^{-1}\big(f(x)\big) \;=\; x$ for all $x \in X$, and $f\big(f^{-1}(y)\big) \;=\; y$ for all $y \in Y$. $\square$

You will find a lot more about functions in ch.5.2 (Functions (Mappings) and Families). Here is just one example. You will learn there that a function f has an inverse f^{-1} if and only if f is "onto": for each $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$, and if f is "one–one": for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$.

Example 2.8. Be sure you understand the following:

(a) $f : \mathbb{R} \to \mathbb{R};\ x \to e^x$ does not have an inverse $f^{-1}(y) = \ln(y)$ since its domain would have to be the codomain $\mathbb{R}$ of f and $\ln(y)$ is not defined for $y \leq 0$.

(b) $g : \mathbb{R} \to]0, \infty[;\ x \to e^x$ has the inverse $g^{-1} :]0, \infty[\to \mathbb{R};\ g^{-1}(y) = \ln(y)$ since

$$\begin{aligned} &Dom_{g^{-1}} = Cod_g =]0, \infty[, \qquad && Cod_{g^{-1}} = Dom_g = \mathbb{R}, \\ &e^{\ln(y)} = y \text{ for } 0 < y < \infty, && \ln(e^x) = x \text{ for all } x \in \mathbb{R}. \ \square \end{aligned}$$

We now briefly discuss (infinite) sequences, subsequences and finite sequences. The exact definition of sequences and their subsequences will be given in Definition 5.22 on p.158, that of finite sequences in Definition 7.2 on p.222.

Definition 2.24. Let $n_\star$ be an integer and let there be a uniquely determined item x_j for each integer $j \geq n_\star$. Such an item can be, e.g., a number or a set (the only items we are looking at for now). In other words:

Assume that a unique item x_j is assigned to each $j \in [n_\star, \infty[_{\mathbb{Z}}$. We write

$$(x_j)_{j \geq n_\star} \quad \text{or} \quad (x_j)_{j \in [n_\star, \infty[_{\mathbb{Z}}} \quad \text{or} \quad (x_j)_{j=n_\star}^{\infty} \quad \text{or} \quad x_{n_\star}, x_{n_\star+1}, x_{n_\star+2}, \dots$$

for such a collection of items, and we call it a **sequence** with **start index** $n_\star$. We call the set $[n_\star, \infty[_{\mathbb{Z}}$ of indices the **index set** of the sequence.

The symbol j is a dummy variable, same as the name x of the argument of a function $f(x)$. See Remark 2.11 on p.31. □

Example 2.9. **(a)** If $u_k = k^2$ for $k \in \mathbb{Z}$, then $(u_k)_{k \geq -2}$ is the sequence of integers

$$4,\ 1,\ 0,\ 1,\ 4,\ 9,\ 16,\ \dots$$

(b) If $A_j = \left[-1 - \frac{1}{j}, 1 + \frac{1}{j}\right] = \left\{x \in \mathbb{R} : -1 - \frac{1}{j} \leq x \leq 1 + \frac{1}{j}\right\}$, then $(A_j)_{j \geq 3}$ is a sequence of sets, the intervals (of real numbers)

$$\left[-\frac{4}{3}, \frac{4}{3}\right], \quad \left[-\frac{5}{4}, \frac{5}{4}\right], \quad \left[-\frac{6}{5}, \frac{6}{5}\right], \quad \left[-\frac{7}{6}, \frac{7}{6}\right], \quad \dots$$

(c) For $j \in [0, \infty]_{\mathbb{Z}}$, let $z_j := (-1)^j$. Then $(z_j)_{j=0}^{\infty}$ is the sequence of integers

$$1,\ -1,\ 1,\ -1,\ 1,\ -1,\ 1,\ -1,\ \dots \ \square$$

(d) The symbols naming a sequence are dummy variables, same as the symbols f and x denoting a function $f(x)$. See Remark 2.11 on p.31. Thus, if $u_m = m^2$ and if $Q_j = j^2$, then $(u_m)_{m \geq -2}$ and $(Q_j)_{j \geq -2}$ are the same sequence of integers as the sequence from **(a)**,

$$(u_k)_{k \geq -2} \quad = \quad 4,\ 1,\ 0,\ 1,\ 4,\ 9,\ 16,\ \dots \ \square$$

Remark 2.13. Sequences can be considered as functions which take the indices as arguments:

One can think of a sequence $(x_i)_{i \geq n_\star}$ in terms of the assignment $i \mapsto x_i$, and the sequence can then be interpreted as the function

$$x(\cdot) : [n_\star, \infty[_{\mathbb{Z}} \longrightarrow \text{ suitable codomain; } \quad i \mapsto x(i) := x_i \,,$$

where that "suitable codomain" depends on the nature of the items x_i. In other words, **Sequences are functions with domain = index set =** $[n_\star, \infty[_{\mathbb{Z}}$.

In Example 2.9**(a)**, we could choose $\mathbb{Z}$ as codomain. We could also choose either of $[0, \infty]_{\mathbb{Z}}$, $\mathbb{Q}$, $\mathbb{R}$, since each of those sets contains all "function" values u_k that belong to the sequence. On the other hand, the set $]0, \infty[$ of all strictly positive real numbers does not qualify since it does not contain the sequence member $u_0 = 0$.

In Example 2.9**(b)**, we could choose $2^{\mathbb{R}}$, the power set of $\mathbb{R}$, as codomain.

In Example 2.9**(c)**, any set that contains the set $\{-1, 1\}$ is a suitable codomain. Observe that the sequence $(z_j)_{j=0}^{\infty}$ is an infinite collection of tagged items, one for each index $j \in [0, \infty]_{\mathbb{Z}}$. However, the set $\{z_j : j \in [0, \infty]_{\mathbb{Z}}\}$ of all values this sequence can attain, only contains two values. We have

$$\{z_j : j \in [0, \infty]_{\mathbb{Z}}\} = \{-1, 1\} \,,$$

since duplicate members of a set are ignored. □

Definition 2.25. We occasionally admit an "ending index" $n^\star$ instead of ∞, i.e., there will be an indexed item x_j for each $j \in [n_\star, n^\star]_{\mathbb{Z}}$. We then talk of a **finite sequence**, and we write

$$(x_n)_{n_\star \leq n \leq n^\star} \quad \text{or} \quad (x_j)_{j=n_\star}^{n^\star} \quad \text{or} \quad x_{n_\star}, \, x_{n_\star+1}, \, \dots, \, x_{n^\star}$$

for such a finite collection of items. If we refer to a sequence $(x_n)_n$ without qualifying it as finite then we imply that we deal with an **infinite sequence**, $(x_n)_{n=n_\star}^{\infty}$.

If one pares down the full set of indices $\{n_\star, \, n_\star + 1, \, n_\star + 2, \dots\}$ to a subset

$$\{n_1, \, n_2, \, n_3, \dots\} \qquad \text{such that} \qquad n_\star \leq n_1 < n_2 < n_3 < \dots$$

then we call the corresponding "thinned out" sequence $(x_{n_j})_{j \in \mathbb{N}}$ a **subsequence** $(x_n)_{n \geq n_\star}$.

If this subset of indices is finite, i.e., we have

$$n_\star \leq n_1 < n_2 < \dots < n_K \qquad \text{for some suitable } K \in \mathbb{N} \,,$$

then we call $(x_{n_j})_{j=1}^{K}$ a **finite subsequence** of the original sequence. □

Remark 2.14. Keep the sequence $\left((-1)^j\right)_{j=0}^{\infty}$ in mind when considering the following which we only state for infinite sequences, but which also applies to subsequences and finite sequences.

Do not confuse a sequence $\left(x_n\right)_{n \geq n_*}$ with the set $\{x_n : n \geq n_*\}$ of its values!
The sequence is a function $n \mapsto x_n$ with domain $[n_*, \infty]_{\mathbb{Z}}$, the set $\{x_n : n \geq n_*\}$ merely is the (smallest possible) codomain of that function.

The sequence $\left(x_n\right)_{n \geq n_*}$ always determines the set $\{x_n : n \geq n_*\}$, but the opposite is not true. For example, if you know that the values belonging to the sequence $\left(x_n\right)_{n\geq 0}$ constitute the set $\{-1, 1\}$ then you do not know whether

$$x_n = (-1)^n \quad \text{or} \quad x_n = (-1)^{n+1} \quad \text{or} \quad x_n = \begin{cases} 1 & \text{if } n \in [0, 100[_{\mathbb{Z}}, \\ -1 & \text{if } n \in [100, \infty[_{\mathbb{Z}}, \end{cases} \quad \text{or} \quad x_n = \dots \quad \square$$

The members x_k of a sequence $\left(x_k\right)_k$ are indexed items in the following sense.

Definition 2.26 (Indexed items)**.** Given is an expression of the form

$$a_i\,.$$

We say that a_i is **indexed by** or **subscripted by** or **tagged by** i. We call i the **index** or **subsript** of a_i, and we call a_i an **indexed item** . □

Remark 2.15. Both a_i and i 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 don't even look like numbers (e.g., α, T, x, t, $\mathscr{A}$ and $\mathbb{R}$). 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: It should be safe to assume that the index $\mathbb{R}$ of $i_{\mathbb{R}}$ denotes the set of all real numbers, since it is hard to imagine that a mathematician would attach a different meaning to that symbol. □

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. 14:

$$\begin{aligned} F &= (a_1, e_1, e_2, i_1, i_2, i_3, o_1, o_2, o_3, o_4, u_3, u_5, u_9, u_{11}, u_{99}) \\ G &= (a_k, e_{-\sqrt{2}}, e_1, i_{-6}, i_{\mathscr{B}}, i_{\mathbb{R}}, o_7, o_{2/3}, o_{-8}, o_3, u_A, u_B, u_C, u_D, u_E) \end{aligned}$$

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 2.15, 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. From now on we ask for the following: The indices of an indexed collection must belong to some set J and each index $i \in J$ must be used exactly once. Remember that this automatically takes care of the duplicate indices problem as a set never contains duplicate values (see Definition 2.1 on p. 14). We also demand that there is a set X such that each indexed item x_i belongs to X.

Those considerations leads us to the definition of a family.

Definition 2.27 (Indexed families)**.** Let J and X be nonempty sets such that

each $i \in J$ is associated with exactly one indexed item $x_i \in X$.

We write $(x_i)_{i \in J}$ for this collection of indexed items and call it an **indexed family** or **family** in X with **index set** J. The indexed items x_j are called the **members of the family**. □

Remark 2.16. Sequences are families with sets of integers as index sets:

(a) Sequences $(x_j)_{j=n_\star}^{\infty}$ are families with index set $[n_\star, \infty[_{\mathbb{Z}}$.
(b) Finite sequences $(x_j)_{j=n_\star}^{n^\star}$ are families with index set $[n_\star, n^\star]_{\mathbb{Z}}$.
(c) Subsequences $\left(x_{n_j}\right)_{j \in \mathbb{N}}$ are families with index set $\{n_1 < n_2 < \dots\}$; $n_j \in \mathbb{Z}$.
(d) Finite subsequences $(x_j)_{j=1}^{K}$ are families with index set $\{n_1 < \dots < n_K\}$; $n_j \in \mathbb{Z}$. □

Example 2.10. Here are some examples of families.

(a) For $r \in \mathbb{R}$, let $B_r := \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$. Then $(B_r)_{r \in \mathbb{R}}$ is a family with index set $\mathbb{R}$ and values in $2^{\mathbb{R}}$. (The indexed items are subsets of $\mathbb{R}$!)

Let $\mathscr{B} := \{B_r : r \in \mathbb{R}\}$ be the set of all tagged items B_r of the above family (thus $\mathscr{B}$ is a set of sets). Do not confuse $\mathscr{B}$ and $(B_r)_{r \in \mathbb{R}}$! The family distinguishes, e.g., between the indexed items B_2 and B_{-2} even though they represent the same set, but $\mathscr{B}$ does not, since sets to not contain any duplicate elements!

We have already seen this behavior for sequences, e.g., in Example 2.9**(c)** on p.33, where we looked at the sequence $\left((-1)^j\right)_{j=0}^{\infty}$. This sequence has infinitely many members since there are infinitely many indices $j = 0, 1, 2, \dots$, but its value set, $\{(-1)^j : j \in [0, \infty[_{\mathbb{Z}}\} = \{-1, 1\}$, only possesses two members.

(b) We take the family from **(a)**, but we shrink the index set to $]-\infty, 0]$, i.e., we consider the family $(B_r)_{r \leq 0}$. Note that $B_0 = \{0\} and\ B_r = \emptyset$ whenever $r < 0$. Hence, $\{B_r : r \leq 0\} = \{\emptyset, \{0\}\}$.

(c) Let the function $h : \mathbb{R} \to \mathbb{R}$ be defined by $h(x) = \sin(x)$, and let $y_x := \sin(x)$ for any real number x. Besides the notation, Is there any real difference between the function h and the family $(y_x)_{x \in \mathbb{R}}$ with values in $\mathbb{R}$? Not really. Both objects do the same; they assign to each $x \in \mathbb{R}$ the real number $\sin(x)$.

(d) This last example generalizes to any function $f : X \to Y$ with arbitrary, nonempty sets X and Y. We can associate with f the function $\big(f(x)\big)_{x \in X}$.

(e) Let $I := [0, 10]$ and $(x_i)_{i \in I}$ the $\mathbb{R}$–valued family defined by $x_i := e^i$. The function

$$\varphi : I \to \mathbb{R}; \qquad i \mapsto e^i$$

is equivalent to that family: both describe the assignment of $i \in I$ to the real number e^i.

(f) This last example generalizes to any X–valued family $(x_i)_{i \in I}$ with arbitrary, nonempty index set I and value set X. We can associate with $(x_i)_{i \in I}$ the function

$$\psi : I \to X; \qquad i \mapsto x_i\,,$$

and both objects convey the same information. □

Remark 2.17. **(a)** Examples 2.10**(c)** through 2.10**(f)** and in particular, examples 2.10**(d)** and 2.10**(f)**, illustrate that families are functions just as sequences are functions: We mentioned in Remark 2.13 on p.34 that a sequence $(x_n)_{n=n_\star}^\infty$ with a suitable codomain X, i.e., $x_n \in X$ for all $n \in [n_\star, \infty[_{\mathbb{Z}}$, can be interpreted as a function with domain $[n_\star, \infty[_{\mathbb{Z}}$ and codomain X. Likewise:

A family $(x_i)_{i \in J}$ can be interpreted as the function

$$x(\cdot) : J \longrightarrow X\,; \quad i \mapsto x(i) := x_i\,.$$

Families in X are functions with domain = index set = J and codomain X .

(b) Same as for sequences, i is a dummy variable: $(x_i)_{i \in J}$ and $(x_k)_{k \in J}$ describe the same family as long as $i \mapsto x_i$ and $k \mapsto x_k$ describe the same function $x(\cdot) : J \to X$. This should not come as a surprise to you if you recall Remark 2.11 on p.31 concerning function arguments and the end of Definition 2.24 on p.33 (sequences).

(c) Do not confuse the family $(x_i)_{i \in J}$ with the set $\{x_i : i \in J\}$ of its function values. We have illustrated this in examples 2.10**(a)** and 2.10**(b)**. □

Example 2.11. For $1 \leq x \leq 10$ let $A_x := [-x, 5x]$. Since $A_x \subseteq \mathbb{R}$, $(A_x)_{x \in [1,10]}$ is a family in $2^{\mathbb{R}}$, the power set of $\mathbb{R}$, with index set $[1, 10]$. If we define $B_z := [-z, 5z]$ and $\mathscr{A}_\alpha := [-\alpha, 5\alpha]$, then both families $(B_z)_{z \in [1,10]}$ and $(\mathscr{A}_\alpha)_{\alpha \in [1,10]}$ are identical to the family $(A_x)_{x \in [1,10]}$ (!) □

This concludes our first look at families. We will have more to say about this topic in Chapter 5.2.8 (Families, Sequences, and Functions as Families).

2.5 Cartesian Products

We next define cartesian products of sets. [13] Those mathematical objects generalize rectangles

$$[a_1, b_1] \times [a_2, b_2] \;=\; \{(x, y) : x, y \in \mathbb{R}, a_1 \leq x \leq b_1 \text{ and } a_2 \leq y \leq b_2\}$$

and quads

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \;=\; \{(x, y, z) : x, y, z \in \mathbb{R}, a_1 \leq x \leq b_1, a_2 \leq y \leq b_2 \text{ and } a_3 \leq z \leq b_3\}.$$

Definition 2.28 (Preliminary definition: Cartesian Product)**.** Let X and Y be two sets The set

$$X \times Y \;:= \{(x, y) : x \in X, y \in Y\} \tag{2.31}$$

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

$$X_1 \times X_2 \cdots \times X_n \;:= \{(x_1, x_2, \dots, x_n) : x_j \in X_j \text{ for each } j = 1, 2, \dots n\} \tag{2.32}$$

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

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

[13] See ch.5.1 (Cartesian Products and Relations) on p.126 for the real thing and examples.

Example 2.12. The graph Γ_f of a function with domain X and codomain Y (see Definition 2.28) is a subset of the cartesian product $X \times Y$. $\square$

Example 2.13. The domains given in **(a)** and **(d)** of the introduction to ch.2.4 (A First Look at Functions, Sequences and Families) are subsets of the cartesian product

$$\mathbb{R}^2 \;=\; \mathbb{R} \times \mathbb{R} \;=\; \{(x,y) : x, y \in \mathbb{R}\} \;\square$$

.

2.6 Arbitrary Unions and Intersections

In Definition 2.5 on p.17 we had defined unions and intersections of finitely many sets $A_1, A_2, \dots, A_n$ as follows:

$$\bigcup_{j=1}^{n} A_j \;= \{x : x \in A_j \text{ for at least one } j = 1, 2, \dots, n\},$$
$$\bigcap_{j=1}^{n} A_j \;= \{x : x \in A_j \text{ for each } j = 1, 2, \dots, n\}.$$

Thus the union of those sets are those items that belong to at least one of those sets, and the intersection of those sets are those items that belong to each one of those sets. This can be generalized to any set of sets [14] or family of sets, finite or not.

Definition 2.29 (Arbitrary unions and intersections)**.**
(A) For a (nonempty) set of sets $\mathscr{A}$, let

$$\bigcup_{B \in \mathscr{A}} B \;:=\; \bigcup \left[B : B \in \mathscr{A}\right] \;:=\; \{x : x \in B \text{ for at least one } B \in \mathscr{A}\}, \tag{2.33}$$

$$\bigcap_{B \in \mathscr{A}} B \;:=\; \bigcap \left[B : B \in \mathscr{A}\right] \;:=\; \{x : x \in B \text{ for each } B \in \mathscr{A}\}. \tag{2.34}$$

We call $\bigcup\limits_{B \in \mathscr{A}} B$ the **union** and $\bigcap\limits_{B \in \mathscr{A}} B$ the **intersection** of the members of $\mathscr{A}$

(B) For a family $(A_i)_{i \in I}$ of sets A_i, let

$$\bigcup_{i \in I} A_i \;:=\; \bigcup \left[A_i : i \in I\right] \;:=\; \{x : x \in A_i \text{ for at least one } i \in I\}, \tag{2.35}$$

$$\bigcap_{i \in I} A_i \;:=\; \bigcap \left[A_i : i \in I\right] \;:=\; \{x : x \in A_i \text{ for each } i \in I\}. \tag{2.36}$$

We call $\bigcup\limits_{i \in I} A_i$ the **union** and $\bigcap\limits_{i \in I} A_i$ the **intersection** of the family $(A_i)_{i \in I}$.

[14] Recall that we encountered, in Example 2.2 on p.15 for example, sets whose elements are sets.

Note that use of the "set style" notation $\bigcup\left[B : B \in \mathscr{A}\right], \bigcup\left[A_i : i \in I\right], \dots$ is less common than that of $\bigcup_{B\in\mathscr{A}} B, \bigcup_{i\in I} A_i, \dots$. We find it advantageous if $\mathscr{A}$ or I consists of a rather lengthy expression.

(C) Let $\mathscr{A}$ be a nonempty set of sets, let $(A_i)_{i\in I}$ be a family of sets.

We call the members of $\mathscr{A}$ **disjoint** , also **mutually disjoint**, if $A, A' \in \mathscr{A}$ and $A \neq A'$ implies $A \cap A' = \emptyset$. We call the family $(A_i)_{i\in I}$ **disjoint** , also **mutually disjoint**, if $A_i \cap A_j = \emptyset$ for all $i, j \in J$ such that $i \neq j$.
As done previously, we allow the use of $\biguplus$ instead of $\bigcup$ to indicate disjoint unions:

$$\biguplus_{B\in\mathscr{A}} B \;:=\; \bigcup_{B\in\mathscr{A}} B\,, \qquad \biguplus_{i\in I} A_i \;:=\; \bigcup_{i\in I} A_i\,. \tag{2.37}$$

(2.38)

Note that disjointness of sets was already defined in Definition 2.6 on p.17, but only for a finite collection of sets.

(D) Assume that there is Ω, $\mathscr{A}$ such that $\mathscr{A} \subseteq \Omega$ and the members of $\mathscr{A}$ are disjoint.
If $\Omega = \biguplus_{B\in\mathscr{A}} B$, then we call $\mathscr{A}$ a **partition** of Ω.
Assume that there is Ω, $(A_i)_{i\in I}$ such that $A_j \subseteq \Omega$ for all $j \in J$ is a disjoint family.
If $\Omega = \biguplus_{i\in I} A_i$, then we call $(A_i)_{i\in I}$ a **partition** of Ω.
Note that being a partition means that each $x \in \Omega$ belongs to exactly one member of $\mathscr{A}$ (of $(A_i)_{i\in I}$ in case of a family).

Since sequences are special kinds of families with index sets

$$[n_*, \infty[_{\mathbb{Z}} = \{\, n_*,\, n_* + 1,\, n_* + 2,\, \dots \}\,,$$

it is natural to write

$$\bigcup_{i=n_*}^{\infty} A_i \;:=\; \bigcup_{i\in[n_*,\infty[_{\mathbb{Z}}} A_i\,, \qquad \bigcap_{i=n_*}^{\infty} A_i \;:=\; \bigcap_{i\in[n_*,\infty[_{\mathbb{Z}}} A_i\,, \quad \square \tag{2.39}$$

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 ($J = \mathbb{N}$).

Remark 2.18. "At least one" can also be expressed as "some" or "there exists", and "for each" can also be expressed as “for all”. Thus one also writes

$$\begin{aligned}
\bigcup_{B\in\mathscr{A}} B \ &= \{x : x \in B \text{ for some } B \in \mathscr{A}\} \ = \ \{x : \text{ there exists } B \in \mathscr{A} \text{ such that } x \in B\},\\
\bigcap_{B\in\mathscr{A}} B \ &= \bigcap\big[B : B \in \mathscr{A}\big] \ := \ \{x : x \in B \text{ for all } B \in \mathscr{A}\}.\\
\bigcup_{i\in I} A_i \ &= \{x : x \in A_i \text{ for some } i \in I\}, \{x : \text{ there exists } i \in I \text{ such that } x \in A_i\},\\
\bigcap_{i\in I} A_i \ &= \{x : x \in A_i \text{ for all } i \in I\}. \ \square
\end{aligned}$$

Example 2.14. In Example 2.2 on p.15 we considered the sets

- **(a)** $\mathscr{A} := \{\,]a, b[: a, b \in \mathbb{R}, 0 < b - a < 2\}$ (all open intervals of length less than 2),
- **(b)** $\mathscr{B} := \{K$:K is a set of integers $\}$ (the power set of $\mathbb{Z}$).

Since each real number x belongs to the set $]x - \frac{1}{2}, x + \frac{1}{2}[$ which is an element of $\mathscr{A}$, it follows that $x \in \bigcup\limits_{B\in\mathscr{A}} B$. Thus $\bigcup\big[B : B \in \mathscr{A}\big] = \mathbb{R}$, the set of all real numbers.

No real number x is an element of $]x + 5, x + 6[$. Since this interval belongs to $\mathscr{A}$, it is not true that $x \in B$ for each $B \in \mathscr{A}$. It follows that $x \notin \bigcap\big[B : B \in \mathscr{A}\big]$. Thus no real number belongs to $\bigcap\limits_{B\in\mathscr{A}} B$; we conclude that $\bigcap\limits_{B\in\mathscr{A}} B = \emptyset$.

Things are just that simple for $\mathscr{B}$. Every integer m is an element of the set $\mathbb{Z}$ of all integers, which in turn is an element of $\mathscr{B}$ It follows that $m \in \bigcup\big[K : K \in \mathscr{B}\big]$. Thus this union equals $\mathbb{Z}$.

To compute the intersection of the members of $\mathscr{B}$ we note that no integer m belongs to the set $\{m\}$ which is an element of $\mathscr{B}$ since it is a set of integers. Thus it is not true that $m \in K$ for all $K \in \mathscr{B}$, thus $x \notin \bigcap\big[K : K \in \mathscr{B}\big]$. Since this is true for all integers m, it follows that $\bigcap\big[K : K \in \mathscr{B}\big] = \emptyset$. $\square$

Example 2.15. Here are two more examples for unions and intersections of sets of sets. The proofs are not as easy as those in the previous example. To understand them you need to be familiar with the properties of limits of real numbers on a beginner's calculus level. [15] Let

$$\begin{aligned}
\mathscr{C} &:= \{\, \big[\,\pi - 3 + \tfrac{1}{n}, \pi + 3 - \tfrac{1}{n}\,\big] \ : \ n = 1, 2, 3, \dots\},\\
\mathscr{D} &:= \{\, \big]\,\pi - 3 - \tfrac{1}{n}, \pi + 3 + \tfrac{1}{n}\,\big[\ : \ n = 1, 2, 3, \dots\}.
\end{aligned}$$

We claim that $\bigcup\limits_{A\in\mathscr{C}} A \ = \ ]\pi - 3, \pi + 3[$ and $\bigcap\limits_{A\in\mathscr{D}} A \ = \ [\pi - 3, \pi + 3]$.

To see this we first observe the following.

- **(1)** The sequence $a_n = (\pi - 3) - \frac{1}{n}$ converges from the left to $\pi - 3$, thus a_n is arbitrarily close to $\pi - 3$ and,

  if $x < \pi - 3$, then $x < a_n < \pi - 3$ is true for all sufficiently large n.
- **(2)** The sequence $b_n = (\pi - 3) + \frac{1}{n}$ converges from the right to $\pi - 3$, thus b_n is arbitrarily close to $\pi - 3$ and,

  if $x > \pi - 3$, then $\pi - 3 < b_n < x$ is true for all sufficiently large n.

[15] Chapter 9.3 (Convergence and Continuity in $\mathbb{R}$) will teach you convergence in a mathematically precise way.

Likewise, we obtain for $c_n = (\pi + 3) - \frac{1}{n}$ and $d_n = (\pi + 3) + \frac{1}{n}$ that

(3) if $x < \pi + 3$, then $x < c_n < \pi + 3$ is true for all sufficiently large n.
(4) if $x > \pi + 3$, then $\pi + 3 < d_n < x$ is true for all sufficiently large n.

Let us choose some more convenient notation. We define

$$C_n := [\, b_n, c_n \,] = \left[\pi - 3 + \frac{1}{n}, \pi + 3 - \frac{1}{n} \right], \quad C := \bigcup_{A \in \mathscr{C}} A\,,$$
$$D_n :=]a_n, d_n[\;=\; \left] \pi - 3 - \frac{1}{n}, \pi + 3 + \frac{1}{n} \right[, \quad D := \bigcap_{A \in \mathscr{D}} A\,.$$

Then

$$\mathscr{C} = \{\, C_n : n \in \mathbb{N} \}, \text{ thus, } C = \bigcup_{n \in \mathbb{N}} C_n\,, \text{ and we must show } C =]\pi - 3, \pi + 3[\,;$$
$$\mathscr{D} = \{\, D_n : n \in \mathbb{N} \}, \text{ thus, } D = \bigcap_{n \in \mathbb{N}} D_n\,, \text{ and we must show } D = [\, \pi - 3, \pi + 3\,]\,.$$

Note that we rewrote C as a union and D as an intersection of a sequence of sets.

(I): We now show that $\bigcup [A : A \in \mathscr{C}] =]\pi - 3, \pi + 3[$.

(I.a) Prove that $C \subseteq]\pi - 3, \pi + 3[$.

Let $x \in C$. We must show that $x \in]\pi - 3, \pi + 3[$. it follows from (2.33) that x belongs to some element of $\mathscr{C}$, i.e., there must be some $n \in \mathbb{N}$ such that $x \in C_n$.

From $C_n = \left[\pi - 3 + \frac{1}{n}, \pi + 3 - \frac{1}{n} \right]$ we obtain $C_n \subseteq]\pi - 3, \pi + 3[$, hence, $x \in]\pi - 3, \pi + 3[$. We have shown $x \in C \Rightarrow x \in]\pi - 3, \pi + 3[$, and this proves **(I.a)**.

(I.b) Prove that $]\pi - 3, \pi + 3[\subseteq C$.

Let $x \in]\pi - 3, \pi + 3[$. We must show that $x \in C$. Since $\pi - 3 < x < \pi + 3$, it follows from **(2)** and **(3)** above that $\pi - 3 < b_n < x < c_n < \pi - 3$ for all sufficiently large n.

$$\text{Thus, } b_n < x < c_n\,, \quad \text{thus, } x \in [\, b_n, c_n], \quad \text{i.e., } x \in C_n$$

is true for all sufficiently large n. Thus there exists an index $n_0 \in \mathbb{N}$ such that $x \in C_{n_0}$. [16] It follows from (2.33) that $x \in C$, and this proves **(I.b)**.

(I.a) and **(I.b)** together yield $\bigcup [A : A \in \mathscr{C}] =]\pi - 3, \pi + 3[$. We have proved **(I)**.

(II): Here is the proof that $\bigcap [A : A \in \mathscr{D}] = [\pi - 3, \pi + 3\,]$.

(II.a) Prove that $[\pi - 3, \pi + 3] \subseteq D$.

Let $x \in [\pi - 3, \pi + 3]$. According to (2.34) we must prove that $x \in A$ for all $A \in \mathscr{D}$, i.e., $x \in]a_n, d_n[$ for all $n \in \mathbb{N}$. This is obviously true, since $[\pi - 3, \pi + 3] \subseteq]\pi - 3 - 1/n, \pi + 3 + 1/n[$ and $a_n = \pi - 3 - 1/n$ and $d_n = \pi + 3 + 1/n$.

(II.b) Prove that $D \subseteq [\pi - 3, \pi + 3]$.

Since $D = \bigcap\limits_{n \in \mathbb{N}}]a_n, d_n[$, we must show, according to (2.34), the following.

(2.40) $$\text{If } x \in]a_n, d_n[\text{ for all } n, \text{ then } x \in [\pi - 3, \pi + 3]\,.$$

[16] Of course there are infinitely many such indices, but that is not important for this proof.

This is a logical statement of the form

(2.41) if P is true, then Q is true, in short, If P, then Q,

where P is the **assumption** "$x \in]a_n, d_n[$ for all n", and Q is the **conclusion** "$x \in [\pi - 3, \pi + 3]$".
This if ... then statement can also be expressed by its **contrapositive**

(2.42) if Q is not true, then P is not true, in short, If not Q, then not P.

We claim that both (2.41) and (2.42) are equivalent logical statements in the following sense:

The validity of "if P, then Q" implies that of "if not Q, then not P", and vice versa.

This can be seen as follows.

- Assume that "if P, then Q" is valid.
- Since the truth of P implies the truth of Q, we cannot have both P true and Q false.
- Thus the falseness of Q implies the falseness of P,
- i.e., if Q is not true then P is not true.

Here is the reverse direction: The validity of (2.42) implies that of (2.41).

- Assume that "if Q is not true then P is not true" is valid.
- Since the falseness of Q implies the falseness of P, we cannot have both Q false and P not false.
- Thus the non-falseness of P implies the non-falseness of Q.
- In other words, the truth of P implies the truth of Q, i.e., if P, then Q.

Thus, to prove that $D \subseteq [\pi - 3, \pi + 3]$, we can replace (2.40) by its contrapositive

If $x \notin [\pi - 3, \pi + 3]$ then it is not true that $x \in]a_n, d_n[$ for all n.

What does it mean that it is not true that $x \in]a_n, d_n[$ for all n? It means that there must exist an index n (at least one) such that $x \notin]a_n, d_n[$. Thus it suffices to prove

(2.43) If $x \notin [\pi - 3, \pi + 3]$ then there is an index $k \in \mathbb{N}$ such that $x \notin]a_k, d_k[$.

So let $x \notin [\pi - 3, \pi + 3]$. Then either $x < \pi - 3$ or $x > \pi + 3$.

First case, $x < \pi - 3$: We have seen in **(1)** [17] that then $x < a_n$ and thus $x \leq a_n$ is true for all sufficiently large n. It follows that $x \notin]a_n, d_n[$ for all such n and, hence, for at least one n.

Second case, $x > \pi + 3$: We have seen in **(4)** that then $x > d_n$ and thus $x \geq d_n$ is true for all sufficiently large n, thus $x \notin]a_n, d_n[$ for all such n and, hence, for at least one n.

In summary, we have shown the validity of (2.43), hence, of (2.43), hence, of $D \subseteq [\pi - 3, \pi + 3]$.

This concludes the proof of **(II.b)** and, thus, **(II)**.

We have learned a few things about logic and proofs which we want to summarize below.

- The statement "if P, then Q" is equivalent to its contrapositive: "if not Q, then not P".
- The method of proving "if P, then Q" by proving the contrapositive is called an **indirect proof by contrapositive**.

[17] near the beginning of the example

Moreover, we used that the opposite of “it” being true for all items is that it is false for at least one item, i.e., that there exists an item for which “it” is false. A moment’s reflection tells us the following: The opposite of “it” being true for at least one item, i.e., the opposite of the existence of an item for which “it” is true, is that “it” is false for all items.

We summarize that as follows.

Let P be some property which can be true or false

- If A is the statement “P is true for all x”, then not A is the statement “there exists some x for which P is false”.
- If B is the statement “there is some x for which P is true”, then not B is the statement “P is false for all x”. □

2.7 Proofs by Induction and Definitions by Recursion

Introduction 2.4. The integers have a property which makes them fundamentally different from the rational numbers (fractions) and the real numbers: Given any two integers $m < n$, there are only finitely many integers between m and n. To be precise, there are exactly $n - m - 1$ of them. For example, there are only 4 integers between 12 and 17: the numbers $13, 14, 15, 16$. [18]

Therefore, given an integer n, we have the concept of its predecessor, $n - 1$, and its successor, $n + 1$. This has some profound consequences. If we know what to do for a certain starting number $k_0 \in \mathbb{Z}$ (we call this number the base case), and if we can figure out for each integer $k \geq k_0$ what to do for $k + 1$ if only we know what to do for k, then we know what to do for **any** $k \geq k_0$! □

We make use of the above when defining a sequence by **recursion**. Here is a simple example.

Example 2.16. Let $k_0 = -2$, $x_{k_0} = 5$ (base case), and $x_{k+1} = x_k + 3$ (i.e., we know how to obtain x_{k+1} just from the knowledge of x_k), then we know how to build the entire sequence

$$x_{-2} = 5,\ x_{-1} = x_{-2} + 3 = 8,\ x_0 = x_{-1} + 3 = 11,\ x_1 = x_0 + 3 = 14,\ \dots,$$

The equation $x_{k+1} = x_k + 3$ which tells us how to obtain the next item from the given one is the **recurrence relation** for that recursive definition. □

Example 2.17. Given is a sequence of sets $A_1, A_2, \dots$. For $n \in \mathbb{N}$ we define $\bigcup\limits_{j=1}^{n} A_j$ and $\bigcap\limits_{j=1}^{n} A_j$ recursively as follows. [19]

$$\bigcup_{j=1}^{1} A_j := A_1, \qquad \bigcup_{j=1}^{n+1} A_j := \Big(\bigcup_{j=1}^{n} A_j\Big) \cup A_{n+1}, \tag{2.44}$$

$$\bigcap_{j=1}^{1} A_j := A_1, \qquad \bigcap_{j=1}^{n+1} A_j := \Big(\bigcap_{j=1}^{n} A_j\Big) \cap A_{n+1}. \tag{2.45}$$

[18] All of this will be made mathematically precise in ch.6.1 on p.167.

[19] An “official” definition for unions and intersections of arbitrarily many sets (not just for finitely many) will be given in Definition 2.29 on p.38.

this tells us the meaning of $\bigcup\limits_{j=1}^{n} A_j$ and $\bigcap\limits_{j=1}^{n} A_j$ for any natural number n. For example, $\bigcap\limits_{j=1}^{4} A_j$ is computed as follows.

$$\begin{aligned}
\bigcap_{j=1}^{1} A_j &= A_1,\\
\bigcap_{j=1}^{2} A_j &= \Big(\bigcap_{j=1}^{1} A_j\Big) \cap A_2 = A_1 \cap A_2,\\
\bigcap_{j=1}^{3} A_j &= \Big(\bigcap_{j=1}^{2} A_j\Big) \cap A_3 = (A_1 \cap A_2) \cap A_3,\\
\bigcap_{j=1}^{4} A_j &= \Big(\bigcap_{j=1}^{3} A_j\Big) \cap A_4 = \big((A_1 \cap A_2) \cap A_3\big) \cap A_4. \ \square
\end{aligned}$$

Remark 2.19. The discrete structure of the integers can also be used as a means to prove a collection of mathematical statements $P(k_0), P(k_0{+}1), P(k_0{+}2), \dots$ which is defined for all integers k, starting at $k_0 \in \mathbb{Z}$. Each $P(k)$ might be an equation or an inequality for two numeric expressions that depend on k. It could also be a relation between sets or it could be something entirely different. For example, $P(k)$ could be the statement $\Big(\bigcup\limits_{j=1}^{k} A_j\Big) \cap B = \bigcup\limits_{j=1}^{k} (A_j \cap B)$. An extremely important tool for proofs of this kind is the following principle. Its mathematical justification will be given later in thm.6.2 on p.169.

Principle of Mathematical Induction
Assume that for each integer $k \geq k_0$ there is an associated statement $P(k)$ such that the following is valid:

A. Base case. The statement $P(k_0)$ is true.
B. Induction Step. Assuming that $P(k)$ is true ("**Induction Assumption**"), it can be shown that $P(k+1)$ also is true.

It then follows that $P(k)$ is true for **each** $k \geq k_0$.

Here is an example which generalizes prop.2.2 on p.20.

Proposition 2.6 (Distributivity of unions and intersections for finitely many sets)**.** *Let $A_1, A_2, \dots$ and B be sets. If $n \in \mathbb{N}$ then*

$$\Big(\bigcup_{j=1}^{n} A_j\Big) \cap B = \bigcup_{j=1}^{n} (A_j \cap B), \tag{2.46}$$

$$\Big(\bigcap_{j=1}^{n} A_j\Big) \cup B = \bigcap_{j=1}^{n} (A_j \cup B). \tag{2.47}$$

PROOF: We only prove (2.46), and this will be done by induction on n, i.e., the number of sets A_j. The proof of (2.47) is left as exercise 2.11

(A) Base case: $k_0 = 1$. The statement $P(1)$ is (2.46) for $n = 1$: $\Big(\bigcup\limits_{j=1}^{1} A_j\Big) \cap B \;=\; \bigcup\limits_{j=1}^{1} (A_j \cap B)$. We must prove that $P(1)$ is true. According to our recursive definition of finite unions which was given in example 2.16 this is the same as $(A_1) \cap B = (A_1 \cap B)$, and this is a true statement. We have proven the base case.

(B) Induction step:

$$\text{(2.48)} \qquad \textbf{Induction assumption: } P(k) : \Big(\bigcup_{j=1}^{k} A_j\Big) \cap B \;=\; \bigcup_{j=1}^{k}(A_j \cap B) \text{ is true for some } k \geq 1.$$

Under this assumption

$$\text{(2.49)} \qquad \text{we must prove the truth of } P(k+1) : \Big(\bigcup_{j=1}^{k+1} A_j\Big) \cap B \;=\; \bigcup_{j=1}^{k+1}(A_j \cap B).$$

The trick is to manipulate $P(k+1)$ in a way that allows us to "plug in" the induction assumption. For (2.49) one way to do this is to take the left–hand side and transform it repeatedly until we end up with the right–hand side, and doing so in such a manner that (2.48) will be used at some point.

$$\begin{aligned}
\Big(\bigcup_{j=1}^{k+1} A_j\Big) \cap B \;&=\; \Big(\Big(\bigcup_{j=1}^{k} A_j\Big) \cup A_{k+1}\Big) \cap B && \text{we used (2.44)}\\
&=\; \Big(\Big(\bigcup_{j=1}^{k} A_j\Big) \cap B\Big) \cup (A_{k+1} \cap B) && \text{we used (2.13) on p. 20}\\
&=\; \bigcup_{j=1}^{k} (A_j \cap B) \cup (A_{k+1} \cap B) && \textbf{we used the induction assumption!}\\
&=\; \bigcup_{j=1}^{k+1} (A_j \cap B) && \text{we used (2.44)}
\end{aligned}$$

We have managed to establish the truth of $P(k+1)$, and this concludes the proof.

Epilogue: It is crucial that you understand in what way the induction assumption was used to get from the left–hand side of (2.49) to the right–hand side, and that you first had to find a base from which to proceed by proving the base case. ■

Proposition 2.7 (The Triangle Inequality for n real numbers). *Let $n \in \mathbb{N}$ such that $n \geq 2$. Let $a_1, a_2, \dots, a_n \in \mathbb{N}$. Then*

$$\text{(2.50)} \qquad |a_1 + a_2 + \ldots + a_n| \;\leq\; |a_1| + |a_2| + \ldots + |a_n|$$

PROOF: Note that this proposition generalizes prop.2.5 on p.28 from 2 terms to n terms. The proof will be done by induction on n, the number of terms in the sum.

(A) Base case: For $k_0 = 2$, inequality 2.50 was already shown (see (2.26) on p.28).

(B) Induction step: Let us assume that 2.50 is true for some $k \geq 2$. This is our induction assumption. We now must prove the inequality for $k+1$ terms $a_1, a_2, \dots, a_k, a_{k+1} \in \mathbb{N}$. 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 $|A| \leq B$. We know from (2.26) that the triangle inequality is valid for the two terms A and a_{k+1}. It follows that $|A + a_{k+1}| \leq |A| + |a_{k+1}|$. We combine those two inequalities and obtain

$$|A + a_{k+1}| \;\leq\; |A| + |a_{k+1}| \;\leq\; B + |a_{k+1}| \tag{2.51}$$

In other words,

$$\left|\big(a_1 + a_2 + \ldots + a_k\big) + a_{k+1}\right| \;\leq\; B + |a_{k+1}| \;=\; \big(|a_1| + |a_2| + \ldots + |a_k|\big) + |a_{k+1}|, \tag{2.52}$$

and this is (2.50) 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 $k \geq k_0 = 2$. ■

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 2.20 (Why induction works)**.** But how can we from all of the above conclude that the distributivity formulas of prop.2.6 and the triangle inequality of prop.2.7 work for all $n \in \mathbb{N}$ such that $n \geq k_0$? We illustrate this for the triangle inequality.

Step 1: We know that the statement is true for $k_0 = 2$ because that was proven in the base case.

Step 2: But according to the induction step, if it is true for $k_0 = 2$, it is also true for the successor $k_0 + 1 = 3$ of 2.

Step 3: But according to the induction step, if it is true for $k_0 + 1$, it is also true for the successor $(k_0 + 1) + 1 = 4$ of $k_0 + 1$.

Step 4: But according to the induction step, if it is true for $k_0 + 2$, it is also true for the successor $(k_0 + 2) + 1 = 5$ of $k_0 + 2$.

. . . .

Step $53,920$: But according to the induction step, if it is true for $k_0 + 53,918$, it is also true for the successor $(k_0 + 53,918) + 1 = 53,921$ of $k_0 + 53,918$.

. . . .

And now we see why the statement is true for any natural number $n \geq k_0$. □

2.8 Some Preliminaries From Calculus

Remark 2.21. To understand this remark 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.

The following is known from calculus (see [15] Stewart, J: Single Variable Calculus): 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".

Let $f :]a, b[\to \mathbb{R}$ be a function which is continuous on $]a, b[$. Then

- **(a)** f is integrable for any $\alpha, \beta \in \mathbb{R}$ such that $a < \alpha < \beta < b$, i.e., the **definite integral** $\int\limits_{\alpha}^{\beta} f(u)du$ exists. For a definition of integrability see, e.g., [15] Stewart, J: Single Variable Calculus.
- **(b)** Integration is "linear", i.e., it is additive: $\displaystyle\int_{\alpha}^{\beta} \big(f(u) + g(u)\big)du \;=\; \int_{\alpha}^{\beta} f(u)du \;+\; \int_{\alpha}^{\beta} g(u)du,$ and you also can "pull out" constant $\lambda \in \mathbb{R}$: $\displaystyle\int_{\alpha}^{\beta} \lambda f(u)du \;=\; \lambda \int_{\alpha}^{\beta} f(u)du.$

(c) Integration is "monotonic":
If $f(x) \leq g(x)$ for all $\alpha \leq x \leq \beta$ then $\displaystyle\int_\alpha^\beta \big(f(u)\big)du \ \leq \ \int_\alpha^\beta g(u)du.$

(d) 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 $x \in]a,b[$.

(e) This antiderivative satisfies $F(\beta) - F(\alpha) \ = \ \displaystyle\int_\alpha^\beta f(u)du$ 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 $\displaystyle\int f(x)dx \ = \ F(x) + C$ for the **indefinite integral** (an integral without integration bounds).

(f) It follows from **(e)** 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., $G = F$. □

2.9 Exercises for Ch.2

2.9.1 Exercises for Sets

Exercise 2.1. Prove (2.14) of prop.2.2 on p.20.

Exercise 2.2. Prove the set identities of prop.2.1.

Exercise 2.3. Prove that for any three sets A, B, C it is true that $(A \setminus B) \setminus C \ = \ A \setminus (B \cup C)$.
Hint: use De Morgan's formula (2.15(**a**)). ■

Exercise 2.4. 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 5.5 for the preliminary definition of the size $|A|$ of a set A and to Definition 5.1 (Cartesian Product of Two Sets) for the definition of Cartesian product. You find both in ch.5.1 (Cartesian Products and Relations) on p.126

Exercise 2.5. Find the size 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 2.6. 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 2.7. Let $X = \{1,2,3,4\}$ and let $Y = \{x,y\}$.
(a) What is $X \times Y$? **(c)** What is $\left|X \times Y\right|$? **(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 $\left|X \times Y\right|$? **(f)** Is $(x,3) \in Y \times X$? **(h)** Is $2 \cdot y \in Y \times X$? □

Exercise 2.8. Let $X = \{8\}$. What is $2^{(2^X)}$?

Exercise 2.9. Let $A = \{1,\{1,2\},2,3,4\}$ and $B = \{\{2,3\},3,\{4\},5\}$. Compute the following.
(a) $A \cap B$ **(b)** $A \cup B$ **(c)** $A \setminus B$ **(d)** $B \setminus A$ **(e)** $A \triangle B$ □

Exercise 2.10. Let A, X be sets such that $A \subseteq X$ and let $x \in X$. Prove the following:
(a) If $a \in A$ then $A = (A \setminus \{a\}) \uplus \{a\}$.
(b) If $a \notin A$ then $A = (A \uplus \{a\}) \setminus \{a\}$.
□

2.9.2 Exercises for Proofs by Induction

Exercise 2.11. Use induction on n to prove (2.47) of prop.2.6 on p.44 of this document: Let $A_1, A_2, \dots$ and B be sets. If $n \in \mathbb{N}$ then $\left(\bigcap\limits_{j=1}^{n} A_j\right) \cup B = \bigcap\limits_{j=1}^{n} (A_j \cup B)$. □

Exercise 2.12. [20]
Let $K \in \mathbb{N}$ such that $K \geq 2$ and $n \in \mathbb{Z}_{\geq 0}$. Prove that $K^n > n$. □

Exercise 2.13. Let $n \in \mathbb{N}$. Then $n^2 + n$ is even, i.e., this expression is an integer multiple of 2. □

PROOF: The proof is given in this instructor's edition.

The proof is done by induction on n.
The base case $(n_0 = 1)$ holds because $1^2 + 1 = 2$, and this is an even number.
Induction step: Let $k \in \mathbb{N}$.

(2.53) Induction assumption: $k^2 + k$ is even, i.e., $k^2 + k = 2j$ for some suitable $j \in \mathbb{Z}$.

We must show that there exists $j' \in \mathbb{Z}$ such that $(k+1)^2 + k + 1 = 2j'$. We have

$$(k+1)^2 + k + 1 = k^2 + 2k + 1 + k + 1 = (k^2 + k) + 2(k+1) \overset{(2.53)}{=} 2j + 2(k+1).$$

Let $j' := j + k + 1$. Then $(k+1)^2 + k + 1 = 2j'$ and this finishes the proof. ■

Exercise 2.14. Use the result from exercise 2.13 above to prove by induction that $n^3 + 5n$ is an integer multiple of 6 for all $n \in \mathbb{N}$. □

[20] Note that this exercise generalizes B/G prop.7.1: If $n \in \mathbb{N}$ then $n < 10^n$. Also note that if you allow K to be a real number rather than an integer then it will not be true for all $K > 1$ and $n \in \mathbb{Z}_{\geq 0}$. For example $K^n > n$ is false for $K = 1.4$ and $n = 2$ (but it is true for $K = 1.5$ and $n = 2$).

PROOF: $\boxed{\text{The proof is given in this instructor's edition.}}$

The proof is done by induction on n.

The base case $(n_0 = 1)$ holds because $1^3 + 5 = 6 = 1 \cdot 6$.

Induction step: Let $k \in \mathbb{N}$.

(2.54)
Induction assumption: $k^3 + 5k$ is an integer multiple of 6, i.e., $k^3 + 5k = 6j$ for some $j \in \mathbb{Z}$.

We must show that there exists $j' \in \mathbb{Z}$ such that $(k+1)^3 + 5(k+1) = 6j'$. We know frome exercise 2.13 that $3(k^2 + k) = 3 \cdot 2 \cdot i$ for a suitable $i \in \mathbb{Z}$, hence

$$\begin{aligned}(k+1)^3 + 5(k+1) &= k^3 + 3k^2 + 3k + 1 + 5k + 5 \;=\; (k^3 + 5k) + 3(k^2 + k) + 6\\ &= (k^3 + 5k) + 6i + 6 \overset{(2.54)}{=} 6(j + i + 1).\end{aligned}$$

Let $j' := j + i + 1$. Then $(k+1)^3 + 5(k+1) = 6j'$ and this finishes the proof. ■

Exercise 2.15. Let $x_1 = 1$ and $x_{n+1} = x_n + 2n + 1$. Prove by induction that $x_n = n^2$ for all $n \in \mathbb{N}$. □

3 The Axiomatic Method

Introduction 3.1. The purpose of this chapter is to familiarize the reader with the axiomatic method, often also called the "proof – theorem" method: How to go about proving a mathematic statement, such as the following:

If m and n both are odd integers then their product mn is odd.

The following is a somewhat simplified description of the axiomatic method. To prove a statement such as the one above one has the following to work with:

(a) Axioms: mathematical statements that are declared to be true and that may be used unquestioningly even though they cannot be proven.
(b) Definitions: declarations that allow you to reference a lengthy sentence or collection of sentences with a convenient short expression. As an example, see definition 3.1 below which allows you to use the words "semigroup" and "monoid" as a short for mathematical objects with certain properties. Thus a definition is not statements, i.e., something that is either true or false, and it makes no sense to ask for the proof of a definition.
(c) Propositions, theorems and lemmata: mathematical statements that may be used because they were previously proven.

Most of this document mainly addresses, besides the general mathematical "plumbing" which consists of sets and functions, topics from the realm of analysis, in particular, convergence and continuity. In contrast, this chapter introduces just enough topics from algebra to provide the foundation for the axiomatic definitions of the integers and the rational and real numbers, as can be found in chapters 1, 2, and 8 of [2] Beck/Geoghegan: The Art of Proof. □

3.1 Semigroups and Groups

Introduction 3.2. To be added later. □

Definition 3.1 (Semigroups and monoids). $\boxed{\star}$
Given is a nonempty set S with a binary operation $\diamond$,
i.e. an "assignment rule" $(s,t) \mapsto s \diamond t$ which assigns to any two elements $s,t \in S$ a third element $u := s \diamond t \in S$.[21] The pair $(S, \diamond)$ is called a **semigroup** if the operation $\diamond$ satisfies

(3.1) **associativity:** $(s \diamond t) \diamond u \ = s \diamond (t \diamond u)$ for all $s,t,u \in S$.

A semigroup for which there exists in addition a **neutral element** with respect to the operation$(s,t) \mapsto s \diamond t$, i.e., some $e \in S$ such that

(3.2) $$s \diamond e \ = \ e \diamond s \ = \ s \text{ for all } s \in S$$

is called a **monoid**.
We can write S instead of $(S, \diamond)$ if it is clear which binary operation on S is represented by $\diamond$. □

[21] In other words, we have a function $\diamond : S \times S \to S, \ (s,t) \mapsto \diamond(s,t) := s \diamond t$ in the sense of Definition 5.7 on p.134. or Definition 2.22 on p.30.

Example 3.1.

- **(a)** $(\mathbb{Z}, +)$ (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.
- **(b)** The following also are monoids: $(\mathbb{N}, \cdot)$, $(\mathbb{Q}, +)$ and $(\mathbb{Q}, \cdot)$, $(\mathbb{R}, +)$ and $(\mathbb{R}, \cdot)$.
- **(c)** In case you have some knowledge about complex numbers: $(\mathbb{C}, +)$ and $(\mathbb{C}, \cdot)$ also are monoids.
- **(d)** Beware: $(\mathbb{N}, +)$ is a semigroup but NOT a monoid since $0 \notin \mathbb{N}$; hence there is no neutral element under addition! □

Example 3.2. If you do not know from linear algebra or ch.11.2 on p.328 about general vector spaces then skip 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?) □

The next example is so important that we state it as a proposition. You should review the (preliminary) definition of a function which was given in Definition 2.22 on p.30 refresh your memory about function composition (chain rule in calculus!) to understand it.

Proposition 3.1. *Let A be a nonempty set and let $S := \{f : f \text{ is a function } A \to A\}$.* [22]

We define a binary operation $\circ$ on S as follows. $(f, g) \mapsto g \circ f$ assigns to two functions $f, g : A \to A$ the function [23]

$$g \circ f : A \to A; \qquad x \mapsto g \circ f(x) := g\big(f(x)\big).$$

$(S, \circ)$ is a monoid.

PROOF:

We need to show that $\circ$ is associative and that S contains a neutral element.

We first prove associativity. For any three functions $f, g, h \in S$ and any $x \in A$ it follows from the definition of $\circ$ that

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

In other words, both the left–hand side $\big((h \circ g) \circ f\big)(x)$ and the right–hand side $\big(h \circ (g \circ f)\big)(x)$ are, for each argument $x \in A$, equal to $h(g(f(x)))$. This shows that those two functions coincide, and we have proven associativity.

We now prove the existence of a neutral element. Let $id_A : A \to A; \quad x \mapsto x$ be the function which does nothing with its arguments. [24] 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 $x \mapsto \big(id_A \circ f\big)(x)$, $x \mapsto \big(f \circ id_A\big)(x)$, and $x \mapsto f(x)$ coincide for all x, i.e., they all represent the same function $x \mapsto f(x)$. This proves (3.2) and hence the existence of a neutral element. ■

[22] 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 $g(x) = 7x + 5e^x$.

[23] See Definition 5.8 (Function composition) on p.136. Example: If $f(x) = 3x^2$ and $g(x) = 7x + 5e^x$ then $g \circ f(x) = g\big(f(x)\big) = g(3x^2) = 21x^2 + 5e^{3x^2}$.

[24] id_A is called the **identity function** or just the **identity** on A.

Theorem 3.1 (Uniqueness of the neutral element in monoids). *Let $(S, \diamond)$ be a monoid and let $e, e' \in S$ such that both*

$$s \diamond e \ = e \diamond s \ = s \tag{3.3}$$

$$s \diamond e' \ = e' \diamond s \ = s \tag{3.4}$$

for all $s \in S$. Then $e = e'$.

PROOF: We have

$$e \overset{(3.4)}{=} e' \diamond e \overset{(3.3)}{=} e'.$$

Here we applied (3.4) with $s = e$ and then (3.3) with $s = e'$. ■

Example 3.3. Here is an example of a binary operation which is not associative. For integers m and n we define $m \diamond n \ := \ |n - m|$, i.e., the distance between m and n. This operation is not associative for all $m, n \in \mathbb{Z}$. To prove that such is the case, we only need to find **one counterexample**, i.e., three specific integers m, n, k such that $(m \diamond n) \diamond k \neq m \diamond (n \diamond k)$. This kind of proof is called a **proof by counterexample**. We choose $m = 5$, $n = 3$, $k = 7$ and obtain

$$(5 \diamond 3) \diamond 7 \ = 2 \diamond 7 \ = \ 5, \quad \text{but} \ \ 5 \diamond (3 \diamond 7) \ = \ 5 \diamond 4 \ = \ 1.$$

It follows that $(\mathbb{Z}, \diamond)$ is not a semigroup. Note that 0 is not a neutral element for $(\mathbb{Z}, \diamond)$, because $n \diamond 0 = |n|$ does not equal n whenever $n < 0$.

What if we replace $\mathbb{Z}$ with the set $\mathbb{Z}_{\geq 0}$ of all nonnegative integers? The counterexample above shows that $(\mathbb{Z}_{\geq 0}, \diamond)$ is not a semigroup either. But in this case 0 is a neutral element for $(\mathbb{Z}_{\geq 0}, \diamond)$, because $n \diamond 0 = |n| = n$ for all $n \in \mathbb{Z}_{\geq 0}$. □

Definition 3.2 (Groups and Abelian groups). Let $(G, \diamond)$ be a monoid with neutral element e which satisfies the following: For each $g \in G$ there exists some $g' \in G$ such that

$$g \diamond g' \ = \ g' \diamond g \ = \ e \ \ \text{for all } g \in G. \tag{3.5}$$

We call such a g' an **inverse element**. of g, and we then call $(G, \diamond)$ a **group**.

Assume moreover that the operation $\diamond$ satisfies

$$\textbf{commutativity:} \quad g \diamond h \ = h \diamond g \ \ \text{for all } g, h \in G. \tag{3.6}$$

Then G is called a **commutative group** or **abelian group**. [25] We write G instead of $(G, \diamond)$ if it is clear which binary operation on G is represented by $\diamond$. □

Groups $(G, \diamond)$ are characterized as follows.

(a)	If $g, h \in G$ then $g \diamond h \in G$	**binary operation**
(b)	If $g, h, k \in G$ then $(g \diamond h) \diamond k \ = \ g \diamond (h \diamond k)$	**associativity**
(c)	There exists $e \in G$ such that $g \diamond e \ = \ e \diamond g \ = \ g$ for all $g \in G$	**neutral element**
(d)	For each $g \in G$ there exists $g' \in G$ such that $g \diamond g' \ = \ g' \diamond g \ = \ e$	**inverse element**

G is a **commutative group (abelian group)** if, in addition,

(e)	$g \diamond h \ = \ h \diamond g$ for all $g, h \in G$	**commutativity**

[25] named so after the Norwegian mathematician Niels Henrik Abel who lived in the first half of the 19th century and died at age 26.

Theorem 3.2 (Uniqueness of the inverse in groups)**.** *Let $(G, \diamond)$ be a group and let $g \in G$. Assume that there exists besides g' another $g'' \in G$ which satisfies (3.5). Then $g'' = g'$.*

PROOF: We have

$$g'' \overset{(3.2)}{=} e \diamond g'' \overset{(3.5)}{=} (g' \diamond g) \diamond g'' \overset{\text{assoc}}{=} g' \diamond (g \diamond g'') \overset{(3.5)}{=} g' \diamond e \overset{(3.2)}{=} g'$$

and this proves uniqueness.

Epilogue: We have taken care in this proof to give for every step a reference. ■

Definition 3.3 (inverse element g^{-1})**.** It is customary to write g^{-1} for the unique element of G that is associated with the given $g \in G$ by means of (3.5). We call g^{-1} <u>the</u> inverse element of g rather than <u>an</u> inverse element of g. □

Example 3.4.

(a) $(\mathbb{R}_{\neq 0}, \cdot)$ (the nonzero real numbers with multiplication) is a commutative group: Multiplication is both associative and commutative, and the number 1 is the neutral element. Let $x \in \mathbb{R}$ not be zero. Then $x^{-1} = \frac{1}{x}$ satisfies $x \cdot x^{-1} = x^{-1} \cdot x = 1$, i.e., (3.5). The notation g^{-1} for the inverse of g comes from this example.

(b) $(\mathbb{Z}, +)$ (the integers with addition) is an abelian group: We have already seen that $(\mathbb{Z}, +)$ is a monoid.

The inverse element to $k \in \mathbb{Z}$ with respect to addition is $-k$ because $k + (-k) = (-k) + k = 0$ for all $k \in \mathbb{Z}$. Note that it would be very confusing to write k^{-1} rather than $-k$ for the inverse element under addition.

This group is abelian because $m + k = k + m$ for all $k, m \in \mathbb{Z}$.

(c) $(\mathbb{Z}_{\neq 0}, \cdot)$ (the nonzero integers with multiplication) is **not** a group: Let $k = 5$. Then $k \in \mathbb{Z}_{\neq 0}$. but $1/5$, the only number m such that $5 \cdot m = m \cdot 5 = 1$ is not an integer and hence does not belong to $\mathbb{Z}_{\neq 0}$. □

Proposition 3.2. *Let $(G, \diamond)$ be a group with neutral element e. Let $g, h \in G$. Then*

$$\left(g^{-1}\right)^{-1} = g, \tag{3.7}$$

$$(h \diamond g)^{-1} = g^{-1} \diamond h^{-1}. \tag{3.8}$$

PROOF of (3.7): By definition of the inverse g^{-1}, we have

$$g \diamond g^{-1} = g^{-1} \diamond g = e.$$

These two equations not only show that g^{-1} is an inverse of g, but also that g is an inverse of g^{-1}. It follows from thm.3.2 that g is the unique inverse $\left(g^{-1}\right)^{-1}$ of g^{-1}. We have shown (3.7).

PROOF of (3.8): We have

$$\begin{aligned}(g^{-1} \diamond h^{-1}) \diamond (h \diamond g) &\overset{(3.1)}{=} g^{-1} \diamond \left(h^{-1} \diamond (h \diamond g)\right) \\ &\overset{(3.1)}{=} g^{-1} \diamond \left((h^{-1} \diamond h) \diamond g\right) \overset{(3.5)}{=} g^{-1} \diamond (e \diamond g) \overset{(3.2)}{=} g^{-1} \diamond g \overset{(3.5)}{=} e.\end{aligned}$$

We substitute h^{-1} for g and g^{-1} for h in the above chain of equations, and we obtain $\big((h^{-1})^{-1} \diamond (g^{-1})^{-1}\big) \diamond (g^{-1} \diamond h^{-1}) \ = \ e$. We apply (3.7) and it follows that $(h \diamond g) \diamond (g^{-1} \diamond h^{-1}) \ = \ e$. It follows that $g^{-1} \diamond h^{-1}$ is an inverse of $h \diamond g$. We have shown (3.8).

Note that it follows from prop.3.2 that $g^{-1} \diamond h^{-1}$ is the unique inverse $(h \diamond g)^{-1}$ of $h \diamond g$. ■

Proposition 3.3. $\boxed{\star}$ *Let $(G, \diamond)$ be a group. Let $g, h \in G$. Then*

$$h \diamond g^{-1} \ = (g \diamond h^{-1})^{-1}. \tag{3.9}$$

FIRST PROOF – doing it the hard way from scratch:

Proof strategy: Let e denote the neutral element G as usual. If we write $x := h \diamond g^{-1}$ and $y := g \diamond h^{-1}$ then our assertion is that $x = y^{-1}$. According to the definition of inverses we thus must prove that $x \diamond y = e$ and $y \diamond x = e$, i.e., we must prove that

$$(h \diamond g^{-1}) \diamond (g \diamond h^{-1}) \ = e \ \ (\star) \qquad \text{and} \qquad (g \diamond h^{-1}) \diamond (h \diamond g^{-1}) \ = \ e \ \ (\star\star).$$

PROOF of $(\star)$:

$$\begin{aligned}
(h \diamond g^{-1}) \diamond (g \diamond h^{-1}) \ &= \ \big[(h \diamond g^{-1}) \diamond g\big] \diamond h^{-1} && \text{(associativity)}\\
&= \ \big[h \diamond (g^{-1} \diamond g)\big] \diamond h^{-1} && \text{(associativity)}\\
&= \ (h \diamond e) \diamond h^{-1} && \text{(def. inverse)}\\
&= \ h \diamond h^{-1} && \text{(def. neutral element)}\\
&= \ e && \text{(def. inverse)}
\end{aligned}$$

The proof of $(\star\star)$ is left as exercise 3.5 (see p.82). ■

Proposition 3.4 (B/G prop.1.9 and B/G prop.8.10). *Let $g, h, h' \in (G, \diamond)$. If $g \diamond h = g \diamond h'$ then $h = h'$.*

PROOF: It follows that the assumption that $g^{-1} \diamond (g \diamond h) \ = \ g^{-1} \diamond (g \diamond h')$.

Thus, by associativity, $(g^{-1} \diamond g) \diamond h \ = \ (g^{-1} \diamond g) \diamond h'$.

It follows from (3.5) in the definition of the group inverse that $g^{-1} \diamond g = e$, thus $e \diamond h = e \diamond h'$.

Since the neutral element acts as a "no–op" we finally obtain $h = h'$. ■

Example 3.5. Let $S \ := \ \{f : f \text{ is a function } \mathbb{R} \to \mathbb{R}\}$ with the operation $(f, g) \mapsto g \circ f$ defined as $g \circ f(x) = g(f(x))$. We have seen in prop.3.1 that $(S, \circ)$ is a monoid. We now show that $(S, \circ)$ is not a group.

Proof strategy: According to Definition 3.2, S is a group if and only if each function $f : \mathbb{R} \to \mathbb{R}$ possesses an inverse element $f^{-1} : \mathbb{R} \to \mathbb{R}$ which satisfies (3.5). Because the neutral element of S is the function $id_{\mathbb{R}} : x \mapsto x$, this inverse f^{-1} must satisfy

$$f \circ f^{-1} \ = f^{-1} \circ f \ = \ id_{\mathbb{R}}, \quad \text{i.e.,} \quad f\big(f^{-1}(x)\big) \ = \ f^{-1}\big(f(x)\big) \ = \ x \text{ for all } x \in \mathbb{R}.$$

Accordingly, to prove that S is not a group, it suffices to produce just one counterexample $f \in S$ for which an inverse f^{-1} does not exist.

Some functions will have an inverse. For example $f(x) = x - 7$ has inverse $f^{-1}(x) = x + 7$.

Recall from calculus that if f has an inverse then f must pass the "horizontal line test": Any parallel to the x–axis may intersect the graph of f (see Definition 2.22 (preliminary definition of a function) on p.30) in at most one point. [26] We We must find a function which does not have an inverse. Here are three.

$f(x) = x^2$:	The horizontal line $y = 4$ intersects the graph of f in $(-2, 4)$ and also in $(2, 4)$.
$f(x) = 21$:	The horizontal line $y = 21$ intersects the graph of f in $(x, 21)$ for each $x \in \mathbb{R}$.
$f(x) = \sin(x)$:	The horizontal line $y = 0$ intersects the graph of f in $(n\pi, 0)$ for each $n \in \mathbb{Z}$. □

Proposition 3.5. *Let G be the set of all polynomials of degree 1. In other words,*

$$G \ = \ \{f : \mathbb{R} \to \mathbb{R} : f(x) = ax + b \textit{ for some } a, b \in \mathbb{R} \textit{ where } a \neq 0\}$$

This is the set of functions whose graph is a straight line in the x, y–plane, which is parallel neither to the x–axis, nor to the y–axis. As in example 3.5, let $(f, g) \mapsto g \circ f$ be defined as $g \circ f(x) = g(f(x))$. Then $(G, \circ)$ is a group.

PROOF:

First, we prove that $\circ$ is a binary operation on G, i.e., if $f, g \in G$ then $g \circ f \in G$ (see the beginning of Definition 3.1 (semigroups and monoids) on p.50). In other words, we will show that the composition of two straight line functions is a straight line function.

So let $f(x) := a_1x + b_1$ and $g(x) := a_2x + b_2$ for suitable $a_1, b_1, a_2, b_2 \in \mathbb{R}$ where moreover $a_1, a_2 \neq 0$. Let $x \in \mathbb{R}$. Then

$$g \circ f(x) \ = g(a_1x + b_1) \ = \ a_2(a_1x + b_1) + b_2 \ = \ (a_1a_2)x + (a_2b_1 + b_2).$$

Hence $g \circ f$ is of the form $x \mapsto ax + b$ with $a = a_1a_2 \in \mathbb{R}_{\neq 0}$ and $b = a_2b_1 + b_2 \in \mathbb{R}$. We have proved that $\circ$ is a binary operation on G.

It follows from from prop.3.1 that $(G, \circ)$ is a monoid. We only have to note that $id_{\mathbb{R}} \in G$ because, if $a = 1$ and $b = 0$, then $id_{\mathbb{R}}(x) = x = ax + b$.

To prove that this monoid is a group, we must prove thefollowing. If $f \in G$, then there exists $g \in G$ such that $g(f(x)) = f(g(x)) = x$, for all $x \in \mathbb{R}$.

We have learned in calculus that, if $y = f(x)$, we must "solve for x" to obtain the inverse function. Let $f(x) = ax + b$ $(a \neq 0)$. Then

$$y \ = \ ax + b \ \Rightarrow ax \ = \ y - b \ \Rightarrow \ x \ = \ \tfrac{1}{a}y + \tfrac{-b}{a}.$$

Let $g(x) := \frac{1}{a}x - \frac{b}{a}$. Then $g \in G$. To prove that $g = f^{-1}$, we must show that

$$g\big(f(x)\big) \ = g\big(f(x)\big) \ = \ x \text{ for all } x \in \mathbb{R}.$$

We have

$$\begin{aligned} g\big(f(x)\big) \ &= g(ax + b) \ = \ \tfrac{1}{a}(ax + b) - \tfrac{b}{a} \ = \ (x + \tfrac{b}{a}) - \tfrac{b}{a} \ = \ x; \\ f\big(g(x)\big) \ &= f(\tfrac{1}{a}x + \tfrac{-b}{a}) \ = \ a(\tfrac{1}{a}x - \tfrac{b}{a}) + b \ = \ (x - b) + b \ = \ x. \end{aligned}$$

Hence $g = f^{-1}$. We have shown that every element f of the monoid $(G, \circ)$ possesses an inverse and it follows that $(G, \circ)$ is a group. ■

The next definition is familiar to you if you have taken a linear algebra course.

[26] Actually, our definition of inverse function demands that any parallel to the x–axis must intersect the graph of f in exactly one point. (see rem. 5.13 (horizontal and vertical line tests) on p.146).

Definition 3.4 (Linear functions on $\mathbb{R}$). $\boxed{\star}$

A function $f : \mathbb{R} \to \mathbb{R}$ is a **linear function on** $\mathbb{R}$ if the following is true for all x, y, λ[27] $\in \mathbb{R}$:

$$f(x+y) = f(x) + f(y) \quad \textbf{(additivity)}, \tag{3.10}$$

$$f(\lambda x) = \lambda f(x) \quad \textbf{(homogeneity)}. \quad \square \tag{3.11}$$

You will learn later about the general definition of a linear function. See Definition 11.9 (linear mappings) on p.335.

Theorem 3.3. *Let $f : \mathbb{R} \to \mathbb{R}$. Then f is linear if and only if there exists $a \in \mathbb{R}$ such that $f(x) = ax$ for all $x \in \mathbb{R}$.*

PROOF:

Proof strategy: The proof of a statement of the form "P is true if and only if Q is true" consists of two parts. We must prove that **a)** if P is true then Q is true and also **b)** if Q is true then P is true. In the context of this theorem we have

P: f is linear,
Q: there exists $a \in \mathbb{R}$ such that $f(x) = ax$ for all $x \in \mathbb{R}$.

a) Proof that if f is linear then there exists $a \in \mathbb{R}$ such that $f(x) = ax$ for all $x \in \mathbb{R}$:

$$f(1) = f\left(\frac{y-x}{y-x}\right) \overset{(3.11)}{=} \frac{f(y-x)}{y-x} \overset{(3.10)}{=} \frac{f(y)-f(x)}{y-x} \quad \text{for all } y \neq x. \tag{3.12}$$

It follows that f represents a straight line in the plane with slope $m = f(1)$.

Next we observe that $f(0) = f(2 \cdot 0) \overset{(3.11)}{=} 2f(0)$, hence $f(0) = 0$.

We substitute $y = 0$ in (3.12) and obtain $f(1) = \frac{-f(x)}{-x}$. It follows with $a := f(1)$ that indeed $f(x) = a \cdot x$ for some $a \in \mathbb{R}$.

b) Proof that if there exists $a \in \mathbb{R}$ such that $f(x) = ax$ for all $x \in \mathbb{R}$ then f is linear:

We show the validity of 3.10 and 3.11 by brute force. Let $x, y, \lambda \in \mathbb{R}$. Then

$$\begin{aligned} f(x+y) &= a(x+y) = ax + ay = f(x) + f(y), \\ f(\lambda x) &= a(\lambda x) = \lambda(ax) = \lambda f(x). \end{aligned}$$

This proves both additivity and homogeneity and hence the linearity of f. ■

Let $(G, \diamond)$ be a group and $H \subseteq G$. Note that if $h, h' \in H$ then $h, h' \in G$ and hence $h \diamond h'$ exists as an element of G, but there is no guarantee that $h \diamond h' \in H$. For example, let $(G, \diamond)$ be the group $(\mathbb{R}, +)$ of all real numbers with addition as its binary operation, and let $H := [-1, 1] = \{x \in \mathbb{R} : -1 \leq x \leq 1\}$. Then $\frac{1}{2}$ and $\frac{3}{4}$ belong to H, but $\frac{1}{2} + \frac{3}{4} \notin H$. Subsets H of G which are "closed" with respect to $\diamond$, i.e., $h \diamond h' \in H$ whenever $h, h' \in H$, deserve a special name.

Definition 3.5 (Subgroup). $\boxed{\star}$ Let $(G, \diamond)$ be a group and $H \subseteq G$.

[27] λ (prounced lambda) is the greek version of the letter l. Chapter 22.1 (Greek Letters) on p.510 contains a list of the most commonly used Greek letters.

We call $(H, \diamond)$ a **subgroup** of G if the following is true:

(3.13) H is not empty,

(3.14) if $h, h' \in H$ then $h \diamond h' \in H$,

(3.15) if $h \in H$ then its inverse element h^{-1} (in G!) belongs to H.

We also write H for $(H, \diamond)$, if there is no confusion about the nature of "$\diamond$". □

Proposition 3.6. *Subgroups are groups.*

PROOF:

Let $(G, \diamond)$ be a group and let H be a subgroup of G. We must prove that $(H, \diamond)$ is a monoid (H is not empty, $\diamond$ is a binary operation on the subset H of G, H has a neutral element e_H, and H satisfies associativity) and that each $h \in H$ possesses $h' \in H$ such that $h \diamond h' = h' \diamond h = e_H$

(a) H is nonempty: This follows from (3.13).

(b) $\diamond$ is a binary operation on H: We must show that if $h, h' \in H$ then $h \diamond h' \in H$ (not just that $h \diamond h' \in G$). But this follows from (3.14).

(c) Existence of a neutral element: Let e be the neutral element of G. H is not empty, hence there exists $h_0 \in H$. h_0 has an inverse h_0^{-1} in the group G which belongs, according to (3.15), to H. It follows from (3.14) that $e = h_0 \diamond h_0^{-1} \in H$.

$g \diamond e = e \diamond g = e$ holds for any $g \in G$ and hence, in particular, for each $g \in H$. This proves that e is a neutral element of H.

(d) We next prove associativity. Let $h_1, h_2, h_3 \in H$. We apply (3.14) four times to obtain

$$h_1 \diamond h_2 \in H, \ (h_1 \diamond h_2) \diamond h_3 \in H, \ h_2 \diamond h_3 \in H, \ \ h_1 \diamond (h_2 \diamond h_3) \in H.$$

It further follows from the associativity of $\diamond$ in G that $(h_1 \diamond h_2) \diamond h_3 \ = \ h_1 \diamond (h_2 \diamond h_3)$. We thus have proven associativity of $\diamond$ in H.

(e) We finally prove that if $h \in H$ then its inverse in G, h^{-1}, is also the inverse of h in H. That follows from the fact that, by (3.15), $h^{-1} \in H$, the neutral element e of G is also the neutral element of H, and

$$h \diamond h^{-1} \ = \ h^{-1} \diamond h \ = \ e. \ \blacksquare$$

Example 3.6.

(a) $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$, because the sum of two integers is an integer and the additive inverse of an integer is an integer.

(b) $(\mathbb{Q}_{\neq 0}, \cdot)$ is a subgroup of $(\mathbb{R}_{\neq 0}, \cdot)$, because the product of two nonzero fractions is a nonzero fraction and the multiplicative inverse of two nonzero fractions is a nonzero fraction.

(c) Let $H := \{x \in \mathbb{R} : 0 < |x| < 1\}$. Then $(H, \cdot)$ is **not** a subgroup of $(\mathbb{R}_{\neq 0}, \cdot)$ (since $1 \notin H$, but also since $0.5^{-1} = -2 \notin H$).

(d) Then $(\mathbb{Z}_{\neq 0}, \cdot)$ is not a subgroup of $(\mathbb{R}_{\neq 0}, \cdot)$ because $\frac{1}{2}$, the multiplicative inverse of $2 \in \mathbb{Z}$, is not an integer.

(e) Let $H := [1, 2]$. Then $(H, \cdot)$ is not a subgroup of $(\mathbb{R}_{\neq 0}, \cdot)$, and $(H, +)$ is not a subgroup of $(\mathbb{R}, +)$.

Proposition 3.7. *Let $(G, \circ)$ be the set of all polynomials of degree 1 with function composition, i.e.,*

$$G \;=\; \{\, \mathbb{R} \xrightarrow{f} \mathbb{R} : \; f(x) = ax + b, \text{ for some } a, b \in \mathbb{R} \text{ such that } a \neq 0 \}\,,$$
$$g \circ f : \; x \mapsto g \circ f(x) \;=\; g(f(x))\,.$$

Further, let

$$H \;:=\; \{\, \mathbb{R} \xrightarrow{f} \mathbb{R} : \; f(x) = ax, \text{ for some nonzero } a \in \mathbb{R} \}\,.$$

Then $(H, \circ)$ is a subgroup of $(G, \circ)$.

PROOF:

It was established in prop.3.5 that $(G, \circ)$ is a group. H is a subset of G because elements of H are those functions $x \mapsto ax + b$ of G for which $b = 0$. To prove that H is a subgroup of G we must show that if $h, h' \in H$ then $h \circ h' \in H$ and that the inverse function h^{-1} in G actually belongs to H.

So let $h(x) := ax$ and $h'(x) := a'x$ ($a, a' \in \mathbb{R}$ and $a, a' \neq 0$). Then

$$h \circ h'(x) \;=\; a(a'x) \;=\; (aa')x$$

shows, because $aa' \neq 0$, that $h \circ h' \in H$. Further, the inverse of h in G is the function $h^{-1} : x \mapsto \frac{1}{a}x$. But $h' \in H$ because $\frac{1}{a} \neq 0$. We have proven that H is a subgroup of G. ■

Note that the above proposition also follows from thm.3.3 on p.56.

Proposition 3.8. *The intersection of two subgroups is a subgroup.*

PROOF:

Let $(G, \diamond)$ be a group and let H_1, H_2 be two subgroups of G. Let $H := H_1 \cap H_2$ and $h, h' \in H$. We must prove (3.14) and (3.15). We conclude from $h, h' \in H \subseteq H_1$ that $h \diamond h' \in H_1$ and $h^{-1} \in H_1$ because H_1 is a subgroup. We further conclude from $h, h' \in H \subseteq H_2$ that $h \diamond h' \in H_2$ and $h^{-1} \in H_2$ because H_2 also is a subgroup. It follows from the definition of an intersection that $h \diamond h' \in H_1 \cap H_2$ and $h^{-1} \in H_1 \cap H_2$. This concludes the proof. ■

We now turn our attention to functions which map from a group to another group in such a way that they are, in a sense, compatible with the binary operations on their domain and codomain.

Example 3.7. Let $(G, \diamond) := (\mathbb{R}, +)$ and $(H, \bullet) := (\,]0, \infty[, \cdot)$. Then both G and H are abelian groups (H is an abelian subgroup of $(\mathbb{R} \setminus \{0\}, \cdot)$ since the product of two strictly positive numbers is again strictly positive and because the neutral element 1 is strictly positive). Let

$$\varphi : (G, \diamond) \to (H, \bullet)\,; \quad x \mapsto e^x\,, \qquad \psi : (H, \bullet) \to (G, \diamond)\,; \quad y \mapsto \ln(y)\,.$$

Note that the following is true for φ:

- $\varphi(x + y) \;=\; \varphi(x) \cdot \varphi(y)$ for all $x, y \in G$,
- $\varphi(0) \;=\; 1$: the image of the neutral element is the neutral element.
- $\varphi(-x) \;=\; e^{-x} \;=\; \frac{1}{\varphi(x)}$: the image of the inverse is the inverse of the image.

> 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.

Mathematicians say that a function function φ that has groups both as its domain and it codomain is **structure compatible** with the algebraic (group) operations on its domain and codomain.

The function $\psi(y) = \ln(y)$ also is structure compatible:

- $\psi(x \cdot y) \; = \; \psi(x) \, + \, \psi(y)$ for all $x, y \in H$,
- $\psi(1) \; = \; 0$: the image of the neutral element is the neutral element.
- $\psi\left(\frac{1}{y}\right) \; = \; \ln\left(\frac{1}{y}\right) \; = \; -\ln(y) \; = \; -\psi(y)$: the function value of the inverse is the inverse of the function value.

Note that those two structure compatible functions φ and ψ are inverses of each other. □

We can generalize the above example as follows.

Definition 3.6 (Homomorphisms and isomorphisms)**.** Let $(G, \diamond)$ and $(H, \bullet)$ be 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. 3.3 on p.53.

Let $\varphi : (G, \diamond) \to (H, \bullet)$ be a function which satisfies the following:

(3.16) $$\varphi(g_1 \diamond g_2) \; = \varphi(g_1) \bullet \varphi(g_2) \, .$$

Then we call φ a **homomorphism**, more specifically, a **group homomorphism**, from the group $(G, \diamond)$ to the group $(H, \bullet)$.
Let $\psi : (H, \bullet) \to (G, \diamond)$ be a group homomorphism from $(H, \bullet)$ to $(G, \diamond)$ such that φ and ψ are inverse to each other. We call such a bijective homomorphism an **isomorphism**, and we call the groups $(G, \diamond)$ and $(H, \bullet)$ **isomorphic**.

For bijectivity, see Definition 5.12 on p.144). □

Theorem 3.4. *Let $(G, \diamond)$ and $(H, \bullet)$ be two groups and let $\varphi : (G, \diamond) \to (H, \bullet)$ be a homomorphism. Let e_G be the neutral element of G and e_H be the neutral element of H. Then*

- ***(a)*** $\varphi(e_G) \; = \; e_H$,
- ***(b)*** *Let $g \in G$. Then* $\varphi(g^{-1}) \; = \; \big(\varphi(g)\big)^{-1}$,
- ***(c)*** *Direct images of subgroups of G are subgroups of H.*
- ***(d)*** *Preimages of subgroups of G are subgroups of H.*

PROOF of **(a)**: It follows from $\varphi(e_G) \; = \; \varphi(e_G \diamond e_g) \; = \; \varphi(e_G) \bullet \varphi(e_G)$ and associativity that

$$\begin{aligned} e_H \; &= \big(\varphi(e_G)\big)^{-1} \bullet \varphi(e_G) \; = \; \big(\varphi(e_G)\big)^{-1} \bullet \varphi(e_G \diamond e_G) \\ &= \big(\varphi(e_G)\big)^{-1} \bullet \big[\varphi(e_G) \bullet \varphi(e_G)\big] \; = \; \big[\big(\varphi(e_G)\big)^{-1} \bullet \varphi(e_G)\big] \bullet \varphi(e_G) \\ &= e_H \bullet \varphi(e_G) \; = \; \varphi(e_G) \, . \end{aligned}$$

PROOF of **(b)**: We apply part **(a)** and (3.16) and obtain

$$e_H \; = \varphi(e_G) \; = \; \varphi(g^{-1} \diamond g) \, . \; = \; \varphi(g^{-1}) \bullet \varphi(g)$$

This proves that $\varphi(g^{-1})$ is the inverse of $\varphi(g)$.

The proof of **(c)** and **(d)** is left as an exercise. ■

The next result which is not hard to prove might surprise you. If a group homomorphism possesses an inverse then this inverse is also a group homomorphism

Theorem 3.5. $\boxed{\star}$ *Let $(G, \diamond)$ and $(H, \bullet)$ be two groups and let $\varphi : (G, \diamond) \to (H, \bullet)$ be a homomorphism which possesses an inverse. Then $\varphi^{-1} : H \to G$ also is a homomorphism and thus φ is an isomorphism*

PROOF:
We recall that the inverse φ^{-1} of φ satisfies $\varphi \circ \varphi^{-1} = id_H$ and $\varphi^{-1} \circ \varphi = id_G$, i.e., $\varphi\big(\varphi^{-1}(h)\big) = h$ for all $h \in H$ and $\varphi^{-1}\big(\varphi(g)\big) = g$ for all $g \in G$.

So let $h, h' \in H$. Since φ is surjective, there exist $g, g' \in G$ such that $\varphi(g) = h$ and $\varphi(g') = h'$. Then,

$$\begin{aligned} \varphi^{-1}(h \bullet h') &= \varphi^{-1}(\varphi(g) \bullet \varphi(g')) \ = \ \varphi^{-1}(\varphi(g \diamond g')) && \text{(since } \varphi \text{ is a homomorphism)} \\ &= g \diamond g' \ = \ \varphi^{-1}\big(\varphi(g)\big) \diamond \varphi^{-1}\big(\varphi(g')\big) && \text{(since } \varphi^{-1} \circ \varphi = id_G) \\ &= \varphi^{-1}(h) \diamond \varphi^{-1}(h'). && \text{(definition of } h \text{ and } h') \ \blacksquare \end{aligned}$$

3.2 Commutative Rings and Integral Domains

Introduction 3.3. The definition of a ring is of great importance in algebra. It will not be given in this document because it is too general for our purposes. We rather restrict ourselves from the outset to so called commutative rings with unit and to integral domains. These definitions not only cover the number systems we all are familiar with, the integers, fractions and real numbers, [28] but also, e.g., the set of all polynomials when considered as functions $p(x)$ of a real variable x. □

Definition 3.7 (Commutative rings with unit)**.** $\boxed{\star}$
Let R be a nonempty set with two binary operations

$\oplus : (a, b) \mapsto a \oplus b$, called **addition**, and $\odot : (a, b) \mapsto a \odot b$, called **multiplication**,

which assign to any two elements $a, b \in R$ uniquely determined $a \oplus b \in R$ and $a \odot b \in R$ such that the following holds:

(a) $(R, \oplus)$ is an abelian (i.e., commutative) group; we denote the neutral element for addition by 0 and the inverse element of $a \in R$ for addition by $\ominus a$.
(b) $(R, \odot)$ is a commutative monoid, i.e., a monoid for which $a \odot b = b \odot a$ for all $a, b \in R$. We denote the neutral element with respect to multiplication by 1.
(c) Multiplication is **distributive** over addition:

(3.17) $$a \odot (b \oplus c) \ = \ (a \odot b) \oplus (a \odot c) \ \text{ for all } a, b, c \in R.$$

(d) $1 \neq 0$.

The triplet $(R, \oplus, \odot)$ is called a **commutative ring with unit**. We may write R instead of $(R, \oplus, \odot)$ if it is clear which binary operations on R are represented by $\oplus$ and by $\odot$. □

[28] See Proposition 3.12 on p.64

Remark 3.1. Recall from thm.3.1 and thm.3.2 that the neutral elements 0 and 1 and the additive inverse $\ominus b$ are uniquely determined ($b \in R$). □

Notation 3.1 (Notation Alert for Commutative Rings With Unit)**.**

(a) It is customary to write ab instead of $a \odot b$ if this does not give rise to confusion.
(b) Multiplication has precedence over (binds stronger than) addition: $a \odot b \oplus c$ means $(a \odot b) \oplus c$, not $a \odot (b \oplus c)$.
(c) Let $a, b, \in R$. Recall from thm.3.1 and thm.3.2 that not only the neutral elements 0 and 1 but also the additive inverse $\ominus b$ are uniquely determined. Accordingly, we can define another binary operation, $\ominus$, on $(R, \oplus, \odot)$ as follows:

$$a \ominus b := a \oplus (\ominus b). \quad \square \tag{3.18}$$

We call $a \ominus b$ the **difference** of a and b.

For a set of numbers A we defined in Definition 2.18 on p.27 the set $\lambda A + b = \{\lambda a + b : a \in A\}$. This generalizes without difficulty to commutative rings with unit.

Definition 3.8 (Translation and dilation of sets)**.** ⋆

Let $R = (R, \oplus, \odot)$ be a commutative ring with unit and $A \subseteq R$. and $\alpha, b \in R$. We define

$$\lambda A \oplus b := \{\lambda a \oplus b : a \in A\}. \tag{3.19}$$

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

$$A \oplus b = \{a \oplus b : a \in A\}, \tag{3.20}$$

$$\ominus A = \{\ominus a : a \in A\}. \quad \square \tag{3.21}$$

Remark 3.2. Note that the above makes sense for any **algebraic structure**, i.e., a set with one or more "algebraic operations", if they have the binary operations "$\oplus$" and/or "$\odot$" of a commutative ring with unit. [29] □

We will now examine the role of the condition $1 \neq 0$. It turns out that it is equivalent to demanding that R is not the trivial "Null ring" $\{0\}$.

Proposition 3.9. *Let $(R, \oplus, \odot)$ be a nonempty set with two binary operations $\oplus$ and $\odot$ which satisfies **(a)**, **(b)**, **(c)** of Definition 3.7, i.e., R satisfies all conditions for a commutative ring with unit except that 1 and 0 need not be different elements of R. Then*

(a) $a \ominus a = 0$ *for all* $a \in R$,
(b) $a \odot 0 = 0$ *for all* $a \in R$.

[29] Ignore this if you are not familiar with vector spaces: In a vector space V the scalar product $(\lambda, a) \mapsto \lambda v$ of a real number (scalar) λ and a vector $v \in V$ (not a binary operation since its domain is $\mathbb{R} \times V$ rather than $V \times V$) would take the place of "$\odot$"

PROOF of **(a)**: This follows from the definitions of inverse and subtraction: $a \ominus a \ = \ a \oplus (\ominus a) \ = \ 0$.
PROOF of **(b)**:

$$a \odot 0 \overset{3.2}{=} a(0 \oplus 0) \overset{3.17}{=} a \odot 0 \oplus a \odot 0, \ \text{ hence} \tag{3.22}$$

$$\begin{aligned} 0 &\overset{3.5}{=} a \odot 0 \oplus \big(\ominus (a \odot 0)\big) \overset{3.22}{=} (a \odot 0 \oplus a \odot 0) \oplus \big(\ominus (a \odot 0)\big) \\ &\overset{3.1}{=} a \odot 0 \oplus \big(a \odot 0 \oplus (\ominus(a \odot 0))\big) \overset{3.5}{=} a \odot 0 \oplus 0 \overset{3.2}{=} a \odot 0. \end{aligned} \tag{3.23}$$

The second chain of equations above proves that $a \odot 0 = 0$. ■

Proposition 3.10.

(a) *The set $R := \{0\}$ satisfies conditions* **(a)**, **(b)**, **(c)** *of Definition 3.7,*

(b) *Let $(R, \oplus, \odot)$ be a nonempty set with two binary operations $\oplus$ and $\odot$ which satisfies* **(a)**, **(b)**, **(c)** *of Definition 3.7. Then the following is true: $1 = 0$ if and only if $R \ = \ \{0\}$*
.

PROOF:
Proof of (a):
Note that because 0 is the only element of R, the operations $\oplus$ and $\odot$ are completely determined by the following:

$$0 \ \oplus \ 0 \ = \ 0; \qquad 0 \ \odot \ 0 \ = \ 0.$$

We only prove here that $(R, \oplus)$ is a monoid. The proofs of the other properties are just as simple.
Let $a, b, c \in R$. Then $a = b = c = 0$ because R does not contain any other elements. We obtain

$$(a \oplus b) \oplus c \ = \ (0 \oplus 0) \oplus 0 \ = \ 0 \oplus 0 \ = \ 0 \oplus (0 \oplus 0) \ = \ a \oplus (b \oplus c),$$

hence $\oplus$ is associative and $(R, \oplus)$ is a semigroup.
Let $a \in R$. Then $a = 0$ because R does not contain any other elements. We obtain

$$a \oplus 0 \ = \ 0 \oplus 0 \ = \ 0 \oplus a,$$

hence 0 is neutral element for $\oplus$ and the semigroup $(R, \oplus)$ is a monoid.
Proof of (b):
■

Definition 3.9 (Zero Divisors and Cancellation Rule).

Let $(R, \oplus, \odot)$ be a commutative ring with unit.

(a) If $a, b \in R$ such that $a \neq 0$ and $b \neq 0$ and $a \odot b = 0$ then we call a and b **zero divisors**.

(b) We say that the **cancellation rule** holds in R if the following is true for all $a, b, c \in R$ such that $a \neq 0$:

$$\text{If } a \odot b \ = \ a \odot c \ \text{ then } b = c. \ \square \tag{3.24}$$

For an example of a commutative ring with unit which contains zero divisors see ch.6.10 (The Integers Modulo n) on p.197.

Definition 3.10 (Integral domains)**.**

Let $(R, \oplus, \odot)$ be a commutative ring with unit which satisfies the

- **no zero divisors condition**: If $a, b \in R$ such that $a \odot b = 0$ then $a = 0$ or $b = 0$ (or both are zero).

The triplet $(R, \oplus, \odot)$ is called an **integral domain**. □

Remark 3.3. We stated the no zero divisors condition in the definition of an integral domain as follows: If $a, b \in R$ such that $a \odot b = 0$ then $a = 0$ or $b = 0$ (or both are zero). We remind you that there was no need to include the "or both are zero" part since "or" is always the inclusive "or". See the 2.2.0.3 section (OR vs. EITHER ... OR) on p.23. □

Remark 3.4. Integral domains $(R, \oplus, \odot)$ are characterized as follows.

(a)	If $a, b \in R$ then $a \oplus b \in R$ and $a \odot b \in R$	**binary operations**
(b)	If $a, b, c \in R$ then $(a \oplus b) \oplus c \ = \ a \oplus (b \oplus c)$	**associativity of** $\oplus$
(c)	If $a, b, c \in R$ then $(a \odot b) \odot c \ = \ a \odot (b \odot c)$	**associativity of** $\odot$
(d)	If $a, b \in R$ then $a \oplus b \ = \ b \oplus a$	**commutativity of** $\oplus$
(e)	If $a, b \in R$ then $a \odot b \ = \ b \odot a$	**commutativity of** $\odot$
(f)	If $a, b, c \in R$ then $a \odot (b \oplus c) \ = \ (a \odot b) \oplus (a \odot c)$	**distributivity**
(g)	There exists $0 \in R$ such that $a \oplus 0 \ = \ a$ for all $a \in R$	**neutral element f.** $\oplus$
(h)	There exists $1 \in R$ such that $1 \neq 0$ and $a \odot 1 \ = \ a$ for all $a \in R$	**neutral element f.** $\odot$
(i)	For each $a \in R$ there exists $a' \in R$ such that $a \oplus a' \ = \ 0$	**inverse element f.** $\oplus$
(j)	If $a, b \in R$ such that $a \neq 0$ and $b \neq 0$ then $a \odot b \ \neq \ 0$	**no zero divisors**

Proposition 3.11. *Let $(R, \oplus, \odot)$ be a commutative ring with unit. Then R satisfies the No zero divisors condition if and only if the cancellation rule holds in R.*

PROOF that the cancellation rule implies the absence of zero divisors:

We assume that the cancellation rule holds in R. Let $a, b \in R$ such that $ab = 0$ and $a \neq 0$. It suffices to show that then $b = 0$. (Why?)

It follows from $ab = 0$ and $b = b \ominus 0$ that $a(b \ominus 0) = 0$, hence $a \odot b = a \odot 0$. The cancellation rule now implies together with $a \neq 0$ that $b = 0$.

PROOF that the absence of zero divisors implies the cancellation rule:

We assume that R has no zero divisors. Let $a, b, c \in R$ such that $ab = ac$ and $a \neq 0$. It suffices to show that $b = c$.

It follows from the distributivity law (3.17) that $0 \ = \ ab \ominus ac \ = \ a(b \ominus c)$. Because R has no zero divisors, at least one of a or $b \ominus c$ must be zero.

We assumed that $a \neq 0$ and it follows that $b \ominus c = 0$, i.e., $b = c$. ■

Corollary 3.1. *A commutative ring with unit is an integral domain if and only if the cancellation rule holds.*

PROOF: Immediate from prop.3.11. ■

Proposition 3.12. *Each of the following algebraic structures is an integral domain:*

(a) $(\mathbb{Z}, +, \cdot)$*: the integers with addition and multiplication,*
(b) $(\mathbb{Q}, +, \cdot)$*: the rational numbers with addition and multiplication,*
(c) $(\mathbb{R}, +, \cdot)$*: the real numbers with addition and multiplication.*
(d) [30] $(\mathbb{C}, +, \cdot)$*: the complex numbers with addition and multiplication.*

PROOF Will not be given here. ■

3.3 Arithmetic in Integral Domains

Notation: In this entire chapter we assume that a fixed integral domain $(R, \oplus, \odot)$ is given and phrases such as "let $x \in R$" refer to that integral domain.
Note also that we will, in accordance with notation 3.1**(a)**, often write $a\,b$ instead of $a \odot b$.

Introduction 3.4. When you look at ch.1.2–1.3 and then again at ch.8.1 of [2] Beck/Geoghegan: The Art of Proof then you notice that identical propositions and theorems are given there: First for integers in ch.1, and then again for real numbers in ch.8. You will not find a third set for the rational numbers, but only because the authors chose instead to define those as a subset of $\mathbb{R}$ instead and, in that manner, inherit their laws of arithmetic from those for the real numbers.

It was mentioned in prop.3.12 on p.64, and it will be proven in ch.6 (The Integers) and ch.9 (The Real Numbers) that not only the integers but also the rational numbers and the real numbers with the binary operations of addition and multiplication are integral domains.

This is the power of mathematical abstraction:

A formula such as, e.g., $(-x)(-y) = xy$ need not be demonstrated separately for $\mathbb{Z}$, for $\mathbb{Q}$, and then again for $\mathbb{R}$. Rather it should be possible to state and prove it once for integral domains and thus have it validated for all three sets of numbers.

This is the route we are going here. The propositions and theorems in this chapter are almost an exact copy of ch.1.2–1.3, and then again of ch.8.1, of [2] Beck/Geoghegan: The Art of Proof, but we use "$\oplus$" instead of "$+$", "$\ominus$" instead of "$-$" and "$\odot$" instead of "$\cdot$" for the binary operations of addition, subtraction and multiplication. This is deliberate. The reader then should more easily remember that these rules also apply to other integral domains such as, e.g., the complex numbers and the so called integers modulo n. (See Definition 6.13 on p.197.) □

We just mentioned that a lot of the following material can, in a sense, also be found in ch.1.2–1.3 and ch.8.1, of [2] Beck/Geoghegan: The Art of Proof. The difference is that those authors represent the material first for the integers (ch.1) and then again for the real numbers (ch.8). In contrast we state the material only once, in the framework of integral domains.

[30] Skip this part of the proposition if you have not learned about complex numbers.

Some of the propositions that only deal with one of the operations $\oplus$ and $\odot$ are immediate consequences of the material of Chapter 3.1 (Semigroups and Groups). We include them here to make it easier to read the Beck/Geoghegan text in parallel.

Where applicable we provide the Beck/Geoghegan references for matching definitions, propositions and theorems. Note that we refer to some of the proofs to that book. On the other hand, if a proof is not given in the B/G book then you will generally also not find it in this document. An exception is the first proposition here, prop.3.13, where we supply the proof to help the readers make the transition between the set $(\mathbb{Z}, +, \cdot)$ and the set $(R, \oplus, \odot)$.

Proposition 3.13 (B/G prop.1.6 and B/G prop.8.8)**.** *Let $a, b, c \in R$. Then* $(a \oplus b) \odot c \ = \ a \odot c \oplus b \odot c$.

PROOF:

$$(a \oplus b) \odot c \ \overset{\text{def.3.7.b}}{=} \ c \odot (a \oplus b) \ \overset{(3.17)}{=} \ c \odot a \oplus c \odot b \ \overset{\text{def.3.7.b}}{=} \ a \odot c \oplus b \odot c. \ \blacksquare$$

Proposition 3.14 (B/G prop.1.7 and B/G prop.8.9)**.** *Let $a \in R$. Then $0 \oplus a = a$ and $1 \odot a = a$.*

PROOF: It follows from Definition 3.7 on p.60 of a commutative ring with unit that $(R, \oplus)$ is a monoid with neutral element 0 and $(R, \odot)$ is a monoid with neutral element 1. The assertion follows from the definition of a monoid. ■

Proposition 3.15 (B/G prop.1.8)**.** *Let $a \in R$. Then* $(\ominus a) \oplus a = 0$.

PROOF: It follows from the definition of inverse elements in groups, even if they are not assumed to be abelian, that $(\ominus a) \oplus a = 0$. See (3.5) on p.52. ■

Proposition 3.16 (B/G prop.1.10 and B/G prop.8.11)**.** *Let $a, b_1, b_2 \in R$. If $a \oplus b_1 = 0$ and $a \oplus b_2 = 0$ then $b_1 = b_2$.*

PROOF: Left as an exercise. ■

Note for the following proposition that parts **(b)** and **(c)** hold true for semigroups (not even commutativity of $\oplus$ is needed) and that part **(d)** holds true for commutative monoids.

Proposition 3.17 (B/G prop.1.11 and B/G prop.8.12)**.** *Let $a, b, c, d \in R$. Then*

(a) $(a \oplus b)\,(c \oplus d) \ = \ (ac \oplus bc) \oplus (ad \oplus bd)$,
(b) $a \oplus \big(b \oplus (c \oplus d)\big) \ = \ (a \oplus b) \oplus (c \oplus d) \ = \ \big((a \oplus b) \oplus c\big) \oplus d$,
(c) $a \oplus (b \oplus c) \ = \ (c \oplus a) \oplus b$,
(d) $a(bc) \ = \ c(ab)$,
(e) $a\big(b \oplus (c \oplus d)\big) \ = \ (ab \oplus ac) \oplus ad$,
(f) $\big(a(b \oplus c)\big)d \ = \ (ab)d \oplus a(cd)$.

The proof is left as an exercise. ■

Proposition 3.18. $\boxed{\star}$ *Let $a, b \in R$. Then* $b \ominus a \ = \ \ominus(a \ominus b)$.

PROOF: Since $(R, \oplus)$ is a group and $\ominus x$ denotes the inverse of $x \in R$ this is a reformulation of prop.3.3 on p.54. ■

The following two propositions state that the neutral element with respect to addition is the unique solution of the equation $a \oplus x = a$.

Proposition 3.19 (B/G prop.1.12 and B/G prop.8.13)**.** *Let $x \in R$ satisfy the following: For each $a \in R$ it is true that $a \oplus x = a$. Then $x = 0$.*

PROOF: Left as an exercise. ■

Proposition 3.20 (B/G prop.1.13 and B/G prop.8.14)**.** *Let $x \in R$ satisfy the following: There exists (at least one) $a \in R$ such that $a \oplus x = a$. Then $x = 0$.*

Left as an exercise. ■

Remark 3.5. Be sure to understand that the last two propositions are different! Both have the same conclusion, $x = 0$, but the assumptions are not the same:

- Prop.3.19 asks for a lot before it allows you to conclude that $x = 0$: **Every** element a of R must satisfy the condition $a \oplus x = a$.
- In contrast prop.3.20 asks for very little so that you may reach the same conclusion: It suffices if you can find **just one** element a of R which satisfies the condition $a \oplus x = a$.

So which of the two propositions is the more powerful one? Of course it is prop.3.20 which allows you to draw the same conclusion under the "weaker" assumption that it suffices to find just one $a \in R$ that satisfies $a \oplus x = a$. □

Proposition 3.21 (B/G prop.1.14 and B/G prop.8.15)**.** *Let $a \in R$. Then $a \odot 0 = 0 = 0 \odot a$.*

PROOF: This is Proposition 3.9**(b)**. ■

[2] Beck/Geoghegan: The Art of Proof gives at this spot the definition of divisibilty. We omit it here and provide it in Definition 6.2 on p.172.

The following two propositions show that the neutral element with respect to multiplication is the unique solution of the equation

$$a \odot x = a.$$

Compare them to prop.3.19 and prop.3.20 above, and also review remark 3.5 which follows them.

Proposition 3.22 (B/G prop.1.18 and B/G prop.8.16)**.** *Let $x \in R$ satisfy the following: For each $a \in R$ it is true that $a \odot x = a$. Then $x = 1$.*

PROOF: Left as an exercise. ■

Proposition 3.23 (B/G prop.1.19 and B/G prop.8.17)**.** *Let $x \in R$ satisfy the following: There exists (at least one) nonzero $a \in R$ such that $a \odot x = a$. Then $x = 1$.*

PROOF: See the proof of B/G prop.1.19. ■

Next we provide more propositions about the additive and multiplicative inverses and about cancellation.

Proposition 3.24 (B/G prop.1.20 and B/G prop.8.18)**.** *Let $a, b \in R$. Then $(\ominus a)(\ominus b) = ab$.*

PROOF: See the proof of B/G prop.1.20. ■

Corollary 3.2 (B/G cor.1.21)**.** $(\ominus 1)(\ominus 1) = 1$.

PROOF: Immediate from prop.3.24. ■

Proposition 3.25 (B/G prop.1.22 and B/G prop.8.19)**.**

(a) *If* $a \in R$ *then* $\ominus(\ominus a) = a$.
(b) $\ominus 0 = 0$.

PROOF of **(a)**: Left as an exercise.

PROOF of **(b)**: Let $x = \ominus 0$ and $a = 0$. Then

$$a \oplus x \;=\; 0 \oplus (\ominus 0) \;=\; 0 \;=\; a\,.$$

According to Proposition 3.20 (B/G prop.1.13 and B/G prop.8.14) on p.66, the existence of $a \in R$ such that $a \oplus x = a$ implies that x = 0. It follows from the above chain of equations that $x = 0$, i.e., $\ominus 0 = 0$. ■

Proposition 3.26 (Unique Solutions of Linear Equations)**.** *Let* $(R, \oplus, \odot)$ *be an integral domain and* $a, b, y \in R$ *such that* $a \neq 0$*. The equation* $y = a \odot x \oplus b$ *possesses at most one solution* $x \in R$*.*

PROOF: Let $x, x' \in R$ satisfy $y = a \odot x \oplus b$ and $y = a \odot x' \oplus b$. We must show that $x = x'$.

It follows from our assumptions that $a \odot x \oplus b \;=\; a \odot x' \oplus b$, hence $a \odot x \;=\; a \odot x'$ by prop.3.4 on p.54. It follows from cor.3.1 on p.64 that $x = x'$ ■

Remark 3.6. Note that the equation $y = a \odot x \oplus b$ need not have a solution. For example, there is no $x \in (\mathbb{Z}, +, \cdot)$ which satisfies the equation $2x + 0 = 1$. □

However the following is true.

Proposition 3.27 (B/G prop.1.23 and B/G prop.8.20)**.** *Let* $a, b \in R$*. Then there exists one and only one* $x \in R$ *such that* $a \oplus x = b$*.*

PROOF: Uniqueness follows from Proposition 3.26.

x exists since we can compute it as follows:

$$x \;=\; (\ominus a \,\oplus\, a) \oplus x \;=\; \ominus a \,\oplus\, (a \oplus x) \;=\; \ominus a \oplus b\,. \;\blacksquare$$

Remark 3.7. Note that "there exists one and only one ..." is the same as "there exists a unique ..." For this reason a statement like the one in the preceding proposition is also called an **existence and uniqueness statement**. □

Proposition 3.28 (B/G prop.1.24 and B/G prop.8.21)**.**

Let $x \in R$*. If* $x \odot x = x$ *then* $x = 0$ *or* $x = 1$*.*

PROOF: Left as an exercise. ■

Proposition 3.29 (B/G prop.1.25 and B/G prop.8.22)**.** *Let $a, b \in R$. Then*

(a) $\ominus(a \oplus b) \;=\; (\ominus a) \oplus (\ominus b)$,
(b) $\ominus a \;=\; (\ominus 1)a$,
(c) $(\ominus a)b \;=\; a(\ominus b) \;=\; \ominus(ab)$.

PROOF: Left as an exercise. ■

Proposition 3.30 (B/G prop.1.26 and B/G prop.8.23)**.** *Let $a, b \in R$. If $ab = 0$ then $a = 0$ or $b = 0$.*

PROOF: This is the no zero divisors condition for integral domains. ■

The next proposition is a collection of properties that involve the difference $a \ominus b = a \oplus (\ominus b)$ of two elements a and b of R.

Proposition 3.31 (B/G prop.1.27 and B/G prop.8.24)**.** *Let $a, b, c, d \in R$. Then*

(a) $(a \ominus b) \oplus (c \ominus d) \;=\; (a \oplus c) \ominus (b \oplus d)$,
(b) $(a \ominus b) \ominus (c \ominus d) \;=\; (a \oplus d) \ominus (b \oplus c)$,
(c) $(a \ominus b)(c \ominus d) \;=\; (ac \oplus bd) \ominus (ad \oplus bc)$,
(d) $a \ominus b \;= c \ominus d$ *if and only if* $a \oplus d \;=\; b \oplus c$,
(e) $(a \ominus b)c \;=\; ac \ominus bc$.

PROOF: For the proof of **(a)** see B/G prop.1.27. The proofs of **(b)** – **(e)** are left as an exercise. ■

3.4 Order Relations in Integral Domains

Introduction 3.5. It is possible to introduce an order $a < b$ on certain integral domains $(R, \oplus, \odot)$ by marking the elements of an appropriate subset P of R as positive and saying that x is less than y if the difference $y \ominus x$ is positive, i.e., if $y \ominus x \in P$. This is how we proceed with integers, real and rational numbers. For each of those three number systems the set $P = \{x : x > 0\}$ plays that role.

For example 7 is less than 12 since $12 - 7 > 0$, and $-12 < -7$ since $-7 - (-12) > 0$. Moreover P satisfies the following: If $x, y \in P$, i.e., $x > 0$ and $y > 0$ then $x + y > 0$ and $xy > 0$, i.e., $x + y \in P$ and $xy \in P$. We also note that the number zero is not positive (not negative either), and that it is true for any number x that (either) $x < 0$ or $x = 0$ $x > 0$, i.e., $x \in P$ or $-x \in P$ or $x = 0$. □

The above motivates the following definition.

Definition 3.11 (Ordered Integral Domains)**. I.** Let $(R, \oplus, \odot)$ be an integral domain. Assume there exists $P \subseteq R$ which satisfies the following:

(a) If $p_1, p_2 \in P$ then $p_1 \oplus p_2 \in P$,
(b) If $p_1, p_2 \in P$ then $p_1 \odot p_2 \in P$,
(c) $0 \notin P$,
(d) Let $a \in R$. Then at least one of the following is true: $a \in P$, $\ominus a \in P$, $a = 0$.

We call P a **positive cone** on the integral domain R.

II. We use P to define on R an "order relation" a < b as follows: Let $a, b \in R$. We define

(3.25) $a \ < b$ if and only if $b \ominus a \in P$ ("a **is less than** b''),
(3.26) $a \ \leq b$ if and only if $a < b$ or $a = b$, ("a **is less than or equal** b''),
(3.27) $a \ > b$ if and only if $b < a$, ("a **is greater than** b''),
(3.28) $a \ \geq b$ if and only if $b \leq a$. ("a **is greater than or equal** b''),

We say that < is the **order induced by** P, and we call the quadruple $(R, \oplus, \odot, P)$ an **ordered integral domain**. Let $a \in R$. If $a \in P$ then we call a a **positive** element of R, and if $\ominus a \in P$ then we call a a **negative** element of R. If a is positive or zero then we call a **nonnegative**, and if a is negative or zero then we call a **nonpositive**. □

Remark 3.8. It may seem obvious to you that property **(d)** of a positive cone implies that an element of an ordered integral domain cannot be both positive and negative, but this requires proof! The above will follow easily from prop.3.33 on p.70. □

The next proposition gives the integers $\mathbb{Z}$, the fractions $\mathbb{Q}$, and the real numbers $\mathbb{R}$ as examples of ordered integral domains. It uses the naive definitions of ch.2 (Preliminaries about Sets, Numbers and Functions) for those sets. We will see in ch.6 and in ch.9 that things are the other way around $\mathbb{Z}$ and $\mathbb{R}$ are defined there in an exact manner as ordered integral domains (with additional properties).

Proposition 3.32.

Each of the following algebraic structures is an ordered integral domain:
- ***(a)*** $(\mathbb{Z}, +, \cdot, \mathbb{N})$: *The integers with addition and multiplication: The positive cone is the subset of all natural numbers.*
- ***(b)*** $(\mathbb{Q}, +, \cdot, \mathbb{Q}_{>0})$: *The rational numbers with addition and multiplication: The positive cone $\mathbb{Q}_{>0}$ is the subset of all fractions $\frac{m}{n}$. where both m, n are positive integers.* [31]
- ***(c)*** $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$: *The real numbers with addition and multiplication. The positive cone here is* $]0, \infty[$.

There is no suitable positive cone to define an order on the complex numbers $(\mathbb{C}, +, \cdot)$ with addition and multiplication. [32]

PROOF: This will be obvious when we see the exact definitions for $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$. ■

Notation: In this entire chapter we assume that a fixed ordered integral domain $(R, \oplus, \odot, P)$ is given and phrases such as "let $a \in R$" refer to elements of that integral domain. We further assume that order relations such as "$a < b$" and "$a \geq b$" refer to the order induced by the positive cone P.

For the following see Definition 2.19 on p.27 and the subsequent notation alert 2.1 concerning intervals of integers, real numbers and rational numbers. Convince yourself that the definitions and

[32] Ignore this example if you have not learned about complex numbers.

notations given there are consistent with the following ones for intervals in ordered integral domains.

Definition 3.12 (Intervals in Ordered Integral Domains)**.**

(A) For the following let $a, b \in (R, \oplus, \odot, P)$.

$[a, b]_R := \{x \in R : a \leq x \leq b\}$ is called the **closed interval** with endpoints a and b.
$]a, b[_R := \{x \in R : a < x < b\}$ is called the **open interval** with endpoints a and b.
$[a, b[_R := \{x \in R : a \leq x < b\}$ and $]a, b]_R := \{x \in R : a < x \leq b\}$ are called **half-open intervals** with endpoints a and b.

(B) We generalize the symbol "∞" from real numbers (see Definition 2.19 on p.27) to arbitrary ordered integral domains as follows. The symbol "∞" stands for an object which itself is not an element of $(R, \oplus, \odot, P)$ but is larger than any of its elements, and the symbol "$\ominus\infty$" stands for an object which itself is not an element of $(R, \oplus, \odot, P)$ but is smaller than any of its elements. We thus have $\ominus\infty < x < \infty$ for any $x \in R$. We write $\overset{\oplus}{\underset{\ominus}{}}\infty$ when we mean "either $\oplus\infty$ or $\ominus\infty$."

We now define

$$]\ominus\infty, a]_R := \{x \in R : x \leq a\} \qquad]\ominus\infty, a[_R := \{x \in R : x < a\}$$
$$]a, \infty[_R := \{x \in R : x > a\} \qquad [a, \infty[_R := \{x \in R : x \geq a\}. \quad \square$$

Remark 3.9. The above definition (part **A**) to be precise) does not assume that $a < b$: If $a > b$ then $[a, b[_R =]a, b[_R =]a, b]_R = [a, b]_R = \emptyset$. If $a = b$ we obtain $[a, a[_R =]a, a[_R =]a, a]_R = \emptyset$, and $[a, a] = \{a\}$. $\square$

Remark 3.10. We are in a very similar situation to that of the introductory remark 3.4 of ch.3.3 on p.64. This time you look at ch.2.1 and 2.2, and then at ch.8.2, of [2] Beck/Geoghegan: The Art of Proof. Again you notice that identical propositions and theorems are given there: First for integers in ch.2, and then again for real numbers in ch.8. Now the reason is prop.3.32 on p.69: Both $(\mathbb{Z}, +, \cdot, \mathbb{N})$ and $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$ are ordered integral domains. By the way, so are the rational numbers when ordered by $\mathbb{Q}_{>0}$. Because of this a proposition involving inequalities, e.g., $x < y \Rightarrow x \oplus z < y \oplus z$, need not be demonstrated separately for $\mathbb{Z}$, for $\mathbb{Q}$, and then again for $\mathbb{R}$. We will state and prove such statements for ordered integral domains, and it follows that they hold for all three sets of numbers.

As in ch.3.3 we will merely rephrase propositions and theorems of [2] Beck/Geoghegan: The Art of Proof. We again use "$\oplus$" instead of "$+$", "$\ominus$" instead of "$-$" and "$\odot$" instead of "$\cdot$" for addition, subtraction and multiplication to remind the reader that these rules apply to any other ordered integral domains. $\square$

We begin with a sharpening of part **(d)** of the definition of an ordered integral domain. It says that an element of an ordered integral domain is either positive or negative or zero.

Proposition 3.33 (B/G prop.2.2 and B/G prop.8.27)**.**

*Let $a \in R$. Then **either** $a \in P$ **or** $\ominus a \in P$ or $a = 0$.*

PROOF: We examine separately the cases $a = 0$ and $a \neq 0$.

Case 1: $a = 0$.

Since $0 = \ominus 0$ (see Exercise 3.18 on p.83) it follows from Definition 3.11(c) that both $a = 0 \notin P$ and $\ominus a = \ominus 0 = 0 \notin P$. The proposition thus is correct.

Case 2: $a \neq 0$.

It follows in this case from Definition 3.11(d) that at least one of $a \in P$ or $\ominus a \in P$ is true. Thus there are only three possibilities:

(2a) $a \in P, \ominus a \notin P$,
(2b) $a \notin P, \ominus a \in P$,
(2c) $a \in P, \ominus a \in P$,

Case 2 holds if a satisfies **(2a)** or **(2b)** but it is false if a satisfies **(2c)**. Thus all that remains to be shown for **Case 2** is that **(2c)** can be ruled out.

To state this positively, we want to prove the following:

(*) At most one $a, \ominus a$ is an element of P.

This will be done by means of an **indirect proof**.

> In an indirect proof we play devil's advocate and assume to the contrary that our assertion is false, (i.e., we assume that both $a \in P$ and $\ominus a \in P$), and we show that this assumption leads to a contradiction. Since the only way to avoid that contradiction is not to assume that our assertion is false. It thus must be true. (In this particular case: the statement **(*)** thus must be true.)

So assume to the contrary that both $a \in P$ and $\ominus a \in P$. It follows from Definition 3.11(a) that

$$a \oplus (\ominus a) \in P \qquad \text{i.e.,} \qquad 0 \in P\,.$$

This contradicts property (c) of the positive cone P and we conclude that the assumption that **(*)** is false is faulty. In other words, **(*)** is correct.

We have thus shown that the case **(2c)** can be ruled out and thus **Case 2** holds.

Since **Case 1** and **Case 2** cover all possibilities the entire proof is completed. ■

Remark 3.11.

(a) Prop.3.33 can be restated as follows: Let $a \in R$. Then either a is positive or a is negative or $a = 0$.

(b) It easily follows from prop.3.33 that 0 is the only element of R which is both nonnegative and nonpositive. □

Prop.3.32 on p.69 had introduced the following three ordered integral domains of number systems that you have been very familiar with even before you entered college: The integers $(\mathbb{Z}, +, \cdot, \mathbb{N})$, the real numbers $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$, and the rational numbers $(\mathbb{Q}, +, \cdot, \mathbb{Q}_{>0})$. The positive cones which induce the "$<$" order relation are $\mathbb{Q}_{>0}$ for the rational numbers and $\mathbb{R}_{>0}$ for the real numbers. This makes perfect sense, as we have been taught to call numbers positive if and only if they are greater than zero. We chose $\mathbb{N}$ as the positive cone for the integers, and that fits the general mold because $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\} = \mathbb{N}_{>0}$. It is remarkable that the equation $P = \{x \in R : x > 0\}$, as the next proposition demonstrates, is true for any ordered integral domain.

Proposition 3.34 (B/G prop.2.13 and B/G prop.8.38)**.**

If $(R, \oplus, \odot, P)$ *is an ordered integral domain, then* $\quad P \ = \ \{ \, x \in R \ : \ x > 0 \, \}.$

Proof strategy:
We will first prove that $P \subseteq \{x \in R : x > 0\}$ and afterwards that $P \supseteq \{x \in R : x > 0\}$.

PROOF of "$\subseteq$": Let $p \in P$. Then $p \ominus 0 = p \in P$, hence $p > 0$, hence $p \in \{x \in R : x > 0\}$.

PROOF of "$\supseteq$": Let $p \in \{x \in R : x > 0\}$. Then $p > 0$, hence $p \ominus 0 \in P$, i.e., $p \in P$. ■

Remark 3.12. Since $P = \{x \in R : x > 0\}$ (see prop.3.34) 3.11**(a)** can be formulated as follows in the language of sets: $R \ = \ P \biguplus (\ominus P) \biguplus \{0\}$. In other words, $\mathfrak{A} := \ \{ \ P, \ominus P, \{0\} \ \}$ is a partition of the set R in the sense of Definition 2.11 on p.22. □

We have established that $\ P \ = \ \{ \ x \in R : \ x > 0\}$. However, it is not self–understood that $1 > 0$ (equivalently, $1 \in P$).

Proposition 3.35 (B/G prop.2.3 and B/G prop.8.28)**.**

The multiplicative unit 1 *of* R *belongs to* P.

PROOF: The proof is left as exercise 3.8 (see p.82). ■

Proposition 3.36. *If* $a \in R$ *then* $a \oplus 1 > a$.

Proof: Left as an exercise. ■

Corollary 3.3. $1 \ > \ 0$.

PROOF: This follows from $1 = 1 \oplus 0$ and prop.3.36, applied to $a := 0$. ■

Proposition 3.37 (B/G prop.2.4 and B/G prop.8.29)**.** *Let* $a, b, c \in R$.

$$\text{(3.29)} \qquad \textit{If} \quad a \ < \ b \quad \textit{and} \quad b \ < \ c, \quad \textit{then} \quad a < c.$$

PROOF: Adopt the proof of B/G prop.2.4. ■

Proposition 3.38. *Let* $a, b, c \in R$.

$$\text{(3.30)} \qquad \textit{If} \quad a \ \leq \ b \quad \textit{and} \quad b \ \leq \ c, \quad \textit{then} \quad a \leq c.$$

PROOF: There are four cases.

(1) $a < b, b < c$: It follows from B/G prop.2.4 (transitivity of "$<$") that $a < c$, in particular, $a \leq c$.
(2) $a < b, b = c$: It follows that $a < c$. This implies $a \leq c$.
(3) $a = b, b < c$: It follows again that $a < c$, hence $a \leq c$.
(4) $a = b, b = c$: It follows that $a = c$. This implies $a \leq c$. ■

Proposition 3.39 (B/G prop.2.5 and B/G prop.8.30)**.** *For each $a \in R$ there exists $p \in P$ such that $a \oplus p > a$.*

PROOF: Left as an exercise. ■

Proposition 3.40 (B/G prop.2.6 and B/G prop.8.31)**.** *Let $a, b \in R$. If $a \leq b \leq a$ then $a = b$.*

PROOF: Left as an exercise. ■

Proposition 3.41 (B/G prop.2.7 and B/G prop.8.32)**.** *Let $a, b, c, d \in R$. Then*

(a) *If $a < b$ then $a \oplus c < b \oplus c$.*
(b) *If $a < b$ and $(c < d)$ then $a \oplus c < b \oplus d$.*
(c) *If $0 < a < b$ and $0 < c \leq d$ then $ac < bd$.*
(d) *If $0 < a \leq b$ and $0 < c \leq d$ then $ac \leq bd$.*
(e) *If $a < b$ and $c < 0$ then $bc < ac$.*

PROOF of **(a)**, **(b)**, and **(e)**: Left as an exercise.

PROOF of **(c)**: Adopt the proof of B/G prop.2.7(iii).

PROOF of **(d)**: If $a < b$ then the proof follows from **(c)**. We thus may assume that $a = b$. But then we also may assume that $c < d$, since otherwise $c = d$, hence $bd = ac$, and nothing remains to prove.

It follows from $a > 0$ that $a = a \ominus 0 \in P$, and it follows from $c < d$ that $d \ominus c \in P$. Thus $a(d \ominus c) \in P$ by Definition 3.11**(b)** on p.68, i.e., $ad \ominus ac \in P$, hence $ad > ac$, hence $ad \geq ac$. ■

Proposition 3.42 (B/G prop.2.8 and B/G prop.8.33)**.** *Let $a, b \in R$. Then either $a < b$ or $a = b$ or $a > b$.*

PROOF: Left as an exercise. ■

Proposition 3.43. *Let $a, b \in R$. Then*

(a) $ab > 0 \Leftrightarrow a, b > 0$ *or* $a, b < 0$,
(b) $ab < 0 \Leftrightarrow$ [*either* $a > 0$ *and* $b < 0$] *or* [$a < 0$ *and* $b > 0$]
(c) $ab = 0 \Leftrightarrow a = 0$ *or* $b = 0$

Proof strategy:
In the current situation it is sufficient to prove the "⇐" direction for each of **(a)**, **(b)**, **(c)** to get the "⇒" direction for free in all three cases. Why? Observe that, on account of prop.3.42, the three left hand sides of **(a)**, **(b)**, **(c)** are mutually exclusive and there is no fourth choice (**either** $ab > 0$ **or** $ab < 0$ **or** $ab = 0$), and that the same is also true for the three right hand sides.

Let us abbreviate the left hand side of **(a)** with LS(a), its right hand side with RS(a), the left hand side of **(b)** with LS(b), etc. Assume that we have proven RS(a) ⇒ LS(a), RS(b) ⇒ LS(b) and RS(c) ⇒ LS(c).

Why is it true that then also LS(a) ⇒ RS(a), LS(b) ⇒ RS(b) and LS(c) ⇒ RS(c)? We will show LS(b) ⇒ RS(b). The proof for the other two cases is obtained by the same reasoning:

Assume to the contrary that LS(b) is true but RS(b) is false. Then one of the other cases RS(a) or RS(c) must be true since there are no other options. But RS(a) is not true since RS(a) ⇒ LS(a) was shown to be correct, thus LS(a) is true. It follows that LS(b) is false since the three left hand expressions are mutually exclusive. We found a contradiction to our assumption that LS(b) is true. We replace "a" with "c" and the same argument show that RS(c) cannot be true either. Since both RS(a) and RS(c) are false and one of RS(a), RS(b), RS(c) must be true it follows that RS(b) is true. We have proven LS(b) ⇒ RS(b).

You should understand that there is nothing magical about the number 3. Assume you have proven the n statements RS(1) ⇒ LS(1), RS(2) ⇒ LS(2), . . . , RS(n) ⇒ LS(n) and that it is the case that **either** LS(1) **or** ... **or** LS(n) is true, and also that **either** RS(1) **or** ... **or** RS(n) is true.

It then follows that LS(1) ⇒ RS(1), LS(2) ⇒ RS(2), . . . , LS(n) ⇒ RS(n).

PROOF of the proposition:

PROOF of RS(a) ⇒ LS(a):

If $a, b > 0$ then the product ab is positive by Definition 3.11**(b)** on p.68 of a positive cone, and if $a, b < 0$ then the product $ab = \big((\ominus 1)a\big),\ \big(\ominus 1)b\big)$ is positive for the same reason.

PROOF of RS(b) ⇒ LS(b):

Assume that $a > 0$ and $b < 0$. It follows from prop.3.41**(e)** (setting $a = 0$) that $a \odot b < 0 \odot b$, i.e., $a \odot b < 0$. We now obtain the proof for $a < 0$ and $b > 0$ by switching the roles of a and b.

PROOF of RS(c) ⇒ LS(c): This is true since $u \odot 0 = 0$ for all $u \in R$. (See prop.3.9 on p,61). ■

Next should be the translation of B/G prop.2.9: If $a \in \mathbb{Z}$ and $a \neq 0$ then $a^2 \in \mathbb{N}$. This following proposition does exactly that if you remember that the positive cone for $(\mathbb{Z}, +, \cdot)$ is the set $\mathbb{N}$ of all natural numbers.

Proposition 3.44 (B/G prop.2.9 and prop.8.34)**.** *Let $a \in R$. If $a \neq 0$ then $a^2 \in P$.*

The proof is left as exercise 3.10 (see p.82). ■

Proposition 3.45 (B/G prop.2.10 and B/G prop.8.35)**.** *The equation $x^2 = \ominus 1$ has no solution (in R).*

The proof is left as exercise 3.11 (see p.82). ■

Proposition 3.46 (B/G prop.2.11 and B/G prop.8.36)**.** *Let $a \in R$ and $p \in P$. If $ap \in P$ then $a \in P$.*

PROOF: Left as an exercise. ■

Proposition 3.47 (B/G prop.2.12 and B/G prop.8.37)**.** *Let $a, b, c \in R$. Then*

(a) $\ominus\, a < \ominus b$ *if and only if* $a > b$.
(b) *If $c > 0$ and $ac < bc$ then* $a < b$.
(c) *If $c < 0$ and $ac < bc$ then* $b < a$.
(d) *If $a \leq b$ and $0 \leq c$ then* $ac \leq bc$.

PROOF: Left as an exercise. ■

The next proposition has been included to stay in sync with [2] Beck/Geoghegan: The Art of Proof, ch.2.

Proposition 3.48 (B/G prop.2.14(ii) and B/G prop.8.39)**.** *If $a \in P$ then $a \oplus 1 \in P$.*

PROOF: This follows from prop.3.35 on p.72 and Definition 3.11**(a)** (ordered integral domains) on p.72. ■

Ordered integral domains have enough structure to define absolute values.

Definition 3.13 (Absolute value)**.**

For an element x of the ordered integral domain R, we define its **absolute value** as
$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ \ominus x & \text{if } x < 0. \end{cases} \qquad \square$$

Here are some properties of squares and absolute values.

Proposition 3.49 (Generalization of B/G prop.10.5)**.** *Let $x, y \in P \cup \{0\}$, i.e., $x, y \geq 0$. Then*

- **(a)** $x \leq y$ *if and only if* $x^2 \leq y^2$,
- **(b)** $x = y$ *if and only if* $x^2 = y^2$,
- **(c)** $x < y$ *if and only if* $x^2 < y^2$.

The proof is left as exercise 3.12 (see p.83). ■

Proposition 3.50 (B/G prop.10.6)**.** *Let $a \in R$. Then $|a|^2 = a^2$.*

The proof is left as exercise 3.13 (see p.83). ■

Proposition 3.51 (B/G prop.10.7)**.** *Let $a, b \in R$. Then $|a| < |b| \quad \Leftrightarrow \quad a^2 < b^2$.*

PROOF: Left as an exercise. ■

The next two propositions are very similar. We will see in ch.11.2.2 (Normed Vector Spaces) that prop.3.52 shows that if $(R, \oplus, \odot, P) = (\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$ then the absolute value satisfies the properties of a norm.

The subsequent proposition 3.53 shows that the assignment $(a, b) \mapsto |b \ominus a|$ turns the ordered integral domain $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$ into a metric space. See Chapter12 (Metric Spaces and Topological Spaces – Part I).

Proposition 3.52 (B/G prop.10.8)**.** *Let $a, b \in R$. Then the following holds:*

- **(a)** $|a| = 0$ *if and only if* $a = 0$,
- **(b)** $|ab| = |a| \odot |b|$,
- **(c)** $\ominus |a| \leq a \leq |a|$,
- **(d)** $|a \oplus b| \leq |a| \oplus |b|$,
- **(e)** *if* $\ominus b < a < b$ *then* $|a| < b$*, in particular,* $b \geq 0$.

PROOF: The proofs of **(a)** and **(d)** can be found in [2] Beck/Geoghegan Art of Proof The proof of the other parts is left as an exercise. ■

Remark: Actually B/G prop.10.8**(e)** states that $\ominus b < a < b$ then $|a| < |b|$. But we see that $b \geq 0$ (and hence $b = |b|$) as follows: $\ominus b < a < b$ implies that $\ominus b < b$. Assume to the contrary that $b < 0$. Then $\ominus b > 0$, thus $\ominus b < a < b < 0 < \ominus b$, thus $\ominus b < \ominus b$. We have reached a contradiction.

Proposition 3.53 (B/G prop.10.10)**.** *If $a, b, c \in R$, then*

(a) $|a \ominus b| = 0 \Leftrightarrow a = b$,
(b) $|a \ominus b| = |b \ominus a|$,
(c) $|a \ominus b| \leq |a \ominus c| \oplus |c \ominus b|$,
(d) $|a \ominus b| \geq \big||a| \ominus |b|\big|$.

PROOF: The proof is left as exercise 3.14 (see p.83). ■

We will give two (very short) proofs for the next proposition. Which one do you prefer?

Proposition 3.54. *This proposition is similar to prop.3.52**(e)**.*
*Let $a, b \in R$ such that both **#1)** $\ominus a \leq b$ and **#2)** $a \leq b$. Then $|a| \leq b$.*

FIRST PROOF:
Case 1) $a \geq 0$: It follows from **#2** that $|a| = a \leq b$, which is what we had to show.
Case 2) $a < 0$: It follows from **#1** that $|a| = \ominus a \leq b$, which is what we had to show.
SECOND PROOF:
Here is an alternate proof which avoids using separate cases.
#1 is equivalent to $a \geq \ominus b$, thus **#1** and **#2** together yield, $\ominus b \leq a \leq b$. It follows from prop.3.52**(e)** above that $|a| < b$. ■.

3.5 Minima, Maxima, Infima and Suprema in Ordered Integral Domains

Notation: In this entire chapter we assume that a fixed ordered integral domain $(R, \oplus, \odot, P)$ is given and phrases such as "let $a \in R$" refer to elements of that integral domain. We further assume that order relations such as "$a < b$" and "$a \geq b$" refer to the order induced by the positive cone P. **Do not confuse** the symbol R for this integral domain with the symbol $\mathbb{R}$ for the real numbers!

We have seen in prop.3.42 that any two elements a, b of R can be compared: Either $a < b$ or $a = b$ or $a > b$. [33] This makes it possible to introduce boundedness, least upper bounds, greates lower bounds, maxima and minima for certain subsets of R. [34]

Definition 3.14 (Upper and lower bounds, maxima and minima)**.** Let $A \subseteq R$ and let $l, u \in R$.

(a) We call l a **lower bound** of A if $l \leq a$ for all $a \in A$.
(b) We call u an **upper bound** of A if $u \geq a$ for all $a \in A$.
(c) We call A **bounded above** if this set has an upper bound.
(d) We call A **bounded below** if A has a lower bound.
(e) We call A **bounded** if A is both bounded above and bounded below.
(f) A **minimum** (min) of A is a lower bound l of A such that $l \in A$.
(g) A **maximum** (max) of A is an upper bound u of A such that $u \in A$. □

[33] In ch.5.1 (Cartesian Products and Relations) we will call sets which carry such an order relation linearly, or totally, ordered. See Definition 5.5 on p.131.

[34] Those concepts will also be introduced for so called partially ordered sets. See Definition 15.1 on p.452.

Remark 3.13. The empty set does not possess a maximum or minimum because either would have to be an element of $\emptyset$. □

Proposition 3.55. *Let $A \subseteq R$. 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 $a \leq u_1$ for all $a \in A$. Hence $u_2 \leq u_1$.
Now we switch the roles of u_1 and u_2 and the same reasoning as above yields $u_1 \leq u_2$.
We thus have equality $u_1 = u_2$. The proof for minima is similar. ■

The last proposition makes it possible to write $\min(A)$ for the minimum of A and $\max(A)$ for the maximum of A in case those items exist for a subset A of R.

Definition 3.15. Let $A \subseteq R$. If A possesses a minimum then we write $\min(A)$ or $\min A$ for this uniquely determined element of R, and if A possesses a maximum then we write $\max(A)$ or $\max A$ for that uniquely determined element of R. □

Definition 3.16. $\boxed{\star}$ Let $A \subseteq R$. We define

$$\begin{aligned} A_{\mathit{lowb}} &:= \{l \in R : l \text{ is lower bound of } A\} \\ A_{\mathit{uppb}} &:= \{u \in R : u \text{ is upper bound of } A\}. \quad \square \end{aligned} \tag{3.31}$$

Remark 3.14. The sets A_{lowb} and/or A_{uppb} may be empty. Examples to that effect are $A = \mathbb{R}$, $A =]0, \infty[$, $A = A =]-\infty, 0[$. □

Remark 3.15. Note that A is bounded above if and only if $A_{\mathit{uppb}} \neq \emptyset$ and bounded below if and only if $A_{\mathit{lowb}} \neq \emptyset$. □

If A is a nonempty subset of R then the set A_{lowb} of its lower bounds need not not necessarily possess a maximum, but if $max(A_{\mathit{lowb}})$ exists then this element of R will be the greatest of all lower bounds of A. This warrants the following definition.

Definition 3.17 (Infimum and supremum in an ordered integral domain)**.** Let A be a nonempty subset of R.

(a) If $\max(A_{\mathit{lowb}})$ exists then it is unique by prop.3.55. We write $\inf(A)$ or g.l.b.(A) for $\max(A_{\mathit{lowb}})$ and call this number the **infimum** or **greatest lower bound** of A.

(b) If $\min(A_{\mathit{uppb}})$ exists then it is unique by prop.3.55. We write $\sup(A)$ or l.u.b.(A) for $\min(A_{\mathit{uppb}})$ and call this element of R the **supremum** or **least upper bound** of A. □

Remark 3.16. If the set A has no upper bounds then A_{uppb} is empty, hence does not possess a minimum (see rem3.13 above), hence $\sup(A)$ does not exist. Likewise, if A has no lower bounds then $\inf(A)$ does not exist. We will introduce infima and suprema for unbounded sets later in this chapter. See Definition 3.18 on p.79. □

Example 3.8.

(a) Let $A := \{\frac{1}{j} : j \in \mathbb{N}\}$. If we consider A as a subset of $\mathbb{Q}$ then $A_{lowb} =]-\infty, 0]_{\mathbb{Q}}$ possesses 0 as it maximum, i.e., $\inf(A) = 0$, but A has no minimum because $0 \notin A$.

(b) Let $A := \{\frac{1}{j} : j \in \mathbb{N}\}$, just as in **(a)**, but now we consider A as a subset of $\mathbb{R}$. Then $A_{lowb} =]-\infty, 0]$ possesses 0 as it maximum, i.e., $\inf(A) = 0$, but A has no minimum because $0 \notin A$. Thus the situation is the same as for $\mathbb{Q}$.

(c) [35] We remind the reader that the real number $\sqrt{2}$ is not rational. [36]

Let $B := \{\frac{n}{d} : n, d \in \mathbb{N} \text{ and } \frac{n^2}{d^2} < 2\}$, i.e., the set of all positive, rational, numbers with a square less than 2. Then $\inf(B) = 0$ and $\min(B)$ does not exist ($0 \notin B$!) regardless whether we consider B a subset of the integral domain $R = \mathbb{Q}$ or $R = \mathbb{R}$. However we have different outcomes for the upper "boundary" of B.

One can prove that $\sup(B)^2 = 2$, i.e., that $\sup(B) = \sqrt{2}$ exists as an element of $\mathbb{R}$. [37] On the other hand it follows from $\sup(B)^2 = 2$ that $\sup(B) \notin B$. Thus the set B has a supremum but not a maximum in $\mathbb{R}$.

In contrast to the above $\sup(B)$ does not exist in $\mathbb{Q}$ because, as mentioned, the square of $\sup(B)$ would have to be 2 and $\sqrt{2}$ is not rational. Thus B possesses neither supremum nor maximum in $\mathbb{Q}$.

(d) Let $(R, +, \cdot)$ be the ordered integral domain of either the rational or the real numbers, and let $C := \{k \in R : 0 < 2k < 7\}$. For both $R = \mathbb{Q}$ and $R = \mathbb{R}$ we have $\inf(C) = 0$, $\sup(C) = 3/2$. However, both $\min(C)$ and $\max(C)$ do not exist since neither 0 nor $3/2$ belongs to C.

(e) Let $R = \mathbb{Z}$ and $C := \{k \in R : 0 < 2k < 7\}$. Then $\min(C) = 1$ and $\max(C) = 3$. The reason: $1 \in C$ is a lower bound of C, $3 \in C$ is an upper bound of C, and 1 and 3 belong to R. □

We are staying away from using functions in the context of integral domains as much as possible because we use them mainly as a generalization of the number systems given by the integers, the rational numbers, and the real numbers. The following is an exception because this material on minima, maxima, infima and suprema is referred to in later chapters.

Notation 3.2.

Notational conveniences:

(a) We may drop the parentheses in expressions like $\max(A)$, $\sup(\{f(x) : x \in B\})$ (here $f : X \to R$ is a function which takes values in an ordered integral domain R and where $B \subseteq X$), etc., if this does not lead to any confusion. We also can write the above as $\max A$ and $\sup\{f(x) : x \in B\}$.

(b) If A consists of two elements $x, y \in R$, i.e., $A = \{x, y\}$ then it is customary to write $\max(x, y)$, $\min(x, y)$, $\sup(x, y)$, and $\inf(x, y)$. □

[35] An example very similar to this one is example 9.1 on p.258.

[36] See rem.2.8 on p.26.

[37] The proof is given in prop.9.25 on p.274.

Proposition 3.56. *Let $A \subseteq R$. If A has a maximum then it also has a supremum, and $\max(A) = \sup(A)$. Likewise, if A has a minimum then it also has an infimum, and $\min(A) = \inf(A)$.*

PROOF: The proof is left as exercise 3.15. ■

Remark 3.17. One can say informally that a supremum is a generalized maximum – generalized in the sense that it need not belong to the set under consideration. Examples for this are given in ch.9.2 when looking at the ordered integral domain of the real numbers. See examples 9.2 and 9.3 on p.259.

Proposition 3.57. *Let $\emptyset \neq A \subseteq B \subseteq R$.*

- ***(a)*** *If both A and B possess an infimum (resp., supremum) then* $\inf(A) \geq \inf(B)$ *(resp.,* $\sup(A) \leq \sup(B)$*).*
- ***(b)*** *If both A and B possess a minimum (resp., maximum) then* $\min(A) \geq \min(B)$ *(resp.,* $\max(A) \leq \max(B)$*).*
- ***(c)*** *If both A and B possess a minimum (resp., maximum) and $\min(B) \notin A$ (resp., $\max(B) \notin A$) then* $\min(A) > \min(B)$ *(resp.,* $\max(A) < \max(B)$*).*

PROOF: We prove **(a)** for suprema. The proof for infima is similar. It follows from $A \subseteq B$ that any upper bound of B also is an upper bound of A. We obtain in particular $\sup(B) \in A_{uppb}$, hence $\sup(A) = \min(A_{uppb}) \leq \sup(B)$.

Note that **(b)** follows from **(a)** because if a set has a minimum then it equals its infimum and if a set has a maximum then it equals its supremum.

We now prove **(c)** for minima. The proof for maxima is similar. If $\min(B) \notin A$ then $\min(A) \in A$ implies $\min(B) \neq \min(A)$. That together with $\min(B) \leq \min(A)$ yields $\min(B) < \min(A)$. ■

Remark 3.18. It is convenient to define $\inf(A)$ if $A \subseteq R$ is empty or has no lower bounds and to define $\sup(A)$ if A is empty or has no upper bounds. If A has no lower bounds (and hence is not empty)

(A) We recall that if $\inf(A)$ exists then it is a lower bound of A, and if $\sup(A)$ exists then it is an upper bound of A. Thus $\inf(A) \leq a \leq \sup(A)$ for all $a \in A$. If A is not bounded below then $\ominus\infty$ is the only object x that satisfies $x \leq a$ for all $a \in A$; if A is not bounded above then ∞ is the only object x that satisfies $x \geq a$ for all $a \in A$. It thus makes sense to define $\inf(A) := \ominus\infty$ if A has no lower bounds and $\sup(A) := \infty$ if A has no upper bounds.

(B) Prop.3.57**(a)** above suggests how to handle the empty set: Since $\emptyset \subseteq B$ for all $B \subseteq R$ and the infimum becomes bigger for smaller sets we would want $\inf(\emptyset)$ to be as large as possible, i.e., $\inf(\emptyset)$ should be ∞. Further $\sup(\emptyset)$ should be as small as possible, i.e., this value should be $\ominus\infty$. □

So we arrive at the following definition.

Definition 3.18 (Supremum and Infimum of unbounded and empty sets). $\boxed{\star}$ Let $A \subseteq R$. If A is not bounded above then we define

$$\sup A = \infty \tag{3.32}$$

If A is not bounded below then we define

$$\inf A = \ominus\infty \tag{3.33}$$

Finally we define [38]

$$\sup\emptyset = \ominus\infty, \qquad \inf\emptyset = \oplus\infty. \quad \square \tag{3.34}$$

Remark 3.19. Be aware that even though we allow $\sup(A) = \genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$ and $\inf(A) = \genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$ we do not allow $\max(A) = \genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$ or $\min(A) = \genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$.
The reason: By definition of, e.g., the maximum, if $\max(A)$ exists then it must satisfy $\max(A) \in A$, hence $\max(A) \in R$. Since the infinity values are not elements of R it is not possible that, e.g., $\max(A) = \genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$. $\square$

Proposition 3.58. *Let $A \subseteq B \subseteq R$.*

(a) *If $\inf(A)$ and $\inf(B)$ both exist then $\inf(A) \geq \inf(B)$.*
(b) *If $\sup(A)$ and $\sup(B)$ both exist then $\sup(A) \leq \sup(B)$.*

PROOF:
The above was proven in prop.3.57 under the condition that $\inf(A)$, $\inf(B)$, $\sup(A)$, $\sup(B)$ exist as elements of R, i.e., we did not permit the values $\genfrac{}{}{0pt}{}{\oplus}{\ominus}\infty$.
We prove this proposition for suprema. The proof for infima is similar. If $A \neq \emptyset$ (hence $B \neq \emptyset$) and B is bounded above (hence A is bounded above) then **(B)** follows from prop.3.57.

Otherwise $A = \emptyset$ in which case $\sup(A) = \ominus\infty \leq \inf(B)$, or B is not bounded above, in which case $\sup(A) \leq \infty = \sup(B)$. In either case **(B)** holds. ■

Recall for the following that $\ominus A = \{\ominus a : a \in A\}$ (see Definition 3.8 on p.61).

Proposition 3.59. *Let $A \subseteq R$ and $x \in R$. Then*

$$x \leq a \ \textit{ for all } a \in A \quad \Leftrightarrow \quad \ominus x \geq a' \ \textit{ for all } a' \in \ominus A, \tag{3.35}$$
$$x \in A_{\mathit{lowb}} \quad \Leftrightarrow \quad \ominus x \in (\ominus A)_{\mathit{uppb}}, \tag{3.36}$$
$$\ominus A_{\mathit{lowb}} \quad = \quad (\ominus A)_{\mathit{uppb}}, \tag{3.37}$$
$$x \geq a \ \textit{ for all } a \in A \quad \Leftrightarrow \quad \ominus x \leq a' \ \textit{ for all } a' \in \ominus A, \tag{3.38}$$
$$x \in A_{\mathit{uppb}} \quad \Leftrightarrow \quad \ominus x \in (\ominus A)_{\mathit{lowb}}, \tag{3.39}$$
$$\ominus A_{\mathit{uppb}} \quad = \quad (\ominus A)_{\mathit{lowb}}, . \tag{3.40}$$

PROOF: We have

$$x \leq a \text{ for all } a \in A \quad \Leftrightarrow \quad \ominus x \geq \ominus a \text{ for all } a \in A \quad \Leftrightarrow \quad \ominus x \geq \ominus a \text{ for all } \ominus a \in \ominus A.$$

Replacing $\ominus a$ with a' yields (3.35). We now obtain (3.36) from (3.35) and (3.31). (3.37) follows from

$$x \in \ominus A_{\mathit{lowb}} \ \Leftrightarrow \ \ominus x \in A_{\mathit{lowb}} \overset{(3.36)}{\Longleftrightarrow} \ominus(\ominus x) \in (\ominus A)_{\mathit{uppb}} \ \Leftrightarrow \ x \in (\ominus A)_{\mathit{uppb}}.$$

We exchange the roles of $\leq$ and $\geq$ and apply similar arguments to obtain (3.38) through (3.40). ■

[38] These definitions for the empty set will also work harmoniously with (8.1) on p.232 in ch.8.1 (More on Set Operations) where $\bigcup_{i\in\emptyset} A_i$ and $\bigcap_{i\in\emptyset} A_i$ are defined.

Proposition 3.60. *Let $\emptyset \neq A \subseteq R$. If the maximum of A_{lowb} exists then A has lower bounds, $\ominus A$ has lower bounds, the minimum of $(\ominus A)_{uppb}$ exists, and we have*

$$\ominus \max(A_{lowb}) = \min\big((\ominus A)_{uppb}\big), \tag{3.41}$$

$$\ominus \min(A_{uppb}) = \max\big((\ominus A)_{lowb}\big). \tag{3.42}$$

PROOF: Let $a^\star := \max(A_{lowb})$. Since a set contains it maximum, $a^\star \in A_{lowb}$, hence $\ominus a^\star \in \ominus A_{uppb}$ by (3.36). To prove that $\ominus a^\star = \min\big((\ominus A)_{uppb}\big)$ we must show that $u \geq \ominus a^\star$ for all $u \in (\ominus A)_{uppb}$.
So let $u \in (\ominus A)_{uppb}$. It follows from (3.31) that $u \geq a'$ for all $a' \in \ominus A$, hence, by (3.41), $\ominus u \leq a$ for all $a \in A$. Thus $\ominus u$ is a lower bound of A, i.e., $\ominus u \in A_{lowb}$. Since $a^\star = \max(A_{lowb})$ it follows that $\ominus u \leq a^\star$, i.e., $u \geq \ominus a^\star$. This is what we needed to show. ■

Corollary 3.4. *The following equations are to be understood in the sense that if the item on the left exists and vice versa, and both sides then are equal.*

$$\ominus \inf(A) = \sup(\ominus A), \tag{3.43}$$

$$\ominus \sup(A) = \inf(\ominus A), \tag{3.44}$$

$$\ominus \min(A) = \max(\ominus A). \tag{3.45}$$

$$\ominus \max(A) = \min(\ominus A), \tag{3.46}$$

PROOF: (3.43) is obtained from (3.41) and (3.44) is obtained from (3.42) by the very definition of suprema as minimal upper bounds and infima as maximal lower bounds. We now obtain (3.45) and (3.46) from prop.3.56 on p.79. ■

Draw a picture for numbers a, b, c to visualize the content of the following proposition and its corollary.

Proposition 3.61. *Let a, b be nonnegative elements of R. Then*

$$|b \ominus a| \leq \max(a,b), \quad \text{i.e.,} \tag{3.47}$$

$$\ominus \max(a,b) \leq b \ominus a \leq \max(a,b). \tag{3.48}$$

PROOF: The proof is left as exercise 3.16 (see p.83). ■

Corollary 3.5. *Let $a, b, c \in R$ such that $0 \leq a, b < c$. Then*

$$\ominus c < b \ominus a < c. \tag{3.49}$$

PROOF: Follows from prop.3.61 with $c := \max(a,b)$. ■

3.6 Exercises for Ch.3

Exercise 3.1. From example 3.1 on p.51:

(a) What laws for the addition of numbers do you need to use to prove that $(\mathbb{Z}, +)$ (the integers with addition) is an abelian group?
(b) What laws for the multiplication of numbers do you need to use to prove that $(\mathbb{Q} \setminus \{0\}, \cdot)$ (the nonzero rational numbers with with multiplication) is an abelian group?
(c) What about $(\mathbb{R}, \cdot)$ Is that a semigroup? a monoid? a group? an abelian group? □

Exercise 3.2. Let $(G, \diamond)$ be a commutative group with neutral element e. Let $g, h_1, h_2 \in G$ such that

$$g \diamond h_1 \ = e \qquad \text{and} \qquad g \diamond h_2 \ = e\,.$$

Use the material before Proposition 3.2 on p.53 to prove that $h_1 = h_2$. □

Exercise 3.3. Let $(S, \diamond)$ be a semigroup. Let $a, b, c, d \in S$. Use the definition of a semigroup only to prove that
(a) $a \diamond \big(b \diamond (c \diamond d)\big) \ = \ (a \diamond b) \diamond (c \diamond d)$,
(b) $\big(a \diamond (b \diamond c)\big) \diamond d \ = \ (a \diamond b) \diamond (c \diamond d)$. □

Exercise 3.4. Let $(S, \diamond)$ be a commutative semigroup, i.e., S is a semigroup which satisfies $s \diamond t \ = \ t \diamond s$ for all $s, t \in S$. Let $a, b, c \in S$. Prove that

$$a \diamond (b \diamond c) \ = \ c \diamond (a \diamond b)$$

Exercise 3.5. Prop.3.3 on p.54 of this document states that if g, h are elements of a group $(G, \diamond)$ then $h \diamond g^{-1} = (g \diamond h^{-1})^{-1}$. The proof only demonstrated that $(h \diamond g^{-1}) \diamond (g \diamond h^{-1}) = e$. Prove what has been omitted, i.e., the equation $(g \diamond h^{-1}) \diamond (h \diamond g^{-1}) = e$. □

Exercise 3.6. Prove part **(b)** of prop.3.10 on p.62 of this document:

Let $(R, \oplus, \odot)$ be a nonempty set with two binary operations $\oplus$ and $\odot$ which satisfies **(a), (b), (c)** of Definition 3.7. Then the following is true:

$1 = 0 \ \Leftrightarrow \ R \ = \ \{0\}\,.$ □

Exercise 3.7. Let $(R, \oplus, \odot)$ be an integral domain and $a, b, c, d \in R$. Prove that $\big(a \oplus (b \oplus c)\big) \oplus d \ = (a \oplus b) \oplus (c \oplus d)$. Do so without using the results of prop.3.17! □

Exercise 3.8. Prove prop.3.35 on p.72: The multiplicative unit 1 of R belongs to P. □

Exercise 3.9. Use everything up to AND including prop.3.35 on p.72 to prove prop.3.36 on p.72 of this document: If R is an ordered integral domain and $a \in R$ then $a \oplus 1 > a$. □

Exercise 3.10. Prove prop.3.44 on p.74 of this document:

If $a \in (R, \oplus, \odot, P)$ and $a \neq 0$ then $a^2 \in P$. □

Exercise 3.11. Prove prop.3.45 on p.74 of this document:

The equation $x^2 = \ominus 1$ has no solution (in R). □

Exercise 3.12. Prove prop.3.49 on p.75 of this document: If $x, y \in P \cup \{0\}$ then

(a) $x \leq y$ if and only if $x^2 \leq y^2$,
(b) $x = y$ if and only if $x^2 = y^2$,
(c) $x < y$ if and only if $x^2 < y^2$.

Hint: Do the proof of **(a)**, $\Leftarrow$) separately for $x^2 = y^2$ and $x^2 < y^2$. $\square$

Exercise 3.13. Prove prop.3.50 on p.75 of this document:
Let $a \in R$. Then $|a|^2 = a^2$. $\square$

Exercise 3.14. Prove prop.3.53 on p.76 of this document: Let $a, b, c \in R$. Then the following are true.

(a) $|a \ominus b| = 0 \;\Leftrightarrow\; a = b$,
(b) $|a \ominus b| = |b \ominus a|$,
(c) $|a \ominus b| \leq |a \ominus c| \oplus |c \ominus b|$,
(d) $|a \ominus b| \geq \big||a| \ominus |b|\big|$. $\square$

Exercise 3.15. Prove prop.3.56 on p.79 of this document: Let $A \subseteq R$. If A has a maximum then it also has a supremum, and $\max(A) = \sup(A)$. Likewise, if A has a minimum then it also has an infimum, and $\min(A) = \inf(A)$. $\square$

Exercise 3.16. Prove prop.3.61 on p.81 of this document: Let a, b be nonnegative elements of R. Then

$$|b \ominus a| \;\leq\; \max(a,b), \quad \text{i.e.,} \quad \ominus \max(a,b) \;\leq\; b \ominus a \;\leq\; \max(a,b).$$

Hint: Handle separately the cases $b \geq a$ and $b < a$. $\square$

Exercise 3.17. Let $R := (R, \oplus, \odot, P)$ be an ordered integral domain and $W \subseteq R$.

(A) Prove that the set W_{uppb} of the upper bounds of W satisfies either of the following:

(a) $W_{\mathit{uppb}} \;=\; [z, \infty[_R$ for some suitable $z \in R$,
(b) $W_{\mathit{uppb}} \;=\,]z, \infty[_R$ for some suitable $z \in R$,
(c) $W_{\mathit{uppb}} \;=\,]\ominus\infty, \infty[_R$ (i.e., $W_{\mathit{uppb}} \;=\; R$).

(B) For what sets W is $W_{\mathit{uppb}} \;=\; R$? $\square$

Exercise 3.18. Prove that if $(G, \diamond)$ is a group with neutral element e then e has itself as an inverse, i.e.,

$$e^{-1} \;=\; e\,. \; \square \tag{3.50}$$

4 Logic $\boxed{\star}$

This chapter uses material presented in ch.2 (Logic) and ch.3 (Methods of Proofs) of [5] Bryant, Kirby Course Notes for MAD 2104.

4.1 Statements and Statement Functions

Note 4.1 (Textual variables)**.** It was mentioned in **(c)** of the introduction to ch.2.4 (A First Look at Functions, Sequences and Families) that the input variables and function values of a function 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 whereas 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." □

Definition 4.1 (Statements)**.** A **statement** [39] 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 4.1. 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 4.2. 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!" □

[39] 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.

For the remainder of the entire chapter on logic we define

$$\mathscr{S} := \text{ the set of all statements} \tag{4.1}$$

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

Be sure to understand the material of ch.2.4 (A First Look at Functions, Sequences and Families) on p.29) before continuing.

Definition 4.2 (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(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 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 (4.1)) just as long as $x_{j0} \in \mathscr{U}_j$. In other words we have a function

$$A : \mathscr{U}_1 \times \mathscr{U}_2 \times \cdots \times \mathscr{U}_n \to \mathscr{S}, \quad (x_1, x_2, \dots, x_n) \mapsto A(x_1, x_2, \dots, x_n) \tag{4.2}$$

in the sense of def. 2.22. 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** [40] or **predicate**. □

Note 4.2 (Relaxed notation for statement functions)**.** You should remember that a statement function is a function in the sense of Definition 2.22 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

[40] 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.

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 4.3. In Definition 4.2 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

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

(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 of the following convention:

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

Example 4.4. 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 4.5. 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 4.6. The following are predicates:

- **(a)** $P :=$ "$2x = 27$" (see example 4.2**(b)**), UoD $\mathscr{U} := \{x \in \mathbb{R} : x > 10\}$
- **(b)** $Q :=$ "$x^2 + y^2 = 34$" (example 4.2**(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 4.7. Let $B :=$ "$x + 7 = 16$ and d is a dog". Let $\mathscr{U}_x := \mathbb{N}$ and $\mathscr{U}_d := \{d : d$ 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)** $B(9,$ 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)** $B(20,$ 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. [41] □

4.2 Logic Operations and their Truth Tables

We now resume our discussion of statements.

4.2.1 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. [42] Their meaning will be explained subsequently, once we define compound statements and compound statement functions.

[41] 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.

[42] 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.

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

Notation 4.1 (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 4 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 4.3 (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 4.8. Statements **(e)** and **(f)** of example 4.1 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**. □

4.2.2 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 4.2.1.

Definition 4.4 (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**.

(4.3) This is expressed in this "truth table" for $\neg A$:[43]

A	¬ A
F	T
T	F

□

Example 4.9. 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.

[43] The definition of a truth table will be given shortly. See Definition 4.6 on p.89.

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

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 “$\neg$” is a **unary operators**, i.e., it has a single input parameter. We will keep referring to “$\neg$” as a connective even though there are no two or more items that can be connected.

Definition 4.5 (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**.

(4.4) 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 $A_1, A_2, \dots, A_n$ in the obvious manner: $A_1 \wedge A_2 \wedge \cdots \wedge A_n$ is **true** if and only if each one of $A_1, A_2, \dots, A_n$ is **true** and **false** otherwise. □

Definition 4.6 (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 4.5 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 which 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 4.7 (Logically impossible)**.** The statements A and B in the truth table of Definition 4.6 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 [44] 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 4.1. 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 4.2. 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.

[44] 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

Example 4.10. Let U, V, W, Z be the statement functions
$U := x \mapsto U(x) :=$ “$x \in [0,4]$”,
$V := x \mapsto V(x) :=$ “$x \notin \emptyset$”,
$W := x \mapsto W(x) :=$ “$x < -1$”,
$Z := x \mapsto Z(x) :=$ “$x > 2$”
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 4.3. As in example 4.10 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 [45] $Q(x) := \neg(U(x) \wedge V(x)) \wedge W(x) \wedge Z(x)$ 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 4.10 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 4.8 (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. □

[45] 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 4.11. Here are some examples of tautologies.

(a) The statements $R(x)$ of remark 4.3 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 4.3 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 4.12. 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 $y > x \geq 1 > 0$, 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

□

4.2.3 Biconditional and Logical Equivalence Operators – Part 1

Definition 4.9 (Double arrow operator (biconditional))**.** The **double arrow operator** [46] 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**.

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

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

□

Definition 4.10 (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.

(4.6) 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.4.2.6 (Biconditional and Logical Equivalence Operators – Part 2) on p.100

4.2.4 Inclusive and Exclusive Or

Definition 4.11 (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**.

(4.7) 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**. □

[46] [5] 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 $A \ \Leftrightarrow \ B$.

Example 4.13. As in example 4.12 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 4.12**(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 4.12**(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 4.12 (Exclusive or)**.** The **exclusive or operator** is represented by the symbol “**xor**”.[47] 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).

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

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

□

Example 4.14. As in example 4.12 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

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

(b) Let $x < y$. We have seen in example 4.12**(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 4.12**(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 Definition 4.8 on p.91. □

[47] Some documents such as [5] Bryant, Kirby Course Notes for MAD 2104. also use the symbol $\oplus$.

Remark 4.4. 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 $\neg[A \textbf{ xor } B]$ is true and false whenever $\neg[A \textbf{ xor } B]$ is false. □

Exercise 4.1. 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. □

4.2.5 Arrow and Implication Operators

Definition 4.13 (Arrow operator)**.** The **arrow operator** [48] is represented by the symbol "$\rightarrow$". We read $A \rightarrow B$ as "A arrow B" but see remark 4.6 below for the interpretation "if A then B".

(4.9) Truth table for $A \rightarrow B$:

A	B	$A \rightarrow B$
F	F	T
F	T	T
T	F	F
T	T	T

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

Definition 4.14 (Implication operator)**.** We say that A **implies** B and we write

$$A \Rightarrow B \tag{4.10}$$

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

(4.11) 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 4.5. 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

□

[48] [5] 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 $A \Rightarrow B$.

Theorem 4.1 (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 4.2 (Transitivity of "$\to$")**.** *Let A, B, C be three statements. Then $\big[(A \to B) \land (B \to C)\big] \;\Rightarrow\; (A \to C)$.*

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

					P :=		
A	B	C	$A \to B$	$B \to C$	$(A \to B) \land (B \to C)$	$A \to C$	$P \to (A \to C)$
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 $\big[(A \to B) \land (B \to C)\big] \;\to\; (A \to C)$ contains **true** everywhere and we have proved that this statement is a tautology. ■

Definition 4.15. In the context of $A \to B$ and $A \Rightarrow B$ we call A the **premise** or the **hypothesis** [49] and we call B the **conclusion**. [50]

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 4.6.

[49] also called the **antecedent**

[50] Another word for conclusion is **consequent** .

(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 Definition 4.8 on p.91.

(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 4.15. 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 (4.37) on p.104. □

Remark 4.7. 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:

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" [51]

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

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

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”

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 Definition 4.13:

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

The truth table (4.11) 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 Definition 4.13 and Definition 4.14 and from there we justified why these operators appropriately model “if A then B”. □

Example 4.16. As in example 4.12 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 4.12**(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 4.12**(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 4.17. 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

$$T : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S} \text{ with } (x,y) \mapsto T(x,y) := \big[\big(A(x) \land C(x,y)\big) \ \to \ B(y)\big] \text{ is a tautology.} \tag{4.12}$$

Note that

(a) The last arrow in (4.12) is the arrow operator $\to$, not the function assignment operator $\mapsto$.

(b) if we can demonstrate that (4.12) is correct then we can replace $\big(A(x) \land C(x,y)\big) \ \to \ B(y)$ with $\big(A(x) \land C(x,y)\big) \ \Rightarrow \ B(y)$. 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 4.12**(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 (4.12). □

Remark 4.8. Let us compare example 4.15**(b)** with example 4.17. 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 4.15**(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 4.17 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 4.6.1 (Building blocks of mathematical theories) on p.114.

4.2.6 Biconditional and Logical Equivalence Operators – Part 2 (Understand this!)

This chapter continues the discussion of the $\leftrightarrow$ and $\Leftrightarrow$ operators from ch.4.2.3 (Biconditional and Logical Equivalence Operators – Part 1) on p.93.

Remark 4.9.

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

$$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B, \tag{4.13}$$
$$\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B. \tag{4.14}$$

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

$$\neg(A_1 \wedge A_2 \wedge \cdots \wedge A_n) \Leftrightarrow \neg A_1 \vee \neg A_2 \vee \cdots \vee \neg A_n, \tag{4.15}$$
$$\neg(A_1 \vee A_2 \vee \cdots \vee A_n) \Leftrightarrow \neg A_1 \wedge \neg A_2 \wedge \cdots \wedge \neg A_n. \tag{4.16}$$

PROOF of 4.13: 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 4.13. 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 4.9.

PROOF of 4.14: Left as an exercise. ■

Example 4.18. As in example 4.12 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 4.12 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 4.12**(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. □

4.2.7 More Examples of Tautologies and Contradictions (Understand this!)

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

Example 4.19. 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 4.12 on p. 92).

(a) Tautologies:

T_0,
$A_1 :=$ "$5 + 7 = 12$",
$A_2 :=$ "Any integer is even **or** odd",
$A_3 := P \vee \neg P$ (Tertium non datur or law of the excluded middle),
$A_4 := P \vee T_0$,
$A_5 := (P \wedge Q) \vee (P \wedge \neg Q)$,
$A_6 := (P \to Q) \leftrightarrow (\neg P \vee Q)$ (Implication is logically equivalent to an **or** statement),
$A_7 := [$"$x < y$" $\wedge A(x)] \to B(y)$ (see 4.15**(b)** on p.97),
$A_8 := A(x) \leftrightarrow B(x)$ (see 4.15**(c)**).

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

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

(b) Contradictions:

F_0,
$B_1 :=$ "$5 + 7 = 15$",
$B_2 :=$ "There are some non-zero numbers x such that $x = 2x$",
$B_3 := P \wedge \neg P$,
$B_3 := P \wedge F_0$,
$B_4 := F_0 \wedge (P \vee \neg P)$,
$B_5 := [\neg P \vee \neg Q] \wedge [P \wedge Q]$,
$B_6 := A(x)$ **xor** $B(x)$ (see 4.14**(c)** on p. 94). □

Proof that A_3 is a tautology:

P	$\neg P$	$P \vee \neg P$
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 4.10. 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. □

4.3 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 [5] Bryant, Kirby Course Notes for MAD 2104.

Theorem 4.4. *Let P, Q, R be statements.*

(a) *Identity Laws:*

$$P \wedge T_0 \Leftrightarrow P \tag{4.17}$$
$$P \vee F_0 \Leftrightarrow P \tag{4.18}$$

(b) *Domination Laws:*

$$P \vee T_0 \Leftrightarrow T_0 \tag{4.19}$$
$$P \wedge F_0 \Leftrightarrow F_0 \tag{4.20}$$

(c) *Idempotent Laws:*

$$P \vee P \Leftrightarrow P \tag{4.21}$$
$$P \wedge P \Leftrightarrow P \tag{4.22}$$

(d) *Double Negation Law:*

$$\neg(\neg P) \Leftrightarrow P \tag{4.23}$$

(e) *Commutative Laws:*

$$P \vee Q \Leftrightarrow Q \vee P \tag{4.24}$$
$$P \wedge Q \Leftrightarrow Q \wedge P \tag{4.25}$$

***(f)** Associative Laws:*

$$(P \lor Q) \lor R \Leftrightarrow P \lor (Q \lor R) \tag{4.26}$$

$$\textit{hence}\ \ (P \lor Q) \lor R \Leftrightarrow P \lor Q \lor R \tag{4.27}$$

$$(P \land Q) \land R \Leftrightarrow P \land (Q \land R) \tag{4.28}$$

$$\textit{hence}\ \ (P \land Q) \land R \Leftrightarrow P \land Q \land R \tag{4.29}$$

***(g)** Distributive Laws:*

$$P \lor (Q \land R) \Leftrightarrow (P \lor Q) \land (P \lor R) \tag{4.30}$$

$$P \land (Q \lor R) \Leftrightarrow (P \land Q) \lor (P \land R) \tag{4.31}$$

***(h)** De Morgan's Laws:* [52]

$$\neg(P \land Q) \Leftrightarrow \neg P \lor \neg Q \tag{4.32}$$

$$\neg(P \lor Q) \Leftrightarrow \neg P \land \neg Q \tag{4.33}$$

***(i)** Absorption Laws:*

$$P \land (P \lor Q) \Leftrightarrow P \tag{4.34}$$

$$P \lor (P \land Q) \Leftrightarrow P \tag{4.35}$$

***(j)** Implication Law:*

$$(P \rightarrow Q) \Leftrightarrow (\neg P \lor Q) \tag{4.36}$$

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:*

$$(P \rightarrow Q) \Leftrightarrow (\neg Q \rightarrow \neg P) \tag{4.37}$$

$$(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P) \tag{4.38}$$

***l.** Tautology:*

$$(P \lor \neg P) \Leftrightarrow T_0 \tag{4.39}$$

***m.** Contradiction:*

$$(P \land \neg P) \Leftrightarrow F_0 \tag{4.40}$$

***n.** Equivalence:*

$$(P \rightarrow Q) \land (Q \rightarrow P) \Leftrightarrow (P \leftrightarrow Q) \tag{4.41}$$

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 $\dots \Leftrightarrow \dots$ statements are indeed logically equivalent.

PROOF of **(h)** (De Morgan's laws):

[52] This is theorem 4.3 (De Morgan's laws for statements).

See theorem 4.3 on p.101.

PROOF of **(j)** (implication law):
We prove (4.36) using a truth table:

We see that the entries T-T-F-T in the $\neg P \vee Q$ column match those given for $P \to Q$ in Definition 4.13 on p.95 of the arrow operator. This proves the logical equivalence of those statements.

P	Q	$\neg P$	$\neg P \vee Q$
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 (4.37) with the help of the previously given laws **(a)** through **(j)**:

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

■

Example 4.20. Use the logical equivalences of thm.4.4 to prove that $\neg\big(\neg A \wedge (A \wedge B)\big)$ is a tautology. □

Solution:

	$\neg\big(\neg A \wedge (A \wedge B)\big)$	
$\Leftrightarrow$	$\neg(\neg A) \vee \neg(A \wedge B)$	De Morgan's Law (4.32)
$\Leftrightarrow$	$A \vee (\neg A \vee \neg B)$	De Morgan (4.32) + Double negation (4.23)
$\Leftrightarrow$	$(A \vee \neg A) \vee \neg B$	Associative law (4.26)
$\Leftrightarrow$	$T_0 \vee \neg B$	Tautology (4.39)
$\Leftrightarrow$	T_0	Commutative Law (4.24) + Domination Law (4.19) ■

Example 4.21. 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:

$$\neg(A \to B) \overset{\textbf{((j))}}{\Leftrightarrow} \neg(\neg A \vee B) \overset{\textbf{((h))}}{\Leftrightarrow} (\neg(\neg A) \wedge \neg B) \overset{\textbf{((d))}}{\Leftrightarrow} (A \wedge \neg B)$$

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

■

Remark 4.11. 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. □

4.4 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:

$$a : \mathscr{U} \to \mathscr{S}, \qquad x \mapsto a(x) =: \text{“}x \in A'' \tag{4.42}$$

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 4.22. 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 4.4 on p.103 of ch.4.3 (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 4.5. *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:*

$$p \wedge T_0 \Leftrightarrow p \qquad P \cap \mathscr{U} = P \tag{4.43}$$
$$p \vee F_0 \Leftrightarrow p \qquad P \cup \emptyset = P \tag{4.44}$$

***(b)** Domination:*

$$p \vee T_0 \Leftrightarrow T_0 \qquad P \cup \mathscr{U} = \mathscr{U} \tag{4.45}$$
$$p \wedge F_0 \Leftrightarrow F_0 \qquad P \cap \emptyset = \emptyset \tag{4.46}$$

***(c)** Idempotency:*

$$p \vee p \Leftrightarrow p \qquad P \cup P = P \tag{4.47}$$
$$p \wedge p \Leftrightarrow p \qquad P \cap P = P \tag{4.48}$$

***(d)** Double Negation:*

$$\neg(\neg p) \Leftrightarrow p \qquad (P^{\complement})^{\complement} = P \tag{4.49}$$

(e) *Commutative:*

(4.50) $p \vee q \;\Leftrightarrow q \vee p$ $\qquad P \cup Q = Q \cup P$

(4.51) $p \wedge q \;\Leftrightarrow q \wedge p$ $\qquad P \cap Q = Q \cap P$

(f) *Associative:*

(4.52) $(p \vee q) \vee r \;\Leftrightarrow\; p \vee (q \vee r)$ $\qquad (P \cup Q) \cup R = P \cup (Q \cup R)$

(4.53) $(p \wedge q) \wedge r \;\Leftrightarrow p \wedge (q \wedge r)$ $\qquad (P \cap Q) \cap R = P \cap (Q \cap R)$

(g) *Distributive:*

(4.54) $p \vee (q \wedge r) \;\Leftrightarrow\; (p \vee q) \wedge (p \vee r)$ $\qquad P \cup (Q \cap R) = (P \cup Q) \cap (P \cup R)$

(4.55) $p \wedge (q \vee r) \;\Leftrightarrow\; (p \wedge q) \vee (p \wedge r)$ $\qquad P \cap (Q \cup R) = (P \cap Q) \cup (P \cap R)$

(h) *De Morgan:*

(4.56) $\neg(p \wedge q) \;\Leftrightarrow \neg p \vee \neg q$ $\qquad (P \cap Q)^{\mathsf{C}} = P^{\mathsf{C}} \cup Q^{\mathsf{C}}$

(4.57) $\neg(p \vee q) \;\Leftrightarrow \neg p \wedge \neg q$ $\qquad (P \cup Q)^{\mathsf{C}} = P^{\mathsf{C}} \cap Q^{\mathsf{C}}$

(i) *Absorption:*

(4.58) $p \wedge (p \vee q) \;\Leftrightarrow p$ $\qquad P \cap (P \cup Q) \;= P$

(4.59) $p \vee (p \wedge q) \;\Leftrightarrow p$ $\qquad P \cup (P \cap Q) \;= P$

j1. *Implication 1:*

(4.60) $(p \to q) \;\Leftrightarrow (\neg p \vee q)$ $\qquad (P \setminus Q)^{\mathsf{C}} \;=\; P^{\mathsf{C}} \cup Q$

*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 $x \notin P \cap Q^{\mathsf{C}} = P \setminus Q$, i.e., $x \in (P \setminus Q)^{\mathsf{C}}$.*

(4.61) $p \Rightarrow q$

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:*

(4.62) $(P \to Q) \;\Leftrightarrow (\neg Q \to \neg P)$ $\qquad P^{\mathsf{C}} \cup Q \;= Q \cup P^{\mathsf{C}}$

(4.63) $(P \Rightarrow Q) \;\Leftrightarrow (\neg Q \Rightarrow \neg P)$ $\qquad P \subseteq Q \;\Leftrightarrow Q^{\mathsf{C}} \subseteq P^{\mathsf{C}}$

l. *Tautology:* (4.64) $(P \vee \neg P) \Leftrightarrow T_0$ $\quad P \cup P^{\mathsf{C}} = \mathscr{U}$

m. *Contradiction:* (4.65) $(P \wedge \neg P) \Leftrightarrow F_0$ $\quad P \cap P^{\mathsf{C}} = \emptyset$

n1. *Equivalence 1:* (4.66) $\begin{aligned}(p \to q) \wedge (q \to p)\\ \Leftrightarrow (p \leftrightarrow q)\end{aligned}$ $\quad \begin{aligned}&(P^{\mathsf{C}} \cup Q) \cap (Q^{\mathsf{C}} \cup P)\\ &= \{x : x \textit{ both in } P, Q \textbf{\textit{ or}}\\ &\quad x \textit{ neither in } P \textit{ nor in } Q\ \}\end{aligned}$

n2. *Equivalence 2:* (4.67) $\begin{aligned}(p \Rightarrow q) \wedge (q \Rightarrow p)\\ \Leftrightarrow (p \Leftrightarrow q)\end{aligned}$ $\quad \begin{aligned}(P \subseteq Q) \textbf{\textit{ and }} (Q \subseteq P)\\ \Leftrightarrow (P = Q)\end{aligned}$

PROOF: The set equalities are evident except for the following:
PROOF of Equivalence 1:

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

■

4.5 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 [5] Bryant, Kirby Course Notes for MAD 2104.

4.5.1 Quantifiers for One–Variable Statement Functions

Definition 4.16 (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

(4.68) "For all x $A(x)''$, 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

(4.69) "For some x $A(x)''$, written $\exists x A(x)$.

The above is a short for "There exists $x \in \mathscr{U}$ such that $A(x)$ is true". [53] We call the symbol $\exists$ the **existential quantifier** symbol.

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

$$\text{"There exists unique } x \text{ such that } A(x)\text{"}, \quad \text{written } \exists!xA(x). \tag{4.70}$$

The above is a short for "There exists a unique $x \in \mathscr{U}$ such that $A(x)$ is true". [54] We call the symbol $\exists!$ the **unique existential quantifier** symbol. □

Example 4.23. Let $A : [-3,3] \to \mathscr{S}$ be the statement function $x \mapsto$ "$x^2 - 4 = 0$".

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

$C =$ "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 zeros of the function $f(x) = x^2 - 4$ in the interval $-3 \leq x \leq 3$ are $x = \pm 2$. It follows that D is a true statement and A and C are false statements. □

Example 4.24. 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:

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

Example 4.25. We continue example 4.24.

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

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 $(S(x) \land C(x)) \to \neg H(x)$.

It follows that

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

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

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

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

[54] Equivalently, "$A(x)$ is true for exactly one $x \in \mathscr{U}$".

We have seen in example 4.21 on p.105 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". □

4.5.2 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

$$B : \mathscr{U}_x \times \mathscr{U}_y \to \mathscr{S}, \quad x \mapsto B(x,y) \tag{4.71}$$

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 4.17 (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 4.26. 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:

(1) The order in which the qualifiers are applied is important.
$\forall x \exists y$ generally does not mean the same as $\exists y \forall x$.
(2) 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 4.1. *Note the following:*

$$\forall x \forall y B(x,y) \;\Leftrightarrow\; \forall y \forall x B(x,y) \tag{4.72}$$
$$\exists x \exists y B(x,y) \;\Leftrightarrow\; \exists y \exists x B(x,y) \tag{4.73}$$
$$\forall x \exists y B(x,y) \;\not\Leftrightarrow\; \exists y \forall x B(x,y) \tag{4.74}$$
$$\exists y \forall x B(x,y) \Rightarrow \forall x \exists y B(x,y) \tag{4.75}$$

PROOF: (4.72) and (4.73) follow from **(a)** and **(e)** in def. 4.17 and we saw an example for (4.72) 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 4.12. The last part of the proof of (4.75) 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 4.13 (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 4.27. 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 $\mathscr{U}_y$". □

4.5.3 Quantifiers for Statement Functions of more than Two Variables

Remark 4.14. Although this document limits its scope to statement functions of one or two variables (see the note before remark 4.4 in ch.4.1 (Statements and statement functions)) we discuss briefly the use of quantifiers for predicates

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

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 4.28 (Continuity vs uniform continuity)**.** This example demonstrates the effect of switching a $\forall$ quantifier with an $\exists$ quantifier for a predicate with four variables. You will learn later that 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. The only point of interest here is the use of the quantifiers.

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 $x \in]a, b[$. [55]

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 $]a, b[$. [56]

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 4.15 (Partially quantified statement functions)**.** What was said in remark 4.13 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 $B : (x, z) \mapsto \big[\forall y \; \neg \exists w \; A(w, x, y, z)\big]$ defines a two-variable predicate B which inherits the UoDs for x and z from the original statement function A. □

4.5.4 Quantifiers and Negation (Understand this!)

Negation of statements involving quantifiers is governed by

[55] See Definition 13.2 (ε-δ continuity) on p.402.

[56] See Definition 13.5 (Uniform continuity of functions) on p.410.

Theorem 4.6 (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 [5] 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 4.29. 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

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

Example 4.30. As in example 4.29, 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:

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

4.6 Proofs (Understand this!)

We have informally discussed proofs in examples 4.15 and 4.17 of chapter 4.2.5 (Arrow and Implication Operators) on p.95 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 [5] Bryant, Kirby Course Notes for MAD 2104.

4.6.1 Building Blocks of Mathematical Theories

Some of the terminology definitions in notations 4.2 and 4.4 were taken almost literally from ch.3 (Methods of Proofs), subchapter 1 (Logical Arguments and Formal Proofs) of [5] Bryant, Kirby Course Notes for MAD 2104.

Notation 4.2 (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 4.31. Let $A :=$ "all continuous functions are differentiable" (known to be false [57]) 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 4.16.

(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 $a^n + b^n = c^n$. [58] 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. □

Notation 4.3 (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

[57] see remark 4.7 on p.97 in ch.4.2.5 (Arrow and Implication Operators).

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

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 4.6.2 (Rules of Inference) on p.117

(b) logical equivalences for statements (see ch.4.2.6 (Biconditional and Logical Equivalence Operators – Part 2) on p.ch.100).

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. □

Notation 4.4 (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 4.17 (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 4 different meanings are attached to those terms:

Each one of **lemma**, **proposition**, **theorem**, **corollary** is a theorem as defined above in notations 4.2, 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. 4.1 on p.84) that what we call a statement, [5] 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 4.32. 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: [59]
(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 [2] 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 mathematical induction (see thm.6.2 on p. 169). □

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 4.18 (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.15 that this lead to problems so serious that they caused some of the leading mathematicians of the time to revisit the foundations of mathematics. □

Example 4.33. 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

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

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 4.32**(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! □

4.6.2 Rules of Inference

Remark 4.19 (Most important rules of inference)**.** In Notations 4.2 on p.114 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

(4.76) the compound statement $\big[A_1 \wedge A_2 \wedge \cdots \wedge A_n\big] \to C$ 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 4.7. *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:*

(4.77) *P_j is **true** for each $j = 1, 2, \dots, n$ and C is **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

(4.78) P is **true** and C is **false** 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

■

Notation 4.5. Rules of inference are commonly written in the following form:

Your explanations go into this area

$$\begin{array}{l} A_1 \\ A_2 \\ \cdots \\ A_n \\ \hline \therefore C \end{array}$$

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

$$\boxed{\begin{array}{l} A_1, A_2, \ldots, A_n \\ \hline \therefore C \end{array}}$$

Theorem 4.8 (The three most important inference rules)**.** *The following lists three inference rules, i.e., those arrow statements are indeed tautoloties:*

(4.79) *Modus Ponens (Law of detachment - the mode that affirms the antecedent (the premise))*

$$\begin{array}{l} A \\ A \to C \\ \hline \therefore C \end{array} \tag{4.79}$$

(4.80) *Modus Tollens (The mode that Denies the consequent (the conclusion))*

$$\begin{array}{l} \neg C \\ A \to C \\ \hline \therefore \neg A \end{array} \tag{4.80}$$

(4.81) *Hypothetical syllogism*

$$\begin{array}{l} A \to B \\ B \to C \\ \hline \therefore A \to C \end{array} \tag{4.81}$$

Here is the compact notation:

Modus Ponens	*Modus Tollens*	*Hypothetical syllogism*
$\begin{array}{l} A,\ A \to C \\ \hline \therefore C \end{array}$	$\begin{array}{l} \neg C,\ A \to C \\ \hline \therefore \neg A \end{array}$	$\begin{array}{l} A \to B, B \to C \\ \hline \therefore A \to C \end{array}$

Note that the proof that the hypothetical syllogism is a tautology was given in thm.4.1 on p.96

PROOF:

■

Example 4.34. Here are five more inference rules.

(4.82) Disjunction Introduction
$$\begin{array}{l} A \\ \hline \therefore A \lor B \end{array}$$

(4.83) Conjunction elimination
$$\begin{array}{l} A \land B \\ \hline \therefore A \end{array}$$

(4.84) Disjunctive syllogism
$$\begin{array}{l} A \lor B \\ \neg A \\ \hline \therefore B \end{array}$$

(4.85) Conjunction introduction
$$\begin{array}{l} A \\ B \\ \hline \therefore A \land B \end{array}$$

(4.86) Constructive dilemma
$$\begin{array}{l} (A \to B) \land (C \to D) \\ A \lor C \\ \hline \therefore B \lor D \end{array}$$

Compact notation:

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

□

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 [5] Bryant, Kirby Course Notes for MAD 2104.

4.6.3 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 $n \in \mathbb{Z}_{\geq 0}$ and we write $\mathscr{D}$ for the set of all differentiable functions. We now can formulate our theorem.

Theorem 4.9. *Given the statements*

$$\begin{aligned}
&\textbf{\textit{a:}}\quad A := \text{“}(n \in \mathbb{Z}_{\geq 0}) \wedge (c_0 \in \mathbb{R}) \wedge (c_1 \in \mathbb{R}) \wedge \dots \wedge (c_n \in \mathbb{R}) \wedge \ \Big(f(x) = \sum_{j=0}^{n} c_j x^j\Big)'',\\
&\textbf{\textit{b:}}\quad B := \text{“}f(x) \in \mathscr{D}'',
\end{aligned}$$

the following is valid: $A \Rightarrow B$. [60]

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.

$$\begin{aligned}
&\textbf{a:}\quad Z_j := \text{“}j \in \mathbb{Z}_{\geq 0}'',\\
&\textbf{b:}\quad C_j := Z_j \wedge \text{“}c_j \in \mathbb{R}'',\\
&\textbf{c:}^{61}\quad X_j := Z_j \wedge \text{“}x^j \in \mathscr{D}'',\\
&\textbf{d:}\quad D_j := Z_j \wedge \text{“}c_j x^j \in \mathscr{D}'',\\
&\textbf{e:}\quad E := Z_n \wedge \text{“}f(x) = \sum_{j=0}^{n} c_j x^j\,'',\\
&\textbf{f:}\quad B := \text{“}f(x) \in \mathscr{D}'' \text{ (repeated for convenient reference)}
\end{aligned}$$

We now can write our theorem as

$$(Z_n \wedge C_0 \wedge C_1 \wedge \dots \wedge C_n \wedge E) \to B. \tag{4.87}$$

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

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

[60] 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.
[61] The expression x^j in **(c)** and **(d)** denotes the function $x \mapsto x^j$.

Then Thm-1 states that $A_1 \Rightarrow B_1$. [62]

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

We rewrite Thm-2 as an implication. Let

$A_2 :=$ “$c \in \mathbb{R}''$ $\land$ “$h(x) \in \mathscr{D}''$ $\land$ “$g(x) = c \cdot h(x)''$,
$B_2 :=$ “$h(x) \in \mathscr{D}''$,

Then Thm-2 states that $A_2 \Rightarrow B_2$.

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

We rewrite Thm-3 as an implication. Let

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

$B_3 :=$ “$g(x) \in \mathscr{D}''$,

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

	Assertion	Reason
a:	$Z_0, Z_1, \dots Z_n$	evident from $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$
b:	$C_0, C_1, \dots C_n$	part of the premise of $A \to B$ (see (4.87))
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:	$(Z_j \land C_j \land X_j) \to D_j \ \ (j = 0, 1, \dots n)$	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:	$(Z_n \land D_0 \land D_1 \land \cdots \land D_n \land E) \to B$	**(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 4.20. 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 $Z_j \ \land$ “$c_j x^j \ \in \mathscr{D}''$ (which itself references the component $Z_j =$ “$j \in \mathbb{Z}''_{\geq 0}$).

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 $\mathbb{Z}_{\geq 0} := \{0, 1, 2, \dots\}$ 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.

[62] 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.

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. □

4.6.4 Invalid Proofs Due to Faulty Arguments

Remark 4.21 (Fallacies in logical arguments)**.** People who are not very analytical often commit the following errors in their argumentation:

(4.88) Affirming the Consequent
(proving the wrong direction)

$$\begin{array}{l} P \to Q \\ Q \\ \hline \therefore P \end{array}$$

(4.89) Denying the Antecedent
(indirect proof in the wrong direction)

$$\begin{array}{l} P \to Q \\ \neg P \\ \hline \therefore \neg Q \end{array}$$

(4.90) 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 4.35 (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: [63]
"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. ■

[63] This is example 1.8.3 in ch.3 (Methods of Proofs) of [5] Bryant, Kirby Course Notes for MAD 2104.

4.7 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:

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

4.7.1 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 4.36 (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. ■

4.7.2 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.15 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 4.10. *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 4.22. 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! □

4.7.3 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 4.6.3 (An example of a direct proof) on p.120 that each polynomial is differentiable (theorem 4.9). That was an example of a direct proof.

4.7.4 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 (4.38) on p.104). 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 4.11. *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 $A \cap B^{\complement} \neq \emptyset$. ■

4.7.5 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 4.12. *Let $A \subseteq \mathbb{Z}$ with the following properties:*

$$m, n \in A \;\Rightarrow\; m + n \in A, \tag{4.91}$$

$$m, n \in A \;\Rightarrow\; mn \in A, \tag{4.92}$$

$$0 \notin A, \tag{4.93}$$

$$\text{if } n \in \mathbb{Z} \text{ then either } n \in A \text{ or } -n \in A \text{ or } n = 0. \tag{4.94}$$

Then $1 \in A$.

Proof by contradiction: Assume that A is a set of integers with properties (4.91) – (4.94) 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 (4.94) and $1 \neq 0$ that $-1 \in A$.
(b) It now follows from (4.92) that $(-1) \cdot (-1) \in A$, i.e., $1 \in A$.

We have reached our contradiction. ■

Remark 4.23. 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. □

4.7.6 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

$$P_1 \vee P_2 \vee \dots \vee P_n \;\Leftrightarrow\; P. \tag{4.95}$$

Proof by cases then rests on the following theorem:

Theorem 4.13. *Let $P, Q, P_1 \vee P_2 \vee \dots \vee P_n$ be statements such that (4.95) is true. Then*

$$\big(P \Rightarrow Q\big) \;\Leftrightarrow\; \big[\,(P_1 \Rightarrow Q) \;\vee\; (P_2 \Rightarrow Q) \;\vee\; \dots\; (P_n \Rightarrow Q)\,\big]. \tag{4.96}$$

Proof (outline): You would do the proof by induction. Prove (4.96) 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 (4.95) becomes $P_1 \vee P_2 \vee \dots \vee P_{n+1} \;\Leftrightarrow\; P$ by setting $A := P_1 \vee P_2 \vee \dots \vee P_n$ and this way reducing the proof of (4.96) for $n+1$ to that of 2 components. You make the validity of $\big(A \Rightarrow Q\big) \;\Leftrightarrow\; \big[\,(P_1 \Rightarrow Q) \;\vee\; (P_2 \Rightarrow Q) \;\vee\; \dots\; (P_n \Rightarrow Q)\,\big]$ the induction assumption. ■

Theorem 4.14. *Prove that for any $x \in \mathbb{R}$ such that $x \neq 5$ we have*

$$\frac{x}{x-5} > 0 \;\Rightarrow\; \big[\,(x < 0) \text{ or } (x > 5)\,\big]. \tag{4.97}$$

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
$P :=$ "$x/(x-5) > 0$", [64] $P_1 := x > 0$ **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 4.13 the validity of (4.97). ■

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

5 Relations, Functions and Families

We now give an in depth presentation of the material of ch.2.4 (A First Look at Functions, Sequences and Families).

5.1 Cartesian Products and Relations

Definition 5.1 (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 $b \in B$.
Let $(a_1,b_1),(a_2,b_2) \in A \times B$. We say they are **equal**, and we write $(a_1,b_1) = (a_2,b_2)$ if and only if $a_1 = a_2$ and $b_1 = b_2$.

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 5.1 (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.e., the 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 5.1 (Function graphs as subsets of cartesian products)**.** We gave the preliminary definition of a function in Definition 2.22, p.30 of ch.2.4 (A First Look at Functions, Sequences and Families). [65] A function

$$f : X \to Y; \quad y = f(x)$$

which assigns each $x \in X$ to a unique function value $f(x) \in Y$, e.g., $f(x) = x^2$, is characterized by its graph

$$\Gamma_f \ := \ \{\big(x, f(x)\big) : x \in X\}$$

which is a subset of the cartesian product $X \times Y$. For example, if $X = [-2,3]$ and $Y = [0,10]$ then $\Gamma_f \ := \ \{(x,x^2) : -2 \leq x \leq 3\}$ is a subset of $[-2,3] \times [0,10]$. We will examine the connection between functions and their graphs in detail later in this chapter. □

[65] The precise definition of a function will be given in section 5.2 on p.131.

Remark 5.2 (Empty cartesian product)**.** Note that $A \times B = \emptyset$ if and only if $A = \emptyset$ or $B = \emptyset$ or both are empty. □

Definition 5.2 (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
(a) **reflexive** if xRx for all $x \in X$,
(b) **symmetric** if x_1Rx_2 implies x_2Rx_1 for all $x_1, x_2 \in X$,
(c) **transitive** if x_1Rx_2 and x_2Rx_3 implies x_1Rx_3 for all $x_1, x_2, x_3 \in X$,
(d) **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.

Example 5.2 (Equality as a relation)**.** Given a set X let $R := \{(x, x) : x \in X\}$ [66]

, i.e., xRy if and only if $x = y$. This defines a relation on X which is reflexive, symmetric, antisymmetric and transitive. □

Example 5.3 (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 on 2^X which is reflexive, antisymmetric and transitive. □

Remark 5.3. Unless a relation on a set X is symmetric there will be at least one pair $x, y \in X$ such that x is related to y whereas y is related to x is false. This is different from how we think of relatedness in a non–mathematical context.

Consider Example 5.3. If A is a proper subset of B then A is related to B but it is not true that B is related to A. □

Example 5.4 (Function graphs as relations)**.** We saw in rem.5.1 on p.126 that functions $f : X \to Y$ are characterized by their graphs $\Gamma_f := \{\big(x, f(x)\big) : x \in X\}$ which are subsets of $X \times Y$, i.e., Γ_f is a relation on $X \times Y$. □

Example 5.5 (Size of sets as a relation)**.** Let X be a set and

$$R := \{(A, B) : A, B \subseteq X \text{ and } \big|A\big| = \big|B\big|\},$$

i.e., ARB if and only if A and B possess the same number of elements. [67] In particular ARB is true if $|A| = |B| = \infty$. This defines a relation on the power set 2^X of X which is reflexive, symmetric and transitive. □

[66] This set is commonly referred to as the **diagonal** of X^2.

[67] See Definition 2.12 (preliminary definition of the size of a set) on p.23.

Example 5.6 (Empty relation)**.** Given two sets X and Y let $R := \emptyset$. This **empty relation** is the only relation which exists on (X, Y) if X or Y is empty. □

Example 5.7. Let $X := \mathbb{R}^2$ be the xy–plane. For any point $\vec{x} = (x_1, x_2)$ in the plane let

$$\|\vec{x}\|\_2 \ := \ \sqrt{x_1^2 \, + \, x_2^2}\,; \qquad R \ := \ \{\, (\vec{x}, \vec{y}) \in \mathbb{R}^2 \times \mathbb{R}^2 : \ \|\vec{x}\|_2 \ = \ \|\vec{y}\|_2 \,\}\,.$$

Let In other words, $\|\vec{x}\|\_2$ is the length of the straight line which extends from the origin of the plane to $\vec{x}$ [68] and two points in the plane are related when they have the same length: they are located on a circle which is centered at the origin and has radius $r = \|\vec{x}\|_2 = \|\vec{y}\|_2$. The relation R is reflexive, symmetric and transitive but not antisymmetric. □

The relations given in examples 5.2, 5.5, 5.6 and 5.7 are reflexive, symmetric and transitive. Such relations are so important that they deserve a special name:

Definition 5.3 (Equivalence relations and equivalence classes)**.** Let R be a relation on a set X.

(a) If R is • reflexive, • symmetric, • transitive, we call R an **equivalence relation** on X.

(b) For an equivalence relation R it is customary to write $x \sim x'$ rather than xRx' (or $(x, x') \in R$). We say in this case that x and x' are **equivalent**.

(c) Given is an equivalence relation "$\sim$" on a set X. For $x \in X$ let

$$(5.1) \qquad [x]_\sim \ := \ \{x' \in X : x' \sim x\} \ = \ \{\text{ all items equivalent to } x\,\}.$$

We call $[x]_\sim$ the **equivalence class** of x. If it is clear from the context what equivalence relation is referred to then we can write $[x]$ instead of $[x]_\sim$. □

Proposition 5.1 (see [2] B/G prop.6.4 & B/G prop.6.5)**.** *Let "$\sim$" be an equivalence relation on a nonempty set X and $x, y \in X$ Then*

(a) $x \in [x]$,
(b) $x \sim y \Leftrightarrow [x] = [y]$,
(c) *either* $[x] = [y]$ *or* $[x] \cap [y] = \emptyset$.

PROOF of **(a)**: This follows from the reflexivity of "$\sim$".

PROOF of **(b)**:

We first show that if $x \sim y$ then $[x] = [y]$. Let $z \in [x]$. It follows from the definition of $[x]$ that $z \sim x$, hence $z \sim y$ (transitivity of "$\sim$"), hence $z \in [y]$. This proves $[x] \subseteq [y]$. We switch the roles of x and y and repeat the above to obtain $[y] \subseteq [x]$.

We now prove that if $[x] = [y]$ then $x \sim y$. It follows from $[x] = [y]$ that $x \in [y]$, hence $x \sim y$ by (5.1).

PROOF of **(c)**: This proof is left as exercise 5.4 (see p.163). ■

For the next proposition recall Definition 2.11 on p.22 of a partition.

[68] See Definition 11.4 on p.327. of the length or Euclidean norm of a vector in n-dimensional space.

Proposition 5.2 (see [2] B/G prop.6.6 for parts **(a)** and **(b)**).

- ***(a)*** *Let “ $\sim$″ be an equivalence relation on a nonempty set X and let $\mathscr{P}_\sim := \{[x] : x \in X\}$ be the set of all its equivalence classes. Then $\mathscr{P}_\sim$ is a partition of X.*
- ***(b)*** *Conversely, let $\mathscr{P}$ be a partition of X and define a relation “$\sim_\mathscr{P}$″ on X as follows: $x \sim_\mathscr{P} y \ \Leftrightarrow$ there is $P \in \mathscr{P}$ such that $x, y \in P$. Then $\sim_\mathscr{P}$ is an equivalence relation on X.*
- ***(c)*** *Let “$\sim$″ be an equivalence relation on X. Let $\mathscr{P}_\sim$ be the associated partition of its equivalence classes. Let “$\sim_{\mathscr{P}_\sim}$″ be the equivalence relation associated with the partition $\mathscr{P}_\sim$. Then “$\sim_{\mathscr{P}_\sim}$″ = “$\sim$″ (i.e., both equivalence relations are equal as subsets of $X \times X$.*
- ***(d)*** *Let $\mathscr{P}$ be a partition of X. Let $\sim_\mathscr{P}$ be the associated equivalence relation defined in part* **(b)***. Let $\mathscr{P}_{\sim_\mathscr{P}}$ be the associated partition of its equivalence classes. Then $\mathscr{P}_{\sim_\mathscr{P}} = \mathscr{P}$.*

PROOF of **(a)**:

We observe that a set contains no duplicates: If $[x] = [y]$, we do not count $[x]$ and $[y]$ as separate members of $\mathscr{P}_\sim$! It follows from prop.5.1**(c)** that those members are mutually disjoint.

It remains to prove that their union is X. Let $x \in X$. Then $x \in [x]$ (reflexivity of “$\sim$”) and $[x] \in \mathscr{P}_\sim$. It follows that $x \in \bigcup\left[P : P \in \mathscr{P}_\sim\right]$ and we obtain that $\bigcup\left[P : P \in \mathscr{P}_\sim\right] \ = \ X$. We have proved that $\mathscr{P}_\sim$ is a partition of X.

PROOF of **(b)**: For the following assume that $x, y, z \in X$.

It follows from $\biguplus\left[P : P \in \mathscr{P}\right] = X$ that for each $x \in X$ there exists $P \in \mathscr{P}$ such that $x \in P$. It further follows from the mutual disjointness of the elements of $\mathscr{P}$ that there exists exactly one such P and we are justfied to write P_x for this uniquely defined set P. In other words, the assignment $x \mapsto P_x$ defines a function $P(\cdot) : X \to \mathscr{P}$.

Reflexivity: $x \in P_x$ implies $x \sim_\mathscr{P} x$.

Symmetry: Let $x \sim_\mathscr{P} y$. This implies $P_x = P_y$, hence $y \in P_x$, hence $y \sim_\mathscr{P} x$.

Transitivity: Let $x \sim_\mathscr{P} y$ and $y \sim_\mathscr{P} z$. $x \sim_\mathscr{P} y$ implies $P_x = P_y$ and $y \sim_\mathscr{P} z$ implies $P_y = P_z$. It follows that $P_z = P_x$, i.e., $x \sim_\mathscr{P} z$.

PROOF of **(c)**: $\boxed{\star}$

Let $x, y \in X$. Then

$$x \sim y \ \Leftrightarrow [x] = [y] \ \Leftrightarrow \ \text{both } x, y \text{ belong to the same element of } \mathscr{P}_\sim \ \Leftrightarrow \ x \sim_{\mathscr{P}_\sim} y.$$

PROOF of **(d)**: $\boxed{\star}$

Let $P \in \mathscr{P}$ and $x, y \in X$. Let $[x]$ and $[y]$ be the equivalence classes of x and y for “$\sim_\mathscr{P}$”. If $x \in P$ then

$$y \in P \ \Leftrightarrow x \sim_{\mathscr{P}_\sim} y \ \Leftrightarrow \ [x] = [y].$$

It follows that $P = [x]$, hence $P \in \ \mathscr{P}_{\sim_\mathscr{P}}$ This true for any $P \in \mathscr{P}$ and it follows that $\mathscr{P} \ \subseteq \ \mathscr{P}_{\sim_\mathscr{P}}$.

Now let $x \in X$. If $y \in X$ then

$$y \in [x] \ \Leftrightarrow y \sim_\mathscr{P} x \ \Leftrightarrow \ y \in P_x.$$

It follows that any equivalence class $[x]$ with respect to “$\sim_\mathscr{P}$” is an element of $\mathscr{P}$, hence $\mathscr{P}_{\sim_\mathscr{P}} \subseteq \mathscr{P}$. We have shown that $\mathscr{P}_{\sim_\mathscr{P}} = \mathscr{P}$ and this finishes the proof of Proof of **(d)**. ■

Relations which are reflexive, antisymmetric and transitive like the relation of example 5.3 (set inclusion) allow to compare items for "bigger" and "smaller" or "before" and "after". They also deserve a special name:

Definition 5.4 (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. [69] 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**. □

Remark 5.4. The properties of a partial ordering can now be phrased as follows:

(5.2) $x \preceq x$ for all $x \in X$ **reflexivity**

(5.3) $x \preceq y$ and $y \preceq x \quad \Rightarrow \quad y = x$ **antisymmetry**

(5.4) $x \preceq y$ and $y \preceq z \quad \Rightarrow \quad x \preceq z$ **transitivity** □

Remark 5.5. 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 $x \leq y$.
2. $X = 2^\Omega$ for some set Ω and $A \preceq B$ if and only if $A \subseteq B$ (example 5.3).
3. $X = \mathbb{R}$ and $x \preceq y$ if and only if $x \geq y$.

(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 $x \preceq y$ 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 $x \preceq y$. □

What makes a partial ordering more general than the "$x \leq y$" order relation on sets of numbers? The answer: You can compare any two numbers: either $x \leq y$ or $y \leq x$ or $x = y$. Set inclusion on 2^Ω on the other hand does not have this property. For example, if $A = [0, 2]$ and $B = [1, 3]$ then neither $A \subseteq B$ nor $B \subseteq A$ nor $A = B$. We call POsets with the property that any two elements can be compared linearly or totally ordered:

[69] Some authors, Dudley among them, do not include reflexivity into the definition of a partial ordering and then distinguish between **strict partial orders** and **reflexive partial orders**.

Definition 5.5 (Linear orderings)**.** $\boxed{\star}$

(a) Let $(X, \preceq)$ be a nonempty POset, i.e., $\preceq$ is a partial ordering on X (see Definition 5.4 on p.130). We say that $\preceq$ 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 $y \preceq x$. We call $(X, \preceq)$ a **linearly ordered set** or a **totally ordered set** set.

(b) Let $(X, \preceq)$ be a nonempty POset and $A, C \subseteq X$. C is a **chain** in X if $(C, \preceq)$ is linearly ordered (with the same ordering). □

Example 5.8.

(a) The real numbers line $(\mathbb{R}, \leq)$ with its usual "$\leq$" ordering is a linearly ordered set. So is $(\mathbb{R}, \geq)$(!)

(b) If X is a set with at least two elements then set inclusion is **not** a linear order on 2^X.

(c) Ordered integral domains $(R, \oplus, \odot, P)$ are totally ordered. □

Definition 5.6 (Inverse Relation)**.** $\boxed{\star}$ 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** of the relation R. □

Example 5.9. Let $R := \{(x, x^3) : x \in \mathbb{R}\}$. Then this relation is the graph Γ_f of 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} is the graph $\Gamma_{f^{-1}}$ of the inverse function $x = f^{-1}(y) = y^{1/3}$. □

5.2 Functions (Mappings) and Families

5.2.1 Some Preliminary Observations about Functions

Remark 5.6 (A layman's definition of a function)**.** We look at the set $\mathbb{R}$ of all real numbers [70] 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.

Let us take a closer look at the function $y = f(x) = \sqrt{4 - x^2}$ and its properties:

[70] Real numbers were defined informally in ch.2.3 (Numbers)

(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.

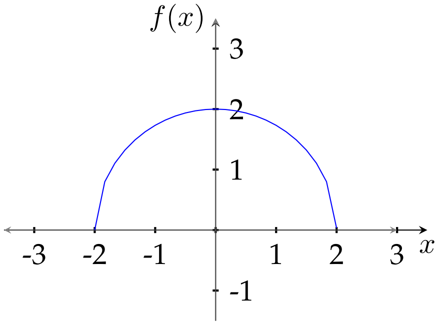
(b) For some other x, e.g., $x = 0$ or $x = 2/3$, there is a function value $f(x)$. A moment's reflection shows that the biggest possible set of potential arguments for our function, called by some authors the **natural domain** of the function (e.g., [3] Brewster/Geoghegan), is the interval $[-2, 2]$. It is customary to write D_f for the natural domain of a function $y = f(x)$.

(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 $y = 0$, which are "hit" more than once by the function: $f(2) = f(-2) = 0$. [71]

(f) Graphs as drawings: We are used to look at the graphs of functions. Here is a picture of the graph of $f(x) = \sqrt{4 - x^2}$.



(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 [72] (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 5.1 on p. 126). □

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 5.7 (A better definition of a function)**.** We make the following alterations to remark 5.6.

[71] 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 $x = \pm\sqrt{4 - y^2}$.

[72] Skip this example on first reading if you do not know about functions of several variables. You will find information about this in chapter 11 ("Vectors and vector spaces") on p.323.

- ▶ 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, e.g., $B := [0, 2]$ 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 is what 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 $x \in A$, i.e., the set $\{x \in A : f(x) = y\}$ can be empty.

(4) On the other hand there may be numbers y which are "hit" more than once by f.
Example: Let $A := \mathbb{N}$, $B := \mathbb{R}$, $f(x) := (-1)^x$. Then both -1 and 1 are mapped to infinitely often by f.

(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.,

$$\Gamma_f \;:=\; \{\big(x, f(x)\big) : x \in A\}. \tag{5.5}$$

Γ_f has the following properties:

(5a) $\Gamma_f \subseteq A \times B$, i.e., Γ_f is a relation on (A, B) (see Definition 5.2 on p.127).

(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: Subtract the square of x from 4, then take the square root of that difference.

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. 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.

5.2.2 Definition of a Function and Some Basic Properties

Introduction 5.1. Remark 5.7 on p.132 made it plausible that a function can be thought of equivalently as an assignment rule $x \mapsto f(x)$ or as a graph $\Gamma_f := \{\big(x, f(x)\big) : x \in A\}$, i.e., as a relation on (X, Y) (see example 5.4 on p.127). Mathematicians prefer the latter approach because "assignment rule" is a rather vague term (an undefined term in the sense of ch. 4.6.1 (Building blocks of mathematical theories) on p.114) whereas "relation" is entirely defined in the language of sets.

Not every relation Γ on $X \times Y$ is can serve as the graph of a function with domain X and codomain Y since we decided that the following is important:

(a) For each $x \in X$ there must be a function value $f(x)$, i.e., some $y \in Y$ such that $(x, y) \in \Gamma$,

(b) There cannot be more than one such function value $f(x)$, i.e., for each $x \in X$ there must be exactly one $y \in Y$ such that $(x, y) \in \Gamma$. □

The above now leads us to the official definition of a function as a relation which satisfies those properties **(a)** and **(b)**.

Definition 5.7 (Mappings (functions))**.** Given are two arbitrary nonempty sets X and Y and a relation Γ on (X, Y) (see 5.2 on p.127) which satisfies the following:

$$\text{for each } x \in X \text{ there exists exactly one } y \in Y \text{such that } (x, y) \in \Gamma. \tag{5.6}$$

We call the triplet $f(\cdot) := (X, Y, \Gamma)$ a **function** or **mapping** from X to Y. The set X is called the **domain** or **source** and Y is called the **codomain** or **target** of the mapping $f(\cdot)$. We will usually use the words "domain" and "codomain" in this document.

Usually mathematicians simply write f instead of $f(\cdot)$ We mostly follow that convention, but sometimes include the "$(\cdot)$" part to emphasize that a function rather than an "ordinary"' element of a set is involved. We write Γ_f or $\Gamma(f)$ if we want to stress that Γ is the relation associated with the function $f = (X, Y, \Gamma)$. Let $x \in X$. We write $f(x)$ for the uniquely determined $y \in Y$ such that $(x, y) \in \Gamma$. It is customary to write

$$f : X \to Y, \qquad x \mapsto f(x) \tag{5.7}$$

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 the expression "$a \mapsto b$" as "a is assigned to b" or "a maps to b".

We call Γ the **graph** of the function f. Clearly

$$\Gamma = \Gamma_f = \Gamma(f) = \{(x, f(x)) : x \in X\}. \tag{5.8}$$

We 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

$$f(X) := \{y \in Y : y = f(x) \text{ for some } x \in X\} = \{f(x) : x \in X\} \tag{5.9}$$

of Y is called the **range** or **image** of the function $f(\cdot)$.

We say "f maps X into Y" and "f maps the domain value x to the function value $f(x)$". □

We say that two functions $f = (X, Y, \Gamma)$ and $f' = (X', Y', \Gamma')$ are **equal** if $X = X'$, $Y = Y'$, and $\Gamma = \Gamma'$. Note that $X = X'$ follows from $\Gamma = \Gamma'$ because

$$x \in X \Leftrightarrow (x, y) \in \Gamma \text{ for some (unique) } y \in Y \Leftrightarrow (x, y) \in \Gamma' \text{ for some } y \in Y \Leftrightarrow x \in X'.$$

Note that the codomain) Y of f and its range $f(X)$ can be vastly different. For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by the assignment $f(x) = \sin(x)$ then $f(\mathbb{R}) = [-1, 1]$ is a very small part of the codomain!

Figure 5.1 on p.135 illustrates the graph of a function as a subset of $X \times Y$.

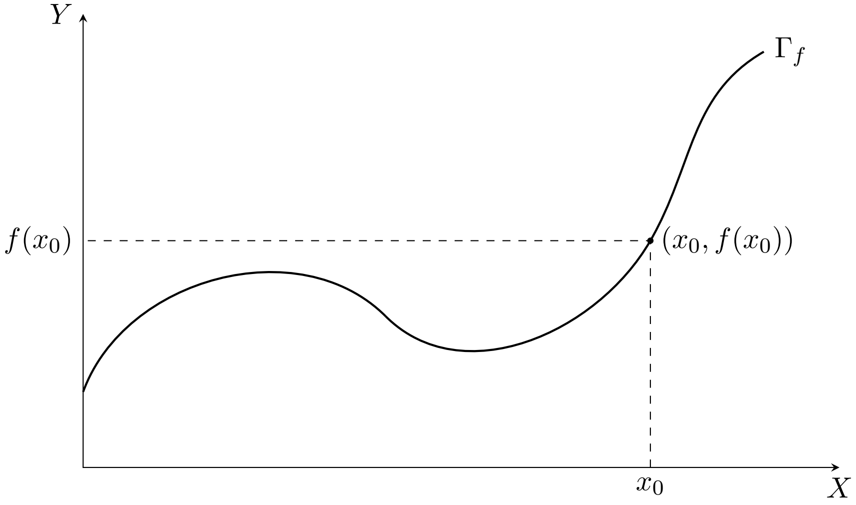


Figure 5.1: Graph of a function.

Remark 5.8. Note that if $Y \subsetneq Y'$ and $f = (X, Y, \Gamma)$ is a function then $f' = (X, Y', \Gamma)$ also is a function: Γ) is a subset of $X \times Y'$ and (5.6) remains valid for Y' in place of Y. But note that the domain X of f is determined by the graph Γ as follows:

$$X = \{x : (x, y) \in \Gamma \text{ for some } y\}. \ \square$$

Remark 5.9 (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, [73]

[73] or if the codomain is a subset of the complex numbers, but we won't discuss complex numbers in this document.

but the majority does what this document tries to adhere to: We use "mapping" and "function" interchangeably and we talk about **real–valued functions** rather than just functions if the codomain is a subset of $\mathbb{R}$ (see (5.16) on p.153). □

Remark 5.10. The symbol x chosen for the argument of the 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 nonnegative real number its (positive) square root:

$$\begin{aligned} f &: [0,\infty[\to\mathbb{R}, && x\mapsto\sqrt{x},\\ f &: [0,\infty[\to\mathbb{R}, && y\mapsto\sqrt{y},\\ f &: [0,\infty[\to\mathbb{R}, && f(\gamma)\ =\ \sqrt{\gamma}. \end{aligned}$$

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:

$$\begin{aligned} &\psi :]0,\infty[\ \to\ \mathbb{R}, && x\mapsto\sqrt{x} \quad \text{(different domain)}.\\ &B : [0,\infty[\ \to]0,\infty[, && x\mapsto\sqrt{x} \quad \text{(different codomain)},\\ &h : [0,1[\ \to\ [0,1[, && x\mapsto\sqrt{x} \quad \text{(different domain and codomain)}. \end{aligned}$$ □

The next topic is function composition. We have already dealt with its associativity in ch.3. See prop.3.1 on p.51.

Definition 5.8 (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 $(x,y)\in\Gamma_f$.

$z:=g(y)=g\big(f(x)\big)$ 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 $\Gamma:=\big\{\big(x,g(f(x)):x\in X\big)\big\}$ is a relation on (X,Z) such that

$$\text{(5.10)}\qquad \text{for each } x\in X \text{ there exists exactly one } z\in Z, \text{namely, } z=g\big(f(x)\big), \text{such that } (x,z)\in\Gamma.$$

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 defined by fairly complex formulas.
□

The following shows how you diagram the composition of two functions. The left picture shows the domains and codomains for each mapping and the right one shows the element assignments.

(5.11) Function composition

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g \circ f"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd} \qquad \begin{tikzcd} x \arrow[r, mapsto, "f"] \arrow[dr, mapsto, "g \circ f"'] & f(x) \arrow[d, mapsto, "g"] \\ & g\big(f(x)\big) \end{tikzcd}$$

The simplest functions are those that map every domain value to one and the same function value.

Definition 5.9 (Constant functions)**.** Let Y be a nonempty set and $y_0 \in Y$. You can think of y_0 as a function from any nonempty set X to Y as follows:

$$y_0(\cdot) : X \to Y; \qquad x \mapsto y_0.$$

In other words, the function $y_0(\cdot)$ assigns to each $x \in X$ one and the same value y_0. We call such a function which only takes a single value a **constant 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. □

We have a special name for the "do nothing function" which assigns each argument to itself:

Definition 5.10 (identity mapping)**.** Given any nonempty 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. □

5.2.3 Examples of Functions

We now give some examples of functions. You might find some of them rather difficult to understand at first reading.

Example 5.10. 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

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^3. \ \square$$

Example 5.11. Let $\Gamma := \{(x, x^2 + 1) : x \in \mathbb{R}\}$. Then $g = (\mathbb{R}, \mathbb{R}, \Gamma)$ is the function

$$g : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^2 + 1. \ \square$$

Example 5.12. Let $\Gamma := \{(a, \ln(a)) : a \in]0, \infty[\}$. Here $\ln(a)$ denotes the natural logarithm of a. Then $h = (]0, \infty[, \mathbb{R}, \Gamma)$ is the function

$$h :]0, \infty[\to \mathbb{R}, \qquad x \mapsto \ln(x). \ \square$$

Example 5.13. Let $\Gamma := \{(x, \sqrt{x}) : x \in [0, \infty[\ \}$. Then $\varphi = ([0, \infty[, \mathbb{R}, \Gamma)$ is the function

$$\varphi : [0, \infty[\to \mathbb{R}, \qquad x \mapsto \sqrt{x}. \ \square$$

Example 5.14. 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

$$\psi : [0, \infty[\to [0, \infty[, \qquad x \mapsto \sqrt{x}. \ \square$$

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 5.15. We define a function which maps two-dimensional vectors to numbers. Let

$$A := \{\big((x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\ \}, \qquad \Gamma := \{\big((x, y), \sqrt{1 - x^2 - y^2}\big) : (x, y) \in A\}.$$

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. (It is the upper half of the surface of the three dimensional ball centered at the origin and with radius 1). $\square$

Example 5.16. 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

$$G : \mathbb{R} \to \mathbb{R}^2, \qquad t \mapsto (\sin t, \cos t).$$

whose image $G(\mathbb{R}$ is the unit circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ Note that the codomain is not a set of real numbers but the Euclidean plane. $\square$

Example 5.17. 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. $\square$

We now reformulate the last example in the framework of linear algebra. Skip this next example if you do not know about matrix multiplication.

Example 5.18. 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 $\mathbb{R} \times \mathbb{R}$ 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 $\vec{x} = (x, y)^\top$ [74] 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

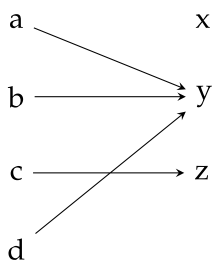
$$H : \mathbb{R}^2 \to \mathbb{R}^2, \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \end{pmatrix}.$$

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 5.19.

Let $X := \{a, b, c, d\}$, $Y := \{x, y, z\}$, $\Gamma := \{ (a, y), (b, y), (c, z), (d, y)\}$. Then $I = (X, Y, \Gamma)$ is the function which maps the elements of X to Y according to the diagram on the right. Note that nothing was said about the nature of the elements of X and Y. One need not know about it to make observations like the following: Examine items **(3)** and **(4)** of remark 5.7 (A better definition of a function) on p.132. 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. □



Example 5.20. 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 $\text{Prob}(A)$ denote the probability that rolling the die results in an outcome $x \in A$.

[74] Here $(x, y)^\top = \begin{pmatrix} x \\ y \end{pmatrix}$ 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.

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 $k \leq 6$ 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. □

Example 5.21. In this example we will define a function $I(\cdot)$ for which the domain $\mathscr{F}$ is a set of functions, and the codomain $\mathscr{G}$ is a set of equivalence classes of functions. For the necessary background on antiderivatives see rem.2.21 on p.46.

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". Let

$$\begin{aligned} \mathscr{F} \;&:= \{f :]a,b[\to \mathbb{R} \text{ such that } f \text{ is continuous on }]a,b[\ \},\\ \mathscr{G} \;&:= \{[g]_\sim : g \text{ is differentiable on }]a,b[\ \ \},\ \text{ where } g \sim g' \;\Leftrightarrow\; g - g' = \text{ const.} \end{aligned}$$

We have seen in rem.2.21 on p.46 that for each $f \in \mathscr{F}$ there exists a differentiable function g, unique up to a constant, such that $g' = f$, i.e., g is an antiderivative of f.

Using "$\sim$" and writing $[g]$ for $[g]_\sim$ this can be rephrased as follows: For each $f \in \mathscr{F}$ there exists a unique $[g]_\sim \in \mathscr{G}$, such that g is an antiderivative of f, i.e., $g' = f$.

We now define a function $I : \mathscr{F} \to \mathscr{G}$ by specifying its graph as the set

$$\Gamma \;:=\; \{\big(f, [g]_\sim\big) : f \in \mathscr{F}, [g] \in \mathscr{G}, g' = f\ \}.\ \square$$

Example 5.22. Compare the following to example 5.21.

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". For example, we could choose $x_0 := \frac{a+b}{2}$. Let

$$\begin{aligned} \mathscr{F} \;&:= \{f : f \text{is a real–valued function with domain }]a,b[\ \},\\ \mathscr{C} \;&:= \{f :]a,b[\to \mathbb{R} \text{ such that } f \text{ is continuous on }]a,b[\ \},\\ \mathscr{D} \;&:= \{f :]a,b[\to \mathbb{R} \text{ such that } g \text{ is differentiable on }]a,b[\ \}. \end{aligned}$$

Note that $\mathscr{D} \subseteq \mathscr{C} \subseteq \mathscr{F}$ because differentiable functions are continuous. We define the following equivalence relation on $\mathscr{D}$: $f \sim g \Leftrightarrow f - g =$ const. [75] Let

$$\mathscr{A} := \{[f] : f \in \mathscr{D}\}$$

be the set of all equivalence classes of differentiable functions on $]a, b[$. Then

$$I : \mathscr{C} \to \mathscr{A}; \quad f \mapsto [I(f)] \text{ where } I(f) :]a, b[\to \mathbb{R} \text{ is the function } x \mapsto I(f)(x) := \int_{x_0}^{x} f(u)du,$$

is a function whose domain $\mathscr{C}$ is a set of functions and whose codomain $\mathscr{A}$ is a set of equivalence classes (i.e., sets(!)) of functions. □

5.2.4 A First Look at Direct Images and Preimages of a Function

Introduction 5.2. We continue with yet another example. It leads to the very important definition of the direct images of subsets of the domain, and of the preimages of subsets of the codomain of a function. □

Example 5.23. Let X and Y be nonempty sets and $f : X \to Y$. We define two functions $f_\star$ and $f^\star$ which are associated with f and for which both arguments and function values are sets(!) as follows.

(a) $f_\star : 2^X \to 2^Y; \qquad A \mapsto f_\star(A) := \{f(a) : a \in A\}$,
(b) $f^\star : 2^Y \to 2^X; \qquad B \mapsto f^\star(B) := \{x \in X : f(x) \in B\}$.

You should convince yourself that indeed $f_\star$ maps any subset of X to a subset of Y, and that $f^\star$ maps any subset of Y to a subset of X. □

The sets $f_\star(A)$ and $f^\star(B)$ are used pervasively in abstract mathematics, but it is much more common nowadays to write $f(A)$ rather than $f_\star(A)$ and $f^{-1}(B)$ rather than $f^\star(B)$. We will follow this convention.

Definition 5.11.

Let X, Y be two nonempty sets and $f : X \to Y$. We associate with f the functions

(5.12) $\qquad f : 2^X \to 2^Y; \qquad A \mapsto f(A) := \{f(a) : a \in A\},$

(5.13) $\qquad f^{-1} : 2^Y \to 2^X; \qquad B \mapsto f^{-1}(B) := \{x \in X : f(x) \in B\}.$

We call $f : 2^X \to 2^Y$ the **direct image function** and $f^{-1} : 2^Y \to 2^X$ the **indirect image function** or **preimage function** associated with $f : X \to Y$.
For each $A \subseteq X$ we call $f(A)$ the **direct image** of A under f, and for each $B \subseteq Y$ we call $f^{-1}(B)$ the **indirect image** or **preimage** of B under f. □

Note that the range $f(X)$ of f (see (5.9) on p.135 is a special case of a direct image.

[75] Note that $f \sim g \Leftrightarrow f - g =$ const also defines equivalence relations on the supersets $\mathscr{C}$ and $\mathscr{F}$.

Notational conveniences I:

> 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 singletons $\{x\} \subseteq X$ and $\{y\} \subseteq Y$ we obtain $f\{x\}$ and $f^{-1}\{y\}$.
> Many mathematicians will write $f^{-1}(y)$ instead of $f^{-1}\{y\}$ but this author 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 are allowed to talk about the latter only in case that the inverse function f^{-1} of f exists.

⚠

> The same symbol f is used for the original function $f : X \to Y$ and the direct image function $f : 2^X \to 2^Y$, and the symbol f^{-1} which is used here for the indirect image function $f^{-1} : 2^Y \to 2^X$ will be used at the start of ch.5.2.5 to define the inverse function $f^{-1} : Y \to X$ of f in case this can be done. [76] Be careful not to let this confuse you! □

Example 5.24 (Direct images). Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

(a) $f(]-4,-2[) = \{ x^2 : x \in]-4,-2[\} = \{ x^2 : -4 < x < -2 \} =]4,16[$.

(b) $f([1,2]) = \{ x^2 : x \in [1,2] \} = \{ x^2 : 1 \leq x \leq 2 \} = [1,4]$.

(c) $f([5,6]) = \{ x^2 : x \in [5,6] \} = \{ x^2 : 5 \leq x \leq 6 \} = [25,36]$.

(d) $f(]-4,-2[\cup [1,2] \cup [5,6]) = \{ x^2 : x \in]-4,-2[\text{ or } x \in [1,2] \text{ or } x \in [5,6] \}$
$=]4,16[\cup [1,4] \cup [25,36] = [1,16[\cup [25,36]$. □

Example 5.25 (Direct images). Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

(a) $f(]-4,2[) = \{ x^2 : x \in]-4,2[\} = \{ x^2 : -4 < x < 2 \} =]4,16[$.

(b) $f([1,3]) = \{ x^2 : x \in [1,3] \} = \{ x^2 : 1 \leq x \leq 3 \} = [1,9]$.

(c) $f(]-4,2[\cap [1,3]) = \{ x^2 : x \in]-4,2[\text{ and } x \in [1,3] \} = \{ x^2 : 1 \leq x < 2 \} = [1,4[$. □

And here are the results for the preimages of the same sets with respect to the same function $x \mapsto x^2$.

Example 5.26 (Preimages). Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$. Determine

a. $f^{-1}(]-4,-2[)$, **b.** $f^{-1}([1,2])$, **c.** $f^{-1}([5,6])$, **d.** $\{-4 < f < -2 \textbf{ or } 1 \leq f \leq 2 \textbf{ or } 5 \leq f < 6]\}$.

Solution:

a. $f^{-1}(]-4,-2[) = \{ x \in \mathbb{R} : x^2 \in]-4,-2[\} = \{ -4 < f < -2 \} = \emptyset$.

b. $f^{-1}([1,2]) = \{ x \in \mathbb{R} : x^2 \in [1,2] \} = \{ 1 \leq f \leq 2 \} = [-\sqrt{2},-1] \cup [1,\sqrt{2}]$.

c. $f^{-1}([5,6]) = \{ x \in \mathbb{R} : x^2 \in [5,6] \} = \{ 5 \leq f \leq 6 \} = [-\sqrt{6},-\sqrt{5}] \cup [\sqrt{5},\sqrt{6}]$.

d. $\{-4 < f < -2 \textbf{ or } 1 \leq f \leq 2 \textbf{ or } 5 \leq f < 6]\} = f^{-1}(]-4,-2[\cup [1,2] \cup [5,6])$
$= \{ x \in \mathbb{R} : x^2 \in]-4,-2[\text{ or } x^2 \in [1,2] \textbf{ or } x^2 \in [5,6] \}$
$= [-\sqrt{2},-1] \cup [1,\sqrt{2}] \cup [-\sqrt{6},-\sqrt{5}] \cup [\sqrt{5},\sqrt{6}]$. □

Example 5.27 (Preimages). Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$. Determine

a. $f^{-1}(]-4,2[)$, **b.** $f^{-1}([1,3])$, **c.** $\{-4 < f < 2 \textbf{ and } 1 \leq f \leq 3\}$.

Solution:

a. $f^{-1}(]-4,2[) \;=\; \{\, x \in \mathbb{R} : x^2 \in\,]-4,2[\, \} \;=\; \{\, x \in \mathbb{R} : -4 < x^2 < 2 \,\} \;=\,]-\sqrt{2},\sqrt{2}[$.

b. $f^{-1}([1,3]) \;=\; \{\, x \in \mathbb{R} : x^2 \in [1,3] \,\} \;=\; \{\, x \in \mathbb{R} : 1 \leq x^2 \leq 3 \,\} \;=\; [-\sqrt{3},-1] \,\cup\, [1,\sqrt{3}]$.

c. $\{-4 < f < 2$ **and** $1 \leq f \leq 3\} \;=\; f^{-1}(]-4,2[\,\cap\,[1,3])$
$= \{\, x \in \mathbb{R} : x^2 \in\,]-4,2[$ **and** $x^2 \in [1,3] \,\}$
$= \{\, x \in \mathbb{R} : 1 \leq x^2 < 2 \,\} \;=\,]-\sqrt{2},-1] \,\cup\, [1,\sqrt{2}[\,.$ □

Remark 5.11 (Notational conveniences II:).

In probability theory the following notation is also very common:
$\{f \in B\} := f^{-1}(B),\ \{f = y\} := f^{-1}\{y\}$.
Let $\mathscr{R}$ be either of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$. Assume that the codomain of f is considered a subset of $\mathscr{R}$. Let $a, b \in \mathscr{R}$ such that $a < b$. We write $\{a \leq f \leq b\} := f^{-1}([a,b]_{\mathscr{R}}),\ \{a < f < b\} := f^{-1}(]a,b[_{\mathscr{R}}),$ $\{a \leq f < b\} := f^{-1}([a,b[_{\mathscr{R}}),\ \{a < f \leq b\} := f^{-1}(]a,b]_{\mathscr{R}}),\ \{f \leq b\} := f^{-1}(]-\infty,b]_{\mathscr{R}})$, etc.
□

Proposition 5.3. *Some simple properties:*

(5.14) $\quad f(\emptyset) \;=\; f^{-1}(\emptyset) \;=\; \emptyset$

(5.15) $\quad A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2)$ $\quad$ (***monotonicity of*** $f\{\dots\}$)

(5.16) $\quad B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2)$ $\quad$ (***monotonicity of*** $f^{-1}\{\dots\}$)

(5.17) $\quad x \in X \Rightarrow f(\{x\}) \;=\; \{f(x)\}$

(5.18) $\quad f(X) \;=\; Y \quad\Leftrightarrow\quad f$ *is "surjective" (see def.5.12 on p.144)*

(5.19) $\quad f^{-1}(Y) \;=\; X \quad$ *always!*

PROOF: Left as exercise 8.9 on p.249. ■

5.2.5 Injective, Surjective and Bijective functions

Introduction 5.3. Given two nonempty sets X and Y we did not find every relation $\Gamma \subseteq X \times Y$ suitable to serve as the graph of a function $X \to Y$: We demanded that for each $x \in X$ there should be one and only one $y \in Y$ suitable as a function value, i.e., there should be one and only one $y \in Y$ such that $y \in \Gamma$. The example $f : \mathbb{R} \to \mathbb{R};\quad x \mapsto x^2$ demonstrates that this relationship between domain elements $x \in X$ and codomain elements $y \in Y$ is not symmetric: One can find $y \in \mathbb{R}$ for which zero elements $x \in \mathbb{R}$ can be found such that $(x,y) \in \Gamma_f$: that would be all negative numbers y. Moreover there also are many $y \in \mathbb{R}$ for which more than one $x \in \mathbb{R}$ exists which is mapped to y: If $y > 0$ then both $\sqrt{y}$ and $-\sqrt{y}$ have y as function value.

Let X, Y be two nonempty sets and $f : X \to Y$ an arbitrary function with domain X and codomain Y. Restricting the domain of f to a small enough subset $A \subseteq X$ may have the effect that the resulting function f' possesses at most one $(x,y) \in \Gamma_{f'}$ whenever $x \in A$. We will call such functions injective. Also, restricting the codomain of f to a small enough subset $B \subseteq Y$ may result in a function f''

which satisfies the following: For each $y \in B$ there exists at least one $x \in X$ such that $(x, y) \in \Gamma_{f''}$. We will call such functions surjective.

We demonstrate this by using again the function $f : \mathbb{R} \to \mathbb{R}; \quad x \mapsto x^2$ as an example. If we restrict its domain $\mathbb{R}$ to $[0, \infty[$ or any nonempty subset thereof then the resulting function will be injective, and if we restrict its codomain $\mathbb{R}$ to $f(\mathbb{R}) = [0, \infty[$ (the range of f) then we wills say of the resulting function that it is surjective. □

The above leads to the following definition.

Definition 5.12 (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

$$\text{(5.20)} \qquad \text{for each } y \in Y \text{ there exists at least one } x \in X \text{ such that } (x, y) \in \Gamma_f.$$

But if $f(X) = Y$, i.e., if (5.20) holds, we call f **surjective** or a **surjection**. We also 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

$$\text{(5.21)} \qquad \text{for each } y \in Y \text{ there exists at most one } x \in X\text{such that } (x, y) \in \Gamma_f.$$

then we call f **injective** or an **injection**. We also say that f is a **one to one** function.

We can express (5.21) 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**, also a . We often write $f : X \xrightarrow{\sim} Y$ for a bijective function f.

It follows from (5.20) and (5.21) that f is bijective if and only if

$$\text{(5.22)} \qquad \text{for each } y \in Y \text{ there exists exactly one } x \in X\text{such that } (x, y) \in \Gamma_f.$$

We rewrite (5.22) by employing Γ_f's inverse relation $\Gamma_f^{-1} = \{(y, x) : (x.y) \in \Gamma\}$ (see def. 5.6 on p.131) and obtain

$$\text{(5.23)} \qquad \text{for each } y \in Y \text{ there exists exactly one } x \in X\text{such that } (y, x) \in \Gamma_f^{-1}.$$

But this implies, according to (5.6), 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

$$\text{(5.24)} \qquad \Gamma_f^{-1} = \Gamma_g.$$

We call g the **inverse mapping** or **inverse function** of f and write f^{-1} instead of g. □

Notation 5.1.

We will occasionally use special arrow symbols to give a visual clue about injectivity, surjectivity and bijectivity of a function.

a) $f : X \twoheadrightarrow Y$ and $X \overset{f}{\twoheadrightarrow} Y$ indicate that the function f is surjective,

b) $f : X \rightarrowtail Y$ and $X \overset{f}{\rightarrowtail} Y$ indicate that the function f is injective,

c) $f : X \xrightarrow{\sim} Y$ and $f : X \xrightarrow{\cong} Y$ indicate that the function f is bijective. □

Moreover, $X \cong Y$ implies that there exists a bijection between the sets X and Y.

Remark 5.12.

(a) It follows from (5.24) that

$$\Gamma_f^{-1} \;=\; \Gamma_{f^{-1}}. \tag{5.25}$$

(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 from **(b)** and **(c)** that

$$\text{if } x \in X, y \in Y \text{ then } f(x) = y \;\Leftrightarrow\; x = f^{-1}(y). \tag{5.26}$$

(e) It also follows from **(b)** and **(c)** 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:

(5.27) Inverse function:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "id_X"'] & Y \arrow[d, "f^{-1}"] \\ & X \end{tikzcd} \qquad \begin{tikzcd} Y \arrow[r, "f^{-1}"] \arrow[dr, "id_Y"'] & X \arrow[d, "f"] \\ & Y \end{tikzcd}$$

□

Theorem 5.1 (Characterization of inverse functions)**.**

Let X and Y be nonempty sets and $f : X \to Y$. 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)** ⇒ **(b)**: We have seen in part **(e)** of remark 5.12 that $g := f^{-1}$ satisfies **(b)**.

PROOF of **(b)** ⇒ **(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., $x_1 = x_2$. This proves injectivity of f. ■

Example 5.28 (Bijective functions)**.**

- **(a)** Let $R = (R, \oplus, \odot)$ be an integral domain and let $a \in R$. Then the function $\varphi : R \to R;\ x \mapsto x \oplus a$ is bijective since it has the function $\varphi^{-1} : R \to R;\ y \mapsto y \ominus a$ as an inverse.
- **(b)** Let $R = (R, \oplus, \odot)$ be an integral domain. Then the function $\psi : R \to R;\ x \mapsto \ominus x$ is bijective since it has the function $\psi^{-1} : R \to R;\ y \mapsto \ominus y$ as an inverse. Note that $\psi^{-1} = \psi$!
- **(c)** Let $G := \{f : \mathbb{R} \to \mathbb{R} : f(x) = ax + b$ for some $a, b \in \mathbb{R}$ where $a \neq 0\}$ of all polynomials of degree 1. We computed in prop.3.5 on p.55 for each element $f : \mathbb{R} \to \mathbb{R};\ f(x) = ax + b$ of G its inverse f^{-1} as the function $y \mapsto \frac{1}{a}y - \frac{b}{a}$. Thus each element $f \in G$ is a bijection $\mathbb{R} \xrightarrow{\sim} \mathbb{R}$.
- **(d)** Let X be a nonempty set and let

$$\mathfrak{E} := \{\sim : \sim \text{ is an equivalence relation on } X\}, \qquad \mathfrak{P} := \{\mathscr{P} : \mathscr{P} \text{ is a partition of } X\}.$$

In prop.5.2 on p.129 we associated with an equivalence relation $\sim$ the partition $\mathscr{P}_\sim = \{[x]_\sim : x \in X$ of its equivalence classes, and we associated with a partition $\mathscr{P}$ of X the equivalence relation $\sim_{\mathscr{P}}$ on X defined as $x \sim_{\mathscr{P}} y \Leftrightarrow x, y$ belong to the same element of $\mathscr{P}$.

With those notations let $\varphi : \mathfrak{E} \to \mathfrak{P}$ be defined as $\varphi(\sim) := \mathscr{P}_\sim$, and let $\psi : \mathfrak{P} \to \mathfrak{E}$ be defined as $\psi(\mathscr{P}) := \sim_{\mathscr{P}}$. We saw in prop.5.2**(c)** that $\sim_{\mathscr{P}_\sim} = \sim$, i.e., that $\psi(\varphi(\sim)) = \sim$ for any $\sim \in \mathfrak{E}$. We further saw in prop.5.2**(d)** that $\mathscr{P}_{\sim_{\mathscr{P}}} = \mathscr{P}$, i.e., that $\varphi(\psi(\mathscr{P})) = \mathscr{P}$ for any $\mathscr{P} \in \mathfrak{P}$. This allows us to restate parts **(c)** and **(d)** of prop.5.2 as follows: The function φ defines a bijection $\mathfrak{E} \xrightarrow{\sim} \mathfrak{P}$, and ψ is the inverse of φ. □

Remark 5.13. [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 need not be sets of real numbers.

Let $R \subseteq X \times Y$.

- **(a)** (5.6) on p.134 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 subset of $X \times Y$ 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)** (5.20) on p.144 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 subset of $X \times Y$ of the form $H(y_0) := \{(x, y_0) : x \in X\}$ for a fixed $y_0 \in Y$ intersects R in **at least one** point.
- **(c)** (5.21) on p.144 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 subset of $X \times Y$ of the form $H(y_0) := \{(x, y_0) : x \in X\}$ for a fixed $y_0 \in Y$ intersects R in **at most one** point.

(d) It follows from (5.22) on p.144 but also from the above that that R is the graph of a bijective 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 subset of $X \times Y$ of the form $H(y_0) := \{(x, y_0) : x \in X\}$ 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. The above is another indication that the inverse graph R^{-1} of a bijective function is a graph of a function (the inverse function f^{-1}). □

Proposition 5.4. *Let $(R, \oplus, \odot, P)$ be an ordered integral domain*

***(A)** Let $b \in R$. Then the function*

$$T : R \to R; \quad x \mapsto x \oplus b,$$

is a bijection.

***(B)** Let $a \in R, a \neq 0$. Then the function*

$$D : R \to a \odot R; \quad x \mapsto a \odot x,$$

is a bijection. (As usual, $a \odot R = aR = \{a \odot r : r \in R\}$.)

The proof is left as exercise 5.11 (see p.164). ■

Remark 5.14. Abstract math is about proving theorems and propositions. Functions are very important tools for many proofs, and in many instances it is 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, and for this reason domain and codomain are very important for the complete specification of a function.

Here is a simple example.

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 □

Proposition 5.5.

Let $X, Y, Z \neq \emptyset$. Let $f : X \to Y$ and $g : Y \to Z$.
(a) *If both f, g are injective then $g \circ f$ is injective.*
(b) *If both f, g are surjective then $g \circ f$ is surjective.*
(c) *If both f, g are bijective then $g \circ f$ is bijective.*

The proof of **(a)** and **(b)** is left as exercise 5.9 on p.164.

PROOF of **(c)**: Follows from **(a)** and **(b)** because bijective = injective + surjective. ■

Corollary 5.1. *Let $X, Y, Z \neq \emptyset$. Let $f : X \to Y$ and $g : Y \to Z$.*

(a) *If f is bijective and g is injective then both $g \circ f$ and $f \circ g$ are injective.*
(b) *If f is bijective and g is surjective then both $g \circ f$ and $f \circ g$ are surjective.*
(c) *If f is bijective and g is bijective then both $g \circ f$ and $f \circ g$ are bijective.*

PROOF:

(a) follows from prop.5.5**(a)** because bijective functions are injective.

(b) follows from prop.5.5**(b)** because bijective functions are surjective.

(c) follows from prop.5.5**(c)**. ■

The following proposition is easy to prove and will be used when we compare the sizes of sets later on.

Proposition 5.6. $\boxed{\star}$ *Let X be an arbitrary set and let A be a nonempty proper subset of X. so that $X = A \biguplus A^{\complement}$ is a partitioning of X into two nonempty subsets A and $A^{\complement}$. Let $a \in A$, $a_0 \in A^{\complement}$ and $A' := (A \setminus \{a\}) \biguplus \{a_0\}$. Then the function*

$$\varphi : A' \xrightarrow{\sim} A; \qquad x \mapsto \begin{cases} x & \text{if } x \neq a_0, \\ a & \text{if } x = a_0 \end{cases}$$

is a bijection.

PROOF: The proof is left as exercise 5.10. ■

We now examine conditions under which there are functions $f : X \to Y$ and $g : Y \to X$ such that $g \circ f = id_X$, i.e.,

(5.28) $g\big(f(x)\big) = x$ for all $x \in X$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "id_X"'] & Y \arrow[d, "g"] \\ & X \end{tikzcd}$$

Proposition 5.7. *Let $X, Y \neq \emptyset$. Let $f : X \to Y$ and $g : Y \to X$ such that $g \circ f = id_X$. Then*

(a) *f is injective,*
(b) *g is surjective.*

PROOF of **(a)**: Let $x_1, x_2 \in X$. If $f(x_1) = f(x_2)$ then

$$x_1 \;=\; id_X(x_1) \;=\; g(f(x_1)) \;=\; g(f(x_2)) \;=\; id_X(x_2) \;=\; x_2.$$

This proves injectivity of f.

PROOF of **(b)**: Let $x_0 \in X$. Let $y := f(x_0)$. Then $g(y) = g(f(x_0)) = g \circ f(x_0) = x_0$. We found for an arbitrary x_0 in the codomain of g some y which maps to x_0. This proves surjectivity of g. ■

Proposition 5.8. *Let $X, Y \neq \emptyset$.*

(a) *Let $f : X \to Y$. If f is injective then there exists $g : Y \to X$ such that $g \circ f = id_X$ and any such function g is necessarily surjective.*
(b) *Let $g : Y \to X$. If g is surjective then there exists $f : X \to Y$ such that $g \circ f = id_X$ and any such function f is necessarily injective.*

PROOF of **(a)**: Let $Y' := f(X)$ and

$$f' : X \to Y', \qquad x \mapsto f(x),$$

i.e., $f(x) = f'(x)$ for all $x \in X$. The only difference between f and f' is that we shrunk the codomain from Y to $f(X)$, thus making f' not only injective but also surjective, hence bijective. It follows that the inverse $(f')^{-1} : Y' \to X$ exists.

Let x_0 be an arbitrary, but fixed, element of X. We define $g : Y \to X$ as follows.

$$g(y) := \begin{cases} (f')^{-1}(y) & \text{if } y \in Y', \\ x_0 & \text{if } y \notin Y'. \end{cases}$$

Let $x \in X$. Then $f(x) \in Y'$, hence $g \circ f(x) = g \circ f'(x) = (f')^{-1}\big(f'(x)\big) = x$. As x was an arbitrary element of x, this proves $g \circ f = id_X$. We observe that g is surjective according to prop.5.7**(b)**.

PROOF of **(b)**: If $x \in X$ then the surjectivity of g implies that $g^{-1}\{x\} \neq \emptyset$. We thus can associate with each $x \in X$ some $y_x \in g^{-1}\{x\}$. [77]

Let $f : X \to Y$ be the function $x \mapsto y_x$ described by the above association. If $x \in X$ then

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

The first equality follows from the definition of f and the second one is true because $y_x \in g^{-1}\{x\}$. It follows from prop.5.7**(a)** that f is injective. ■

There are special names for functions f and g which are related by (5.28).

Definition 5.13 (Left inverses and right inverses)**.** Let $X, Y \neq \emptyset$.

Let $f : X \to Y$ and $g : Y \to X$ such that $g \circ f = id_X$. We say that

(a) f possesses a **left inverse**,
(b) g is a **left inverse** of f,
(c) g possesses a **right inverse**,
(d) f is a **right inverse** of g. □

Remark 5.15. There is no good way to remember which function in the composition of f and g is/has a left inverse and which one is/has a right inverse since the order of f and g in the expression $g \circ f$ is reversed in the expression $X \xrightarrow{f} Y \xrightarrow{g} Z$. The author's suggestion:

- f has (g as) a left inverse since g is to the left of f in the expression $g \circ f$,
- g has (f as) a right inverse since f is to the right of g in the expression $g \circ f$. □

We combine the definition of left/right inverses with the preceding two proposition and obtain

Theorem 5.2. *Let $X, Y \neq \emptyset$.*

(a) *Let $f : X \to Y$. Then f is injective $\Leftrightarrow$ f has a left inverse (which is necessarily surjective).*
(b) *Let $g : Y \to X$. Then g is surjective $\Leftrightarrow$ g has a right inverse (which is necessarily injective).*
(c) *An injection $X \to Y$ exists $\Leftrightarrow$ a surjection $Y \to X$ exists.*

[77] The ability to do such selections $y_x \in g^{-1}\{x\}$ regardless of the nature of X, Y and of the surjective function $g : Y \to X$ is not something one can prove. It requires acceptance of the **Axiom of Choice**. See Chapter 5.3 (optional) in which a complete proof is given that the Axiom of Choice is equivalent to the existence of $f : X \to Y$ such that $g \circ f = id_X$ for any surjective $g : Y \to X$. See also Remark 15.1 on p.453 in ch.15 (Applications of Zorn's Lemma).

PROOF of **(a)** $\Rightarrow$): prop.5.8**(a)**.
PROOF of **(a)** $\Leftarrow$): prop.5.7**(a)**.

PROOF of **(b)** $\Rightarrow$): prop.5.8**(b)**.
PROOF of **(b)** $\Leftarrow$): prop.5.7**(b)**.

PROOF of **(c)** $\Rightarrow$): Let $f : X \to Y$ be injective. According to part **(a)** there exists a left inverse $g : Y \to X$ and this function is surjective
PROOF of **(c)** $\Leftarrow$): Let $g : Y \to X$ be surjective. According to part **(b)** there exists a right inverse $f : X \to Y$ and this function is injective ■

Remark 5.16. Let X and Y be two nonempty sets. No assumptions are made concerning how X and Y might be related.

(a) Let $y_0 \in Y$. Then the function

(5.29) $$f : X \xrightarrow{\sim} \{y_0\} \times X; \quad x \mapsto (y_0, x)$$

is bijective because f has the function $(y_0, x) \mapsto x$ as an inverse.

(b) Let u, v elements of some set. An injection/surjection/bijection $X \to Y$ exists if and only if an injection/surjection/bijection $\{u\} \times X \to \{v\} \times Y$ exists.

(c) Let u, v elements of some set such that $u \neq v$. Then the sets $\{u\} \times X$ and $\{v\} \times Y$ are disjoint. □

5.2.6 Binary Operations and Restrictions and Extensions of Functions

Introduction 5.4. When we defined groups, integral domains and other algebraic structures in ch.3 (The Axiomatic Method) we made use of binary operations such as "$\odot$" which assign to any two elements x and y of such an algebraic structure $\mathfrak{A}$ a third element $z \in \mathfrak{A}$, and also of the "unary operations $\ominus$ and $\cdot^{-1}$" which assign to $x \in \mathfrak{A}$ its inverse $\ominus x \in \mathfrak{A}$ or $x^{-1} \in \mathfrak{A}$ if it exists.

Beside formalizing these notions we will also define restrictions of functions to subsets of their domain and extensions of functions to supersets of their domain. We have previously discussed in the introduction to ch.5.2.5 (Injective, Surjective and Bijective functions) that confining a function to a smaller domain may make that restriction injective. □

We start with the formal definition of unary and binary operations as functions.

Definition 5.14 (Binary and unary operations). $\boxed{\star}$ Let X be a nonempty set. A **binary operation** on X is a function

(5.30) $$\diamond :\ X \times X \longrightarrow X; \quad (x, y) \ \mapsto\ x \diamond y \ := \diamond(x, y).$$

A **unary operation**, on X is a function

(5.31) $$\bullet :\ X \longrightarrow X; \quad x \ \mapsto\ \bullet(x).\ \square$$

One often writes $x^\bullet$ or $\bullet x$ instead of $\bullet(x)$. For example, $-x$ instead of $-(x)$ and x^{-1} rather than ${}^{-1}(x)$.

Example 5.29. The following are examples of binary operations.

(a) Addition on $X = \mathbb{N}$ or $X = \mathbb{Z}$ or $X = \mathbb{Q}$ or $X = \mathbb{R}$ is a binary operation

$$+ :\ X \times X \longrightarrow X; \quad (x, y) \ \mapsto \ x + y. \tag{5.32}$$

(b) Multiplication on $X = \mathbb{N}$ or $X = \mathbb{Z}$ or $X = \mathbb{Q}$ or $X = \mathbb{R}$ is a binary operation

$$\cdot :\ X \times X \longrightarrow X; \quad (x, y) \ \mapsto \ x \cdot y. \tag{5.33}$$

(c) Let X be a nonempty set and $\mathscr{F} \ := \ \{$ functions $f : X \to X\}$. Function composition

$$\circ :\ \mathscr{F} \times \mathscr{F} \longrightarrow \mathscr{F}; \quad (f, g) \ \mapsto \ g \circ f \tag{5.34}$$

where $g \circ f : X \to X$ is the function defined by $x \mapsto g \circ f(x) \ := \ g\big(f(x)\big)$ $(x \in X)$.

Here are some examples of unary operations.

(d) Negative number: Let $X = \mathbb{N}$ or $X = \mathbb{Z}$ or $X = \mathbb{Q}$ or $X = \mathbb{R}$. Then

$$- :\ X \to X; \quad x \ \mapsto \ -x. \tag{5.35}$$

is a unary operation.

(e) Reciprocal: Let $X = \mathbb{Q}_{\neq 0}$ or $X = \mathbb{R}_{\neq 0}$

$$\cdot^{-1} :\ X \to X; \quad x \ \mapsto \ x^{-1} \ = \ 1/x. \tag{5.36}$$

is a unary operation.

(f) Let X be a nonempty set and $\mathscr{B} \ := \ \{$ bijective functions $f : X \to X\}$. Let

$$\cdot^{-1} :\ \mathscr{B} \to \mathscr{B}; f(\cdot) \ \mapsto \ f^{-1}(\cdot) \tag{5.37}$$

be the function which assigns to the function $x \mapsto f(x)$ its (uniquely determined) inverse function $y \mapsto f^{-1}(y))$. Then this assignment is a unary operation on $\mathscr{B}$.

Note that assignment of the reciprocal number and assignment of the inverse function both are denoted by the symbol "$\cdot^{-1}$". There is no danger of confusing the two unary operations because one of them operates on a set of numbers and the other one on a set of functions. □

Definition 5.15 (Restriction/Extension of a function)**.** Given are three nonempty sets A, X and Y such that $A \subseteq X$ and a function $f : X \to Y$ with domain X.

(a) We define the **restriction of f to A** as the function

(5.38) $$f\big|_A : A \to Y \quad \text{defined as} \quad f\big|_A(x) := f(x) \text{ for all } x \in A.$$

(b) 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 5.30. 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 $0 \leq x \leq 1$ and αx otherwise defines a different extension of f to $\mathbb{R}$. □

Notation 5.2. 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 5.17. The restriction $f\big|_A$ is always uniquely determined by f. Such is not the case for extensions if A is a strict subset of X unless some conditions are imposed on the nature of the extension.

For example, if we had asked for continuity [78] of the extension φ_α of f in example 5.30 above, only $\varphi_1(x) = 3x^2$ if $0 \leq x \leq 1$ and x otherwise would qualify. □

Proposition 5.9. *Let X, Y be nonempty sets. Let $f : X \xrightarrow{\sim} Y$ be bijective*

- *(a)* *Let $\emptyset \neq A \subseteq X$, $B := f\big|_A(A) = \{f(a) : a \in A\}$.* [79] *Let $f' : A \to B;\ \ x \mapsto f(x)$, i.e., $f' = f\big|_{A'}$ except that we have shrunken the codomain Y to B. Then f' is bijective.*
- *(b)* *Let $\emptyset \neq V \subseteq Y$. Let $U := \{x \in X : f(x) \in V\}$.* [80] *Let $f'' : U \to V;\ \ x \mapsto f(x)$, i.e., $f'' = f\big|_{U'}$ except that we have shrunken the domain X to U. Then f'' is bijective.*

The proof of **(a)** is left as exercise 5.17. See p.166.

PROOF of **(b)**:

We first prove injectivity. Let $u, u' \in U$ such that $u \neq u'$. Then $f(u) \neq f(u')$ because f is injective. But then $f'(u) = f(u) \neq f(u') = f'(u')$. It follows that f' is injective.

Let $b \in V$. Since $U = \{x \in X : f(x) \in V\}$, the set U contains all items with function values in V, hence there exists $u \in U$ such that $f(u) = v$. We have proven surjectivity. ■

Example 5.31. For example let $f : [0, \infty[\to [0, \infty[;\ \ x \to x^2$. Then f is bijective.

- **(a)** Let $A := [0,2]$. Then $f(A) = f\big|_A(A) = [0,4]$, and $f\big|_{[0,2]}[0,2] \xrightarrow{\sim} [0,4]$ is bijective.
- **(b)** Let $V := [1,9]$. Then $f^{-1}(V) = [1,3]$, and $f\big|_{[1,3]}[1,3] \xrightarrow{\sim} [1,9]$ is bijective. □

[78] Continuity of functions $y = f(x)$ with real numbers x and y will be defined in ch.9.3 (Convergence and Continuity in $\mathbb{R}$). See Definition 9.12 on p.271. Until then use your knowledge from calculus.

[79] i.e., $B = f(A)$

[80] i.e., $U = f^{-1}(B)$

5.2.7 Real–Valued Functions and Polynomials

Introduction 5.5. If we deal with functions such as $f(x) = sin(2x) - 3x^3$ or $g(x,y,z) = \sqrt{x^2+y^2+z^2}$ or $h(x,y,z) = (x^y)^z$ for which the codomain is (a subset of) $\mathbb{R}$, i.e., each function value is a real number, then we can add those function values or multiply them or do anything else one can do with real numbers. In particular we can define for two functions $f_1, f_2 : X \to \mathbb{R}$ with matching domain X their sum $(f_1+f_2)(x) = f_1(x)+f_2(x)$ and their product $(f_1 \cdot f_2)(x) = f_1(x)f_2(x)$. Thus we have $g+h(x,y,z) = \sqrt{x^2+y^2+z^2} + (x^y)^z$ and $gh(x,y,z) = \sqrt{x^2+y^2+z^2}(x^y)^z$.

Does it matter at all what kind of structure if any the domain has been endowed with? Not for the subject matter that will be discussed here. For example let $C := \{$ all inhabitants of Chicago $\}$, let $a : X \to [0,\infty[$ be the function which assigns to each person x who lives in Chicago her/his age in days $a(x)$, and let $s : C \to [0,\infty[$ be the function which assigns to x the number of days $s(x)$ s/he has been severely ill so far. Then we can build the function $s/a : C \to [0,\infty[\,; \quad x \mapsto 100\big(s(x)/a(x)\big)$ which assigns to each inhabitant of Chicago the percentage of time they have been sick so far.

We just mention in passing that we can apply this principle to codomains which carry any kind of structure. For example if $(G,\diamond)$ is a group and X is a nonempty set then we can associate with $f, gX \to G$ the functions $f \diamond g : x \mapsto f(x) \diamond g(x)$ and $f^{-1} : x \mapsto \big(f(x)\big)^{-1}$.

Here is an example where each function value is a set. Let C again denote the inhabitants of Chicago. We assign to each $x \in C$ the set $P(x) \subseteq C$ of all people who live in Chicago and whom x knows professionally, and the set $F(x)$ of all friends that x has in Chicago. We thus have defined two functions $P, F : C \to 2^C$. We cannot build the sum $P+F$ or the quotient P/F, but we can construct functions such as $P \cap F : x \mapsto P(x) \cap F(x)$, the set of all friends who live in Chicago whom x knows professionally and $P^{\complement} : x \mapsto P(x)^{\complement}$, the set of all Chicagoans with whom x does not have a professional relationship. □

We start with the definition of a real–valued function.

Definition 5.16 (Real–Valued Function)**.** Let X be an arbitrary, nonempty set. If the codomain Y of a mapping

$$f : X \to Y; \qquad x \mapsto f(x)$$

is a subset of $\mathbb{R}$, then we call $f(\cdot)$ a **real function** or **real–valued function**. □

Note that the above definition does not exclude the case $Y = \mathbb{R}$ because $Y \subseteq \mathbb{R}$ is in particular true if both sets are equal.

As we mentioned in the introduction to this section real–valued 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–valued functions. Of course, division by zero is not allowed:

Definition 5.17 (Operations on real–valued functions)**.** $\boxed{\star}$ Let X be an arbitrary nonempty set. Given are two real–valued functions $f(\cdot), g(\cdot) : X \to \mathbb{R}$ and a real number α. The **sum** $f+g$, **difference** $f-g$, **product** fg or $f \cdot g$, **quotient** 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:

$$
\begin{aligned}
(f+g)(x) &:= f(x)+g(x),\\
(f-g)(x) &:= f(x)-g(x),\\
(fg)(x) &:= f(x)g(x),\\
(f/g)(x) &:= f(x)/g(x) \qquad \text{for all } x \in X \text{ where } g(x) \neq 0,\\
(\alpha f)(x) &:= \alpha \cdot g(x). \;\square
\end{aligned} \tag{5.39}
$$

Remark 5.18. Note that scalar multiplication $(\alpha f)(x) = \alpha \cdot f(x)$ is a special case of multiplying two functions $(gf)(x) = g(x)f(x)$, namely the case where $g(x) = \alpha$ for all $x \in X$ (constant function α). □

Definition 5.18 (Negative function)**.** $\boxed{\star}$ Let X be an arbitrary, nonempty set and let $f : X \to \mathbb{R}$. The function

$$-f(\cdot) : X \to \mathbb{R}; \qquad x \mapsto -f(x).$$

is called **negative** f or **minus** f. We usually write $-f$ for $-f(\cdot)$. □

Note 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.

All those last definitions about sums, products, scalar products, . . . of real–valued functions are very easy to understand if you remember that, for any fixed $x \in X$, you just deal with ordinary numbers!

Example 5.32 (Arithmetic operations on real–valued functions)**.**
For simplicity, let $X := \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. Let

$$
\begin{array}{ll}
f : \mathbb{R}_{\geq 0} \to \mathbb{R}; & x \mapsto (x-1)(x+1)\\
g : \mathbb{R}_{\geq 0} \to \mathbb{R}; & x \mapsto (x-1)\\
h : \mathbb{R}_{\geq 0} \to \mathbb{R}; & x \mapsto (x+1)
\end{array}
$$

Then

$$
\begin{aligned}
(f+h)(x) &= (x-1)(x+1)+x+1 \;= x^2-1+x+1 \;=\; x(x+1)\ \forall x \in \mathbb{R}_{\geq 0},\\
(f-g)(x) &= (x-1)(x+1)-(x-1) \;= x^2-1-x+1 \;=\; x(x-1)\ \forall x \in \mathbb{R}_{\geq 0},\\
(gh)(x) &= (x-1)(x+1) \;= f(x)\ \ \forall x \in \mathbb{R}_{\geq 0},\\
(f/h)(x) &= (x-1)(x+1)/(x+1) \;= x-1 \;=\; g(x)\ \forall x \in \mathbb{R}_{\geq 0},\\
(f/g)(x) &= (x-1)(x+1)/(x-1) \;= x+1 \;=\; h(x)\ \forall x \in \mathbb{R}_{\geq 0} \setminus \{1\}
\end{aligned}
$$

It is really, really important for you to understand that f/g and h are **not the same functions**. Here is the reason. f/g is not defined for $x = 1$ because $\frac{(1-1)(1+1)}{1-1} =$ "$0/0$". The domain of f/g is $\mathbb{R}_{\geq 0} \setminus \{1\}$. It is different from $\mathbb{R}_{\geq 0}$, the domain of h. It follows that both functions are different. □

Definition 5.19 (Polynomials)**.** Let A be subset of the real numbers and let $p(\cdot) : A \to \mathbb{R}$ be a real–valued function on A. $p(\cdot)$ is called a **polynomial**. if there is an integer $n \geq 0$ 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

$$p(x) \ = \ a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \ = \ \sum_{j=0}^{n} a_jx^j. \tag{5.40}$$

In other words, polynomials are linear combinations of the **monomials** $x \to x^k \quad (k \in (Z)_{\geq 0}$. If $a_n \neq 0$ then we call n the **degree** of p. The zero function $x \mapsto 0 = 0 \cdot x^0$ is a polynomial which we call the **zero polynomial**. Note that it has no degree because we cannot represent it in the form (5.40) with a non–zero coefficient a_n. We call $z \in A$ a **root** of the polynomial p if $p(z) = 0$.

If we talk about polynomials without explicitly specifying the domain then it is implied that the domain is $\mathbb{R}$. □

Proposition 5.10. *If p_1 and p_2 are polynomials and if $\lambda \in \mathbb{R}$ then*

- ***(a)*** *The sum $x \mapsto p_1(x) + p_2(x)$ is a polynomial.*
- ***(b)*** *The "scalar product" $x \mapsto \lambda p_1(x)$ is a polynomial.*

The proof is left as exercise 5.19 (see p.166). ■

Example 5.33. Polnomials may not always be given in their "normalized form" (5.40) on p.155. For $x \in \mathbb{R}$ let

$$\begin{aligned} p(x) \ &:= b_0x^0(1-x)^n + b_1x^1(1-x)^{n-1} + \ldots + b_{n-1}x^{n-1}(1-x)^1 + b_nx^n \\ &= \sum_{k=0}^{n} b_kx^k(1-x)^{n-k} \end{aligned} \tag{5.41}$$

We have $b_kx^k(1-x)^{n-k} \ = \ b_kx^k + (-b_k)x^n$. Let $b := \sum\limits_{k=0}^{n} b_k$. Then $p(x) \ = \ \sum\limits_{k=0}^{n} b_kx^k \ - \ bx^n$ is of the form (5.40) (define $a_k := b_k$ for $0 \leq k < n$ and $a_n := b_n - b$) and hence is a polynomial.

The so called Bernstein polynomials which we will examine in ch.6.7 are of the form (5.41). □

Many more properties of functions will be discussed later. Now we look at families, sequences and some additional properties of sets.

5.2.8 Families, Sequences, and Functions as Families

Introduction 5.6. In Chapter 2.4 (A First Look at Functions, Sequences and Families) We were introduced to the notion of a family as collection $(x_i)_{i \in I}$ of items x_i which are subscripted or indexed by the elements i of an arbitrary index set I. We saw that any such family can be thought of as a function with index set I, and that sequences are families with index sets of integers that contain a smallest element, the start index. We also noticed that families (in particular, sequences) are functions in disguise shich associate with each index $i \in I$ the corresponding indexed item x_i.

We never gave a reason why one would introduce families if, whatever one can do with them, also can be done with functions. It's really just convenience: The expression

$$\Big(\bigcup_{j \in J} (A_i \cap B_j)\Big)_{i \in I}$$

is significantly shorter than the expression

$$\varphi : I \to 2^{\Omega}; \quad \varphi(i) = \bigcup_{j \in J} \big(A(i) \cap B(j)\big) \qquad \text{where} \quad A : I \to 2^{\Omega} \ \text{ and } \ B : J \to 2^{\Omega}\,.$$

Again, we define sequences as special families, those with index set $J = J = [k_0, \infty[_{\mathbb{Z}}$ for some initial index $k_0 \in \mathbb{Z}$. We also give in this chapter the definition of an infinite subsequence which is consistent with the one given in Chapter 2.4,

Even though finite sequences and their subsequences were already defined in Chapter 2.4, you will not find the exact definitions here. That must wait until Chapter 7 (Cardinality I: Finite and Countable Sets) when the precise definition of finiteness is available. □

We now are ready to give the definition of a family.

Definition 5.20 (Indexed families)**.** Let J and X be nonempty sets and assume that

for each $\imath \in J$ there exists exactly one indexed item $x_\imath \in X$.

(a) We write $(x_\imath)_{\imath \in J}$ for this collection of indexed elements and call it an **indexed family** or simply a **family** in X.
(b) J is called the **index set** of the family.
(c) For each $\jmath \in J$, $x_\jmath$ is called a **member of the family** $(x_\imath)_{\imath \in J}$.

□

Remark 5.19. **(a)** The index $\imath$ 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 $x(\cdot) : J \to X$. This should not surprise you if you recall remark 5.10 on p.136.

(b) Let $R := \{(\imath, x_\imath) : \imath \in J\}$. Then R is a relation on (J, X) which satisfies (5.6) of the definition of a function

$$x(\cdot) : \ J \longrightarrow X, \qquad \imath \mapsto x(\imath) := x_\imath$$

(see Definition 5.7 on p.134), whose graph $\Gamma_{x(\cdot)}$ equals R. In other words,

- Every family $(x_\imath)_{\imath \in J}$ in X can be interpreted as a function
$$x(\cdot) : J \longrightarrow X; \quad \imath \mapsto x_\imath\,. \ \square$$

Remark 5.20. The codomain X does not occur in the notation $(x_\imath)_{\imath \in J}$. This is not a problem because we do not care about surjectivity or injectivity of families. The only thing that matters about the set X is that it is big enough to contain each indexed item. Here are two natural choices for a codomain.

(a) If there is a universal set X which contains all tagged items of the family then selecting X as codomain makes perfect sense.

(b) If there is no universal set then you can think of

$$X = \bigcup \left[x_\imath : \imath \in J \right] := \{ x : x = x_{\imath_0} \text{ for some } \imath_0 \in I \}$$

as the codomain. [81] □

Definition 5.21 (Equality of families).

Two families $(x_i)_{i \in I}$ and $(y_j)_{j \in J}$ are equal if

(a) $I = J$,

(b) $x_i = y_i$ for all $i \in I$. □

Remark 5.21. Equality of families and equality of functions are not identical concepts, since no demand is made in the latter that both families are families in the same set, say, X. But of course, if $(x_i)_{i \in I}$ is a family in X and $(y_j)_{j \in J}$ is a family in Y and those two families are equal then this necessitates

$$\{x_i : i \in I\} = \{y_j : j \in J\} \subseteq X \cap Y. \quad \square \tag{5.42}$$

Note 5.1 (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 $(x_\imath)_{\imath \in J}$. We even may shorten this to $(x_\imath)$ if doing so does not lead do confusion.

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 the index set, 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 5.34. 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 $0 \leq x \leq 1$ 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 5.22. If a family is just some kind of function, why bother with yet another definition? We already gave an answer in the introduction to this section: There we saw an example where writing something as a collection of indexed items rather than as a function is a notational convenience. Here is another example. Take a peek at theorem 8.1 (De Morgan's Law) on p.234. 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

$$\Big(\bigcup_\alpha A_\alpha\Big)^{\complement} = \bigcap_\alpha A_\alpha^{\complement}.$$

[81] General unions and intersections will be defined in ch.8.1 (More on set operations). See Definition 2.29 on p.38.

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 5.35 (Sequences as families)**.** We have worked with special families before: those where the index set is $J = \mathbb{N} = [1, \infty[_{\mathbb{Z}}$ or $J = [0, \infty[_{\mathbb{Z}}$ or, more generally $J = [k_0, \infty[_{\mathbb{Z}}$ for some "start index" $k_0 \in \mathbb{Z}$, and where X is a subset of the real numbers. Example: $x_n := 1/n$. Here

$$(x_n)_{n\in\mathbb{N}} \text{ corresponds to the indexed collection } 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots \quad \square$$

The families from the last example will be called sequences. Preliminary definitions for sequences, subsequences, finite sequences and finite subsequences were given in Definition 2.24 on p.33 and Definition 2.25 on p.34. We will now give precise definitions for sequences and subsequences. Those for finite sequences and finite subsequences will have to wait until ch.7.3 (Finite Sequences and Subsequences and Eventually True Properties).

Definition 5.22 (Sequences and subsequences)**.** Let $n_\star \in \mathbb{Z}$, let $J := [n_\star, \infty[_{\mathbb{Z}} = \{k \in \mathbb{Z} : k \geq n_\star\}$. Let X be an arbitrary nonempty set. An indexed family $(x_n)_{n\in J}$ in X with index set J is called a **sequence** in X with **start index** $n_\star$. We will also write

$$(x_n)_{n\geq n_\star} \quad \text{or} \quad (x_n)_{n=n_\star}^{\infty} \quad \text{or} \quad x_{n_\star}, x_{n_\star+1}, x_{n_\star+2}, \dots$$

for this sequence. As for families, the name of the index variable of a sequence is unimportant as long as it is applied consistently. It does not matter whether one writes, e.g.,

$$(x_n)_{n\geq n_\star} \quad \text{or} \quad (x_j)_{j\geq n_\star} \quad \text{or} \quad (x_\beta)_{\beta\geq n_\star} \quad \text{or} \quad (x_A)_{A=n_\star}^{\infty}.$$ [82]

Let $(n_j)_{j=1}^{\infty}$ be a sequence of integers n_j such that

1) $n_j \in J$ (i.e., a sequence of indices for the above sequence $(x_j)_{j=n_\star}^{\infty}$)
2) $i < j \Rightarrow n_i < n_j$ for all $i, j \in \mathbb{N}$.

Note that $n_j \in J$ for all $j \in \mathbb{N}$ implies $n_\star \leq n_1 < n_2 < \dots$. If we write $I := \{n_j : j \in \mathbb{N}\}$ then we see that $(x_n)_{n\in I} = (x_{n_j})_{j\in\mathbb{N}}$, thus this object is an indexed family whose index set I is a subset of the original index set J. We call $(x_{n_j})_{j\in\mathbb{N}} = (x_{n_j})_{j=1}^{\infty}$ a **subsequence** of the sequence $(x_j)_{j=n_\star}^{\infty}$. This is an appropriate name since we obtain $(x_{n_j})_{j=1}^{\infty}$ from $(x_j)_{j\in J}$ by removing all members x_n such that none of the n_j equals n. Be sure to understand that, according to this definition, the sequence $(n_j)_{j\in\mathbb{N}}$ is a subsequence of the full sequence of indices $(n)_{n=n_\star}^{\infty}$. We will also write

$$(x_{n_j})_{j\in\mathbb{N}} \quad \text{or} \quad (x_{n_j})_{j\geq 1} \quad \text{or} \quad (x_{n_j})_{j=1}^{\infty} \quad \text{or} \quad x_{n_1}, x_{n_2}, x_{n_3}, \dots$$

for this subsequence. □

[82] see Definition 5.20 (indexed families) on p.156.

Note 5.2 (Simplified notation for sequences)**.**

- **(a)** 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.
- **(b)** By default the index set for a sequence is $\mathbb{N} = \{1, 2, 3, 4, \dots\}$.
- **(c)** We are allowed to write $(x_n)_n$ or just (x_n) if there is no confusion about the value of $n_\star$ or if this value is irrelevant for the statement at hand.
- **(d)** Customary simplified notation for subsequences is either of $(x_{n_j})_{j\in\mathbb{N}}$, $(x_{n_j})_{j\geq 1}$, $(x_{n_j})_j$ or simply (x_{n_j}).

Compare this to note 5.1 about simplified notation for families. □

Part **(b)** of the above note deserves repeating:

Assumption 5.1 (indices of sequences)**.**

Unless explicitly stated otherwise, sequences are always indexed $1, 2, 3, \dots$, i.e., the first index is 1, there is no largest index and, given any index, you obtain the next one by adding 1 to it. □

Example 5.36. For $j \in \mathbb{N}$ let $x_j := (-1)^j$. Then $\big((-1)^n\big)_{n=1}^\infty$ is the sequence

$$x_1 = -1, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 1, \quad x_5 = -1, \ \dots$$

With the notations of Definitions 2.24 and 2.25 we have $X = \mathbb{Z}$ and $n_\star = 1$ (i.e., $J = \mathbb{N}$). If we choose $n_j := 2j$, then the corresponding index set $\{2, 4, 6, \dots\}$ is the set of all even indices, and we obtain the subsequence

$$(x_{n_j})_{j=1}^\infty \ = \ \big((-1)^{2j}\big)_{j=1}^\infty \ = \ 1, 1, 1, 1, \dots.$$

If we choose $n_j := 2j - 1$ then we obtain as index set the subset $\{1, 3, 5, \dots\}$ of all odd indices, and thus the subsequence

$$(x_{n_j})_{j=1}^\infty \ = \ \big((-1)^{2j-1}\big)_{j=1}^\infty \ = \ -1, -1, -1, -1, \dots \quad \square$$

Here is another example of a sequence.

Example 5.37 (Series (summation sequence))**.** Let $s_k := 1 + 2^{-1} + 2^{-2} + \ldots + 2^{-k}$ $(k = 1, 2, 3, \dots)$:

$$\begin{aligned} s_1 =& 1, \quad s_2 = 1 + 1/2 = 2 - 1/2, \quad s_3 = 1 + 1/2 + 1/4 = 2 - 1/4, \quad \dots, \\ s_k =& 1 + 1/2 + \ldots + 2^{k-1} = 2 - 2^{k-1}; \quad s = 1 + 1/2 + 1/4 + 1/8 + \ldots \quad \text{“infinite sum”}. \end{aligned}$$

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**. [83] □

[83] The precise definition of a series will be given in ch.13.2 (Function Sequences and Infinite Series) on p.414.

Remark 5.23. 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

$$\text{“}\big((\imath, x_\imath)\big)_{\imath\in J}\text{''}. \tag{5.43}$$

Doing so should also make it much easier to see the equivalence of functions and families: (5.43) looks at its core very much like the graph $\{(\imath, x_\imath) : \imath \in J\}$ of the function $\imath \mapsto x_\imath$. □

Remark 5.24 (Families and sequences can contain duplicates)**.** One of the important properties of sets is that they do not contain any duplicates (see Definition 2.1 (sets) on p.14). On the other hand, remark 5.23 casually mentions that families, and hence sequences as special kinds of families, can contain duplicates. Let us look at this 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 $\Gamma \subseteq J \times A = J \times B$ [84] are different from each other because two pairs (a, b) and (x, y) are equal only if $x = a$ and $y = b$. □

In contrast to sets, families and sequences allow us to incorporate duplicates.

A family $(x_\imath)_{\imath\in J}$ in X is specified by 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. Thus we have

Proposition 5.11 (Functions are families and families are functions)**.**

The following two ways of specifying a function $f : X \to Y$, $x \mapsto f(x)$ are equivalent:
(a) *f is defined by its graph* $\{(x, f(x)) : x \in X\}$.
(b) *f is defined by the following family in Y:* $(f(x))_{x\in X}$

Note that the above is one case where a family needed explicit mention of the codomain Y.

PROOF: This follows from the material leading to the above proposition. ■

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.

5.3 Right Inverses and the Axiom of Choice ⋆

The following is a greatly expanded version of the online article http://planetmath.org/surjectionandaxiomofchoice about the equivalence of the Axiom of Choice and the existence of right inverses for arbitrary, surjective functions.

[84] Be sure to understand that $J \times A = J \times B$!

Definition 5.23 (Choice function)**.** Let $\mathscr{A}$ be a collection of nonempty sets and let Ω be a set such that $\bigcup[A : A \in \mathscr{A}] \subseteq \Omega$. Let the function

$$c : \mathscr{A} \longrightarrow \Omega \qquad \text{satisfy} \qquad c(A) \in A \;\; \text{for all } A \in \mathscr{A}$$

Then we call c a **choice function** [85] on $\mathscr{A}$. □

The following is a repeat of Proposition 5.8**(b)** and its proof, with emphasis on the use of a choice function. It shows that the acceptance of the Axiom of Choice implies that right inverses exist for any choice of nonempty Y and X and surjective $g : Y \to X$.

Proposition 5.12. *Let $X, Y \neq \emptyset$. Let $g : Y \to X$. If g is surjective then there exists $f : X \to Y$ such that $g \circ f = id_X$.*

PROOF: Let $\mathscr{A} := \{g^{-1}\{x\} : x \in X\}$. The surjectivity of g implies that $g^{-1}\{x\} \neq \emptyset$ for all $x \in X$. According to the Axiom of Choice there exists a choice function $c : \mathscr{A} \to Y$.

Let $f : X \to Y$ be the function $x \mapsto y_x := c(g^{-1}\{x\})$ and let $x \in X$.

Since c is a choice function, $y_x \in g^{-1}\{x\}$ and thus $g(y_x) \in \{x\}$, i.e., $g(y_x) = x$. Thus

$$g \circ f(x) \;=\; g \circ c(g^{-1}\{x\}) \;=\; g(y_x) \;=\; x.$$

The first equality follows from the definition of f and the second one from that of y_x. ■

It is not as easy to show the other direction: Accepting that surjective functions $g : Y \to X$ have right inverses for any choice of nonempty X, Y and $y \mapsto g(y)$ implies the existence of choice functions $\mathscr{A} \to \Omega$ for arbitrary, nonempty Ω and $\mathscr{A} \subseteq 2^\Omega \setminus \emptyset$. This will be shown in the next lemma and subsequent proposition.

Lemma 5.1. *Assume that each surjective function possesses a right inverse, i.e., if Y and X are nonempty and $g : Y \to X$ is surjective then there exists $f : X \to Y$ (necessarily injective) such that $g \circ f = id_X$. See Definition 5.13 (Left inverses and right inverses) on p.149 and the subsequent material. Assume further that $\mathscr{A}$ is a collection of nonempty and disjoint sets.*

Then there exists a choice function on $\mathscr{A}$.

PROOF: Let $\Omega := \biguplus[A : A \in \mathscr{A}]$. Since the elements of $\mathscr{A}$ are disjoint there exists for each $\omega \in \Omega$ a unique A_ω such that $\omega \in A_\omega$. Thus the association

$$\omega \mapsto A_\omega \;\; \text{defines a function} \;\; g : \Omega \to \mathscr{A} \;\; \text{such that } \omega \in g(\omega) = A_\omega\,. \tag{5.44}$$

(A): We show that g is surjective and thus possesses a right inverse.

Let $A \in \mathscr{A}$ and $\omega \in A$. Such ω exists because $A \neq \emptyset$. Let

$$A_\omega \;:=\; g(\omega)\,. \tag{5.45}$$

Then $\omega \in A_\omega$ by the definitions of g and A_ω. Since we assumed $\omega \in A$, it follows that $\omega \in A \cap g(\omega)$. Since the elements of $\mathscr{A}$ are disjoint, $A = g(\omega)$. We have found for an arbitrary $A \in \mathscr{A}$ an $\omega \in \Omega$ such that $\omega \in g(\omega)$, thus g is surjective.

[85] denoted so since this function chooses from each of its arguments A an item $\omega = c(A)$.

(B): By assumption g possesses a right inverse, i.e., there is

$$c : \mathscr{A} \to \Omega \quad \text{such that } c \text{ is injective and} \quad g \circ c \;=\; id_{\mathscr{A}}\,. \tag{5.46}$$

Let $A \in \mathscr{A}$ and $\omega := c(A)$. Then

$$g(\omega) \;=\; g\bigl(c(A)\bigr) \;=\; id_{\mathscr{A}}(A) \;=\; A\,. \tag{5.47}$$

Since $c(A) = \omega$, $\omega \in g(\omega)$ by (5.44) and $g(\omega) = A$ by (5.47), it follows that $c(A) \in A$. This holds for arbitrary $A \in \mathscr{A}$, thus c is a choice function on $\mathscr{A}$. ■

We now remove the restrictive assumption that the members of $\mathscr{A}$ must be mutually disjoint.

Proposition 5.13. *Assume that each surjective function possesses a right inverse. Assume further that $\mathscr{A}$ is a collection of nonempty sets. Then there exists a choice function on $\mathscr{A}$.*

PROOF:
Let $\Omega := \bigcup[A : A \in \mathscr{A}]$. Let

$$\jmath : \mathscr{A} \longrightarrow \Omega \times \mathscr{A}\,; \qquad A \;\mapsto\; \{(\omega', A) : \omega' \in A\}\,.$$

Let $A, A' \in \mathscr{A}$ such that $A \neq A'$. Then $\jmath(A) \cap \jmath(A') = \emptyset$ since all elements $(\omega, A) \in \jmath(A)$ have different second coordinate from the elements $(\omega', A') \in \jmath(A')$, and two elements (x, y) and (x', y') of a cartesian product $X \times Y$ are different unless both $x = x'$ and $y = y'$.
In particular, $A \neq A' \;\Rightarrow\; \jmath(A) \neq \jmath(A')$. It follows that the function

$$\imath : \mathscr{A} \xrightarrow{\sim} \jmath(\mathscr{A})\,; \qquad A \;\mapsto\; \imath(A) \tag{5.48}$$

bijects $\mathscr{A}$ to a collection of disjoint subsets of $\Omega \times \mathscr{A}$.
We infer from Lemma 5.1 the existence of a choice function $c : \imath(\mathscr{A}) \to \Omega \times \mathscr{A}$.
Let $A \in \mathscr{A}$. Since c is a choice function, $c \circ \imath(A) \in \imath(A)$, i.e.,

$$c \circ \imath(A) \;=\; c \circ \jmath(A) \;\in\; \{(\omega', A) : \omega' \in A\}$$

according to the definition of $\jmath(A)$. Thus

$$\text{there exists } \omega \in A \;\text{ such that } c \circ \imath(A) \;=\; (\omega, A)\,. \tag{5.49}$$

$$\text{Let } \;\pi_\Omega : \imath(\mathscr{A}) \longrightarrow \Omega\,; \qquad (\omega', A') \;\mapsto\; \omega'\,,$$

be the projection to the first coordinate. Then

$$\pi_\Omega \circ c \circ \imath(A) \;=\; \pi_\Omega\bigl((\omega, A)\bigr) \;=\; \omega$$

where ω satisfies $\omega \in A$ according to (5.49). We have shown that the function

$$c^* \;:=\; \pi_\Omega \circ c \circ \imath : \mathscr{A} \longrightarrow \Omega \qquad A \;\mapsto\; \omega := \pi_\Omega \circ c \circ \imath : \mathscr{A}(A)$$

maps any $A \in \mathscr{A}$ to an element $\omega \in A$. We conclude that c^* is a choice function on $\mathscr{A}$. ■

We state the content of Proposition 5.12 and Proposition 5.13 as follows.

Theorem 5.3. *The following are equivalent.*

> ***(a)*** *For any sets $X, Y \neq \emptyset$ and surjective $g : Y \to X$ there exists a right inverse for g, i.e., a function $f : X \to Y$ such that $g \circ f = id_X$.*
>
> ***(b)*** *The Axiom of Choice holds: For any collection $\mathscr{A}$ of nonempty sets there exists a choice function on $\mathscr{A}$, i.e., a function $c : \mathscr{A} \to \bigcup[A : A \in \mathscr{A}]$ such that $c(A) \in A$ for all $A \in \mathscr{A}$.*

PROOF: See Proposition 5.12 and Proposition 5.13. ■

5.4 Exercises for Ch.5

5.4.1 Exercises for Functions and Relations

Exercise 5.1. Prove that $A \times B = \emptyset \;\Leftrightarrow\; A = \emptyset$ or $B = \emptyset$ or both are empty. □

Exercise 5.2.

(a) Which of the following is an equivalence relation? a partial ordering? on $\mathbb{R}$?
a1. $xRy \Leftrightarrow x < y$, **a2.** $xRy \Leftrightarrow x \leq y$, **a3.** $xRy \Leftrightarrow x = y$, **a4.** $xRy \Leftrightarrow x \neq y$.

(b) Define $xRy \Leftrightarrow xy > 0$. Is this an equivalence relation on $\mathbb{R}$? on $\mathbb{R}_{\neq 0}$? on $\mathbb{R}_{>0}$? on $\mathbb{R}_{<0}$? □

Exercise 5.3. It was stated in example 5.8 on p.131 that $(\mathbb{R}, \geq)$ is a linearly ordered set. Prove it. (Prove first that this is a POset.) □

Exercise 5.4. Prove prop.5.1(c) on p.128 of this document: If "$\sim$" is an equivalence relation on a nonempty set X and $x, y \in X$ then either $[x] = [y]$ or $[x] \cap [y] = \emptyset$. □

Exercise 5.5. 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 its domain is 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 5.6 (Excercise 5.5 continued)**.** Let $A \subseteq \mathbb{R}$.

Part 1.

- Let $F_1 : A \to [-2, 20[; \quad x \mapsto x^2$.
- Let $F_2 : A \to [2, 20[; \quad x \mapsto x^2$.

What choice of A makes

(a) F_1 surjective? **(c)** F_2 surjective?
(b) F_1 injective? **(d)** F_2 injective?

Part 2.

- Let $G_1 : A \to [-2, 20[; \quad x \mapsto \sqrt{x}$.
- Let $G_2 : A \to [2, 20[; \quad x \mapsto \sqrt{x}$.

What choice of A makes

(e) G_1 surjective?
(f) G_1 injective?
(g) G_2 surjective?
(h) G_2 injective?

Part 3.

- 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

(i) G_3 surjective?
(j) G_3 injective?
k. G_4 surjective?
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 5.7. 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 5.19 on p.139. □

Exercise 5.8.

(a) Prove prop.5.5**(a)**: The composition of two injective functions is injective.
(b) Prove prop.5.5**(b)**: The composition of two surjective functions is surjective. □

Exercise 5.9. You proved in the previous exercise that
injective $\circ$ injective $=$ injective,
surjective $\circ$ surjective $=$ surjective.

This exercise 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. Think simple and look at example 5.19 on p.139. □

Exercise 5.10. Prove prop.5.6 on p.148: Let X be an arbitrary set and let A be a nonempty proper subset of X. so that $X = A \biguplus A^{\complement}$ is a partitioning of X into two nonempty subsets A and $A^{\complement}$. Let $a \in A$, $a_0 \in A^{\complement}$ and $A' := (A \setminus \{a\}) \biguplus \{a_0\}$. Then the function $\varphi : A' \xrightarrow{\sim} A$; $\varphi(x) = a$ if $x = a_0$ and $\varphi(x) = x$ else is a bijection. □

Exercise 5.11. Prove prop.5.4 on p.147 of this document: Let $(R, \oplus, \odot, P)$ be an ordered integral domain

(A) Let $b \in R$. Then the function

$$T : R \to R; \quad x \mapsto x \oplus b,$$

is a bijection.

(B) Let $a \in R, a \neq 0$. Then the function

$$D : R \to a \odot R; \quad x \mapsto a \odot x,$$

is a bijection. (As usual, $a \odot R = aR = \{a \odot r : r \in R\}$.)

Hint: Find the inverses of T (obvious) and D (tricky: you cannot write $a^{-1}y$ since the inverse of a may not exist). □

Only the group structure of R was used in part **A** of the previous exercise:

Exercise 5.12. If $(G, \diamond)$ is a group and and $h \in G$ then the function

$$T : G \to G; \quad g \mapsto g \diamond h,$$

is a bijection. □

Exercise 5.13. Prove **(c)** of remark 5.16 on p.150: Let X and Y be two nonempty sets and $u \neq v$ arbitrary items. Then the sets $\{u\} \times X$ and $\{v\} \times Y$ are disjoint. □

Exercise 5.14. Prove **(b)** of remark 5.16 on p.150: Let X and Y be two nonempty sets and u, v arbitrary.
Then an injection/surjection/bijection $X \to Y$ exists if and only if an injection/surjection/bijection $\{u\} \times X \ \to \ \{v\} \times Y$ exists. □

Exercise 5.15. Let $f : \mathbb{R} \to \mathbb{R}$ be the function $x \mapsto 2x - 4$. Let the relation Γ_f be defined as the graph of f.

(a) Compute the inverse relation $(\Gamma_f)^{-1}$.

(b) Is $(\Gamma_f)^{-1}$ the graph of a function? If yes, what function? Don't forget to include domain and codomain. □

Exercise 5.16. B/G Project 6.9.:
On $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ we define the relation $\sim$ as follows.

$$(m_1, n_1) \ \sim \ (m_2, n_2) \ \Leftrightarrow m_1 \cdot n_2 = n_1 \cdot m_2. \tag{5.50}$$

(a) Prove that $\sim$ defines an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.
Let

$$\mathfrak{Q} \ := \{[(m, n)] \ : \ m, n \in \mathbb{Z} \text{ and } n \neq 0\} \tag{5.51}$$

be the set of all equivalence classes of $\sim$. We define two binary operations $\oplus$ and $\otimes$ on $\mathfrak{Q}$ as follows;

$$[(m_1, n_1)] \oplus [(m_2, n_2)] \ := [(m_1 n_2 \ + \ m_2 n_1, n_1 n_2)], \tag{5.52}$$

$$[(m_1, n_1)] \otimes [(m_2, n_2)] \ := [(m_1 m_2, n_1 n_2)] \tag{5.53}$$

(b) Prove that these binary operations are defined consistently: the right–hand sides of (5.52) and (5.53) 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

$$[(m_1 n_2 \ + \ m_2 n_1, n_1 n_2)] \ = [(p_1 q_2 \ + \ p_2 q_1, q_1 q_2)], \tag{5.54}$$

$$[(m_1 m_2, n_1 n_2)] \ = [(p_1 p_2, q_1 q_2)]. \tag{5.55}$$

or, equivalently, then

$$(m_1 n_2 \ + \ m_2 n_1, n_1 n_2) \ \sim (p_1 q_2 \ + \ p_2 q_1, q_1 q_2), \tag{5.56}$$

$$(m_1 m_2, n_1 n_2) \ \sim (p_1 p_2, q_1 q_2). \ \square \tag{5.57}$$

Exercise 5.17. Prove prop.5.9**(a)** on p.152: Let A, X, Y be nonempty sets and $A \subseteq X$. Let $f : X \xrightarrow{\sim} Y$ be bijective. Let $B := \{f(a) : a \in A\}$. Let $f' : A \to B;\quad x \mapsto f(x)$. Then f' is bijective.

Hint: Study the proof of prop.5.9**(b)** Your proof is very similar. □

Exercise 5.18. What are the graphs $\Gamma_{f_\star}$ and $\Gamma_{f^\star}$ of the functions $f_\star$ and $f^\star$ of example 5.23 on p.141? Do not use the symbols $f_\star$ and $f^\star$ when you write the formulas!

Exercise 5.19. Prove prop.5.10 on p.155 of this document: If p_1 and p_2 are polynomials and if $\lambda \in \mathbb{R}$ then

- **(a)** The sum $x \mapsto p_1(x) + p_2(x)$ is a polynomial.
- **(b)** The "scalar product" $x \mapsto \lambda p_1(x)$ is a polynomial. □

Exercise 5.20. Prove (5.42) of rem.5.21 on p.157 of this document: If $(x_i)_{i \in I}$ is a family in X, $(y_j)_{j \in J}$ is a family in Y, and those two families are equal then

$$\{x_i : i \in I\} \ = \{y_j : j \in J\} \ \subseteq \ X \cap Y. \ \square$$

6 The Integers

Note to Math 330 students: This chapter contains a lot of, but by no means all of the material of chapters 2.3, 2.4, 4, 6 and 7 of [2] Beck/Geoghegan: Art of Proof. On the other hand this chapter also contains some generalizations which cannot be found in that book, and I have given alternate versions of some proofs which I found difficult to follow. An other reason to duplicate that material here is that doing so allows this author to give internal references which you can click on rather than having to go back and forth between two different sources.

Note to Math 330 students: You should read this chapter in parallel with chapters 2, 4, 6 and 7 of [2] Beck/Geoghegan Art of Proof

6.1 The Integers, the Induction Axiom, and the Induction Principles

In ch.2.3 (Numbers) on p.24 we informally defined the integers $\mathbb{Z}$ as those numbers n which can be expressed as finite strings of decimal digits, possibly preceded by a minus sign. This is problematic from a very unexpected perspective: We will need a precise definition of the integers as a prerequisite for a precise definition of finiteness.

We also defined the natural numbers just as informally as the set $\mathbb{N} = \{1, 2, 3, \dots\}$. We will now give precise, axiomatic, definitions of those sets by using as a starting point prop.3.32**(a)** on p.69, which asserts that $(\mathbb{Z}, +, \cdot, \mathbb{N})$ is an ordered integral domain. This "proposition" was stated at a point where the exact definition of $\mathbb{Z}$ and $\mathbb{N}$ was not provided yet. The next axiomatic definition will close that gap.

Since addition and multiplication are associative in integrals domains $(R, \oplus, \odot)$ we will heneforth write $a \oplus b \oplus c$ for either of $(a \oplus b) \oplus c$, $a \oplus (b \oplus c)$, and $a \odot b \odot c$ for either of $(a \odot b) \odot c$, $a \odot (b \odot c)$. Here we assumed that $a, b, c \in R$.
The case of more than three operands will be taken care of later by Theorem 6.6 (Generalized Law of Associativity) on p.181.

Axiom 6.1 (Integers and Natural Numbers)**.** We postulate the existence of two sets $\mathbb{Z}$ and $\mathbb{N}$ which satisfy the following:

(a) $\mathbb{Z}$ is endowed with two binary operations "$+$" (called addition) and "$\cdot$" (called multiplication) and with a positive cone $\mathbb{N}$ such that $(\mathbb{Z}, +, \cdot, \mathbb{N})$ is an ordered integral domain. We denote the additive unit of this integral domain by 0 and its multiplicative unit by 1.

(b) **Induction Axiom**: Let $A \subseteq \mathbb{Z}$ such that

- **(1)** $1 \in A$,
- **(2)** $k \in A$ implies $k + 1 \in A$.

Then $A \supseteq \mathbb{N}$.

We call $\mathbb{Z}$ the set of **integers**, and we call $\mathbb{N}$ the set of **natural numbers**. □

Definition 6.1 (Decimal Digits)**.** So far the only two integers we know are 0 (the neutral element for "$+$"), and 1 (the neutral element for "$\cdot$"). We define the following integers:

$$2 := 1+1,\ 3 := 2+1,\ 4 := 3+1,\ 5 := 4+1,\ 6 := 5+1,\ 7 := 6+1,\ 8 := 7+1,\ 9 := 8+1.$$

We call the elements of the set $\{0,1,2,3,4,5,6,7,8,9\}$ **digits** or **decimal digits**. [86] □

Remark 6.1. Let $m,n \in \mathbb{Z}$. We remind the reader that precise definitions were given at the beginning of ch.3.4 on p.68 about statements like, e.g., $m < n$ (it means that $n - m \in \mathbb{N}$), about positivity (n is positive if and only if $n \in \mathbb{N}$), and about negativity (n is negative if and only if $-n \in \mathbb{N}$).

The following simple proposition and its corollary will allow us to generalize the induction axiom to sets of the form $[k_0, \infty[_{\mathbb{Z}} = \{k \in \mathbb{Z} : j \geq k_0\}$ where k_0 is an arbitrary integer.

Proposition 6.1. *Let $i,j,n \in \mathbb{Z}$. Then $n + i \in [i, \infty[_{\mathbb{Z}} \Leftrightarrow\ n + j \in [j, \infty[_{\mathbb{Z}}$.*

The proof is left as exercise 6.1 (see p.209). ■

Corollary 6.1. *Let $k_0, n \in \mathbb{Z}$. Then $n \in [k_0, \infty[_{\mathbb{Z}}$ if and only if $n - k_0 + 1 \in \mathbb{N}$.*

PROOF: We apply prop.6.1 with $n - k_0$ instead of n, k_0 instead of i, and 1 instead of j:

$$(n - k_0) + k_0 \in [k_0, \infty[_{\mathbb{Z}} \Leftrightarrow\ (n - k_0) + 1 \in [1, \infty[_{\mathbb{Z}}, \quad \text{i.e.,} \quad n \in [k_0, \infty[_{\mathbb{Z}} \Leftrightarrow\ n - k_0 + 1 \in \mathbb{N}. \ \blacksquare$$

Theorem 6.1 (Generalization of the Induction Axiom)**.**

Let $k_0 \in \mathbb{Z}$ and let

$$A_{k_0} \ :=\ \{k \in \mathbb{Z} : j \geq k_0\} \ =\ [k_0, \infty[_{\mathbb{Z}}$$

be the set of all integers at least as big as k_0. Let $A \subseteq \mathbb{Z}$ such that

(a) $k_0 \in A$,

(b) $k \in A$ *implies* $k + 1 \in A$.

Then $A \supseteq A_{k_0}$.

Proof strategy: We will shift everything by the amount $-k_0 + 1$: Let

(A) $$B \ :=\ 1 - k_0 + A \ =\ \{\, a - k_0 + 1 \ :\ a \in A \,\}\,.$$

Our proof then proceeds as follows.

(1) Show that $1 \in B$.

(2) For an arbitrary $b \in B$ let $a := b + k_0 - 1$ be the corresponding item in A. By assumption, $a + 1 \in A$. Use this to show that $b + 1 \in B$.

(3) It follows from **(1)** and **(2)** that B satisfies both properties **(1)** and **(2)** of the induction axiom, thus $B \supseteq \mathbb{N}$.

(4) We complete the proof by adding $k_0 - 1$ to both B and $\mathbb{N}$ and obtaining $A \supseteq A_{k_0}$.

[86] Note that all we needed to define those decimal digits were the existence of $0, 1$, and the "$+$" operation, thus we could have defined the above for any integral domain, even for any commutative ring with unit.

PROOF of Theorem 6.1:
So let $B := 1 - k_0 + A$. Since $k_0 \in [k_0, \infty[_{\mathbb{Z}}$ and $[k_0, \infty[_{\mathbb{Z}} \subseteq A$, $k_0 \in A$. Thus $1 = k_0 - k_0 + 1 \in B$. This proves step **(1)**.

Let $b \in B$ and $a := b + k_0 - 1$. Then $a \in A$ by definition of B. From assumption **(b)** of this theorem we obtain $a + 1 \in A$ and thus $b + 1 = (a + 1) - k_0 + 1 \in B$ by definition of B. This proves step **(2)**.

We have proven for the set B that $1 \in B$ and that this set contains with each element b also the integer $b + 1$. It follows from the induction axiom that $B \supseteq \mathbb{N}$. This proves step **(3)**.

It follows from $B \supseteq \mathbb{N}$ that $(k_0 - 1) + B \supseteq (k_0 - 1) + \mathbb{N}$, i.e., $A \supseteq [k_0, \infty[_{\mathbb{Z}} = A_{k_0}$. ■

The generalized induction axiom allows us to give a proof for the **principle of mathematical induction** which was introduced in rem.2.19 (p.44) of ch.2.7.

Theorem 6.2 (Principle of Mathematical Induction)**.**

Assume that for each integer $k \geq k_0$ there is an associated statement $P(k)$ such that

A. Base case. *The statement $P(k_0)$ is true.*

B. Induction Step. *For each $k \geq k_0$ we have the following: Assuming that $P(k)$ is true ("**Induction Assumption**"), it can be shown that $P(k+1)$ also is true.*

*It then follows that $P(k)$ is true for **each** $k \geq k_0$.*

PROOF: Let $A_{k_0} := \{k \in \mathbb{Z} : j \geq k_0\}$, and let $A := \{k \in A_{k_0} : P(k)$ is true $\}$. It follows from **A** that $k_0 \in A$, and it follows from **B** that if $k \in A$ then $k + 1 \in A$. We conclude from thm.6.1 above that $A_{k_0} \subseteq A$. Thus $P(k)$ is true for all $k \in A_{k_0}$. ■

Remark 6.2. The above theorem 6.2 is often stated for the special case $k_0 = 1$. [87]

We remind the reader that several examples for proofs by induction were given in ch.2.7.

Theorem 6.3 (Principle of Strong Mathematical Induction)**.**

Let $k_0 \in \mathbb{Z}$ and assume that for each integer $k \geq k_0$ there is an associated statement $P(k)$ such that the following is valid:

A. Base case. *The statement $P(k_0)$ is true.*

B. Induction Step. *For each $k \geq k_0$ we have the following: Assuming that $P(j)$ is true for <u>all</u> $j \in \mathbb{Z}$ such that $k_0 \leq j \leq k$ ("**Induction Assumption**"), it can be shown that $P(k+1)$ also is true.*

*It then follows that $P(k)$ is true for **each** $k \geq k_0$.*

PROOF: Let

$$A_{k_0} := \{k \in \mathbb{Z} : j \geq k_0\}; \quad A := \{k \in A_{k_0} : P(j) \text{ is true for all } j \in [k_0, k]_{\mathbb{Z}}\}.$$

It follows from **A** that $k_0 \in A$, and it follows from **B** that if $k \in A$, i.e., $P(j)$ is true for all $k_0 \leq j \leq k$, then also $P(k+1)$ is true, hence $P(j)$ is true for all $k_0 \leq j \leq k+1$, i.e., $k + 1 \in A$. We conclude from thm.6.1 on p.168 that $A_{k_0} \subseteq A$. Thus $P(k)$ is true for all $k \in A_{k_0}$. ■

The following is an example for a proof that is best done with strong induction.

[87] [2] Beck/Geoghegan refers to the case $k_0 = 1$ as the Principle of mathematical induction — first form and to the general case as the Principle of mathematical induction — first form revisited.

Example 6.1. [88] Let $(x_n)_{n\in\mathbb{N}}$ be the sequence $\ x_1 := 2,\ x_2 := 8,\ x_n := 4(x_{n-1} - x_{n-2})\ (n \geq 3)$. Prove that $x_n = n2^n$ for all $n \in \mathbb{N}$.

Solution strategy: This is a two-step recursion: To know the value of the sequence at "time" k we must know x_n for both $n = k-1$ and $n = k-2$. We need strong induction rather than ordinary induction on n, and we must "anchor" the proof with two base cases: $n = 2$ and $n = 1$ so that we can bootstrap ourselves and conclude the validity of $x_n = n2^n$ for $n = 3$ from that of the base cases and the recursion formula $x_n := 4(x_{n-1} - x_{n-2})$.

SOLUTION (by strong induction on n):

Base cases:

$n = 1 \ \Rightarrow\ n2^n = 1\cdot 2^1 = 2 = x_1$,

$n = 2 \ \Rightarrow\ n2^n = 2\cdot 2^2 = 8 = x_2$.

Induction step:

Induction assumption: We have some $n \in \mathbb{N}$ such that $x_j = j2^j$ for all $j \leq n$. $\quad(\star)$

We **must show under this assumption** that $x_{n+1} = (n+1)2^{n+1}$. $\quad(\star\star)$

$$\begin{aligned} x_{n+1} &= 4(x_n - x_{n-1}) && \text{(recursion formula for } x_n\text{)}\\ &= 4\big(n2^n \ - \ (n-1)2^{n-1}\big) && \text{(induction assumption } (\star)\text{)}\\ &= 2n2^{n+1} \ - \ (n-1)2^{n+1}\big)\\ &= (2n - n + 1)2^{n+1}\\ &= (n+1)2^{n+1}. \end{aligned}$$

We have shown the validity of $(\star\star)$, and this completes the proof by strong induction. □

Here is another example for a proof by strong induction.

Example 6.2. Example [89] Let $(x_n)_{n\in\mathbb{N}}$ be the following sequence of real numbers: $x_1 := x_2 := 1,\ x_n := \frac{1}{2}(x_{n-1} + 2/x_{n-2})\ (n \geq 3)$. Prove that $1 \leq x_n \leq 2$ for all $n \in \mathbb{N}$.

SOLUTION (by strong induction on n):

Base cases:

We need both $n = 1$ and $n = 2$ as base cases for the same reason as in example 6.1. Their validity is obvious from $x_1 = x_2 = 1$, since then $1 \leq x_1 \leq 2$ and $1 \leq x_2 \leq 2$.

Induction step:

Induction assumption: We have some $n \in \mathbb{N}$ such that $1 \leq x_j \leq 2$ for all $j \leq n$. $\quad(\star)$

We **must show under this assumption** that $1 \leq x_{n+1} \leq 2$. $\quad(\star\star)$ for all $(n \geq 3)$.

It follows from $(\star)$ that

- **(i)** $\quad x_n \leq 2$, hence $\frac{1}{2}x_n \leq \frac{1}{2}\cdot 2 = 1$,
- **(ii)** $\quad x_{n-1} \geq 1$, hence $\frac{1}{2}\cdot\frac{2}{x_{n-1}} = \frac{1}{x_{n-1}} \leq 1$.

It follows from the recursion formula $x_n = \frac{1}{2}(x_{n-1} + 2/x_{n-2})$ that $x_n \leq 1 + 1 = 2$ for $(n \geq 3)$.

It also follows from $(\star)$ that

[88] This is example 3.40 of D'Angelo and West [8].

[89] This is exercise 3.57 of D'Angelo and West [8].

(iii) $x_n \geq 1$, hence $\frac{1}{2}x_n \geq \frac{1}{2}$,

(iv) $x_{n-1} \leq 2$, hence $\frac{1}{2} \cdot \frac{2}{x_{n-1}} = \frac{1}{x_{n-1}} \geq \frac{1}{2}$.

It follows from $x_n := \frac{1}{2}(x_{n-1} + 2/x_{n-2})$ that $x_n \geq \frac{1}{2} + \frac{1}{2} = 1$ for $(n \geq 3)$.

We have shown the validity of $(\star\star)$, and this completes the proof by strong induction. □

Remark 6.3. How do thm.6.2 and thm.6.3 compare? The base case "P_{k_0} is true" is the same for both, and so is the conclusion "P_{k+1} is true". The difference is in the induction assumptions. Strong induction allows you to assume a lot more (the validity of P_j for **a]]** $j \leq k$) than ordinary induction where you only may assume the validity of P_k. □

6.2 The Discrete Structure of the Integers

In the previous section the induction principles were derived from the induction axiom and they are very powerful tools for the design of a proof. Accordingly, you encounter such proofs quite frequently. The induction axiom itself is used rather seldom in comparison, but the next theorem is a nice example where this is the case.

Theorem 6.4 (B/G Prop.2.20)**.** *If $k \in \mathbb{N}$ then*

$$k \geq 1. \tag{6.1}$$

PROOF:

Let $A := \{n \in \mathbb{N} : n \geq 1$. It suffices to show that $A \supseteq \mathbb{N}$.

(a) Clearly, $1 \geq 1$. Hence, $1 \in A$.

(b) Next, let $a \in A$. Then $a \geq 1$. Hence, $a + 1 \geq 1 + 1 > 1$ (use Proposition 3.36 on p.72).

(c) Thus, $a + 1 \geq 1$ and thus, $a + 1 \in A$.

We have shown in **(a)** and **(c)** that $1 \in A$ and also, that $\in A \Rightarrow a + 1 \in A$. It follows from the induction axiom that $A \supseteq \mathbb{N}$ ■

Proposition 6.2 (B/G Prop.2.21)**.** *There exists no $x \in \mathbb{Z}$ such that $0 < x < 1$.*

The proof is left as exercise 6.12 (see p.210). ■

Corollary 6.2 (B/G Cor.2.22)**.** *Let $n \in \mathbb{Z}$. There exists no $x \in \mathbb{Z}$ such that $n < x < n + 1$.*

The proof is left as exercise 6.13 (see p.210). ■

Proposition 6.3 (sharpening of B/G Prop.2.13)**.** $\mathbb{N} = \{k \in \mathbb{Z} : k \geq 1\}$.

PROOF: This follows from $1 > 0$ (cor.3.3 on p.72) and prop.6.2 above. ■

It follows from the last proposition that $\mathbb{N} = [1, \infty[$ and thus $\min(\mathbb{N}) = 1$. We will see in subchapter 6.8 (The Well–Ordering Principle) on p.193 that this is a special case of the following: All nonempty subsets of $\mathbb{N}$ possess a minimum.

6.3 Divisibility

For any two real numbers a and b such that $b \neq 0$ one can construct the quotient $\frac{a}{b}$, and the result is again a real number. The same situation exists for fractions. In contrast there are integers $m, n \neq 0$ for which the quotient $\frac{m}{n}$ is not an integer, e.g., if $m = 12$ and $n = 5$. One says in this case that m is not divisible by n. Questions of divisibility are of great interest in the mathematical discipline of number theory, and we will examine divisibility at various times.

Definition 6.2 (Divisibility)**.**

(a) Let $m, n \in \mathbb{Z}$. We say that n **divides** m or, equivalently, that m **is divisible by** n if there exists $j \in \mathbb{Z}$ such that $m = jn$. We then write $n \mid m$, and we write $n \nmid m$ if n does not divide m, i.e., there is no $k \in \mathbb{Z}$ that satisfies $m = kn$.

(b) Let $m \in \mathbb{Z}$. We say that m is an **even** integer if $2 \mid m$. We say that m is an **odd** integer if m is not even, i.e., $2 \nmid m$. □

Proposition 6.4. *Let $m, n \in \mathbb{Z}$ such that $m \neq 0$ and $m \mid n$, i.e., there exists $j \in \mathbb{Z}$ be such that $n = j \cdot m$. Then j is uniquely determined.*

PROOF: Assume that there is $j' \in \mathbb{Z}$ such that also $n = j' \cdot m$. Then

$$n \;=\; j' \cdot m \;=\; j \cdot m, \quad \text{hence } (j' - j)m \;=\; 0.$$

The integral domain $\mathbb{Z}$ has no zero divisors, hence $a \cdot b = 0$ implies $a = 0$ or $b = 0$.
It follows from $m \neq 0$ that $j' - j = 0$ and hence $j' = j$. ■

The above result implies that the integer j is unique determined by m and n and hence allows us to make the following definition.

Definition 6.3 (Quotients)**.** Let $d, n \in \mathbb{Z}$ such that $d \mid n$ and $d, n \neq 0$.

Let $q \in \mathbb{Z}$ be the unique integer for which $n = q \cdot d$. We write either of

$$\frac{n}{d}, \quad n/d, \quad n \div d \quad \text{instead of } q\,,$$

and we call n the **dividend** or **numerator**, d the **divisor** or **denominator**, and q the **quotient** of the expression n/d. We also define $\frac{0}{d} := 0$ if $d \neq 0$, but we leave $\frac{n}{0}$ undefined for any integer n. □

Note 6.1. Note that the assignment $(d, n) \mapsto n/d$ is **not a "binary operation"** on $\mathbb{Z}$ as is the case for $(m, n) \mapsto m + n$ and $(m, n) \mapsto m \cdot n$. The reason: $m + n$ and $m \cdot n$ are defined for **any** two $m, n \in \mathbb{Z}$ whereas n/d is only defined for those $d, n \in \mathbb{Z}$ which satisfy the condition $d \mid n$.

Also note that the order of n and d is reversed in these two expressions:
We write $d \mid n$ to indicate that the quotient n/d exists! □

Proposition 6.5 (B/G prop.1.16)**.** *If m and n are even integers, then so are $m+n$ and mn.*

PROOF: Left as an exercise. ■

Proposition 6.6 (B/G prop.1.17)**.**

(a) *If m is an integer then $m \mid 0$. In particular, $0 \mid 0$.*
(b) *If m is a nonzero integer then $0 \nmid m$.*

PROOF: Left as an exercise. ■

Proposition 6.7 (B/G prop.2.18)**.** *Let $n \in \mathbb{N}$. Then*

(a) *n^3+2n is divisible by 3,*
(b) *$n^4-6n^3+11n^2-6n$ is divisible by 4,*
(c) *n^2+n is even, i.e., divisible by 2,*
(d) *n^3+5n is divisible by 6.*

PROOF of **(a)**: See [2] B/G (Beck/Geoghegan), prop.2.18(i).
PROOF of **(b)**, **(c)**, **(d)**: Left as an excercise. Note that **(c)** will be helpful for the proof of **(d)**. ■

The following example shows how to structure a proof by induction of divisibility. Note that it makes use of Proposition 6.7**(c)**.

Example 6.3 (Divisibility)**.** Prove by induction that $\quad 6 \mid (5n^3 + 7n) \quad$ for $n \in [0,\infty[_{\mathbb{Z}}$.
PROOF:
We need to find $j \in \mathbb{Z}$ such that $5n^3+7n = 6j$.
Base case: $n=0$.
Then $5n^3+7n = 0+0 = 0 = 0 \cdot 1$. The base case holds since we may choose $j=1$.
Induction assumption: Assume that $n \in [0,\infty[_{\mathbb{Z}}$ is such that there exists $j \in \mathbb{Z}$ such that

(IA) $$5n^3 + 7n = 6j\,.$$

We need to show that

(NTS) there exists $j' \in \mathbb{Z}$ such that $5(n+1)^3 + 7(n+1) = 6j'$.

We transform the left side of that equation as follows:

$$\begin{aligned} 5(n+1)^3 + 7(n+1) &= (5n^3 + 15n^2 + 15n + 5) + 7n + 7 \\ &= (5n^3 + 7n) + (15n^2 + 15n + 5) + 7 \\ &= (5n^3 + 7n) + 15(n^2 + n) + 12\,. \end{aligned}$$

According to Proposition 6.7**(c)**, $n^2 + n$ is even and thus equals $2j''$ for a suitable integer j''. We further apply **(IA)** and obtain

$$5(n+1)^3 \,+\, 7(n+1) \,=(6j) \,+\, 15(2j'') \,+\, 12 \,=\, 6(j \,+\, 5j'' \,+\, 2)\,.$$

Let $j' := j + 5j'' + 2$. Then **(NTS)** is satisfied and the proof by induction is finished ■

Proposition 6.8 (B/G Prop.2.24)**.** *Let $n \in \mathbb{N}$. Then $n^2 + 1 > n$.*

PROOF: The proof is left as exercise 6.14 (see p.210). ■

Proposition 6.9 (B/G prop.2.23)**.** *Let $m, n \in \mathbb{N}$. If $m \mid n$ then $m \leq n$*

The proof is left as exercise 6.11 (see p.210). ■

6.4 Embedding the Integers Into an Ordered Integral Domain

The presentation of this material follows ch.9 of [2] B/G (Beck/Geoghegan).

Introduction 6.1. Allowing integers to be viewed as certain elements of an ordered integral domain $R = (R, \oplus, \odot, P)$ makes it possible to look at products na of integers n and $a \in R$.
In particular we can multiply binomial coefficients $\binom{n}{k}$ with elements of R and thus formulate and prove the binomial theorem

$$(a \oplus b)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

for elements a and b of such an arbitrary ordered integral domain.
We will "embed" the ordered integral domain $\mathbb{Z} = (\mathbb{Z}, +, \cdot, \mathbb{N})$ into the ordered integral domain $R = (R, \oplus, \odot, P)$ by means of a function $e : \mathbb{Z} \to R$ which respects their algebraic operations (addition and multiplication) and also the order relation induced by their positive cones in a sense which is specified in theorem 6.5 further down (p.178). □

We assume in the following that $R = (R, \oplus, \odot, P)$ is a fixed ordered integral domain. We will distinguish the additive units of the domain $\mathbb{Z}$ and the codomain R by tagging them as $0_{\mathbb{Z}}$ and 0_R, and we will distinguish their multiplicatitive units by tagging them as $1_{\mathbb{Z}}$ and 1_R. We also will distinguish order relations $x < y, x \leq y, ...$ by writing $m < n, m \leq n, \ldots$ if we deal with elements $m, n \in \mathbb{Z}$, and we will write $a \prec b, a \preceq b, \ldots$ for $a, b \in R$.

We will see examples for the above notational conventions in the following definition.

Definition 6.4 (Natural Embedding of the Integers Into $(R, \oplus, \odot, P)$)**.** $\boxed{\star}$
We define a function $e : \mathbb{Z} \to R$, partially by recursion, as follows.

$$e(0_{\mathbb{Z}}) := 0_R\,, \tag{6.2}$$
$$e(n+1_{\mathbb{Z}}) := e(n) \oplus 1_R \quad \text{for } n \in [0,\infty[_{\mathbb{Z}}\,, \tag{6.3}$$
$$e(n) := \ominus\, e(-n) \quad \text{for } n \in\,]-\infty,-1\,]_{\mathbb{Z}}\,. \tag{6.4}$$

We call e the **natural embedding of $\mathbb{Z}$ into** $(R,\oplus,\odot,P)$. □

We establish some properties of the natural embedding.

Lemma 6.1.

$$e(1_{\mathbb{Z}}) = 1_R, \tag{6.5a}$$
$$e(-k) = \ominus\, e(k) \quad \textit{for any } k \in \mathbb{Z}. \tag{6.5b}$$

PROOF of (6.5a):

$$e(1_{\mathbb{Z}}) = e(0_{\mathbb{Z}}+1_{\mathbb{Z}}) \overset{(6.3)}{=} e(0_R) \oplus 1_R \overset{(6.2)}{=} 1_R.$$

PROOF (6.5b): (6.5b) is obviously true if $k=0$. If $k<0$ then this equation follows from (6.4), and if $k>0$ then we also obtain it from (6.4) since then $-k<0$, thus

$$e(-k) = \ominus\, e\big(-(-k)\big) = \ominus\, e(k). \blacksquare$$

Note that (6.5a) expresses that the image of the multiplicatve unit in $\mathbb{Z}$ is the multiplicative unit in R, and (6.5b) expresses that the image of the additive inverse is the additive inverse of the image.

We now show that the natural embedding e is compatible with the algebraic operations "+" of $\mathbb{Z}$ and "$\oplus$" of R in the sense that the image of the sum is the sum of the images.

Proposition 6.10. *Let $m,n \in \mathbb{Z}$. Then* $e(m+n) = e(m) \oplus e(n)$.

Proof strategy: We will do a proof by cases:
Case 1: $m=0_{\mathbb{Z}}$ or $n=0_{\mathbb{Z}}$,
Case 2: $m>0_{\mathbb{Z}}$ and $n>0_{\mathbb{Z}}$,
Case 3: $m<0_{\mathbb{Z}}$ and $n<0_{\mathbb{Z}}$.
The remaining case that either $m>0_{\mathbb{Z}}, n<0_{\mathbb{Z}}$ or $n>0_{\mathbb{Z}}, m<0_{\mathbb{Z}}$ only needs to be shown for one of those two possibilities, say, $n>0_{\mathbb{Z}}$ and $m<0_{\mathbb{Z}}$. We subdivide this case into two separate cases as follows:
Case 4: $n>0_{\mathbb{Z}}$ and $m<0_{\mathbb{Z}}$ and $n \geq -m$,
Case 5: $n>0_{\mathbb{Z}}$ and $m<0_{\mathbb{Z}}$ and $n<-m$.
Only the second case needs a proof by induction.
PROOF of **case 1**: $m=0_{\mathbb{Z}}$ or $n=0_{\mathbb{Z}}$: This is trivial since if, say, $n=0_{\mathbb{Z}}$ then

$$e(m+n) = e(m) \overset{(6.2)}{=} e(m) \oplus e(0_R) = e(m) \oplus e(n).$$

PROOF of **case 2**: $m>0_{\mathbb{Z}}$ and $n>0_{\mathbb{Z}}$:

We consider $m > 0$ as fixed but arbitrary and do the proof by induction on n.
Base case: $n = 1_{\mathbb{Z}}$. The assertion $e(m + 1_{\mathbb{Z}}) \;=\; e(m) \oplus 1_R$ is just (6.3).
Induction assumption (IA): Assume that $e(m + n) \;=\; e(m) \oplus e(n)$ for some $n \in \mathbb{N}$.
We must show that $e\big(m + (n + 1_{\mathbb{Z}})\big) \;=\; e(m) \oplus e(n + 1_{\mathbb{Z}})$. This follows from

$$\begin{aligned} e\big(m + (n + 1_{\mathbb{Z}})\big) \;&=\; e\big((m + n) + 1_{\mathbb{Z}}\big) \overset{(6.3)}{=} e(m + n) \oplus 1_R \\ &\overset{\text{(IA)}}{=} e(m) \oplus \big(e(n) \oplus 1_R\big) \overset{(6.3)}{=} e(m) \oplus e(n + 1_{\mathbb{Z}}). \end{aligned}$$

PROOF of **case 3**: $m < 0_{\mathbb{Z}}$ and $n < 0_{\mathbb{Z}}$:
Since $-m > 0$ and $-n > 0$ we may apply what we already proved in case 2 above:

$$\begin{aligned} e(m + n) &\overset{(6.4)}{=} \ominus\, e\big(-(m + n)\big) \;=\; \ominus\, e\big((-m) + (-n)\big) \\ &= \ominus\, \big(e(-m) \oplus e(-n)\big) \;=\; \big(\ominus\, e(-m)\big) \oplus \big(\ominus\, e(-n)\big) \overset{(6.4)}{=} e(m) \oplus e(n). \end{aligned}$$

PROOF of **case 4**: $n > 0_{\mathbb{Z}}$ and $m < 0_{\mathbb{Z}}$ and $n \geq -m$.
It follows from the assumptions of this case that $n = (n + m) + (-m)$ is the sum of the natural numbers $n + m$ and $-m$. We apply what we proved in case 2 and obtain

$$e(n) \;=\; e(n + m) \oplus e(-m) \overset{(6.5b)}{=} e(n + m) \ominus e(m),$$

thus $e(m + n) = e(m) \oplus e(n)$.
PROOF of **case 5**: $n > 0_{\mathbb{Z}}$ and $m < 0_{\mathbb{Z}}$ and $n < -m$:
It follows from the assumptions of this case that $-m = -(m + n) + n$ is the sum of the natural numbers $-(n + m)$ and n. We apply what we proved in case 2 and obtain with repeated use of (6.5b) that

$$\ominus\, e(m) \;=\; e(-m) \;=\; e\big(-(m + n) + n\big) \;=\; e\big(-(m + n)\big) \oplus e(n) \;=\; \ominus\, e(m + n) \oplus e(n),$$

thus $e(m + n) = e(m) \oplus e(n)$. ■

We will prove next that the natural embedding e also is compatible with the algebraic operations "$\cdot$" of $\mathbb{Z}$ and "$\odot$" of R in the sense that the image of the product is the product of the images.

Proposition 6.11. *Let $m, n \in \mathbb{Z}$. Then* $e(m \cdot n) \;=\; e(m) \odot e(n)$.

Proof strategy: We do a proof by cases just as we did for prop.6.10, but we only need four cases:
Case 1: $m = 0_{\mathbb{Z}}$ or $n = 0_{\mathbb{Z}}$.
Case 2: $m > 0_{\mathbb{Z}}$ and $n > 0_{\mathbb{Z}}$.
Case 3: $m < 0_{\mathbb{Z}}$ and $n < 0_{\mathbb{Z}}$.
The remaining case that either $m > 0_{\mathbb{Z}}, n < 0_{\mathbb{Z}}$ or $n > 0_{\mathbb{Z}}, m < 0_{\mathbb{Z}}$ only needs to be shown for one of those two, say, $n > 0_{\mathbb{Z}}$ and $m < 0_{\mathbb{Z}}$ since multiplication is commutative and we obtain the proof for the other case by switching the roles of m and n. Thus we are left with
Case 4: $m < 0_{\mathbb{Z}}$ and $n > 0_{\mathbb{Z}}$.
Only the second case needs a proof by induction.

PROOF of **case 1**: $m = 0_{\mathbb{Z}}$ or $n = 0_{\mathbb{Z}}$: This is trivial since if, say, $m = 0_{\mathbb{Z}}$ then

$$e(m \cdot n) = e(0_{\mathbb{Z}}) \overset{(6.2)}{=} 0_R = e(m) \odot 0_R \overset{(6.2)}{=} e(m) \odot e(0_{\mathbb{Z}}).$$

PROOF of **case 2**: $m > 0_{\mathbb{Z}}$ and $n > 0_{\mathbb{Z}}$:
We consider $m > 0$ as fixed but arbitrary and do the proof by induction on n.
Base case: $n = 1_{\mathbb{Z}}$. The assertion $e(m \cdot 1_{\mathbb{Z}}) = e(m) \odot e(1_{\mathbb{Z}})$ follows from (6.5b).
Induction assumption (IA): Assume that $e(m \cdot n) = e(m) \odot e(n)$ for some $n \in \mathbb{N}$.
We must show that $e\big(m \cdot (n + 1_{\mathbb{Z}})\big) = e(m) \odot e(n + 1_{\mathbb{Z}})$. This follows from

$$\begin{aligned} e\big(m(n + 1_{\mathbb{Z}})\big) &= e(mn + m) \overset{\text{prop.6.10}}{=} e(mn) \oplus e(m) \\ &\overset{\text{(IA)}}{=} e(m) \odot e(n) \oplus e(m) = e(m)(e(n) \oplus 1_R) \overset{(6.3)}{=} e(m) \odot e(n + 1_{\mathbb{Z}}). \end{aligned}$$

PROOF of **case 3**: $m < 0_{\mathbb{Z}}$ and $n < 0_{\mathbb{Z}}$:
Since $-m > 0$ and $-n > 0$ we may apply what we already proved in case 2 above:

$$e(m \cdot n) = e\big((-m)(-n)\big) = e(-m) \odot e(-n) \overset{(6.5b)}{=} \big(\ominus e(m)\big) \cdot \big(\ominus e(n)\big) = e(m) \odot e(n).$$

PROOF of **case 4**: $n > 0_{\mathbb{Z}}$ and $m < 0_{\mathbb{Z}}$: We again apply what we proved in case 2 to the natural numbers $-m$ and n and obtain with the use of (6.5b) that

$$\ominus e(mn) = e(-mn) = e\big((-m)n\big) = e(-m) \odot e(n) = \ominus e(m) \odot e(n),$$

thus $e(mn) = e(m) \odot e(n)$. ■

We now turn to the relationship which the natural embedding e establishes between the order relation $m < n$ on $\mathbb{Z}$ (induced by its positive cone $\mathbb{N}$) on the one hand, and the order relation $a < b$ on R (induced by its positive cone P) on the other hand.

Proposition 6.12 (B/G Prop.9.15)**.** *Let $n \in \mathbb{N}$. Then $e(n) \in P$, i.e., $e(n)$ is positive.*

The proof is left as exercise 6.2 (see p.209). ■

The natural embedding e is order preserving both ways in the sense specified in the next proposition.

Proposition 6.13 (B/G Prop.9.19)**.** *Let $m, n \in \mathbb{Z}$. Then*

$$m < n \Leftrightarrow e(m) \prec e(n)\,, \tag{6.6}$$
$$m \leq n \Leftrightarrow e(m) \preceq e(n)\,. \tag{6.7}$$

PROOF of ⇒ of (6.7):
Let us assume that $e(m) \preceq e(n)$. We must show that $m \leq n$.
Case 1 – $e(m) \prec e(n)$: then $m < n$ according to the already proven part (6.7), thus $m \leq n$.
Case 2 – $e(m) = e(n)$: We prove that $m = n$, hence $m \leq n$ by showing that both $m < n$ and $n < m$ lead to a contradiction.

If $m < n$ then $n - m \in \mathbb{N}$, thus $e(n) \ominus e(m) = e(n-m) \in P$ according to Proposition 6.12, i.e., $e(n) \ominus e(m) \in P$ since $e(n) \ominus e(m) = e(n-m)$ according to Proposition 6.10 on p.175.

It follows from $0_R \notin P$ that $e(m) \ominus e(n) \neq 0_R$, i.e., $e(m) \neq e(n)$. This contradicts the assumption $e(m) = e(n)$ we made at the beginning of this case 2.

We have shown that it is not possible that $m < n$, i.e., that $m \geq n$. The proof that $n \geq n$ is obtained by switching the roles of m and n in the above reasoning.

We have completed the proof of **case 2** of $\Rightarrow$ of (6.7) and thus the proof of the entire proposition. ■

Corollary 6.3. *The natural embedding $e : \mathbb{Z} \to R$ is injective.*

PROOF:

The proof was already done as part of **case 2** of $\Rightarrow$ of **(B)** of Proposition 6.13 where we saw that equality $e(m) = e(n)$ of function values implies that of the arguments m and n. This is the very definition of injectivity. We give here a streamlined proof that builds on Proposition 6.13.

Let $m, n \in \mathbb{Z}$ such that $m \neq n$. Then either $m < n$ or $m > n$. In the first case it follows from prop.6.13 that $e(m) \prec e(n)$ and thus $e(m) \neq e(n)$, in the second case it follows from **(A)** of prop.6.13 that $e(m) \succ e(n)$ and thus $e(m) \neq e(n)$. ■

We summarize everything said in this subchapter in the following theorem.

Theorem 6.5. *Let $R = (R, \oplus, \odot, P)$ be an ordered integral domain.*

The natural embedding $e : (\mathbb{Z}, +, \cdot, \mathbb{N}) \longrightarrow (R, \oplus, \odot, P)$ which is defined as follows:

$$e(0_{\mathbb{Z}}) = 0_R, \quad e(n + 1_{\mathbb{Z}}) = e(n) \oplus 1_R \ \text{ if } n \in \mathbb{N}, \quad e(n) = \ominus e(-n) \ \text{ if } n < 0$$

is an injective function with the following structure preserving properties ($m, n \in \mathbb{Z}$):

(a) *e maps neutral element to neutral element: $e(0_{\mathbb{Z}}) = 0_R$ and $e(1_{\mathbb{Z}}) = 1_R$.*
(b) *Image of the sum = sum of the images: $e(m + n) = e(m) \oplus e(n)$.*
(c) *Image of the product = product of the images: $\Rightarrow$ $e(m \cdot n) = e(m) \odot e(n)$.*
(d) *Image of the additive inverse = additive inverse of the image: $e(-m) = \ominus e(m)$.*
(e) *e preserves the order: $m < n \Leftrightarrow e(m) \prec e(n)$ and $m \leq n \Leftrightarrow e(m) \preceq e(n)$.*

PROOF: follows from the material presented prior to this theorem. ■

Remark 6.4. The function e does such nice job of embedding, i.e., injecting the integers into R that it is justified to "identify" the integers with their images in R. One thus does not distinguish between $n \in \mathbb{Z}$ and $e(n) \in R$. □

Functions which prefer algebraic structures play a very important role in abstract algebra, where they are called **homomorphisms**. The group homomorphisms we briefly discussed in Chapter 3 (The Axiomatic Method) are an example of such homomorphisms. Note that **(b)**, **(d)** and the formula $e(0_{\mathbb{Z}}) = 0_R$ in **(a)** of Theorem 6.5 imply that the natural embedding is a group homomorphism $(\mathbb{Z}, +) \to (R, \oplus)$.

Definition 6.5 (Ring homomorphism). $\boxed{\star}$

A function $h : (R, \oplus, \odot) \longrightarrow (R', \oplus', \odot')$ between two commutative rings with unit and in particular between two ordered integral domains [90] which satisfies Theorem6.5.**a–d** is called a **ring homomorphism**.
Note that ring homomorphisms play for commutative rings with unit the role which group homomorphisms play for groups. □

Theorem:

Let $R = (R, \oplus, \odot, P)$ be an ordered integral domain which satisfies the induction axiom. See Axiom 6.1 (Integers and Natural Numbers) on p.167.
Then the natural embedding $e : (\mathbb{Z}, +, \cdot, \mathbb{N}) \longrightarrow (R, \oplus, \odot, P)$ is an isomorphism of ordered integral domains, i.e., e is bijective and it's inverse, e^{-1}, also satisfies **(a)–(e)** of Theorem 6.5.

PROOF:

Let $A := P \cap e(\mathbb{Z})$

Step **(1)** We show that $A = P$.

Since $1_R = e(1)$ and $1_R > 0$, $1_R \in A$.

Next, let $p \in A$. Since $p \in e(\mathbb{Z})$, there is some $k \in \mathbb{Z}$ such that $p = e(k)$. Then

$$a \oplus 1_R \ = \ e(k) \oplus e(1) \ = \ e(k+1)\,.$$

Thus, $a \oplus 1_R \in e(\mathbb{Z})$. Further, $a \in P$ implies $a \oplus 1_R \in P$.

We have shown that A satisfies the induction axiom and it follows that $A \subseteq P$. By definition, $A \subseteq P$. It follows that $A = P$.

Step **(2)** We show that $-P \subseteq e(\mathbb{Z})$.

Let $x \in \ominus P$ and $a := \ominus x$. Then $a \in P = A$. Thus, there exists $k \in \mathbb{Z}$ such that $= e(k)$. Further, $e(-k) = \ominus a = b$. Thus, $\ominus P \subseteq e(\mathbb{Z})$.

Step **(3)** We show that $R \subseteq e(\mathbb{Z})$.

Since $0_R = e(0) \in e(\mathbb{Z})$, it follows from steps **(1)** and **(2)** that $\{0_R\} \uplus (P) \uplus (-P) \subseteq e(\mathbb{Z})$. As an ordered integral domain, R satisfies $R = \{0_R\} \uplus (P) \uplus (-P)$.

We have shown that e is surjective, hence, bijective.

Step **(3)** We show that e^{-1}, also satisfies **(a)–(e)** of Theorem 6.5.

(a) We apply e^{-1} to $e(0_{\mathbb{Z}}) = 0_R$ and $e(1_{\mathbb{Z}}) = 1_R$ and obtain $0_{\mathbb{Z}} = e^{-1}(0_R)$ and $1_{\mathbb{Z}} = e^{-1}(1_R)$.

(b) follows from Theorem 3.5 on p.60.

(c) is obtained by applying the proof of Theorem 3.5 to "$\odot$" instead of "$\oplus$".

(d) We show that $e^{-1}(\ominus a) = -e^{-1}(a)$ as follows.

Let $a \in R$ and $n := e^{-1}(a)$, i.e., $a = e(n)$. Then

$$e^{-1}(\ominus a) \ = \ e^{-1}\big(\ominus e(n)\big) \ \overset{(6.5b)}{=} \ e^{-1}\big(e(-n)\big) \ = \ -n \ = \ -e^{-1}(a)\,.$$

(e) That e^{-1} preserves both "$\prec$" and "$\preceq$" follows from Theorem6.5.**e** by setting

$m := e^{-1}(a)$ and $n := e^{-1}(b)$ for arbitrary $a, b \in R$,

and noting that the "$\Leftrightarrow$" arrows allow us to go back and forth between $\mathbb{Z}$ and R. ■

6.5 Recursive Definitions of Sums, Products and Powers in Integral Domains

We start this chapter with the generalizations of some definitions and several of the propositions that you will find in ch.4 of [2] Beck/Geoghegan Art of Proof for the specific ordered integral domain $(\mathbb{Z},+,\cdot,\mathbb{N})$ and the specific "start index $j = 1$. Except for those generalizations almost all of the material in this subchapter has been copied almost literally from that book.

Assume in this entire chapter that $R = (R, \oplus, \odot, P)$ is an ordered integral domain

The following definition is a generalization of "Σ" in B/G p.34, 35.

Definition 6.6. Let $k \in \mathbb{Z}$ and let $(x_j)_{j=k}^{\infty} \in R$.

For each $n \in \mathbb{Z}$ such that $k \leq n$, we define an element of R called $\displaystyle\sum_{j=k}^{n} x_j$ as follows. [91]

$$\text{(6.8)} \qquad \text{(i)} \quad \sum_{j=k}^{k} x_j \;=\; x_k, \qquad \text{(ii)} \quad \sum_{j=k}^{n+1} x_j \;=\; \sum_{j=k}^{n} x_j \;\oplus\; x_{n+1}.$$

We call $\displaystyle\sum_{j=k}^{n} x_j$ the **sum** of $x_k, x_{k+1}, \dots, x_{n-1}, x_n$. □

The following definition is a generalization of "$\prod$" (B/G p.34, 35).

Definition 6.7 (Definition of $\prod_{j=k}^{n} x_j$)**.** Let $k \in \mathbb{Z}$ and let $(x_j)_{j=k}^{\infty} \in R$.

For each $n \in \mathbb{Z}$ such that $k \leq n$, we define an element of R called $\displaystyle\prod_{j=k}^{n} x_j$ as follows. [92]

$$\text{(6.9)} \qquad \text{(i)} \quad \prod_{j=k}^{k} x_j \;=\; x_k, \qquad \text{(ii)} \quad \prod_{j=k}^{n+1} x_j \;=\; \prod_{j=k}^{n} x_j \;\odot\; x_{n+1}.$$

We call $\displaystyle\prod_{j=k}^{n} x_j$ the **product** of $x_k, x_{k+1}, \dots, x_{n-1}, x_n$. □

Note that in the following proposition we make use of the results of ch.6.4 (Embedding the Integers Into an Ordered Integral Domain). It was shown there that we can identify integers k as certain

[91] $\displaystyle\sum_{j=k}^{n} x_j$ can also be written $x_k \oplus x_{k+1} \oplus \cdots \oplus x_n$.

[92] $\displaystyle\prod_{j=k}^{n} x_j$ can also be written $x_k \odot x_{k+1} \cdots \odot x_n$ or $x_k x_{k+1} \dots x_n$.

elements $e(k)$ of R by means of the embedding function $e : \mathbb{Z} \to R$. For example the equation $\sum\limits_{j=k}^{n} x_j = n \ominus k \oplus 1$ in part **(b)** of Proposition 6.14 below is to be understood as $\sum\limits_{j=k}^{n} x_j = e(n-k+1)$, an equation between elements of R.

Proposition 6.14 (B/G prop.4.15)**.** *Let $m, n, k \in \mathbb{Z}$, $c \in R$, and let $(x_j)_{j=k}^{\infty}$ be a sequence in R. Then*

(a) $c \odot \left(\sum\limits_{j=k}^{n} x_j \right) = \sum\limits_{j=k}^{n} (c \odot x_j)$.

(b) *If $x_j = 1$ for all $j \in [k,n]_{\mathbb{Z}}$ then* $\sum\limits_{j=k}^{n} x_j = n \ominus k \oplus 1$.

(c) *If $x_j = c$ for all $j \in [k,n]_{\mathbb{Z}}$ then* $\sum\limits_{j=k}^{n} x_j = (n \ominus k \oplus 1) \odot c$.

PROOF: Left as an exercise. ■

Proposition 6.15 (B/G prop.4.16)**.**
Let $m, n, p \in \mathbb{Z}$ such that $m \leq n < p$, and let $(x_j)_{j=m}^{p}$ and $(y_j)_{j=m}^{p}$ be sequences in R. Then

(a) $\sum\limits_{j=m}^{p} x_j = \sum\limits_{j=m}^{n} x_j \oplus \sum\limits_{j=n+1}^{p} x_j$,

(b) $\sum\limits_{j=m}^{p} (x_j \oplus y_j) = \sum\limits_{j=m}^{p} x_j \oplus \sum\limits_{j=m}^{p} y_j$.

PROOF: Left as an exercise. ■

Proposition 6.16 (B/G prop.4.17)**.**

Let $m, n, p \in \mathbb{Z}$ such that $m \leq n$, and let $(x_j)_{j=m}^{n}$ be a sequence in R. Then $\sum\limits_{j=m}^{n} x_j = \sum\limits_{j=m+p}^{n+p} x_{j-p}$.

PROOF: Left as an exercise. ■

Proposition 6.17 (B/G prop.4.18)**.**
Let $m, n \in \mathbb{Z}$ such that $m \leq n$, and let $(x_j)_{j=m}^{n}$ and $(y_j)_{j=m}^{n}$ be sequences in R such that $x_j \leq y_j$ for all $m \leq j \leq n$. Then $\sum\limits_{j=m}^{n} x_j \leq \sum\limits_{j=m}^{n} y_j$.

PROOF: Left as an exercise. ■

We established earlier the convention to write

$$\begin{aligned} &x \oplus y \oplus z \quad \text{for either of} \quad (x \oplus y) \oplus z \ \text{ and } \ x \oplus (y \oplus z)\,, \\ &x \odot y \odot z \quad \text{for either of} \quad (x \odot y) \odot z \ \text{ and } \ x \odot (y \odot z)\,, \end{aligned}$$

since the operations $\oplus$ and $\odot$ are associative and we announced that we will be able to dispense with parentheses in expressions that involve sums or products of more than three items

Theorem 6.6 (Generalized Law of Associativity)**.** *Let $x_1, x_2, \dots, x_n \in R$. Then the formulas for associativity $x_1 \oplus (x_2 \oplus x_3) = (x_1 \oplus x_2) \oplus x_3$ for sums and $x_1(x_2x_3) = (x_1x_2)x_3$ for products extend to*

$x_1, x_2, \dots, x_n$ in the sense that it does not matter how parentheses are used to control the order how the sum of those n items is evaluated. Matter of fact, the value of any such grouping is $\sum\limits_{j=1}^{n} x_n$ for summation and $\prod\limits_{j=1}^{n} x_n$ for products.

PROOF: The proof is given for summation only because the proof for products is similar. It is done by induction on the size n of a list of elements of R.

Base case: The proof is obvious for $n = 1, 2, 3$. (We use associativity for $k = 3$).

Induction assumption:

If $k \leq n$ and $y_1, \dots y_k \in R$ then any grouping with parentheses of $y_1 \oplus \dots \oplus y_k$ equals $\sum\limits_{j=1}^{k} y_j$.

Let us now assume that we have a sum of $x_1, \dots x_n, x_{n+1}$, grouped by parentheses. Let A denote that sum. We may assume that $n \geq 4$ and that there are no superfluous parentheses, i.e., no parentheses of the form **(a)** (x_j), **(b)** $((\dots))$, and **(c)** pairs of parentheses that enclose the entire list of $n+1$ elements. Because parentheses of type **(a)** and **(c)** are excluded, we have either of

$$\text{case 1: } x_1 \oplus (\dots) \qquad \text{or} \qquad \text{case 2: } (\dots) \oplus x_{n+1} \qquad \text{or} \qquad \text{case 3: } (\dots) \oplus (\dots),$$

where $(\dots)$ contains at most n of the $n+1$ items.

Case 1: $(\dots)$ is a sum of the items $x_2, \dots, x_{n+1}$, grouped by parentheses. It follows from the induction assumption that $A = x_1 \oplus \sum\limits_{j=2}^{n+1} x_j = \sum\limits_{j=1}^{1} x_j \oplus \sum\limits_{j=2}^{n+1} x_j$.

Case 2: $(\dots)$ is a sum of the items $x_1, \dots, x_n$, grouped by parentheses. It follows from the induction assumption that $A = \sum\limits_{j=1}^{n} x_j \oplus x_{n+1} = \sum\limits_{j=1}^{n} x_j \oplus \sum\limits_{j=n+1}^{n+1} x_j$.

Case 3: There will be some $K \in \mathbb{N}$ such that $2 \leq K < n$ and such that the left grouping $(\dots)$ consists of $x_1, \dots, x_K$ and the right grouping $(\dots)$ consists of $x_{K+1}, \dots, x_{n+1}$. It follows from the induction assumption that $A = \sum\limits_{j=1}^{K} x_j \oplus \sum\limits_{j=K+1}^{n+1} x_j$.

In either case we conclude with the help of prop.6.15**(a)** that $A = \sum\limits_{j=1}^{n} x_{n+1}$. ■

Definition 6.8. Let $\beta \in R$. For any $n \in \mathbb{Z}_{\geq 0}$ we define $\beta^n \in R$ recursively as follows:

(6.10) (i) $\beta^0 := 1$, (ii) $\beta^{n+1} = \beta^n \odot \beta$.

In an expression of the form β^n we call β the **basis**, we call n the **exponent**, and we call β^n the n–th **power** of β. □

Remark 6.5. Note that the above definition implies that $0^0 = 1$.

Proposition 6.18 (B/G prop.4.6: Arithmetic Rules for Exponentiation)**.** *Let $\beta \in R$ and $k, m \in \mathbb{Z}_{\geq 0}$. We have the following:*

(a) *If $\beta > 0$ then $\beta^k > 0$,*
(b) $\beta^m \odot \beta^k \ = \ \beta^{m+k}$*,*
(c) $(\beta^m)^k \ = \ \beta^{mk}$*.*

PROOF: The proof of **(a)** is given (for $R = \mathbb{Z}$) in [2] Beck/Geoghegan Art of Proof, ch.4. The proofs of **(b)** and **(c)** is left as exercise 6.5 (see p.209). ■

Proposition 6.19 (B/G prop.10.9)**.** *Let $a \in R$ such that $0 \leq a \leq 1$, and let $m, n \in \mathbb{N}$ such that $m \geq n$. Then $a^m \leq a^n$.*

The proof is left as exercise 6.6 (see p.209). ■

Proposition 6.20 (B/G prop.8.41)**.** *Let $a \in R$. Then $a^2 < a^3$ if and only if $a > 1$.*

PROOF: Left as an exercise. ■

Definition 6.9 (Finite Geometric Series)**.** Let $q \in R$ and $n \in \mathbb{Z}_{\geq 0}$.

We call a sum of the form $\displaystyle\sum_{j=0}^{n} q^j$ a **finite geometric series.** □

Proposition 6.21 (Finite Geometric Series Formula (B/G prop.4.13))**.**

Let $q \in R$. If $n \in \mathbb{Z}_{\geq 0}$ then

$$(1 \ominus q) \odot \sum_{j=0}^{n} q^j \ = \ 1 \ominus q^{n+1} \, .$$

PROOF: Left for as exercise 6.8 (see p.209). That exercise is stated for $R = \mathbb{Z}$, but the proof for general R is no different. ■

Remark 6.6.

Except for prop.6.17 there are no inequalities involved in the formulas for generalized sums, products and powers, and we did not take advantage of the absence of zero divisors either. Thus we could have worked everywhere else with a commutative ring with unit instead of an ordered integral domain. □

6.6 Binomial Coefficients

The material here follows very closely ch.4.4 (The Binomial Theorem) of [2] Beck/Geoghegan.

The recursive definitions of sums $\sum x_j$, products $\prod x_j$, and powers x^n can be generalized to more general objects x_j and x (see ch.6.5 on p.180), but the following definition cannot be generalized to objects n more general than nonnegative integers.

Definition 6.10 (Definition of Factorials)**.** For any $n \in \mathbb{Z}_{\geq 0}$ we define a natural number $n!$ recursively as follows:

$$(6.11) \qquad \text{(i)} \quad 0! := 1, \qquad \text{(ii)} \quad (n+1)! = n! \cdot (n+1).$$

We pronounce $n!$ as n **factorial**. □

Remark 6.7. If $n \geq 1$ then $n!$ represents the product of the "first n natural numbers":

$$\begin{aligned} 1! &= 0! \cdot 1 = 1, \\ 2! &= 1! \cdot 2 = 1 \cdot 2, \\ 3! &= 2! \cdot 3 = 1 \cdot 2 \cdot 3, \\ &-- - - - - - - - - - - - -- \\ n! &= 1 \cdot 2 \cdot 3 \cdot \dots \cdot n. \ \square \end{aligned}$$

Definition 6.11 (Binomial coefficients)**.** Let $n, k \in \mathbb{Z}$ such that $0 \leq k \leq n$. We define the **binomial coefficient** $\begin{pmatrix} n \\ k \end{pmatrix}$ (pronounced "n choose k") as follows:

$$(6.12) \qquad \begin{pmatrix} n \\ k \end{pmatrix} := \begin{cases} 1 & \text{if k = 0 or k = n,} \\ \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} + \begin{pmatrix} n-1 \\ k \end{pmatrix} & \text{otherwise, i.e., } n \geq 2 \text{ and } 0 < k < n. \ \square \end{cases}$$

Remark 6.8. The binomial coefficients form a triangle, the so called **Pascal triangle**. The recursive formula above states that a binomial coefficient $\binom{n}{k}$ which is neither at the left side ($k = 0$) nor at the right side ($k = n$) of that triangle,is computed as the sum of its upper left and upper right ancestors. Here are the binomial coefficients for $n = 0, 1, 2, 3, 4$.

```
                1
            1       1
        1       2       1
    1       3       3       1
1       4       6       4       1
```

Note that binomial coefficients are natural numbers since they are either 1 or the sum of natural numbers. In particular they are integers and not fractions. □

Proposition 6.22.

Let $n, k \in \mathbb{Z}$ such that $0 \le k \le n$. Then

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{6.13}$$

PROOF: We do the proof by strong induction on n.
(6.13) is obvious for $k = 0$ or $k = n$ since $\frac{n!}{0!\cdot n!} = 1$. This takes care of the base cases $n = 0$ and $n = 1$. Let $n \ge 2$ and $1 \le k \le n-1$. Our induction assumption is that

$$\binom{m}{j} = \frac{m!}{j!(m-j)!} \quad \text{if } m < n \text{ and } 0 \le j \le m. \tag{6.14}$$

We prove the validity of (6.13) as follows.

$$\begin{aligned}
\binom{n}{k} &= \binom{n-1}{k-1} + \binom{n-1}{k} \\
&= \frac{(n-1)!}{(k-1)!\cdot((n-1)-(k-1))!} + \frac{(n-1)!}{k!\cdot((n-1)-k)!} \\
&= \frac{(n-1)!}{(k-1)!\cdot(n-k)!} + \frac{(n-1)!}{k!\cdot(n-1-k)!} \\
&= \frac{k\cdot(n-1)!}{k!\cdot(n-k)!} + \frac{(n-1)!\cdot(n-k)}{k!\cdot(n-k)!} \\
&= (k+n-k)\cdot\frac{(n-1)!}{k!\cdot(n-k)!} = \frac{n\cdot(n-1)!}{k!\cdot(n-k)!} = \frac{n!}{k!\cdot(n-k)!}.
\end{aligned}$$

Equation #1 above follows from (6.12), #2 follows from the induction assumption (6.14). ■

Note that, in contrast to our approach, B/G uses prop.6.22 above as the definition of the binomial coefficients, and (6.12) of this document then becomes a proposition.

The reduction formula in the following lemma allows to express a binomial coefficient in terms of another with smaller numbers.

Lemma 6.2 (Symmetry and reduction lemma)**.**

$$\binom{n}{k} = \binom{n}{n-k} \quad (0 \le k \le n) \qquad \textit{symmetry} \tag{6.15a}$$

$$\binom{n}{k} = \frac{n}{k}\cdot\binom{n-1}{k-1} \quad (1 \le k \le n) \qquad \textit{reduction} \tag{6.15b}$$

PROOF of (6.15a):

$$\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!} = \frac{n!}{(n-k)!\cdot k!} = \binom{n}{n-k}$$

PROOF (6.15b):

$$\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!} = \frac{n\cdot(n-1)!}{k\cdot(k-1)!\cdot((n-1)-(k-1))!} = \frac{n}{k}\cdot\binom{n-1}{k-1} \ ■$$

We recall the binomial formula for squares

$$(a + b)^2 = 1 \cdot a^2 + 2 \cdot ab + b^2.$$

and the one for cubes:

$$(a + b)^3 = 1 \cdot a^3 + 3 \cdot a^2 b + 3 \cdot ab^2 + 1 \cdot b^3.$$

We see that the coefficients of the terms $a^i b^j$ match the numbers of the Pascal triangle: They are the binomial coefficients $\binom{n}{k}$ for $n = 2$ and $n = 3$. Here is the generalization to compute $(a + b)^n$ for arbitrary n.

Let $R = (R, \oplus, \odot)$ be an integral domain. We also remember that exponentials x^n and products kx are defined for $n \in [0, \infty[$, $k \in \mathbb{Z}$, $x \in R$, as follows:

$$\begin{aligned} x^n &= 1 \text{ if } n = 0 \quad \text{and} \quad x^n = x \odot x^{n-1} \text{ if } n > 0\,, \\ kx &= e(k) \odot x \quad \text{where} \quad k \mapsto e(k) \text{ is the embedding } \mathbb{Z} \to R \text{ defined in Chapter 6.4}\,. \end{aligned}$$

Theorem 6.7 (Binomial theorem)**.** *Let $R = (R, \oplus, \odot)$ be an integral domain.*

For any $n \in \mathbb{Z}_{\geq 0}$ and $a, b \in R$,

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{6.16}$$

PROOF:

The proof is done by induction over n

Base case $n = 0$: Follows from $\binom{0}{0} = 1$ and $a^0 = b^0 = (a + b)^0 = 1$.

Induction assumption: For some $n \geq 0$ it is true that

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{6.17}$$

We need to show that

$$(a + b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}. \tag{6.18}$$

It follows from the formulas in ch.6.5 (Recursive Definitions of Sums, Products and Powers in Integral Domains) and prop.6.22 that

$$\begin{aligned}
\sum_{k=0}^{n+1}\binom{n+1}{k}a^k b^{n+1-k} &= \binom{n+1}{0}a^0 b^{n+1} \oplus \sum_{k=1}^{n}\binom{n+1}{k}a^k b^{n+1-k} \oplus \binom{n+1}{n+1}a^{n+1}b^0 \\
&= b^{n+1} \oplus \sum_{k=1}^{n}\left(\binom{n}{k-1}+\binom{n}{k}\right)a^k b^{n+1-k} \oplus a^{n+1} \\
&= b^{n+1} \oplus \sum_{k=1}^{n}\binom{n}{k-1}a^k b^{n+1-k} \oplus \sum_{k=1}^{n}\binom{n}{k}a^k b^{n+1-k} \oplus a^{n+1} \\
&= b^{n+1} \oplus \sum_{k=0}^{n-1}\binom{n}{k}a^{k+1} b^{n+1-(k+1)} \oplus \sum_{k=1}^{n}\binom{n}{k}a^k b^{n+1-k} \oplus a^{n+1}.
\end{aligned}$$

The last equation above results from application of prop.6.16 to the first summation term. We continue by pulling a^{n+1} into the first sum and b^{n+1} into the second sum, using prop.6.15**(a)**. This yields

$$\begin{aligned}
\sum_{k=0}^{n+1}\binom{n+1}{k}a^k b^{n+1-k} &= \sum_{k=0}^{n}\binom{n}{k}a^{k+1} b^{n-k} \oplus \sum_{k=0}^{n}\binom{n}{k}a^k b^{n+1-k} \\
&\overset{\text{prop.6.14(a)}}{=} a\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} \oplus b\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}.
\end{aligned}$$

We apply the induction assumption (6.17) to each sum in the last expression and obtain

$$\sum_{k=0}^{n+1}\binom{n+1}{k}a^k b^{n+1-k} = a(a\oplus b)^n + b(a\oplus b)^n = (a\oplus b)(a\oplus b)^n = (a\oplus b)^{n+1}.$$

We have proven (6.18), and this completes the induction step. ■

Corollary 6.4. *Let* $n \in \mathbb{Z}_{\geq 0}$. *Then* $\sum_{k=0}^{n}\binom{n}{k} = 2^n$.

PROOF: We apply the binomial theorem with $a = b = 1$. Since $1^j = 1$ for all $j \in \mathbb{Z}_{\geq 0}$,

$$2^b = (1+1)^n = \sum_{k=0}^{n}\binom{n}{k}a^j b^{n-k} = \sum_{k=0}^{n}\binom{n}{k}. \quad ■ \tag{6.19}$$

6.7 Bernstein Polynomials $\boxed{\star}$

Note that this chapter is starred, hence optional,

but also note that we will cite the results of this chapter later in this document.

The material in this chapter makes extensive use of the properties of binomial coefficients.

Definition 6.12 (Bernstein Polynomials). ⋆

Let $f : [0,1] \to \mathbb{R}$ be a real–valued function on the unit interval which need not necessarily be continuous. If $n \in \mathbb{N}$ then

$$(6.20) \qquad B_n^f : \mathbb{R} \to \mathbb{R}; \qquad x \mapsto B_n^f(x) \ := \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

defines a function of the form (5.41) (see p.155), thus B_n^f is a polynomial which we call the n–th **Bernstein polynomial** associated with $f(\cdot)$. □

Remark 6.9. Note that the degree of B_n^f need not be n. For example, any function f such that B_n^f is the zero polynomial has no degree. Obviously such is the case for the zero function $0 : x \to 0$ where $0 \leq x \leq 1$. Here is a less trivial example. Consider the function

$$(6.21) \qquad g : \mathbb{R} \to [0,1]; \qquad g(x) := \begin{cases} 0 & \text{if } x \in \mathbb{Q}, \\ 1 & \text{else.} \end{cases}$$

Since $\frac{k}{n} \in \mathbb{Q}$ for all $n \in \mathbb{N}$ and all integers k such that $0 \leq k \leq n$ it follows that $g(\frac{k}{n}) = 0$ for such k and n, thus $B_n^g = 0$. □

We now compute the Bernstein polynomials for some specific functions. The proof makes extensive use of the properties of binomial coefficients. Note that all this takes place on the unit interval $[0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$.

Proposition 6.23 (The Bernstein polynomials for $1, id(\cdot), id^2(\cdot)$). *Let*

$$(6.22) \qquad 1 : x \mapsto 1; \qquad id : x \mapsto x; \qquad id^2 : x \mapsto x^2; \qquad (0 \leq x \leq 1)$$

be the constant function 1*, the identity function and the square function on the unit interval* $[0,1]$*. Then*

$$(6.23a) \qquad B_n^1 = 1,$$

$$(6.23b) \qquad B_n^{id} = id,$$

$$(6.23c) \qquad B_n^{id^2} = \frac{1}{n} id + \frac{n-1}{n} id^2.$$

In other words, for any real number x *we have*

$$\begin{aligned} B_n^1(x) &= 1 \\ B_n^{id}(x) &= id(x) = x \\ B_n^{id^2}(x) &= \frac{1}{n} id(x) + \frac{n-1}{n} id^2(x) = \frac{1}{n} x + \frac{n-1}{n} x^2. \end{aligned}$$

PROOF of (6.23**(a)**): If $x \in \mathbb{R}$ then

$$B_n^1(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = (x + (1-x))^n = 1^n = 1$$

The second equality results from (6.16) (binomial theorem) on p.186 applied to $a = x$ and $b = 1 - x$.

PROOF of (6.23**(b)**): Let $x \in \mathbb{R}$. We must show that $B_n^{id}(x) = x$. Observe that

$$B_n^{id}(x) \ = \ \sum_{k=0}^{n} \binom{n}{k} \frac{k}{n}\, x^k (1-x)^{n-k} \ = \ \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k}\, x^k (1-x)^{n-k}.$$

We were able to discard the $k = 0$ term because $\frac{k}{n} = 0$. The reduction formula (6.15b) on p.185 yields

$$\frac{k}{n}\binom{n}{k} \ = \ \binom{n-1}{k-1} \qquad (1 \leq k \leq n),$$

thus

$$B_n^{id}(x) \ = \ \sum_{k=1}^{n} \binom{n-1}{k-1}\, x^k (1-x)^{n-k}.$$

We change the summation index to $j := k - 1$, i.e., $k = j + 1$. Then

$$B_n^{id}(x) \ = \ \sum_{j=0}^{n-1} \binom{n-1}{j}\, x \cdot x^j (1-x)^{n-(j+1)}.$$

We rewrite $n - (j+1) = n - j - 1 = (n-1) - j$. This yields (6.23**(b)**):

$$B_n^{id}(x) \ = \ x \sum_{j=0}^{n-1} \binom{n-1}{j}\, x^j (1-x)^{(n-1)-j} \ = \ x\left(x + (1-x)\right)^{n-1} \ = \ x \cdot 1^{n-1} \ = \ x. \tag{6.24}$$

PROOF of (6.23**(c)**): The proof of this formula is significantly more complicated than that of 6.23**(b)**. We have

$$B_n^{id^2}(x) \ = \ \sum_{k=0}^{n} \binom{n}{k} \frac{k^2}{n^2}\, x^k (1-x)^{n-k} \ = \ \sum_{k=1}^{n} \frac{k^2}{n^2} \binom{n}{k}\, x^k (1-x)^{n-k}. \tag{6.25}$$

We were able to throw away the $k = 0$ term because this term is the product of k^2/n^2 with some other stuff and $k^2/n^2 = 0$. As in the proof of part b, we'll use the symmetry and reduction lemma. Moreover, the definition formula for binomial coefficients

$$\binom{n}{k} \ = \ \binom{n-1}{k-1} + \binom{n-1}{k} \quad \text{for } 1 \leq k \leq n-1$$

will be used in this proof. This formula is not valid for $k = n$ and we must split off the corresponding summation term

$$\binom{n}{n} \frac{n^2}{n^2}\, x^n (1-x)^{n-n} \ = \ 1 \cdot 1 \cdot\ x^n (1-x)^0 \ = \ x^n. \tag{6.26}$$

before applying the triangle formula. For $k < n$ we obtain

$$\frac{k^2}{n^2}\binom{n}{k} \ = \ \frac{k^2}{n^2}\,\frac{n}{k}\binom{n-1}{k-1} \ = \ \frac{k}{n} \cdot \left(\binom{n}{k} - \binom{n-1}{k}\right) \tag{6.27}$$

by applying the reduction formula to the first equation and the Pascal triangle formula to the second one. Hence, remembering from (6.26) that the n^{th} term is x^n,

$$\begin{aligned} B_n^{id^2}(x) \;&=\; \sum_{k=1}^{n-1} \frac{k}{n} \cdot \left(\binom{n}{k} - \binom{n-1}{k} \right) \cdot x^k(1-x)^{n-k} \;+\; x^n \\ &=\; \sum_{k=1}^{n-1} \left(\frac{k}{n}\binom{n}{k} - \frac{k}{n}\binom{n-1}{k} \right) \cdot x^k(1-x)^{n-k} \;+\; x^n. \end{aligned} \tag{6.28}$$

We use the symmetry and reduction lemma again and substitute

$$\frac{k}{n}\binom{n}{k} \;=\; \binom{n-1}{k-1}$$

in the left hand side of the difference. This yields

$$\begin{aligned} B_n^{id^2}(x) \;&=\; \sum_{k=1}^{n-1} \binom{n-1}{k-1} x^k(1-x)^{n-k} \;+\; x^n \\ &\quad - \sum_{k=1}^{n-1} \frac{k}{n}\binom{n-1}{k} x^k(1-x)^{n-k}. \end{aligned} \tag{6.29}$$

To make the proof easier to follow we abbreviate

$$\Phi_1 \;:= \sum_{k=1}^{n-1} \binom{n-1}{k-1} x^k(1-x)^{n-k} \;+\; x^n, \tag{6.30a}$$

$$\Phi_2 \;:= \sum_{k=1}^{n-1} \frac{k}{n}\binom{n-1}{k} x^k(1-x)^{n-k}, \tag{6.30b}$$

thus

$$B_n^{id^2}(x) \;=\; \Phi_1 \;-\; \Phi_2. \tag{6.31}$$

We will transform Φ_1 and Φ_2 separately.
First we simplify Φ_1. We substitute the summation index k the same way we did before in part b:

$$j \;:=\; k-1; \qquad \text{i.e.,} \qquad k \;=\; j+1.$$

Since

$$n-k \;=\; (n-1) \;-\; (k-1) \;=\; (n-1) \;-\; j$$

and

$$\binom{n-1}{n-1} \;= 1 \;=\; (1-x)^0$$

we conclude that

$$\begin{aligned}\Phi_1 &= \sum_{j=0}^{n-2} \binom{n-1}{j} x^{j+1}(1-x)^{(n-1)-j} \;+\; x^n \\ &= \sum_{j=0}^{n-2} \binom{n-1}{j} x \cdot x^{j}(1-x)^{(n-1)-j} \;+\; \binom{n-1}{n-1} x \cdot x^{n-1}(1-x)^{0} \\ &= x\sum_{j=0}^{n-1} \binom{n-1}{j} x^{j}(1-x)^{(n-1)-j}.\end{aligned}$$

We apply the binomial theorem to the last term and obtain

(6.32) $$\Phi_1 \;=\; x\big(x+(1-x)\big)^{n-1} \;=\; x\cdot 1^{n-1} \;=\; x.$$

Now we simplify Φ_2. Since $\frac{k}{n} = \frac{n-1}{n}\cdot\frac{k}{n-1}$ we can write

$$\begin{aligned}\Phi_2 &:= \sum_{k=1}^{n-1} \frac{k}{n}\binom{n-1}{k} x^{k}(1-x)^{n-k} \\ &= \frac{n-1}{n}\sum_{k=1}^{n-1} \frac{k}{n-1}\binom{n-1}{k} x^{k}(1-x)^{n-k} \\ &= \frac{n-1}{n}\sum_{k=1}^{n-1} \binom{n-2}{k-1} x\cdot x^{k-1}(1-x)(1-x)^{(n-1)-k}.\end{aligned}$$

The last equality follows from the reduction formula $\frac{k}{n-1}\binom{n-1}{k} = \binom{n-2}{k-1}$ See (6.15b) on p.185. Since $(n-1)-k = (n-2)-(k-1)$ we conclude that

$$\begin{aligned}\Phi_2 &= \frac{n-1}{n}\sum_{k=1}^{n-1} \binom{n-2}{k-1} x\cdot x^{k-1}(1-x)(1-x)^{(n-2)-(k-1)} \\ &= \frac{n-1}{n}x(1-x)\sum_{j=0}^{n-2} \binom{n-2}{j} x^{j}(1-x)^{(n-2)-j}.\end{aligned}$$

We obtained the last equality by substituting again $j := k-1$. Another application of the binomial theorem yields

$$\sum_{j=0}^{n-2} \binom{n-2}{j} x^{j}(1-x)^{(n-2)-j} \;=\; (x+(1-x))^{n-2} \;=\; 1^{n-2} \;=\; 1.$$

Thus

(6.33) $$\Phi_2 \;=\; \frac{n-1}{n}x(1-x).$$

We finally use the expressions (6.32) for Φ_1 and (6.33) for Φ_2 in the equation $B_n^{id^2}(\cdot) = \Phi_1 - \Phi_2$:

$$\begin{aligned} B_n^{id^2}(\cdot) &= \Phi_1 - \Phi_2 \;=\; x \;-\; \frac{n-1}{n}x(1-x) \\ &= x \;-\; \frac{n-1}{n}x \;+\; \frac{n-1}{n}x^2 \\ &= x\Big(1 - \frac{n-1}{n}\Big) \;+\; \frac{n-1}{n}x^2 \\ &= \frac{x}{n} \;+\; \frac{n-1}{n}x^2. \end{aligned} \tag{6.34}$$

This concludes the proof of equation (6.23c). ■

We finish this chapter with the interpretation of the Bernstein polynomials B_n^f as expected values of the discretizations $f_n(k) = f(k/n)$ of a real–valued function defined on the unit interval. You may find it difficult to follow the next remark without some background in probability theory.

Remark 6.10 (Connection between Bernstein polynomials and probability theory)**.** Let $f : [0,1] \to \mathbb{R}$ be a nonnegative function. For each $n \in \mathbb{N}$ we define

$$f_n : [0,n]_{\mathbb{Z}} \to \mathbb{R}; \quad k \mapsto f(k/n). \tag{6.35}$$

In other words we obtain f_n by "digitizing" or "sampling" f at the points $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$.

It is well known to those who have had some exposure to probability theory that for fixed $p \in [0,1]$ the formula

$$P_p\{k\} \;:=\; \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{6.36}$$

defines a probability on the set $[0,n]_{\mathbb{Z}}$. This is the binomial distribution with parameters n and p, and its meaning is as follows.

Assume that the items in some population Ω possess a certain property B of interest, and that the probability of choosing "at random" an $\omega \in \Omega$ which possesses that property is p. For example let Ω be a box which contains 500 marbles of different colors which are well shuffled, that 100 of those marbles are of green color, that the person who picks a marble is blindfolded, and that the property of interest is B: "A green marble was picked". Then $p = \frac{100}{500} = 0.2$.

Assume now that

- n times in a row an item ω_j is chosen at random from Ω $(j = 1, 2, \dots, n)$,
- it is recorded after pick j whether or not ω_j possesses that property,
- ω_j is put back into Ω in such a way that the probabilistic situation is no different from the one we had before ω_j was chosen. In the example with the marbles this means that the marbles will be thoroughly reshuffled before each pick).

Let $S = S(\omega_1, \omega_2, \dots, \omega_n)$ denote how many of the n chosen items $\omega_1, \dots, \omega_n$ satisfy B. We can think of S as a function

$$S : \Omega^n \to [0,n]_{\mathbb{Z}}; \quad (\omega_1, \dots, \omega_n) \mapsto S(\omega_1, \dots, \omega_n). \tag{6.37}$$

Consider the preimage of the set $\{k\}$ under the function S, i.e., the set

$$\{S = k\} \;=\; S^{-1}\{k\} \;=\; \{(\omega_1, \dots, \omega_n) \in \Omega^n : \text{ exactly } k \text{ of the } \omega_j \text{ have property B}\}. \tag{6.38}$$

It is a well known fact that under the above conditions the "random variable" S follows a binomial distribution with parameters n and p as described in (6.36) above, i.e.,

$$\text{(6.39)} \qquad \text{probability of } \{S = k\} \ = P_p\{k\} \ = \ \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

After this brief excursion into binomial distributions we go back to the function f which is defined on the codomain of the random variable S. We will subsequently write $\vec{\omega}$ for $(\omega_1, \dots, \omega_n)$. The compositions $\omega \mapsto f \circ S(\omega)$ and $\omega \mapsto f_n \circ S(\omega)$ itself can be thought of as random variables since their values $f\big(S(\vec{\omega})\big)$ and $f_n\big(S(\vec{\omega})\big)$ depend on the randomly selected argument $\vec{\omega}$. The so–called **expectation** or **expected value** of the random variable $f_n \circ S$ under the probability distribution P_p defined by (6.39) is then given as

$$\text{(6.40)} \qquad E_p\big(f_n \circ S\big) \ = \ \sum_{k=0}^{n} f_n(k) \cdot P_p\{k\} \ = \ \sum_{k=0}^{n} f(k/n) \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ = \ B_n^{f_n}(p). \ \square$$

6.8 The Well–Ordering Principle

Theorem 6.8 (Well–Ordering Principle)**.**

Every nonempty subset of $\mathbb{N}$ *possesses a minimum, i.e., a smallest element.*

PROOF:
Let $A \subseteq \mathbb{N}$ such that A does not possess a minimum. We claim that it suffices to prove that

$$[1, k]_{\mathbb{Z}} \subseteq A^{\complement} \quad \text{for all } k \in \mathbb{N}\,.$$

Here, $A^{\complement} = \mathbb{N} \setminus A$ denotes the complement of A in $\mathbb{N}$. This is so because, according to Proposition 6.3 on p.171, $1 = \min(\mathbb{N})$ and thus, $\mathbb{N} \ = \ [1, \infty[_{\mathbb{Z}} = \bigcup_{k=1}^{\infty} [1, k]_{\mathbb{Z}}$.

For $k \in \mathbb{N}$ let $p(k)$ be the statement $[1, k]_{\mathbb{Z}} \subseteq A^{\complement}$.

We will use induction on k to prove that $p(k)$ is true for all $k \in \mathbb{N}$.

Base case $k = 1$: Note that $1 \notin A$, because, as the minimum of $\mathbb{N}$, 1 would also be the minimum of A. Hence, $1 \in A^{\complement}$, hence $[1, 1]_{\mathbb{Z}} = \{1\} \subseteq A^{\complement}$, and this proves the base case.

(IA) Induction assumption: $p(k)$ is true, i.e., $[1, k]_{\mathbb{Z}} \ \subseteq \ A^{\complement}$.

$(\star)$ We must prove under this assumption that $[1, k+1]_{\mathbb{Z}} \ \subseteq \ A^{\complement}$.

- Since $A \subseteq \mathbb{N}$, the induction assumption $[1, k]_{\mathbb{Z}} \subseteq A^{\complement}$ implies that $k < a$ for all $a \in A$.
- Since $]k, k+1[_{\mathbb{Z}} = \emptyset$ by Corollary 6.2 on p.171, it follows that $k + 1 \leq a$ for all $a \in A$. Thus, $k + 1$ is a lower bound of A.
- If it was true that $k+1 \in A$, then $k+1 = \min(A)$ by definition of a minimum. Since we assumed that A does not possess a minimum, we conclude that $k + 1 \notin A$, i.e., $k + 1 \in A^{\complement}$.
- Since we assumed $[1, k]_{\mathbb{Z}} \subseteq A^{\complement}$, it follows from $k + 1 \notin A$ that $[1, k+1]_{\mathbb{Z}} \subseteq A^{\complement}$.

We have shown $(\star)$ and this finishes the proof by induction. ■

Alternate proof: $\boxed{\star}$

(A) Let $\Gamma \ := \ \{k \in \mathbb{N} \, : \text{ if } B \subseteq \mathbb{N} \text{ such that } [1, k]_{\mathbb{Z}} \cap B \neq \emptyset \text{ then } \min(B) \text{ exists } \}\,.$

Proof strategy:

(1) We will show that $1 \in \Gamma$.
(2) We will show that $n+1 \in \Gamma$ whenever $n \in \Gamma$.
(3) It follows from the induction axiom that $\Gamma = \mathbb{N}$. We will show how this implies that any nonempty $B \subseteq \mathbb{N}$ possesses a minimum.

Proof of (1): Let $B \subseteq \mathbb{N}$ such that $[1,1]_{\mathbb{Z}} \cap B \neq \emptyset$, i.e., $1 \in B$. (Thus B is nonempty.) It follows from Proposition 6.3 (sharpening of B/G Prop.2.13) on p.171 and $B \subseteq \mathbb{N}$ that 1 is a lower bound of B. Since $1 \in B$, $1 = \min(B)$. We have proved that $1 \in \Gamma$, and we are done with step **(1)**.

Proof of (2): Now assume that $n \in \Gamma$. To prove that $n+1 \in \Gamma$ we proceed as follows.

Let $B \subseteq \mathbb{N}$ such that $[1, n+1]_{\mathbb{Z}} \cap B \neq \emptyset$ (which implies that B had to be nonempty to begin with). Since the set B satisfying this was arbitrarily chosen,

(**B**) to prove that $n+1 \in \Gamma$ it suffices to show that $\min(B)$ exists.

We consider the two separate cases $[1,n] \cap B = \emptyset$ and $[1,n] \cap B \neq \emptyset$.

Case 1: $[1,n] \cap B \neq \emptyset$.

Since we assume $n \in \Gamma$, we obtain from the definition of Γ that B possesses a minimum.

Case 2: $[1,n] \cap B = \emptyset$. We assumed that $[1,n+1]_{\mathbb{Z}} \cap B \neq \emptyset$, thus $n+1 \in B$. Note that $[1,n] \cap B = \emptyset$ implies that $n+1$ is a lower bound of B. Since $n+1 \in B$, $n+1 = \min(B)$.

Case 1 and **case 2** together prove **(B)**, and we are done with step **(2)**.

Proof of (3): Let $\emptyset \neq A \subseteq \mathbb{N}$ and $a \in A$. Such a exists since A is not empty. We finish the proof of the well–ordering principle by showing that A possesses a minimum.

It follows from **(1)** and **(2)** and the induction axiom that $\Gamma = \mathbb{N}$. From $A \subseteq \mathbb{N}$ we obtain $a \in \Gamma$. Since $[1,a]_{\mathbb{Z}} \cap A \neq \emptyset$, A possesses a minimum by definition of the set Γ. ■

Theorem 6.9 (Extended 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 $a' \in A'$, i.e., $A' \subseteq \mathbb{N}$. 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 $\min(B') \leq b'$ for all $b' \in B'$ that $m \geq b$ for all $b \in B$. But then m must be the maximum of A.

(c) is a trivial consequence of **(a)** and **(b)** ■

We have not yet given a precise definition of the real and rational numbers. That will be done in axiom 9.1 on p.255 and Definition 9.4 on p.255. For now we have make do with the informal definitions of ch.2.3 (Numbers) on p.24.

Example 6.4 (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 the greatest lower bound or infimum $\inf(A)$ of A, and $\sqrt{2}$ is the least upper bound or supremum $\sup(A)$ 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 possesses neither min, max, inf, sup! □

Proposition 6.24. *Let $\emptyset \neq A \subseteq B \subseteq \mathbb{Z}$.*

(a) *If B is bounded below (resp., above), then $\min(A) \geq \min(B)$ (resp., $\max(A) \leq \max(B)$).*
(b) *If also $\min(B) \notin A$ (resp., $\max(B) \notin A$), then $\min(A) > \min(B)$ (resp., $\max(A) < \max(B)$).*

PROOF:
This follows from prop.3.57 on p.79 and the extended well-ordering principle. ■

Proposition 6.25 ($\mathbb{N}$ is unbounded in $\mathbb{Z}$)**.** *For any $k \in \mathbb{Z}$ there exists $n \in \mathbb{N}$ such that $n > k$, i.e., there are no upper bounds for $\mathbb{N}$ in $\mathbb{Z}$.*

PROOF: Assume to the contrary that there exists an upper bound of $\mathbb{N}$. According to thm.6.9 (extended well-ordering principle) on p.194 $\mathbb{N}$ has a maximum. Let $u^\star := \max(\mathbb{N})$. Then $u^\star + 1$ belongs to $\mathbb{N}$ as the sum of two natural numbers. It follows from $u^\star + 1 > u^\star$ that $u^\star$ is not an upper bound of $\mathbb{N}$ and we have reached a contradiction. ■

6.9 The Division Algorithm

You will find a more complete treatment of this subject in [2] Beck/Geoghegan Art of Proof, ch.6.2.

Theorem 6.10 (Division Algorithm for Integers (B/G thm.6.13))**.**

Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ Then there exists a unique pair of integers q and r such that

$$m = qn + r \qquad \text{and} \qquad 0 \leq r < n. \tag{6.41}$$

*We call q the **quotient** and r the **remainder** when dividing n into m.*

PROOF: The proof is left as exercise 6.16 (see p.210). ■

The next two propositions are easy to prove with help of the division algorithm. Hint: What is n?

Proposition 6.26 (B/G prop.6.15)**.** *Let $m \in \mathbb{Z}$. Then m is odd if and only if there exists $q \in \mathbb{Z}$ such that $m = 2q + 1$.*

PROOF: Left as an exercise. ■

Proposition 6.27. *Any product of odd numbers is odd.*

The proof is left as exercise 6.17 (see p.210). ■

Proposition 6.28 (B/G prop.6.16)**.** *Let $n \in \mathbb{Z}$. Then n is even or $n + 1$ is even.*

PROOF: Left as an exercise. Hint: It suffices to show that if n is odd then $n + 1$ is even. WHY? ■

Proposition 6.29 (B/G prop.6.17)**.** *Let $n \in \mathbb{Z}$. Then n is even if and only if n^2 is even.*

PROOF: Left as an exercise. Hint: It suffices to show that if n is odd then n^2 is odd, and if n is even then n^2 is even: See the proof strategy of the proof of prop.3.43 on p.73. ■

Proposition 6.30 (Division Algorithm for Polynomials (B/G prop. 6.18))**.**

Let $\alpha, \beta \in \mathbb{Z}_{\geq 0}$ and let

$$n(x) \ := \sum_{j=0}^{\alpha} a_j x^j, \qquad m(x) \ := \sum_{j=0}^{\beta} b_j x^j, \tag{6.42}$$

be two polynomials with real coefficients a_j, b_j such that $n(x)$ is not the null polynomial $p(x) = 0$. Then there exist polynomials $q(x)$ and $r(x)$ such that $r(x)$ has degree less than α or $r(x) = 0$ (and hence has no degree), such that

$$m(x) \ := q(x)n(x) + r(x). \tag{6.43}$$

PROOF:

Case 1: $\alpha = 0$, i.e., $n(x) = a_0 = \text{const}$.

Note that $n \neq 0$ because we assume that $n(x)$ is not the null polynomial. Let $q(x) := \frac{m(x)}{a_0}$ and $r(x) := 0$. Then $m(x) = n(x)q(x) + r(x)$ yields the desired decomposition 6.43 of $m(x)$.

Case 2: $\alpha > 0$ and $\beta < \alpha$.

Let $q(x) := 0$ and $r(x) := m(x)$. Then $m(x) = n(x)q(x) + r(x)$ satisfies 6.43.

Case 3: $\alpha > 0$ and $\beta \geq \alpha$.

We handle this case using strong induction on β. First some notation. Let

$$A \ := \sum_{j=0}^{\alpha-1} \frac{b_\beta a_j}{a_\alpha} x^{j+\beta-\alpha}, \qquad B \ := \sum_{j=0}^{\beta-1} b_j x^j, \tag{6.44}$$

$$p(x) \ := m(x) \ - \ \frac{b_\beta}{a_\alpha} x^{\beta-\alpha} n(x) \ = \ b_\beta x^\beta \ + \ B \ - \ \frac{b_\beta}{a_\alpha} x^{\beta-\alpha} a_\alpha x^\alpha \ - \ A \ = \ B - A. \tag{6.45}$$

Note that none of the terms of A and B has a power of x with an exponent larger than $\beta - 1$ and, hence, that any constant multiple of $p(x)$ would be a suitable candidate for $r(x)$ as far as the degree is concerned.

Base case: $\beta = \alpha$.

(6.45) yields $p(x) = m(x) - \frac{b_\alpha}{a_\alpha} n(x)$. We have $m(x) = p(x) + \frac{b_\alpha}{a_\alpha} n(x)$. Let $q(x) := \frac{b_\alpha}{a_\alpha}$ and $r(x) := p(x)$. Then $m(x) = n(x)q(x) + r(x)$ yields the desired decomposition 6.43 of $m(x)$.

Induction assumption: Any polynomial $m'(x)$ with degree less than β can be written as $m'(x) = q'(x)n(x) + r'(x)$ such that $r'(x)$ has degree less than α or $r'(x) = 0$.

We use again the notation of (6.44) and (6.45). We have seen that the degree of $p(x)$ is less than β (unless it has no degree because $p(x) = 0$). It follows from (6.45) that $m(x) = p(x) + \frac{b_\beta}{a_\alpha} x^{\beta-\alpha} n(x)$. Let $q(x) := \frac{b_\beta}{a_\alpha} x^{\beta-\alpha}$ and $r(x) := p(x)$. It follows that $m(x) = n(x)q(x) + r(x)$ satisfies 6.43. ■

For the next proposition recall that a root of a polynomial $p(x)$ is a number z such that $p(z) = 0$ (see Definition 5.19 on p.154).

Proposition 6.31 (B/G prop.6.19)**.**

Let $p(x)$ be a polynomial and $z \in \mathbb{R}$. Then z is a root of p if and only if there exists a polynomial $q(x)$ such that

$$p(x) = (x - z)q(x) \text{ for all } x \in \mathbb{R}. \tag{6.46}$$

PROOF: Left as an exercise. Hint: Use the division algorithm for polynomials for the proof of "only if". ■

6.10 The Integers Modulo n

In this chapter we assume that $n \in \mathbb{N}$ is fixed.

Proposition 6.32 (B/G prop.6.24)**.**

For two integers a and b we define

$$a \sim b \quad \text{if and only if} \quad n \mid (a - b). \tag{6.47}$$

Then

(a) (6.47) *defines an equivalence relation on $\mathbb{Z}$,*

(b) *The equivalence class for $m \in \mathbb{Z}$ is $[m] = [r]$, where r is the remainder of m modulo n. See thm.6.10 (division algorithm for integers) on p.195.*

(c) *If $r \in [0, n-1]_{\mathbb{Z}}$ then $[r] = \{qn + r : q \in \mathbb{Z}\}$.*

(d) *This equivalence relation has exactly n distinct equivalence classes $[0], [1], \dots, [n-1]$.*

PROOF: See [2] Beck/Geoghegan Art of Proof, ch.6.3. ■

Definition 6.13 (Equivalence Modulo n)**.**

(a) We write $a \equiv b \mod n$ for $a \sim b$. We call n the **modulus**, and we say that a **equals** b **modulo** n.

(b) We write

$$\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z} := \{ [0], [1], \dots, [n-1] \} \tag{6.48}$$

for the set of equivalence classes resulting from the equivalence relation $a \sim b$. (See prop.6.32**(b)** above.) We call $\mathbb{Z}_n$ the set of **integers modulo** n. □

Remark 6.11. ⋆

It is beyond the scope of this document to discuss the reason why mathematicians choose to write the set $\mathbb{Z}_n$ set as a "quotient" $\mathbb{Z}$ divided by $n\mathbb{Z}$. If you take a course in abstract algebra then you will learn that the subset $n\mathbb{Z}$ of $\mathbb{Z}$ is what one calls an **ideal** in the commutative ring with unit $\mathbb{Z}$. Given a commutative ring with unit R and an ideal $\mathfrak{r} \subseteq R$ one can define an equivalence relation $a \sim b$ on R whose set of equivalence classes $R/\mathfrak{r} := \{[a] : a \in R\}$ is called the **quotient ring** of R with respect to the ideal $\mathfrak{r}$ The reason: One can define operations $[a] \oplus [b]$ and $[a] \odot [b]$ on $R/\mathfrak{r}$ which render this set into a commutative ring with unit. In the special case of $R = \mathbb{Z}$ and $\mathfrak{r} = n\mathbb{Z}$ the equivalence relation $a \sim b$ turns out to be (6.47) above, the quotient ring of equivalence classes is $\mathbb{Z}_n$, and addition and multiplication will be the operations defined in Definition 6.14 below. □

Remark 6.12. We want to introduce

$$\text{addition } [a] \oplus [b] := [a+b] \quad \text{and multiplication } [a] \odot [b] := [a \cdot b] \tag{6.49}$$

as binary operations on the mod n equivalence classes.
Since binary operations on $\mathbb{Z}_n$ are functions

$$\oplus, \odot : \mathbb{Z}_n \times \mathbb{Z}_n \longrightarrow \mathbb{Z}_n\,,$$

they must define the definition of a function. For example, $\oplus$ is such a function, if and only if for each $([a],[b]) \in \mathbb{Z}_n \times \mathbb{Z}_n$ there is a UNIQUE $[c] \in \mathbb{Z}_n$ such that

$$\big(([a],[b]),[c]\big) \in \Gamma_\oplus\,, \quad \text{the graph of the function } \oplus. \tag{6.50}$$

Considering (6.49), this can be equivalently expressed as

$$[a] \oplus [b] = [c] \quad \text{and this in turn is equivalent to } a + b \sim c\,. \tag{6.51}$$

Why would that assertion not be obvious? Note that the integers a and b are not the only elements of their equivalence classes

$$[a] = \{qn + a : q \in \mathbb{Z}\}\,; \qquad [b] = \{q'n + b : q' \in \mathbb{Z}\} \tag{6.52}$$

Now, if we select some arbitrary $q, q' \in \mathbb{Z}$ in (6.52) and define

$$a' := qn + a\,; \quad b' := q'n + b$$

then the uniqueness of $[c]$ in (6.50) with respect to the definition of $\oplus$ given in (6.49) and the equivalent formulation given in (6.51) demands that also

$$[a' \odot b'] \ = \ [c] \quad \text{i.e.,} \ \ a' \ + \ b' \ \sim \ c\,. \tag{6.53}$$

Here, we smade use of the definitions $a' = qn + a$, $b' = q'n + b$. We remove the middleman $[c]$ in (6.51) and (6.53) and obtain that being able to define $[a] \oplus [b] := [a + b]$, requires that

$$a + b \ \sim \ (qn + a) \ + \ (q'n + b) \quad \text{for all } q, q' \in \mathbb{Z}\ .$$

In a likewise manner, one shows that

$$a \cdot b \ \sim \ (qn + a) \cdot (q'n + b)$$

must be satisfied for all $q, q' \in \mathbb{Z}$ if one wants to define $[a] \odot [b] := [a \cdot b]$. □

Proposition 6.33 (B/G prop.6.25). *Let $a, a', b, b' \in \mathbb{Z}$ such that $a \sim a'$ and $a \sim a'$, i.e., $n \mid (a - a')$ and $n \mid (b - b')$. Then $a + b \sim a' + b'$ and $ab \sim a'b'$.*

PROOF: Left as an exercise. ■

According to Remark 6.12, that last proposition allows it to define operations $[a] \oplus [b]$ and $[a] \odot [b]$ on $\mathbb{Z}_n$.

Definition 6.14. Let $a, b \in \mathbb{Z}$. We define addition $[a] \oplus [b]$ and multiplication $[a] \odot [b]$ for the corresponding equivalence classes $[a], [b] \in \mathbb{Z}_n$ in terms of ordinary addition and multiplication in $\mathbb{Z}$ as follows.

$$[a] \oplus [b] \ := \ [a + b]; \qquad [a] \odot [b] \ := \ [ab]. \tag{6.54}$$

We further define $[a]^0 := [1]$. □

Proposition 6.34 (B/G prop.6.26 and B/G project 6.27).

- ***(a)*** *The operations $\oplus$ and $\odot$ on $\mathbb{Z}_n$ of Definition 6.14 above turn $(\mathbb{Z}_n, \oplus, \odot)$ into a commutative ring with unit.*
- ***(b)*** *$(\mathbb{Z}_n, \oplus, \odot)$ is an integral domain, i.e., there are no zero divisors, if and only if n is prime.* [93]

PROOF: The proof is left as exercise 6.18 (see p.210). ■

The following cannot be found in the B/G text.

Proposition 6.35 (Arithmetic mod n). *Let $m_1, m_2, \dots m_k,\ a_1, a_2, \dots a_k \in \mathbb{Z}$. Then*

$$[m_1 + m_2 + \cdots + m_k] \ = [m_1] \oplus [m_2] \oplus \cdots \oplus [m_k], \tag{6.55}$$

$$[m_1 \cdot m_2 \cdots m_k] \ = [m_1] \odot [m_2] \odot \cdots \odot [m_k], \tag{6.56}$$

$$[\ \sum_{j=1}^{k} a_j x^j\ ] \ = \sum_{j=1}^{k} [a_j] \odot [x]^j\,. \tag{6.57}$$

[93] A prime number is an integer $p \geq 2$ which can be divided evenly only by ± 1 or $\pm p$ We will discuss prime numbers in ch.6.12 (Prime Numbers) on p.202.

PROOF: We only give the proof of (6.56). It is done by induction on the number of factors k. The proof of (6.55) is similar and (6.57) is a simple consequence of the two first equations.

Basis: The proof is obvious for $k = 1$. We note that (6.56) is true for two factors (prop.6.33 and Definition 6.14 above.

Induction assumption: We assume that (6.56) holds for some $k \in \mathbb{N}$. We then obtain for $k+1$ that

$$\begin{aligned} [m_1 \cdot m_2 \cdots m_{k+1}] &= [(m_1 \cdot m_2 \cdots m_k) \cdot m_{k+1}] \\ &= \big([m_1 \cdot m_2 \cdots m_k]\big) \odot [m_{k+1}] \quad \text{(B/G def. of “}a \odot b\text{”)} \\ &= \big([m_1] \odot [m_2] \odot \cdots \odot [m_k]\big) \odot [m_{k+1}]. \quad \text{(Induction assumption (6.56))} \ \blacksquare \end{aligned}$$

6.11 The Greatest Common Divisor

We follow [2] Beck/Geoghegan Art of Proof ch.2 and ch.6.

We learned long before college about computing the greatest common divisor of two integers. To find, e.g., $\gcd(90, 225)$ we factor both 90 and 225 into their primes: $90 = 2\cdot3\cdot3\cdot5$, $225 = 3\cdot3\cdot5\cdot5$, and we extract the factors both “prime factorizations” have in common. In the case above that would be two factors 3 and one factor 5, so $\gcd(90, 225) = 3 \cdot 3 \cdot 5 = 45$.

Unfortunately we need to prove certain properties of the greatest common divisor as a stepping stone for the proof that any natural number greater than 1 can be factored uniquely, up to permutations of the factors, into prime numbers. We will get to that in ch.6.12 (Prime Numbers) on p.202, but first we must learn a few things about gcds.

We start with the following lemma [94]

Lemma 6.3 (B/G prop.2.34)**.** *For $m, n \in \mathbb{Z}$ let*

$$S := S(m,n) \ := \ \{k \in \mathbb{N} : k = mx + ny \text{ for some } x, y, \in \mathbb{Z}\}. \tag{6.58}$$

Then S is empty if and only if $m = n = 0$.

The proof is left as exercise 6.19 (see p.211). $\blacksquare$

Lemma 6.4. *For $m, n \in \mathbb{Z}$ let $S(m,n)$ be defined as in (6.58). Then*

(a) $S(m,n) = S(n,m)$,

(b) $S(m,n) = S(-m,n) = S(m,-n) = S(-m,-n)$,

(c) $S(m,n) = S(|m|,|n|)$.

PROOF of **(a)**: $S(m,n) = S(n,m)$ follows from the symmetry of the expression $mx+ny$ with respect to m and n.

PROOF of **(b)**:

It suffices to prove that $S(m,n) = S(-m,n)$ for all $m,n \in \mathbb{Z}$, because then $S(-n,m) = S(n,m)$, and it follows from **(a)** that $S(m,-n) = S(-n,m)$. Thus $S(m,-n) = S(-n,m) = S(n,m)$, and we have proven the first equation of **(b)**. The remaining equations are shown in a similar fashion.

Now to the proof that $S(m,n) = S(-m,n)$. Let $k \in S(m,n)$, i.e., there exist $x,y \in \mathbb{Z}$ such that $k = xm + yn$ and $k > 0$. Let $x' := -x$ and $y' := y$. Then $x', y' \in \mathbb{Z}$ and $x'(-m) + y'n = k$. It

[94] a lemma is a “proof subroutine” which is not remarkable on its own but very useful as a reference for other proofs

follows from $k > 0$ that $k \in S(-m,n)$. Since k is an arbitrary element of $S(m,n)$, it follows that $S(m,n) \subseteq S(-m,n)$. We apply this result to $-m$ instead of m and obtain, since $-(-m) = m$, the reverse inclusion $S(-m,n) \subseteq S\big(-(-m),n\big) = S(m,n)$.

PROOF of **(c)**: This follows from **(b)** since $|m| = m$ or $|m| = -m$, and $|n| = n$ or $|n| = -n$. ■

Definition 6.15 (Greatest Common Divisor). $\boxed{\star}$

For $m,n \in \mathbb{Z}$ let $S = S(m,n)$ be the set defined in (6.58). It follows from lemma 6.3 that if $m \neq 0$ or $n \neq 0$ then $S \neq \emptyset$, hence S possesses a minimum according to the extended well ordering principle. We thus are allowed to define the following. Let

$$\gcd(m,n) := \begin{cases} 0 & \text{if m = n = 0,} \\ \min(S) & \text{otherwise.} \end{cases} \tag{6.59}$$

We call $\gcd(m,n)$ the **greatest common divisor** of m and n. □

Proposition 6.36 (B/G prop.6.29). *Let $m,n \in \mathbb{Z}$. Then*

> ***(a)*** *$\gcd(m,n) \mid m$ and $\gcd(m,n) \mid n$,*
> ***(b)*** *If $m \neq 0$ or $n \neq 0$ then $\gcd(m,n) > 0$,*
> ***(c)*** *Let $k \in \mathbb{Z}$ such that $k \mid m$ and $k \mid n$. Then $k \mid \gcd(m,n)$.*

PROOF: The proof given here is that of B/G prop.6.29 with some minor cosmetic changes. In the following we abbreviate $g := \gcd(m,n)$.

Case 1: $m = n = 0$.

Then $g = 0$ according to Lemma 6.3 on p.200. Thus **(a)** holds since $0 \mid 0$ and **(c)** holds since $k \mid 0$ is true for any integer k.

Case 2: $m \neq 0$ or $n \neq 0$.

Then $g = \min(S)$, thus $g \in S$, thus $g \in \mathbb{N}$, and this proves **(b)**. We divide the proof of **(a)** and **(c)** into two cases as follows.

Case 2a: Either $m = 0$ and $n \neq 0$ or $n = 0$ and $m \neq 0$.

We may assume $m = 0$ since the set S is defined symmetrically with respect to m and n. Then

$$S = \{ny : y \in \mathbb{Z} \text{ and } ny > 0\} = \{|n|y : y \in \mathbb{N}\},$$

thus $g = \min(S) = |n|$. It follows that **(a)** holds since $|n| \mid n$ and $|n| \mid 0$. Further, **(c)** holds since $k \mid n \Rightarrow k \mid |n|$, i.e., $k \mid g$.

Case 2b: Both $m \neq 0$ and $n \neq 0$.

Since $S(m,n) = S(|m|,|n|)$, we may assume that $m > 0$ and $n > 0$. Since $g \in \mathbb{N}$ by the already proven part **(b)** we may apply the Division Algorithm (Theorem 6.10 on p.195) and obtain integers q, q', r, r' such that

$$m = qg + r \quad \text{and} \quad n = q'g + r' \quad \text{and} \quad 0 \leq r, r' < g\,. \tag{6.60}$$

We now prove that $r = 0$. Since $g = mx + ny$ for suitable $x, y \in \mathbb{Z}$ it follows from (6.60) that

$$r = m - qg = m - q(mx + ny) = m(1 - qx) + n(-qy),$$

thus $r > 0$ would imply $r \in S$. Since $r < g$ and $g = \min(S)$ it would not be true that $\min(S)$ is a lower bound of S. Thus the assumption that $r > 0$ contradicts the definition of a minimum, thus $r = 0$. It follows that $m = qg$, i.e., $g|m$. We obtain in likewise manner that $g|n$ and the proof of **(a)** is done.

Proving **(c)** is much simpler: Assume that $k \in \mathbb{Z}$ satisfies both $k|m$ and $k|n$. By definition of divisibility there exist integers j, j' such that $m = jk$ and $n = j'k$. Hence, for any $x, y \in \mathbb{Z}$,

$$mx + ny = (jk)x + (j'k)y = (jx + j'y)k,$$

hence $k \mid mx + ny$. Since $g \in S$, $g = mx + ny$ for suitable $x, y \in \mathbb{Z}$. Thus $k \mid g$ and **(c)** follows. ■

Remark 6.13.

- Proposition 6.36**(a)** justifies us calling gcd(m, n) a common divisor of m and n.
- If $i, j \in \mathbb{N}$ such that $i \mid j$ then $i \leq j$ according to Proposition 6.9 (B/G prop.2.23) on p.174. Thus Proposition 6.36**(c)** shows that gcd(m, n) is in fact the largest possible of all common divisors of m and n. □

Proposition 6.37 (B/G prop.6.30)**.** *Let $k, m, n \in \mathbb{Z}$. Then gcd(km, kn) = $|k| \cdot$ gcd(m, n).*

PROOF: Left as exercise 6.24 on p.211. ■

6.12 Prime Numbers

Definition 6.16 (Prime numbers and prime factorizations)**.**

(a) Let $p \in \mathbb{N}, p \geq 2$. p is a **prime number** or p is **prime** if $q \in \mathbb{Z}$ and $q \mid p$ implies that $q = \pm 1$ or $q = \pm p$. We note that 1 is **not** prime.

(b) Let $p \in \mathbb{N}, p \geq 2$. p is called a **composite number** or just a **composite** if p is not prime.

(c) Let $m \in \mathbb{N}, m \geq 2$. If there are primes $p_1, \dots, p_k$ such that $m = p_1 \cdot p_2 \cdots p_k$ then $p_1, \dots, p_k$ are called **factors** or **prime factors** of m and $p_1 \cdot p_2 \cdots p_k$ is called a **prime factorization** or just a **factorization** of m.

(d) If the prime factorizations of $m, n \in \mathbb{N}$ both contain the prime number p then we call p a **common factor** of m and n.

(e) If $m \in \mathbb{Z}$ satisfies $m \leq -2$ and if $p_1 \cdot p_2 \cdots p_k$ is a prime factorization of the positive(!) integer $-m$ then we call $-(p_1 \cdot p_2 \cdots p_k)$ a prime factorization of m. □

Remark 6.14. Note the following for the previous definition.

- no need for minus signs anywhere in part **(c)** since we assume that m is positive.
- It follows from Proposition 6.36 (B/G prop.6.29) on p.201 that $p \mid$ gcd(m, n). □

Proposition 6.38 (B/G prop.6.28)**.** *Let $n \in \mathbb{N}$ such that $n > 1$. Then n has a prime factorization.*

PROOF: The proof is left as exercise 6.20. See p.211.

Lemma 6.5. *Let p be prime and let $n \in \mathbb{N}$. We have the following:*

(a) *Either gcd$(p, n) = 1$ or gcd$(p, n) = p$.*
(b) *If $p \nmid n$ (p does not divide n) then gcd$(p, n) = 1$.*

The proof is left as exercise 6.21 (see p.211). ■

Definition 6.17 (relatively prime)**.** Let $m, n \in \mathbb{Z}$. We say that m and n are **relatively prime** if their greatest common divisor satisfies

$$\text{gcd}(m, n) = 1. \;\square$$

Proposition 6.39. *Two natural numbers m and n are relatively prime if and only if they possess no common factors.*

PROOF: The proof is left as exercise 6.22 (see p.211). ■

Remark 6.15. Lemma 6.5 above can now also be formulated this way: If p is prime and $n \in \mathbb{N}$ then

(a) Either p and n are relatively prime or gcd$(p, n) = p$.
(b) If $p \nmid n$ then p and n are relatively prime.

We next look at Euclids Lemma and the unqueness of prime number factorizations. Thm.6.11 below states the following: Every integer ≥ 2 can be factored uniquely (i.e. up to permutation) into primes. The proof of that theorem requires Euclid's lemma which in turn uses lemma 6.5 above.

Proposition 6.40 (B/G prop.6.31: Euclid's Lemma for Two Factors)**.** *Let p be prime and $m, n \in \mathbb{N}$. If $p \mid (mn)$ then $p \mid m$ or $p \mid n$.*

PROOF: The proof is left as exercise 6.25 (see p.211). ■

The generalization of Euclid's lemma to more than two factors is a straightforward proof by induction.

Proposition 6.41 (Euclid's Lemma for more than two factors)**.**

Let p be prime and $m_1, m_2, \dots, m_k \in \mathbb{N}$. If $p \mid (m_1 m_2 \cdots m_k)$ then $p \mid m_j$ for some $1 \leq j \leq k$.

PROOF: Done by strong induction on the number of factors k.

Basis: There is nothing prove for $k = 1$ and prop.6.40 (Euclid's lemma for two factors) shows the validity for $k = 2$.

Induction assumption: We assume that if p divides a product $n = n_1 n_2 \cdots n_j)$ of less than k factors then $p \mid n_i$ for some $1 \leq i \leq j$.

To prove that $p \mid m_i$ for some $1 \leq i \leq k$ we write $m_1 m_2 \cdots m_k = (m_1 m_2 \cdots m_{k-1}) m_k$. It follows from prop.6.40 that $p \mid m_k$ or $p \mid (m_1 m_2 \cdots m_{k-1})$. If p divides m_k then we are done. Otherwise we apply the induction assumption to the product $m_1 m_2 \cdots m_{k-1}$ of less than k factors and obtain that there is some $1 \leq i < k$ such that p divides m_i. ■

Theorem 6.11 (B/G thm 6.32: Uniqueness of prime factorizations)**.**

Every integer ≥ 2 can be factored uniquely (i.e., up to permutation) into primes.

PROOF: by strong induction on n.

Base case: $n = 2$: 2 is its own and obviously unique prime factorization.

Induction assumption: assume that if $2 \leq j < n$ then j has a unique PF (up to reordering).

We now show that n has a unique PF (up to reordering).

Case 1: n is prime: then n is the only and hence unique PF of itself.

Case 2: Else let $n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$ be two PFs for n. $p_1 \mid q_1 q_2 \cdots q_l$, hence $p_1 \mid q_{j_0}$ for some j_{j_0} by the generalized form of Euclid's lemma.

But then $p_1 = q_{j_0}$ because $p_1 \neq 1$ and q_{j_0} is the only integer bigger than 1 that divides the prime q_{j_0}.

Let us reorder the q_j in such a way that $j_0 = 1$. Then

$$n \; = \; p_1 n_2 \qquad \text{where} \qquad n_2 \; = \; p_2 p_3 \cdots p_k \; = \; q_2 q_3 \cdots q_l$$

is an integer less than n. It follows from the induction assumption that $q_2 \cdots q_l$ is just a reordering of $p_2 \cdots p_k$. ■

As an easy corollary of the uniqueness of prime factorizations up to the order in which they occur we obtain the following proposition which will be used in ch.7 (Cardinality I: Finite and Countable Sets) to show the existence of a bijection $\mathbb{N} \to \mathbb{N}^2$.

Notation 6.1 ("The" prime factorization of an integer greater than 1)**.**

> When we talk about prime factorizations of some $n \in [2, \infty[$ it usually does not matter in which order the prime factors of n occur. We will in such instances talk about **<u>the</u>** prime factorization of n. For example, We might say, "The prime factorization of n does not contain the number 2." □

Remark 6.16. Let $m, n \in [2, \infty[_\infty$ and p prime. Let the prime factorizations of m and n be

$$m \; = \; p_1 \cdot p_2 \cdots p_i, \qquad n = q_1 \cdot q_2 \cdots q_j$$

for suitable $i, j \in \mathbb{N}$. The following are immediate consequences of the uniqueness of prime factorizations up to reordering of the factors.

> **(a)** $p_1 \cdots p_i \cdot q_1 \cdots q_j$ is the prime factorization of $m \cdot n$.
>
> **(b)** If p is a prime factor of m then $p = p_k$ for some suitable $1 \leq k \leq i$.
>
> **(c)** It follows from **(b)** that if $p > p_k$ for each $1 \leq k \leq i$ then p is not a prime factor of m.
>
> **(d)** If p is prime and $p \mid mn$ then p is a prime factor of mn. If p is not a prime factor of m then it follows from **(a)** that p is a prime factor of n. That is of course just a reformulation of Euclid's lemma, but note that we used the uniqueness of prime factorizations to deduce this. □

The next proposition essentially states the same as **(c)** above.

Proposition 6.42 (B/G Prop.6.33)**.** *Let $a, b \in \mathbb{N}$, and assume that $a \mid b$. Assume that p is a prime factor of b that is not a prime factor of a. Then $a \mid \frac{b}{p}$.*

PROOF: The proof is left as exercise 6.26 (see p.211). ■

Proposition 6.43 (B/G Prop.6.34)**.** *Let p be a prime and $k \in \mathbb{N}$ such that $0 < k < p$. Then $p \mid \dbinom{p}{k}$.*

PROOF:

Since $k! = 2 \cdot 3 \cdots k$, $(p-k)! = 2 \cdot 3 \cdots (p-k)$ and $(p-1)!$ are product of integers that belong to $[2, p-1[$, it follows from item **(b)** of the previous remark that p is not a prime factor of $k!(p-k)!$.

Obviously p divides $p! = \binom{p}{k} \cdot \big(k!(p-k)!\big)$ It follows from item **(c)** of that remark that p is a prime factor of $\binom{p}{k}$, in particular that $p \mid \binom{p}{k}$ ■

Since $\binom{p}{k} \cdot k! \ = \ p(p-1)\cdots(p-k+1)\,,$ it follows that $p \,\big|\, \binom{p}{k} \cdot (k!)\,.$

Thus p is a prime factor of $\binom{p}{k}(k!)$. Since all prime factors of $k! = 1 \cdot 2 \cdots k$ are bounded by k and we assume $k < p$, the number p is not a prime factor of $k!$.

Hence p must be a prime factor of $\binom{p}{k}$. In particular, $p \mid \binom{p}{k}$. ■

Theorem 6.12 (Fermat's Little Theorem (B/G thm 6.35))**.**

If $m \in \mathbb{Z}$ and p is prime, then $m^p \equiv m \mod p$.

The proof is left as exercise **??** (see p.**??**). ■

Remark 6.17. We note that if $p = 2$ then Fermat's Little Theorem states that either both m^2 and m have remainder zero (both are even) or both have remainder 1, i.e., both are odd. This is true according to prop.6.29 on p.196.

Proposition 6.44 (Corollary to Fermat's Little Theorem (B/G cor.6.36))**.** *Let p be prime and let $m \in \mathbb{N}$ such that $p \nmid m$. Then*

$$m^{p-1} \equiv 1 \mod p.$$

The proof is left as exercise 6.23 (see p.211). ■

6.13 The Base–β Representation of the Integers

We have learned in school that any nonnegative integer n which is written as a string of digits $d_\mu d_{\mu-1} \dots d_1 d_0$ represents the number $n = \sum\limits_{j=0}^{\mu} d_j 10^j$. Example: $8375 = 5 \cdot 10^0 + 7 \cdot 10^1 + 3 \cdot 10^2 + 8 \cdot 10^3$. What is the difference between the 'string" or "word" 8375 and the mathematical expression $5 \cdot 10^0 + 7 \cdot 10^1 + 3 \cdot 10^2 + 8 \cdot 10^3$? None whatsoever for the mathematician who DEFINES the string $d_n d_{n-1} \dots d_1 d_0$ of decimal digits d_j, i.e., integers between 0 and 9 (see Definition 6.1 on p.167), to be the integer $\sum\limits_{j=0}^{n} d_j 10^j$.

Especially if you have done some programming you know that besides "base" 10 one also can expess n as a sum $n = \sum\limits_{j=0}^{\mu} d_j \beta^j$ where the base β is an integer 2 or bigger and each d_j is now an integer between 0 and $\beta - 1$.

If $\beta = 2$, then each d_j is either zero or one, and one speaks of a binary representation. For example, the word 10001010 which we will write as $10001010_{(2)}$, i.e., we will tag it with the base in parentheses, is a binary representation of the integer

$$0 \cdot 2^0 + 1 \cdot 2^1 + 0 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 + 0 \cdot 2^5 + 0 \cdot 2^6 + 1 \cdot 2^7$$

which equals the word $138_{(10)} = 8 \cdot 10^0 + 3 \cdot 10^1 + 1 \cdot 10^2$ when one chooses 10 as a base.

If $\beta = 16$, then $n = \sum\limits_{j=0}^{\mu} d_j 16^j$ is the hexadecimal representation of n. Here each "hexadecimal digit" d_j is an integer between 0 and 15. It is customary to write the additionally needed hex digits as

$$A := 10,\ B := 11,\ C := 12,\ D := 13,\ E := 14,\ F := 15\,.$$[95]

For example the word $(60)_{(10)}$ in base 10 becomes, since $60 = 3 \cdot 16 + 1 \cdot 12 = 3 \cdot 16^1 + C \cdot 16^0$, the word $3C_{(16)}$ in hexadecimal representation.

To show that, given a fixed base β, we can replace a nonegative integer n with its equivalent word of base β digits, we must do some work. First we must show that each for each such n there exists an index μ and base β digits $d_0, \dots, d_\mu$ such that $n = \sum\limits_{j=0}^{\mu} d_j \beta^j$ Second we must show that the association of n with $d_0 d_1 \dots d_\mu$ is unique in the following sense: If $n = \sum\limits_{j=0}^{\mu'} d'_j \beta^j$ yields a second collection of base β digits $d'_0, \dots, d'_{\mu'}$ and if both representations are minimal, i.e., $d_\mu > 0$ and $d'_{\mu'} > 0$, then $\mu = \mu'$ and $d_j = d'_j$ for all $0 \leq j \leq \mu$.
We will now set out to do that.

Definition 6.18. $\boxed{\star}$ If $\beta \in \mathbb{Z}_{\geq 2}$ then we mean by a set of **base β digits** a set of $\beta - 1$ distinct symbols $\{d_i : i \in \mathbb{Z}, 0 \leq i < \beta$ such that each d_i represents the integer i. As an example, when we talked above about hexadecimal representations, we had defined the hex digits as

$$\begin{cases} d_j \ := & \text{decimal digit for } j \in \mathbb{Z} \text{ if } 0 \leq j \leq 9, \\ d_{9+1} := & A,\ d_{9+2} := B,\ d_{9+3} := C,\ d_{9+4} := D,\ d_{9+5} := E,\ d_{9+6} := F. \end{cases}$$

No further digits are needed because $9 + 7 = \beta = 10_{(\beta)}$. $\square$

Proposition 6.45 (B/G thm.7.7: Existence of base–β representations)**.**

Let $n \in \mathbb{N}$ and $\beta \in \mathbb{N}$ such that $\beta \geq 2$. Then there exists a nonnegative integer $\mu = \mu(n)$, and there exist integers d_j $(0 \leq j \leq \mu)$ such that $0 \leq d_j < \beta$ for each j and $d_\mu > 0$, and also

$$n = \sum_{j=0}^{\mu} d_j \beta^j. \tag{6.61}$$

[95] To be picky, the right–hand sides of the equations of this line are base 10 representations $A = 10 = 9 + 1$, $B = 11 = 9 + 2$, etc.

PROOF: $\boxed{\star}$ The proof is done by strong induction on n.

Base case $n = 1$: Let $\mu = 0$ and $d_0 = 1$. Then $1 = d_0 \cdot \beta^0$. This proves the base case.

Induction assumption. If k is a natural number such that $k < n$ then there exists $\mu(k) \in \mathbb{Z}_{\geq 0}$ and integers c_j $\big(0 \leq j \leq \mu(k)\big)$ such that $0 \leq c_j < \beta$ for each j and $c_{\mu(k)} > 0$, and such that $k = \sum\limits_{j=0}^{\mu(k)} c_j \beta^j$.

Step 1: The function $j \mapsto \beta^j$ is strictly increasing: If $i, j \in \mathbb{Z}_{\geq 0}$ and $i < j$ then $\beta^{j-i} \geq 2$, hence $\beta^i < (\beta^i)(\beta^{j-i}) = \beta^j$. Moreover it follows from exercise 2.12 on p.48 that $\beta^n > n$ for $n \in \mathbb{Z}_{\geq 0}$.
All this implies that the set $A := \{j \in \mathbb{N} : \beta^j \leq n\}$ has n as an upper bound, hence it possesses a maximum $\mu := \max(A)$ (extended well–ordering principle). Clearly $\beta^\mu \leq n < \beta^{\mu+1}$. If $\beta^\mu = n$ then we have a representation (6.61) for n because we can choose $d_j = 0$ for $0 \leq j < \mu$ and $d_\mu = 1$. So we rule out this case and assume from now on that

$$\beta^\mu \; < \; n \; < \; \beta^{\mu+1}. \tag{6.62}$$

Step 2: Let $n' := n - \beta^\mu$. It follows from (6.62) that $n' \in \mathbb{N}$. Since $n' < n$ the induction assumption yields $\mu(n') \in \mathbb{Z}_{\geq 0}$ and integers a_j $\big(0 \leq j \leq \mu(n')\big)$ such that $0 \leq a_j < \beta$ for each j and $a_{\mu(n')} > 0$, and such that $n' = \sum\limits_{j=0}^{\mu(n')} a_j \beta^j$.

Step 3: We show that $\mu(n') \leq \mu$. Otherwise we would have $\mu(n') \geq \mu + 1$. Since $a_{\mu(n')} \geq 1$,

$$n - \beta^\mu \; = n' \; = \; \sum_{j=0}^{\mu(n')} a_j \beta^j \; \geq \; 1 \cdot \beta^{\mu(n')} \; \geq \; \beta^{\mu+1} \; > \; n$$

(the last inequality results from (6.62)), and we would have reached a contradiction. If $\mu(n') \neq \mu$, i.e., $\mu(n') < \mu$, we define $\alpha_j = 0$ for $\mu(n') \leq j \leq \mu$. It follows that $n' = \sum\limits_{j=0}^{\mu} a_j \beta^j$.

Step 4: We show that $a_\mu < \beta - 1$. Otherwise we would have

$$n - \beta^\mu \; = n' \; = \; \sum_{j=0}^{\mu} a_j \beta^j \; \geq \; (\beta - 1) \cdot \beta^\mu \; = \; \beta^{\mu+1} - \; \beta^\mu,$$

hence $n \geq \beta^{\mu+1}$, a contradiction to (6.62).

Step 5: It follows from $n - \beta^\mu = \sum\limits_{j=0}^{\mu} a_j \beta^j$ and $a_\mu < \beta - 1$ that

$$n \; = \; (n - \beta^\mu) + \beta^\mu \; = \; \sum_{j=0}^{\mu} a_j \beta^j + \beta^\mu \; = \; \sum_{j=0}^{\mu} d_j \beta^j \quad \text{where} \;\; d_j \; = \; \begin{cases} a_j & \text{if } j < \mu, \\ a_j + 1 & \text{if } j = \mu. \end{cases}$$

Since $d_\mu = a_\mu + 1 \neq 0$ we have found a representation of the form (6.61) for n. ■

Remark 6.18. The proof of prop.6.45 shows that if $n = \sum\limits_{j=0}^{K} d_j \beta^j$ then the maximal index i for a nonzero d_i is $i = \mu = \max\{j \in \mathbb{N} : \beta^j \leq n\}$. In other words, if $n < \beta^j$ then $d_j = 0$. □

Proposition 6.46 (B/G prop.7.9: Uniqueness of base–β representations)**.**

Let $n \in \mathbb{N}$ and $\beta \in \mathbb{N}$ such that $\beta \geq 2$. Assume that

$$n = \sum_{j=0}^{\mu} d_j \beta^j = \sum_{j=0}^{\mu'} d'_j \beta^j \tag{6.63}$$

where $\mu, \mu' \in \mathbb{Z}_{\geq 0}$, each d_i and each d'_i is a base β digit, $d_\mu \neq 0$ and $d'_{\mu'} \neq 0$. Then $\mu = \mu'$ and $d_i = d'_i$ for all i.

PROOF: The proof is left as exercise 6.27 on p.**??**. ■

We learn at a very young age that an integer is divisible by 3 if and only if the sum of its digits is divisible by 3. For example, the number $3 \mid 2,784$ since $2+7+8+4$ is divisible by 3, and $528 \nmid 3$ since $5+2+8$ is not divisible by 3. We will prove this as an application of the base–10 Representation of the Integers.

Proposition 6.47 (B/G Prop.7.11)**.** *Let $n := \sum\limits_{j=0}^{\mu} x_j 10^j$, where each x_j is a digit and $x_\mu \neq 0$. Then*

$$n = x_0 + x_1 + \cdots + x_\mu \mod 3. \tag{6.64}$$

The proof is left as exercise 6.28 (see p.211). ■

6.14 The Addition Algorithm for Two Nonnegative Numbers (Base 10)

We give a simpler version of the addition algorithm than the one found in ch.7.2 of B/G.

Remark 6.19 (Addition subroutine)**.** Given are

$$x := \sum_{n=0}^{K} x_n \cdot 10^n, \quad y := \sum_{n=0}^{K} y_n \cdot 10^n$$

in base–10 representation, i.e., x_n, y_n are digits $0, 1, 2 \ldots, 9$. We may choose the same ending index K for both x and y by "filling up" the number with less digits with leading zeros.
Here is the pseudocode for a subroutine, Add(n, x_n, y_n, z_n), whose task it is to compute the n–th digits z_n for the sum $z := \sum\limits_{n=0}^{K+1} z_n \cdot 10^n := x + y$.

```
Subroutine Add( $n, x_n, y_n, i_{n-1}, z_n, i_n$ ) :
//*
//* Inputs: $n, x_n, y_n, i_{n-1}$
//* Output: $z_n, i_n$
//*
If $n = 0$ then {
    $i_{-1} := 0$;
```

}

$$\begin{aligned} i_n \; &:= \begin{cases} 0 & \text{if } x_n + y_n + i_{n-1} < 10 \\ 1 & \text{if } x_n + y_n + i_{n-1} \geq 10 \end{cases} \\ z_n \; &:= (x_n + y_n + i_{n-1}) \;-\; i_n \cdot 10 \end{aligned}$$

end-of-Subroutine

Note that the "sum digit" z_n and the "carry" i_n are associated with the "Euclidean division algorithm decomposition"

$$x_n + y_n + i_{n-1} \;=\; 10 \cdot q + r$$

for the integer $x_n + y_n + i_{n-1}$ as follows:

$$i_n = q \quad \text{and} \quad z_n = r. \;\square$$

6.15 Exercises for Ch.6

Exercise 6.1. Prove prop.6.1 on p.168 of this document: If $i, j, n \in \mathbb{Z}$ and $A_i \;=\; \{k \in \mathbb{Z} : k \geq i\}$ then $n + i \in A_i \;\Leftrightarrow\; n + j \in A_j$. $\square$

Exercise 6.2. Prove prop.6.12 on p.177 of this document: If $n \in \mathbb{N}$ then $e(n) \in P$. $\square$

Exercise 6.3. Prove the following part of prop.6.13 on p.177 of this document:
Let $m, n \in \mathbb{N}$. Then $e(n) \prec e(m) \;\Rightarrow\; n < m$. $\square$

Exercise 6.4. Let $x_0 = 8,\; x_1 = 16,\; x_{n+1} = 6x_{n-1} - x_n$ for $n \in \mathbb{N}$. Prove that $x_n = 2^{n+3}$ for every integer $n \geq 0$. Hint: Use strong induction. $\square$

Exercise 6.5. Prove parts **(b)** and **(c)** of prop.6.18 on p.183 of this document:
Let $\beta \in \mathbb{Z}$ and $k, m \in \mathbb{Z}_{\geq 0}$. Then

(b) $\beta^m \odot \beta^k \;=\; \beta^{m+k}$,
(c) $(\beta^m)^k \;=\; \beta^{mk}$. $\square$

Exercise 6.6. Prove prop.6.19 on p.183 of this document: Let $a \in R$ such that $0 \leq a \leq 1$, and let $m, n \in \mathbb{N}$ such that $m \geq n$. Then $a^m \leq a^n$. $\square$

Exercise 6.7. Let $R = (R, \oplus, \odot, P)$ be an ordered integral domain, let $n \in [2, \infty[_{\mathbb{Z}}$, and let $x_j \in R$ for $j \in \mathbb{N}$. Prove by induction that

$$\prod_{j=1}^{n} |x_j| \;=\; \left| \prod_{j=1}^{n} x_j \right|.$$

You may use that: for any two $x, y \in R$ it is true that $|a| \odot |b| \;=\; |a \odot b|$. $\square$

Exercise 6.8. Prove prop.6.21 on p.183 of this document:

Let $q \in \mathbb{Z}$. If $n \in \mathbb{Z}_{\geq 0}$ then $(1 - q) \sum\limits_{j=0}^{n} q^j \;=\; 1 - q^{n+1}$.

Hint: Prove the case $q \neq 1$ by induction on n. $\square$

Exercise 6.9. Let $R \subseteq \mathbb{Z}^2$ be the divisibility relation on $\mathbb{Z}$: $mRn \Leftrightarrow m \mid n$.

(a) Prove that R is reflexive and transitive.

(b) Prove that R is not antisymmetric (and hence not a partial order relation) by finding two different integers m, n such that $m|n$ and $n|m$. □

Exercise 6.10. Prove Theorem 6.4 on p.171 of this document: If $k \in \mathbb{N}$ then $k \geq 1$. □

Exercise 6.11. Prove prop.6.9 on p.174 of this document: Let $m, n \in \mathbb{N}$. If $m \mid n$ then $m \leq n$ □

Exercise 6.12. Prove prop.6.2 on p.171 of this document: There exists no $x \in \mathbb{Z}$ such that $0 < x < 1$. □

Exercise 6.13. Prove cor.6.2 on p.171 of this document: If $n \in \mathbb{Z}$ then there exists no $x \in \mathbb{Z}$ such that $n < x < n + 1$. □

Exercise 6.14. Prove prop.6.8 on p.174: Let $n \in \mathbb{N}$. Then $n^2 + 1 > n$. Try to prove this with and without the use of induction. □

Exercise 6.15. Let $a \in \mathbb{Z}$. Prove that there exists $b \in]-\infty, 0[$ such that $b < a$, i.e., there are no lower bounds for $]-\infty, 0[$ in $\mathbb{Z}$. Do so without making use of prop.6.25 ....

Hint: ..... But base your proof on that of prop.6.25. □

Exercise 6.16. Prove prop.6.10 on p.195 of this document: Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ Then there exists a unique pair of integers q and r such that

$$m = qn + r \qquad \text{and} \qquad 0 \leq r < n.$$

(a) Prove uniqueness of the "decomposition" $m = qn+r$: If you have a second such decomposition $m = \tilde{q}n + \tilde{r}$ then show that this implies $q = \tilde{q}$ and $r = \tilde{r}$. Start by assuming that $r \neq \tilde{r}$ which means that one of them is smaller than the other and take it from there.

(b) Prove the existence of q and r. This is much harder than **(a)**.

Hint for **(b)**: Review the extended well–ordering principle (thm.6.9 on p.194). Its use will give the easiest way to prove this theorem. Apply it to the set

$$A := A(m,n) := \{x \in \mathbb{Z}_{\geq 0} : x = m - kn \text{ for some } k \in \mathbb{Z}\}$$

. Hint for both **(a)** and **(b)**: Prop.3.61 and cor.3.5 on p.81 from ch.3.5 (Minima, Maxima, Infima and Suprema in Ordered Integral Domains) in their formulation for $(R, \oplus, \odot, P) = (\mathbb{Z}, +, \cdot, \mathbb{N})$ will come in handy in connection with the condition $0 \leq r < n$. □

Exercise 6.17. Prove prop.6.27 on p.196 of this document:

Any product of odd numbers is odd.

Hint: Use induction on k to prove that the product $n_1 n_2 \cdots n_k$ of k odd numbers x_j is odd. □

Exercise 6.18. Prove prop.6.34 on p.199 of this document:

(a) The operations $\oplus$ and $\odot$ on $\mathbb{Z}_n$ of Definition 6.14 above turn $(\mathbb{Z}_n, \oplus, \odot)$ into a commutative ring with unit.

(b) $(\mathbb{Z}_n, \oplus, \odot)$ is an integral domain, i.e., there are no zero divisors, if and only if n is prime. □

Exercise 6.19. Prove lemma 6.3 on p.200 of this document: For $m, n \in \mathbb{Z}$ let

$$S := S(m,n) \ := \ \{k \in \mathbb{N} : k = mx + ny \text{ for some } x, y, \in \mathbb{Z}\}.$$

Then S is empty if and only if $m = n = 0$.

Hint: The difficult part is proving that S is not empty if at least one of m, n is not zero. What does S look like if $m = 0$ and $n \neq 0$? Do that case first, then do the case where both m and n are not zero. Play around with specific number to see what happens before you attempt to do the proof. □

Exercise 6.20. Prove prop.6.38 on p.202: Any integer ≥ 2 has a prime factorization. □

Exercise 6.21. Prove lemma 6.5 on p.203 of this document: Let p be prime and let $n \in \mathbb{N}$. We have the following:

(a) Either $\gcd(p,n) = 1$ or $\gcd(p,n) = p$.
(b) If $p \nmid n$ (p does not divide n) then $\gcd(p,n) = 1$. □

Exercise 6.22. Prove that two natural numbers m and n are relatively prime if and only if they possess no common factors: □

Exercise 6.23. Prove prop.6.44 on p.205 of this document: Let p be prime and let $m \in \mathbb{N}$ such that $p \nmid m$. Then $m^{p-1} \equiv 1 \mod p$.

Hint: Modify the problem so that you can apply Fermat's Little Theorem to it. □

Exercise 6.24. Prove prop.6.37 on p.202 of this document:

Let $k, m, n \in \mathbb{Z}$. Then $\gcd(km, kn) \ = \ |k| \cdot \gcd(m,n)$.

Hint: You must distinguish the cases where $S(m,n)$ and/or $S(km,kn)$ is empty from the others, and you want to work with nonnegative k, m, n as much as possible. Do the following cases in the sequence given:

Case 1: $k = 0$ • **Case 2:** $m = n = 0$ • **Case 3:** $m \geq 0, n \geq 0$. At least one of $m, n \neq 0$, $k > 0$ • **Case 4:** $k < 0$, at least one of $m, n \neq 0$.

Cases 1 and 2 are trivial

For case 3 abbreviate $g := \gcd(m,n)$, $g' := \gcd(km,kn)$, $S := S(m,n)$, $S' := S(km,kn)$.

(i) Show that $kg \in S'$ and use that to prove that $g \leq kg'$.

(ii) Show that there exists $z \in S$ such that $g' = kz$. Use that to prove that $g \geq kg'$.

It is easy to prove case 4 using case 3 and lemma 6.4**(c)**: $S(a,b) = S(|a|,|b|)$. □

Exercise 6.25. Prove prop.6.40 on p.203 of this document: Let p be prime and $m, n \in \mathbb{N}$. If $p \mid (mn)$ then $p \mid m$ or $p \mid n$. □

Exercise 6.26. Prove prop.6.42 on p.204 of this document: Let $a, b \in \mathbb{N}$, and assume that $a \mid b$. Assume that p is a prime factor of b that is not a prime factor of a. Then $a \mid \frac{b}{p}$. □

Exercise 6.27. Prove prop.6.46 on p.208:

Let $n \in \mathbb{N}$ and $\beta \in \mathbb{N}$ such that $\beta \geq 2$. Assume that $n \ = \ \sum_{j=0}^{\mu} d_j \beta^j \ = \ \sum_{j=0}^{\mu'} d'_j \beta^j$ where $\mu, \mu' \in \mathbb{Z}_{\geq 0}$, each d_i and each d'_i is a base β digit, $d_\mu \neq 0$ and $d'_{\mu'} \neq 0$. Then $\mu = \mu'$ and $d_i = d'_i$ for all i. □

Exercise 6.28. Prove prop.6.47 on p.208 of this document: Let $n := \sum\limits_{j=0}^{\mu} x_j 10^j$, where each x_j is a digit and $x_\mu \neq 0$. Then $n \ = \ x_0 + x_1 + \cdots + x_{\nu(n)-1} \mod 3$. □

6.16 Blank Page after Ch.6

This page is intentionally left blank.

7 Cardinality I: Finite and Countable Sets

Notation: In this entire chapter, if $n \in \mathbb{N}$, the symbol $[n]$ does not denote an equivalence class of any kind but the set $[1, n]_{\mathbb{Z}} = \{1, 2, \dots, n\}$ of the first n natural numbers. we further define $[0] := \emptyset$.

7.1 The Size of a Set

We said in the preliminary definition of the size of a set (Definition 2.12 on p.23) that the size $|X|$ of a set X is the number of its elements. It is surprisingly difficult to make this definition precise. The following proposition will help us in this endeavor.

Proposition 7.1. *Let $n \in \mathbb{N}$. Let $\emptyset \neq A \subsetneq [n]$ be a proper, nonempty subset of $[n]$. Then there is no surjection from A onto $[n]$.*

PROOF: There is nothing to prove for $n = 1$ since the set $[1] = \{1\}$ does not strictly contain any sets other than the empty set. The proof for $n \geq 2$ will be done by induction on n.

Base case: Let $n = 2$. The only proper subsets of $[2]$ are the singleton sets $\{1\}$ and $\{2\}$. For any function $f : \{1\} \to [2]$ we have either $f(1) = 1$ in which case $2 \notin f(\{1\})$ or $f(1) = 2$ in which case $1 \notin f(\{1\})$. It follows in either case that f is not surjective. The proof for functions $\{2\} \to [2]$ is similar. This proves the base case.

Induction assumption: Let $n \in \mathbb{N}$ such that if $\emptyset \neq \Gamma \subsetneq [n-1]$ then there is no surjection $\Gamma \to [n-1]$.

We must prove that if $\emptyset \neq A \subsetneq [n]$ then there is no surjection from A to $[n]$.

We assume to the contrary that a surjective $f : A \to [n]$ exists.

case 1: $n \notin A$**:**

Then $A \subseteq [n-1]$. Let $\Gamma := A \setminus \{f = n\}$. Because f is surjective, $\{f = n\}$ is not empty, hence $\Gamma \subsetneq A \subseteq [n-1]$, hence Γ is a strict subset of $[n-1]$.

From $n \geq 3$ we obtain $n \neq 1$. Thus $\{f = 1\} \cap \{f = n\} = \emptyset$, thus $\{f = 1\} \subseteq \{f = n\}^{\complement}$, thus

$$(*) \qquad A \cap \{f = 1\} \subseteq A \cap \{f = n\}^{\complement} = A \setminus \{f = n\} = \Gamma\,.$$

Since f is surjective, $\{f = 1\} \cap A \neq \emptyset$. From this and **(*)** we obtain $\Gamma \neq \emptyset$.

We have seen earlier that Γ is a strict subset of $[n-1]$, thus $\emptyset \neq \Gamma \subsetneq [n-1]$. It follows that the induction assumption applies to Γ; hence there is no surjective $\psi : \Gamma \to [n-1]$.

We will obtain a contradiction to the above and thus finish the proof of **case 1** by showing that if we restrict the domain of f to Γ and its codomain to $[n-1]$ then $f\big|_{\Gamma} : \Gamma \to [n-1]$ is surjective.

So let $y \in [n-1]$. f is surjective, hence there exists $x \in A$ such that $f(x) = y$. Since $y \leq n-1$, $y \neq n$, hence $x \notin \{f = n\}$, hence $x \in \Gamma$. We found for arbitrary $y \in [n-1]$ some x in the domain of $f\big|_{\Gamma}$, hence this function is surjective. This completes the proof of **case 1**.

case 2: $n \in A$**:**

Because $A \subsetneq [n]$ there exists $j \in [n-1]$ such that $j \notin A$. Let $A' := (A \setminus \{n\}) \uplus \{j\}$. It follows from prop.5.6 on p.148 that there is a bijection $g : A' \xrightarrow{\sim} A$. Hence $f \circ g : A' \to [n]$ is surjective as the composition of a surjection with a bijection (see cor.5.1.b on p.147).

But A' satisfies the conditions of **case 1** since $n \notin A'$, and we have already proven that such a surjection cannot exist. Again we have reached a contradiction. ■

Corollary 7.1. *The following contains B/G thm.13.4 and B/G cor.13.5. Let $m, n \in \mathbb{N}$.*

- ***(a)*** *If $m < n$ then there exists no surjective function $f : [m] \to [n]$.*
- ***(b)*** *If $m > n$ then there exists no injective function $g : [m] \to [n]$. This is commonly referred to as the **pigeonhole principle**.*
- ***(c)*** *If $m \neq n$ then there exists no bijective function $f : [m] \to [n]$.*
- ***(d)*** *There exists no surjective function $h : [m] \to \mathbb{N}$.*

The proof of **(a)**–**(c)** is left as exercise 7.1 (see p.230). ■

PROOF of **(d)**: Clearly the function

$$\psi : \mathbb{N} \to [n+1]; \qquad m \mapsto \psi(m) := \begin{cases} m & \text{if } m \leq n+1, \\ 1 & \text{if } m > n+1 \end{cases}$$

is surjective. If there were a surjective function $h : [n] \to \mathbb{N}$ then $\psi \circ h : [n] \to [n+1]$ would be surjective by cor.5.5**(b)** on p.147. But we have established in part **(a)** of this corollary that no surjection $[n] \to [n+1]$ exists. ■

Remark 7.1. The fact that there is no surjective function $[m] \to [n]$ if $m < n$ can be expressed as follows: If a flock of m pigeons flies toward n pigeonholes for shelter then not all of those pigeonholes will be occupied. The pigeonhole principle states the other side of the coin: If a flock of n pigeons flies toward m pigeonholes for shelter then at least one of those pigeonholes will be occupied by more than one pigeon. □

Cor.7.1 yields yet another important benefit: We can now make precise the preliminary definition 2.12 of the size of a set which was given on p.23, at the end of ch.2.1 (Sets and Basic Set Operations).

Definition 7.1 (Finite and infinite sets)**.**

- **(a)** Let X be a nonempty set. Let $n \in \mathbb{N}$ such that there exists a bijective mapping $F : [n] \to X$. It follows from cor.7.1**(c)** above that if $n' \in \mathbb{N}$ such that there exists another bijective mapping $F' : [n'] \to X$ then $n = n'$, i.e., n is uniquely defined by the property that $[n]$ can be bijected to X. This allows us to call n the **size** of X, and we write $|X| = n$.
  If we write x_j for $F(j)$, we see that X is of the form
  $$X = F([n]) = \{F(j) : j \in [n]\} = \{x_j : 1 \leq j \leq n\},$$
  i.e., its elements can be enumerated as $x_1, x_2, \dots, x_n$. This is the mathematician's way of stating that $|X| = n$ is the number of elements of X.
- **(b)** We say that the empty set $\emptyset$ has size $|\emptyset| = 0$.
- **(c)** We call a set X **finite**, if there exists $n \in [0, \infty[_{\mathbb{Z}}$ such that X has size n. Note that this implies that the empty set is finite. We say that X is **infinite** and we write $|X| = \infty$ if X is not finite.

As strange as this may seem, 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. We have a special name for infinite sets whose elements can be mapped bijectively to $\mathbb{N}$.

(d) Let X be a set such that there is a bijection $f : \mathbb{N} \xrightarrow{\sim} X$. In other words, all of the elements of X can be arranged in a sequence $(x_n)_{n \in \mathbb{N}}$ such that

$$X = \{ x_n : n \in \mathbb{N}, x_n = f(n) \}$$

Then we call X a **countably infinite** set.

(e) We call a set that is either finite or countably infinite a **countable** set.

(f) A set that is neither finite nor countably infinite is called **uncountable** or **not countable**.

We use the phrase **"finitely many"** items, **"countably many"** items, **"infinitely many"** items, etc., if they would constitute a finite set, a countable set, an infinite set, etc. □

Example 7.1. Let $X := \{2n : n \in \mathbb{N}\}$ be the set of all even natural numbers. Then X is countably infinite because the function

$$f : \mathbb{N} \longrightarrow X; \quad n \mapsto 2n \qquad \text{has the function} \quad g : X \longrightarrow \mathbb{N}; \quad k \mapsto \frac{k}{2}$$

as its inverse and hence is bijective.

Note that $\frac{k}{2}$ exists (as an integer) for all $k \in X$ since the even natural numbers are divisible by 2. □

Proposition 7.2. *A countably infinite set is infinite (and not finite).*

Proof: Assume to the contrary that a set A is both finite and countably infinite. Thus we have bijections $g : A \xrightarrow{\sim} \mathbb{N}$ and, for a suitable $n \in \mathbb{N}$, $f : [n] \xrightarrow{\sim} A$. By cor.5.1.b on p.147 the composition $g \circ f : [n] \xrightarrow{\sim} \mathbb{N}$ is bijective, thus surjective. This contradicts cor.7.1**(d)** on p.214. ■

Proposition 7.3.

Let X and Y be two nonempty sets with a bijection $f : X \xrightarrow{\sim} Y$. Then

(a) *Y is finite if and only if X is finite,*

(b) *Y is countably infinite if and only if X is countably infinite,*

(c) *Y is countable if and only if X is countable,*

(d) *Y is uncountable if and only if X is uncountable.*

(e) $|Y| = |X|$.

PROOF: The proof of **(a)**, **(b)** and **(e)** is based on prop.5.5.**(c)** on p.147 which states that the composition of two bijective functions is bijective.

We only need to prove the " ⇒" directions because we obtain " ⇐" by switching the roles of X and Y.

PROOF of **(a)** and **(e)**. If X is finite then there exists $n \in \mathbb{N}$ and a bijection $g : X \xrightarrow{\sim} [n]$. $Y \xrightarrow{g \circ f^{-1}} [n]$ is bijective according to prop.5.5.**(c)** on p.147. This proves both that Y is finite and $|Y| = n = |X|$.

PROOF of **(b)**. If X is countably infinite then there exists a bijection $g : X \xrightarrow{\sim} \mathbb{N}$. Because $Y \xrightarrow{g \circ f^{-1}} \mathbb{N}$ is bijective, Y also is countably infinite.

PROOF of **(c)**. If X is countable then this set is either finite or countably infinite. If X is finite then Y is finite according to part **(a)**; if X is countably infinite then Y is countably infinite according to part **(b)**. This proves that Y is countable.

PROOF of **(d)**. Assume to the contrary that X is uncountable and Y is countable. It follows from part **(c)** that X is countable and we have reached a contradiction. Thus, Y is uncountable. ■

Proposition 7.4.

let A and B two mutually disjoint, finite sets. Then $A \uplus B$ is finite and

$$|A \uplus B| \;=\; |A| \,+\, |B|\,.$$

PROOF: There are $m, n \in \mathbb{N}$ such that $|A| = m$ and $|B| = n$, i.e., there exist bijections $\varphi : [1, m]_{\mathbb{Z}} \to A$ and $\psi : [1, n]_{\mathbb{Z}} \to B$. The function

$$f : [1, n]_{\mathbb{Z}} \to [m+1, m+n]_{\mathbb{Z}}; \qquad j \mapsto m + j$$

is a bijection since the function

$$g : [m+1, m+n]_{\mathbb{Z}} \to [1, n]_{\mathbb{Z}}; \qquad i \mapsto i - m$$

satisfies $g \circ f = id_{[1,n]_{\mathbb{Z}}}$ and $f \circ g = id_{[m+1,m+n]_{\mathbb{Z}}}$, and thus g is the inverse of f.

Thus the function $\psi \circ g$ is a bijection $[m+1, m+n]_{\mathbb{Z}} \xrightarrow{\sim} B$ as the composition of the two bijections g and ψ.

$$\begin{tikzcd}
{[1,n]_{\mathbb{Z}}} \arrow[r, "\psi"] & B \\
{[m+1,m+n]_{\mathbb{Z}}} \arrow[u, "g"] \arrow[ur, "\psi \circ g"'] &
\end{tikzcd}$$

Next, let $F : [1, m+n]_{\mathbb{Z}} \longrightarrow A \uplus B$ be defined as $F(k) \;:=\; \begin{cases} \varphi(k) & \text{if } 1 \leq k \leq m, \\ \psi\big(g(k)\big) & \text{if } m+1 \leq k \leq m+n. \end{cases}$

We claim that F is bijective.

To prove injectivity we assume that $k, k' \in [1, m+n]_{\mathbb{Z}}$ such that $k \neq k'$. We separately examine the three cases $k, k' \leq m$, $k, k' > m$, and $k \leq m < k'$.

- **(a)** If $k, k' \leq m$ then $F(k) = \varphi(k) \neq \varphi(k') = F(k')$ since φ is injective.
- **(b)** If $k, k' > m$ then $F(k) = \psi\big(g(k)\big) \neq \psi\big(g(k')\big) = F(k')$ since $\psi \circ g$ is injective.
- **(c)** If $k \leq m < k'$ then $F(k) = \varphi(k) \in A$ and $F(k') = \psi\big(g(k')\big) \in B$. Since A and B are disjoint we conclude that $F(k) \neq F(k')$.

To prove surjectivity let $x \in A \uplus B$. Then

- **(a)** either $x \in A$, and the surjectivity of φ allows us to conclude that there exists $k \in [1, m]_{\mathbb{Z}}$ such that $\varphi(k) = x$, i.e., $F(k) = x$;
- **(b)** or $x \in B$, and there exists $k \in [m+1, m+n]_{\mathbb{Z}}$ such that $F(k) = \psi \circ g(k) = x$ since the function $\psi \circ g$ is surjective.

The existence of a bijection $F : [1, m+n]_{\mathbb{Z}} \xrightarrow{\sim} A \uplus B$ proves that $|A \uplus B| = m + n = |a| + |B|$. ■

The assertion of the next lemma is intuitively clear. If one adds a new element ω to X then one obtains for each existing $A \subseteq X$ an additional subset $A \cup \{\omega\}$ of $X \cup \{\omega\}$.

Lemma 7.1. *Let X, Ω be sets such that $X \subsetneq \Omega$ and $\omega \in X^{\complement}$, and let $\mathfrak{B} := \{A \uplus \{\omega\} : A \in 2^X\}$. Then the function $F : 2^X \to \mathfrak{B}$; $A \mapsto A \uplus \{\omega\}$ is a bijection.*

The proof is left as exercise **??** (see p.**??**). ■

Proposition 7.5.

> *Let $n \in \mathbb{Z}_{\geq 0}$. Let Ω be a set such that $|\Omega| = n$. Then its power set has size $|2^{\Omega}| = 2^n$.*

PROOF:
The proof is done by induction on $n = |\Omega|$.
Base case $n = 0$: $|\Omega| = 0$ means that Ω is empty. Since $|\emptyset| = 0$ and $2^{\emptyset} = \{\emptyset\}$ has size $1 = 2^0$ the base case is proven.
Induction assumption: Let $n \in \mathbb{Z}_{\geq 0}$ such that $|2^Y| = 2^n$ for all sets Y such that $|Y| = n$. $(\star)$
Let $|\Omega| = n + 1$. We need to show that $|2^{\Omega}| = 2^{n+1}$. $(\star\star)$
Let $\omega \in \Omega$ and $X := \Omega \setminus \{\omega\}$. Let $\mathfrak{B} := \{A \uplus \{\omega\} : A \in 2^X\}$. Note that $2^{\Omega} = 2^X \cup \mathfrak{B}$ since 2^X contains all subsets A of Ω such that $\omega \notin A$, and $\mathfrak{B}$ contains all subsets B of Ω such that $\omega \in B$. This characterization of 2^X and $\mathfrak{B}$ also implies that subsets of Ω which are elements of 2^X do not belong to $\mathfrak{B}$ and vice versa, i.e., 2^X and $\mathfrak{B}$ are mutually disjoint, thus $2^{\Omega} = 2^X \uplus \mathfrak{B}$.
We obtain from lemma 7.1 on p.217 that there is a bijection $2^X \xrightarrow{\sim} \mathfrak{B}$, hence $|\mathfrak{B}| = |2^X|$.
It follows from $\Omega = X \uplus \{\omega\}$ and $|\{\omega\}| = 1$ and prop.7.4 on p.216 that

$$n + 1 \;=\; |\Omega| \;=\; |X| + |\{\omega\}| \;=\; |X| + 1\,, \quad \text{i.e.,}\ |X| = n\,.$$

The induction assumption $(\star)$ thus applies to the set X, and it yields $|2^X| = 2^n$. We apply once more prop.7.4 to $2^{\Omega} = 2^X \uplus \mathfrak{B}$ and obtain

$$|2^{\Omega}| \;=\; |2^X \uplus \mathfrak{B}| \;=\; |2^X| + |\mathfrak{B}| \;=\; 2 \cdot 2^n \;=\; 2^{n+1}.$$

We have shown $(\star\star)$, and the proof by induction is completed. ■

7.2 The Subsets of $\mathbb{N}$ and Their Size

Proposition 7.6. *Let $\emptyset \neq A \subseteq \mathbb{N}$ and let $A_j \subseteq A$ and $a_j \in A$ $(j \in \mathbb{N})$ be recursively defined as follows.*

$$A_1 := A, \qquad a_1 := \min(A_1); \tag{7.1}$$

$$A_{n+1} := A \setminus \{a_j : j \in \mathbb{N}, j \leq n\}; \qquad a_{n+1} := \begin{cases} \min(A_{n+1}) & \textit{if } A_{n+1} \neq \emptyset, \\ a_n & \textit{else.} \end{cases} \tag{7.2}$$

The following is true for all $i, j, n \in \mathbb{N}$.

- ***(a)*** *The sequence of sets $A_1, A_2, A_3 \ldots$ is nonincreasing: if $i < j$ then $A_i \supseteq A_j$.*
- ***(b)*** *If $j < n$ and $A_n \neq \emptyset$ then $a_j < a_n$.*
- ***(c)*** *If $A_n \neq \emptyset$ then $a_n \geq n$.*
- ***(d)*** *Let $n \geq 2$. If $a \in A$ and $a < a_n$ then $a = a_j$ for some $j < n$.*
- ***(e)*** *Let $n \in \mathbb{N}$. There is no $a \in A$ such that $a_n < a < a_{n+1}$.*

(f) *If $A_n = \emptyset$ for some $n \in \mathbb{N}$ then A is bounded. Let $K := \max\{j \in \mathbb{N} : A_j \neq \emptyset\}$. Then $\max(A) = a_K$. Figure 7.1 illustrates this for the case $K = 4$. Moreover,*

$$A \;=\; \{a_j : j \in \mathbb{N}, j \leq K\} \;=\; \{\min(A_j) : j \in \mathbb{N}, j \leq K\}, \tag{7.3}$$

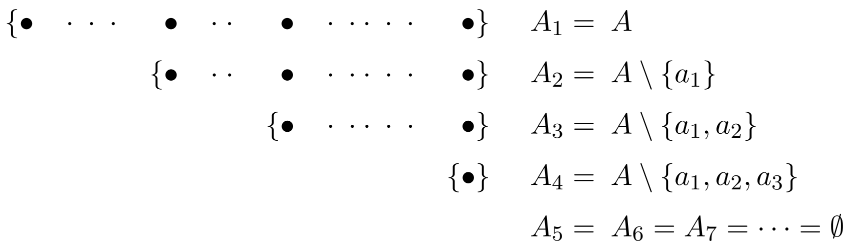
$$\textit{If } n \geq K \;\; \textit{then } a_n = a_K. \tag{7.4}$$

(g) *The sequence $a_j : j \in \mathbb{N}$ is nondecreasing: if $i < j$ then $a_i \leq a_j$.*

(h) *If $A_n \neq \emptyset$ for all $n \in \mathbb{N}$ then A is unbounded and*

$$A \;=\; \{a_j : j \in \mathbb{N}\} \;=\; \{\min(A_j) : j \in \mathbb{N}\}.$$

Figure 7.1: $K = 4$: $a_4 = \max(A)$.


$\{\bullet \cdots \bullet \cdot\cdot \bullet \cdots\cdots \bullet\}$ $A_1 = A$
$\{\bullet \cdot\cdot \bullet \cdots\cdots \bullet\}$ $A_2 = A \setminus \{a_1\}$
$\{\bullet \cdots\cdots \bullet\}$ $A_3 = A \setminus \{a_1, a_2\}$
$\{\bullet\}$ $A_4 = A \setminus \{a_1, a_2, a_3\}$
$A_5 = A_6 = A_7 = \cdots = \emptyset$

PROOF of **(a)**: Let $i < j$. Then $\{a_k : k \in \mathbb{N}, k < i\} \;\subseteq\; \{a_k : k \in \mathbb{N}, k < j\}$, hence $A \;\setminus\; \{a_k : k \in \mathbb{N}, k < i\} \;\supseteq\; A \;\setminus\; \{a_k : k \in \mathbb{N}, k < j\}$, i.e., $A_i \supseteq A_j$.

PROOF of **(b)**: A_n is not empty and $j < n$, hence $A_j \supseteq A_n$, hence $A_j \neq \emptyset$. Thus $a_j = \min(A_j)$ and $a_n = \min(A_n)$. Since $a_j \in \{a_i : i \in \mathbb{N}, i \leq j\}$, it follows that $\min(A_j) = a_j \notin A_n$. We obtain from prop.6.24**(b)** on p.195 that $\min(A_j) < \min(A_n)$, i.e., $a_j < a_n$.

PROOF of **(c)**: This is a simple proof by induction on n.

Base case $n = 1$: $A_1 = A \subseteq \mathbb{N}$, hence A_1 is not empty, hence $a_1 = \min(A_1) \geq \min(\mathbb{N}) = 1$ by prop.6.24**(a)**. This proves the base case.

Induction assumption: Assume that $A_n \neq \emptyset$ and $a_n \geq n$.

We must show that, if $A_{n+1} \neq \emptyset$, then $a_{n+1} \geq n + 1$. It follows from **(b)** that $a_n + 1 \leq a_{n+1}$ and from the induction assumption that $n + 1 \leq a_n + 1$. Thus $n + 1 \leq a_n + 1 \leq a_{n+1}$.

PROOF of **(d)**: We prove this by induction on n.

Base case $n = 2$: $a \in A = A_1$ and $a < a_2 = \min(A_2)$, hence $a \notin A_2$, hence $a \in A_1 \setminus A_2 = \{a_1\}$, hence $a = a_1$. This proves the base case.

Induction assumption: If $n > 2$ and $a \in A$ and $a < a_n$ then $a = a_j$ for some $j < n$.

We must show that if $a \in A$ and $a < a_{n+1}$ then $a = a_j$ for some $j < n + 1$. There are three cases.

Case 1: $a = a_n$. We may choose $j = n$ and we are done.

Case 2: $a < a_n$. The induction assumption implies that there is some $j < n < n + 1$ such that $a = a_j$ and we are done.

Case 3: $a_n < a < a_{n+1}$. We will show that this is not possible. Note that $a_n < a_{n+1}$ implies that

A_{n+1} is not empty. Thus part **(b)** implies that $a_j < a_n < a$ for all $j \in [1, n[_{\mathbb{Z}}$, hence $a \neq a_i$ for all $i \in [1, n]_{\mathbb{Z}}$, hence $a \in A \setminus \{a_i : i \in \mathbb{N}, i \leq n\}$, i.e., $a \in A_{n+1}$. It follows that $a \geq a_{n+1}$, and we have reached a contradiction to the assumption $a_n < a < a_{n+1}$. This finishes the proof of **(d)**.

PROOF of **(e)**: If $a_n = a_{n+1}$ this is obviously true. If $a_n < a_{n+1}$ it follows from **(d)** that if $a \in A$ and $a < a_{n+1}$ then $a \leq a_n$. This proves **(e)**.

PROOF of **(f)**: Assume that $A_n = \emptyset$. It follows from **(a)** that A_j is empty for all $j > n$, hence n is an upper bound for the set $J = \{i \in \mathbb{N} : A_i \neq \emptyset\}$. It follows from thm.6.9 (Generalization of the Well-Ordering Principle) on p.194 that J has a maximum, which we denote by K. Note that it follows from $A_K \neq \emptyset$ and $A_{K+1} = \emptyset$ and (7.2) that $A \setminus \{a_j : j \in \mathbb{N}, j \leq K\} = A_{K+1} = \emptyset$. This proves 7.3 and also that $a_n = a_K = \min(A_K)$ for all $n \geq K$.

PROOF of **(g)**: We examine three cases separately. Let K be as defined in part **(f)**. If $i < j \leq K$ then $a_i < a_j$, according to **(b)**. If $K \leq i < j$ then $a_i = a_j$, according to (7.4). If $i < K < j$ then $a_i < a_K = a_j$, according to **(b)** and (7.4).

PROOF of **(h)**: Let $n \in \mathbb{N}$. From $A_n \neq \emptyset$ we obtain $a_n = \min(A_n)$. Also, $n \leq a_n$ by **(c)**.

$$\text{Thus } A_n \subseteq [a_n, \infty[_{\mathbb{Z}} \subseteq [n, \infty[_{\mathbb{Z}}, \quad \text{thus } \bigcap_{n\in\mathbb{N}} A_n \subseteq \bigcap_{n\in\mathbb{N}} [n, \infty[_{\mathbb{Z}} = \emptyset\,.$$

Let $B := A \setminus \{a_j : j \in \mathbb{N}\}$ By definition, $A_n = A \setminus \{a_j : j \in \mathbb{N}, j \leq n-1\}$.

Thus $B \subseteq A_n$ for all n, thus $B \subseteq \bigcap\limits_{n\in\mathbb{N}} A_n$, thus $B = \emptyset$, thus $A = \{a_j : j \in \mathbb{N}\}$. This proves **(h)**. ■

Proposition 7.7. *Let A be a nonempty subset of $\mathbb{N}$. Let $A_j \subseteq A$ and $a_j \in A$ $(j \in \mathbb{N})$ be defined as in prop. 7.6 on p.217. Then*

- *either $A_n \neq \emptyset$ for all $n \in \mathbb{N}$. In this case A is not bounded and there exists a bijection $\mathbb{N} \xrightarrow{\sim} A$. Further $A = \{a_n : n \in \mathbb{N}\}$*
- *or A_n is empty for some $n \in \mathbb{N}$. In this case A is bounded and there exists a bijection $[1, K]_{\mathbb{Z}} \xrightarrow{\sim} A$ for some suitable $K \in \mathbb{N}$. Further $A = \{a_n : n \in \mathbb{N}$ such that $1 \leq n \leq K\}$*

In both cases the integers a_n and a_{n+1} are adjacent for each index n in the sense that there is no $a \in A$ such that $a_n < a < a_{n+1}$.

PROOF: It is immediate that either **(a)** is true or **(b)** is true, since, either $A_n \neq \emptyset$ for all n or A_n is empty for some n. But we still must prove, e.g., that $A_n \neq \emptyset$ for all $n \in \mathbb{N}$ implies that A is not bounded and one can biject $\mathbb{N}$ to A.

PROOF of the statements in **(a)**: Let $F : \mathbb{N} \longrightarrow A$, defined by $F(n) = a_n$. It follows from prop.7.6**(c)** that no $n \in \mathbb{N}$ is an upper bound of A because $a_{n+1} \geq n+1 > n$. This proves that A is unbounded.

Let $n \in A$. We just saw that $a_{n+1} \geq n+1 > n$. It follows from prop.7.6**(d)** that there exists $j \in \mathbb{N}$ such that $F(j) = a_j = n$. This proves that F is surjective. Injectivity of F follows from prop.7.6**(b)**.

Since $F(n) = a_n$ for all $n \in \mathbb{N}$ we obtain $A = F(\mathbb{N}) = \{F(n) : n \in \mathbb{N}\} = \{a_n : n \in \mathbb{N}\}$. This proves **(a)**.

PROOF of the statements in **(b)**: Let

$$K := \max\{j \in \mathbb{N} : A_j \neq \emptyset\}\,.$$

Let $F : \{j \in \mathbb{N} : j \leq K\} \to A$, defined by $F(n) = a_n$. According to (7.3), $A = \{F(j) : j \in \mathbb{N}, j \leq K\}$, hence F is surjective. It follows from $A_K \neq \emptyset$ and prop.7.6**(b)** that

$$F(i) \ < \ F(j) \quad \text{whenever} \ \ 1 \ \leq \ i \ < \ j \ \leq K \,, \quad \text{hence, (i) } F \text{ is injective, and (ii), } F(K) = \max(A)\,.$$

Thus F is bijective, and bounded by 1 and $F(K)$. We have shown **(b)**.

Finally, the adjacency of a_n and a_{n+1} follows for both **(a)** and **(b)** from prop.7.6**(e)** ■

We extend the results of the last proposition to subsets of integers which are bounded below.

Proposition 7.8. *Let J be a nonempty set of integers which is bounded below.Then*

- ***(a)*** *If J is bounded above then there exists $K \in \mathbb{N}$ and integers n_j $(1 \leq j \leq K)$ such that $J = \{n_j : 1 \leq j \leq K\}$.*
- ***(b)*** *If J is not bounded above then there exist integers n_j $(j \in \mathbb{N})$ such that $J = \{n_j : j \in \mathbb{N}\}$.*
- ***(c)*** *In both cases **(a)** and **(b)** the integers n_j satisfy $i < j \Rightarrow n_i < n_j$, and n_j and n_{j+1} are adjacent for each index j: There is no $n \in J$ such that $n_j < n < n_{j+1}$.*

PROOF: If $\min(J) \geq 1$, i.e., $J \subseteq \mathbb{N}$ then the above is a direct consequence of prop.7.7, so we may assume that $\min(J) \leq 0$. The function $n \mapsto n + 1 - \min(J)$ is a bijection $\varphi : J \xrightarrow{\sim} \varphi(J)$ since it has the function $m \mapsto m - 1 + \min(J)$ as its inverse. Further $\varphi(J) \subseteq \mathbb{N}$ since $n \in J \ \Rightarrow n \geq \min(J)$ we obtain $\varphi(n) \geq \varphi\big(\min(J)\big) = 1$ for all $n \in J$ and thus $\varphi(J) \subseteq \mathbb{N}$.

We apply prop.7.7 to $\varphi(J)$ and obtain

$$\text{either} \ \ \varphi(J) \ = \ \{m_j : 1 \ \leq j \ \leq K\} \ (\text{case } \textbf{(a)}) \quad \text{or} \ \ \varphi(J) \ = \ \{m_j : j \in \mathbb{N}\} \ (\text{case } \textbf{(b)})$$

for suitable $m_j \in \varphi(J)$ and $K \in \mathbb{N}$ which satisfy the properties stated in **(2)**. We look at the inverse images $n_j := \varphi^{-1}(m_j)$ and obtain $J = \varphi^{-1}\big(\varphi(J)\big)$. Thus $J = \{n_j : 1 \leq j \leq K\}$ (case **(a)**) or $J = \{n_j : j \in \mathbb{N}\}$ (case **(b)**). We have proven parts **(a)** and **(b)** of the proposition.

We still need to prove **(c)**. Since the function φ^{-1} shifts its arguments by a constant number to the left, it follows that $i < j \Rightarrow n_i < n_j$. Finally, assume to the contrary that there exists some index j and $n \in J$ such that $n_j < n < n_{j+1}$. It follows that the φ–images satisfy $\varphi(n_j) < \varphi(n) < \varphi(n_{j+1})$, i.e., $m_j < \varphi(n) < m_{j+1}$. But $\varphi(n) \in \varphi(J)$, and this contradicts the adjacency of the m_j in $\varphi(J)$. ■

Notation 7.1 (Notation Alert for bounded below subsets of the integers)**.**

If J is a nonempty subset of the integers which is bounded below then the last proposition makes it natural to introduce the following notation:

- **(a)** If J is finite, i.e., bounded above and hence of the form $J = \{n_j : 1 \leq j \leq K\}$ then we also say that J consists of the numbers $n_1 < n_2 < ... < n_K$.
- **(b)** If J is infinite, i.e., not bounded above and hence of the form $J = \{n_j : j \in \mathbb{N}\}$ then we also say that J consists of the numbers $n_1 < n_2 < ....$ □

Proposition 7.9. *Let A be a nonempty, finite subset of $\mathbb{N}$. Then A is bounded.*

The proof is left as exercise 7.3 (see p.230). ■

We will see later that the reverse also is true: Bounded subsets of $\mathbb{N}$ are finite. But first we must prove

Proposition 7.10. *Let $B \subseteq A \subseteq \mathbb{N}$ and assume that A is finite. Then B is finite.*

The proof is left as exercise 7.4 (see p.230). ■

Theorem 7.1.

Let A be a nonempty subset of the natural numbers. Then

(a) *A is finite if and only if A is bounded,*

(b) *A is countably infinite if and only if A is not bounded.*

(c) *All subsets of $\mathbb{N}$ are countable.*

PROOF of **(a)**: It follows from prop.7.9 above that nonempty, finite subsets of $\mathbb{N}$ are bounded. We now prove the reverse. If A is bounded then $\max(A)$ exists according to the extended well-ordering principle (thm.6.9 on p.194). A is a subset of the finite set $[\max(A)]$, hence A is finite according to the previous proposition. This proves **(a)**.

PROOF of **(b)**: First assume that A is countably infinite. It follows from **(a)** and prop.6.25 ($\mathbb{N}$ is unbounded in $\mathbb{Z}$) on p.195 that $\mathbb{N}$ is infinite, hence A is infinite according to prop.7.3**(a)** on p.215. We employ once more the already proven part **(a)** to conclude that A is not bounded.

Now assume that A is not bounded. Then A does not satisfy prop.7.7**(b)** above, hence A satifies prop.7.7**(a)**. Thus there exists a bijection $\mathbb{N} \xrightarrow{\sim} A$, i.e., A is countably infinite. This proves **(b)**.

PROOF of **(c)**: Either A is bounded or A is not bounded. In the first case it follows from **(a)** that A is finite, hence countable. In the second case it follows from **(b)** that A is countably infinite, hence countable. This proves **(c)**. ■

Theorem 7.2.

(a) *Let X be a finite set and $A \subseteq X$. Then A is finite.*

(b) *Let $X_1, X_2, \dots, X_n$ be finite sets. Then $\bigcup\limits_{j=1}^{n} X_j$ is finite.*

PROOF of **(a)**: We may assume that $A \neq \emptyset$ because the empty set is finite. Since X is finite there exists a bijection $\phi : X \to [1,n]_{\mathbb{Z}}$ for some suitable $n \in \mathbb{N}$. Consider $\phi\big|_A : A \to \phi(A)$, i.e., the restriction of ϕ to A with a codomain which is shrunken from $[1,n]_{\mathbb{Z}}$ to $\phi(A)$.

Then $\phi\big|_A$ is bijective according to prop.5.9**(a)** on p.152. Moreover, since $\phi(A) \subseteq [1,n]_{\mathbb{Z}}$, it follows from prop.7.10 above that $\phi(A)$ is finite. Since A is the bijective image $\left(\phi\big|_A\right)^{-1}\left(\phi(A)\right)$ of the finite set $\phi(A)$, It follows from prop.7.3**(a)** on p. 215 that $|A|$ is finite.

PROOF of **(b)**: We give the proof for $n = 2$. The proof for arbitrary $n \in \mathbb{N}$ is done by induction on n and left as an easy exercise.

Let $X := X_1 \cup X_2$ and $A := X_2 \setminus X_1$. Then $X := X_1 \biguplus A$. Since $A \subseteq X_2$ and X_2 is finite, A is finite according to the already proven part **(a)**. By Proposition 7.4 on p.216, X is finite. ■

We saw in thm.7.1 on p.221 that subsets of the natural numbers are finite if they are bounded and that they are countably infinite otherwise. We had to establish that subsets of finite subsets are finite to extend this theorem to subsets of the integers.

Theorem 7.3.

Let A be a nonempty set of integers. Then
(a) *A is finite if and only if A is bounded,*
(b) *A is countably infinite if and only if A is not bounded.*

PROOF:

Part 1: bounded $\Rightarrow$ finite:

Let $A' := (1 - \min(A)) + A = \{a - \min(A) + 1 : a \in A\}$. Then $A' \subseteq \mathbb{N}$ and is bounded above by $\max(A) - \min(A) + 1$, hence bounded in $\mathbb{N}$, hence finite by thm.7.1 above. Moreover the function $a \mapsto a - \min(A) + 1$ is a bijection $A \xrightarrow{\sim} A'$, hence A is finite by prop.7.3**(a)** on p.215.

Part 2: not bounded above $\Rightarrow$ infinite:

Let $A' := A \cap \mathbb{N}$. If A has no upper bounds then neither does A'. It follows from thm.7.1 on p.221 that A' is not finite, and from prop.7.10 that A is not finite.

Part 3: not bounded below $\Rightarrow$ infinite:

Let $A' := -A = \{-m : m \in A\}$. The function $\varphi : x \mapsto -x$ is a bijection $A \xrightarrow{\sim} A'$ because it has $y \mapsto -y$ as an inverse. It follows from (3.36) on p.80 that φ maps lower bounds of A to upper bounds of A', thus A' is not bounded above.

We have proven in part 2 that A' is not finite, hence its bijective image $A = \varphi^{-1}(A')$ is not finite. We are done with the proof of part 3. ■

Remark 7.2. It follows from the above proposition that subsets of the integers are finite if and only if bounded and infinite otherwise. We will see in ch.7.4 (Countable Sets) that we also can extend thm.7.1**(c)** to the integers: All subsets of $\mathbb{Z}$ are countable. □

7.3 Finite Sequences and Subsequences and Eventually True Properties

Definition 5.22 (p.158) of ch.5 (Relations, Functions and Families) gave the exact definition of sequences and subsequences, more precisely, only of infinite sequences and subsequences: We assumed there that the index set of a sequence $(x_n)_n$ was of the form $[n_\star, \infty[$ for some $n_\star \in \mathbb{Z}$, i.e., we assumed that the index set was not bounded above and hence infinite. Now that we understand the structure of the subsets of $\mathbb{Z}$ which are bounded below we are ready to study finite sequences and finite subsequences.

Definition 7.2 (Finite sequences)**.** Let $n_\star, n^\star \in \mathbb{Z}$ such that $n_\star \leq n^\star$, let $J := [n_\star, n^\star]_{\mathbb{Z}}$. Then J is a finite set of integers since it is bounded below by $n_\star$ and above by $n^\star$. Let X be an arbitrary nonempty set. We call an indexed family $(x_n)_{n \in J}$ in X with index set J a **finite sequence**. We write

$$(x_n)_{n_\star \leq n \leq n^\star} \quad \text{or} \quad (x_n)_{n=n_\star}^{n^\star} \quad \text{or} \quad x_{n_\star}, x_{n_\star+1}, \dots, x_{n^\star-1}, x_{n^\star} \quad \text{or} \quad (x_{n_\star}, x_{n_\star+1}, \dots, x_{n^\star-1}, x_{n^\star})$$

for such a finite sequence. We will sometimes call a sequence $(y_n)_{n=n_\star}^{\infty}$ an **infinite sequence** if we want to stress that its set of indices $[n_\star, \infty[$ is infinite.

If all members x_j of the finite sequence are (real) numbers then we also talk about a **vector** [96] of dimension $\big|[n_\star, n^\star]_{\mathbb{Z}}\big| = n^\star - n_\star + 1$. In this case we always must surround the members of that

[96] Vectors can be of a more general nature than just being a finite sequence of numbers. See ch.11.2 (General Vector Spaces) on p.328 (General Vector Spaces).

finite sequence with parentheses, and we will often use a symbol with "arrow notation"

$$\vec{x} \ = \ \big(x_1, x_2, x_3, \dots, x_{n-1}, x_n\big) \tag{7.5}$$

to refer to such a vector. □

Example 7.2. Here are some examples of finite sequences.

(a) $\big(3.5, -97\pi, 4, \sqrt{8}\big)$ is both a finite sequence and a vector of dimension 4. We have $x_1 = 3.5,\ x_2 - 97\pi,\ x_3 = 4,\ x_4 = \sqrt{8}$.

(b) $\big(3k+2\big)_{k=-2}^{3} \ = \ -4, -1, 2, 5, 8, 11 \ = \ \big(-4, -1, 2, 5, 8, 11\big)$ is both a finite sequence and a vector of size/dimension 6.

(c) Joe, 5, -6.8, Dolores is a finite sequence of size 4. This is not a vector because not all of its members are numeric. □

Definition 7.3 (Finite subsequences)**.** Assume that either $J := [n_\star, \infty[_{\mathbb{Z}}$ or $J := [n_\star, n^\star]_{\mathbb{Z}}$ ($n_\star, n^\star \in \mathbb{Z}$ and $n_\star \leq n^\star$). Let $(n_j)_{j=1}^K (K \in \mathbb{N})$ be a finite sequence of integers $n_j \in J$ such that $i < j \Rightarrow n_i < n_j$ for all $i, j \in \mathbb{N}$. Note that if $J = [n_\star, \infty[_{\mathbb{Z}}$ then $n_j \in J$ for all j implies $n_\star \leq n_1 < n_2 < \dots < n_K$, and if $J = [n_\star, n^\star]_{\mathbb{Z}}$ then this implies $n_\star \leq n_1 < n_2 < \dots < n_K \leq n^\star$.

Let $(x_n)_{n \in J}$ be a sequence in a nonempty set X. We call $(x_{n_j})_{j=1}^K$ a **finite subsequence** of the original sequence since its index set $\{n_j : 1 \leq j \leq K\}$ is finite and we obtain $(x_{n_j})_{j=1}^K$ from $(x_n)_{n \in J}$ by omitting all members x_n for which there is no n_j which equals n. □

Example 7.3. Let $y_k := 2k + 10$. Then $(y_k)_{k=-3}^{2}$ is the finite sequence $4, 6, 8, 10, 12, 14$. It is a finite subsequence not only of the finite sequences $(y_i)_{i=-10}^{10}$ and $(y_i)_{i=-5}^{2}$, but also of the (infinite) sequences $(y_m)_{m \geq -10}$ and $(y_j)_{j=-3}^{\infty}$. □

Definition 7.4. Let X be a nonempty set, $n_\star \in \mathbb{Z}$, $J := \{k \in \mathbb{Z} : k \geq n_\star\}$, and let $(x_n)_{n=n_\star}^{\infty}$ be a sequence in X. If the set of indices $n \in J$ for which a certain property does not hold is empty or bounded then we say that the sequence $(x_n)_n$ satisfies this property **eventually** or that it satisfies this property for **eventually all indices** n. [97] □

Proposition 7.11.

We have the following equivalent ways to state that a sequence (x_n) satisfies a property P eventually:

(a) *There is $K \in J$ such that if P is false for some x_j then $j \leq K$.*

(b) *There is $K \in J$ such that P is true for all x_j such that $j > K$.*

(c) *The set of all indices j such that P is false for x_j is finite.*

PROOF: Let $F := \{j \in J : P$ is false for $x_j\}$ and $T := \{j \in J : P$ is true for $x_j\}$ Then Definition 7.4 states that $(x_n)_{n=n_\star}^{\infty}$ satisfies P eventually if and only if F is empty or bounded. If F is empty then $(x_n)_{n=n_\star}^{\infty}$ satisfies P eventually and each statement **(a)**, **(b)**, **(c)** is true, so the proposition is proven.

[97] You will also find in the mathematical literature the notation "for **almost all indices** n. We prefer not to use this notation in this context because "almost all" is of central importance in measure and probability theory, and it means something very different there.

We thus may assume that F is not empty. Then $(x_n)_{n=n_\star}^\infty$ satisfies P eventually if and only if F is bounded. Since F is always bounded below (by $n_\star$) we conclude that

$$
\begin{aligned}
&(x_n)_{n=n_\star}^\infty \text{ satisfies } P \text{ eventually } \Leftrightarrow\ F \text{ is bounded above}\\
\Leftrightarrow\ &\text{there exists } K \geq n_\star \text{ such that } \big[j > K \Rightarrow j \notin F\big]\\
\Leftrightarrow\ &\text{there exists } K \geq n_\star \text{ such that } \big[j \in F \Rightarrow j \leq K\big]\ \Leftrightarrow\ \textbf{(a)}.
\end{aligned}
$$

This proves the equivalence of Definition 7.4 and **(a)**.

Since the opposite of “there is $K \in J$ such that P is false for some x_j” is “there is $K \in J$ such that P is true for all x_j”, and the opposite of “$j \leq K$” is “$j > K$” we conclude that **(a)** ⇔ **(b)**.

By thm.7.3 on p.221 F is bounded (above) if and only if F is finite. Thus

$$
\begin{aligned}
&\textbf{(a)}\ \Leftrightarrow\ \text{there exists } K \geq n_\star \text{ such that } \big[j \in F \Rightarrow j \leq K\big]\\
\Leftrightarrow\ &\text{there exists } K \geq n_\star \text{ such that } F \subseteq [n_\star, K]\\
\Leftrightarrow\ &\text{there exists } K \geq n_\star \text{ such that } F \text{ is bounded}\\
\Leftrightarrow\ &\text{there exists } K \geq n_\star \text{ such that } F \text{ is finite}\ \Leftrightarrow\ \textbf{(c)}.
\end{aligned}
$$

This proves the proposition. ■

7.4 Countable Sets

In the last chapter we studied the sizes of subsets of natural numbers, and we were able to obtain a complete answer in the last theorem (thm.7.1): All subsets of $\mathbb{N}$ are countable, and they are finite if and only if they are bounded.

Most of the results in this chapter are for general sets and their subsets: No assumption is made about their nature. We may not deal with natural numbers or any other kind of numbers. They might, e.g., be sets of functions or sets of sets.

Now that we know from thm.7.1 on p.221 that all subsets of $\mathbb{N}$ are countable we are able to characterize countable sets by means of injective and surjective functions.

Proposition 7.12 (Countability Criterion)**.** *Let $X \neq \emptyset$.*

The following are equivalent:
- ***(a)*** *X is countable.*
- ***(b)*** *There exists an injective function $f : X \to \mathbb{N}$.*
- ***(c)*** *There exists a surjective function $g : \mathbb{N} \to X$.*

Proof: It follows from thm.5.2**(c)** on p.149 that **(b)** and **(c)** are equivalent, hence it suffices to show that **(a)** and **(b)** are equivalent.

PROOF of **(a)** $\Rightarrow$ **(b)**:

Case 1: If X is finite, then there exists $n \in \mathbb{N}$ and bijective $\varphi : X \xrightarrow{\sim} [1, n]_{\mathbb{Z}}$. Let

$$f : X \to \mathbb{N}; \qquad x \mapsto \varphi(x)$$

be the “same” function as φ, except that we enlarge the codomain to $\mathbb{N}$. f inherits injectivity from φ, and we are done.

Case 2: Otherwise, if X is infinite, i.e., countably infinite, there exists a bijective, hence injective, function $f : X \xrightarrow{\sim} \mathbb{N}$. We are done.

PROOF of **(b)** $\Rightarrow$ **(a)**:

We modify f by shrinking its codomain from $\mathbb{N}$ to the range $f(X)$ of f: Let

$$f' : X \to f(X); \qquad x \mapsto f(x).$$

Then f' inherits injectivity from f, and it also is surjective: If $k \in f(X)$ then there is, by definition of the direct image function, some $x_k \in X$ such that $f(x_k) = k$, i.e., $f'(x_k) = f(x_k) = k$. This proves surjectivity, hence bijectivity, of f'.

According to thm.7.1**(c)** on p.221 $f(X)$ is countable as a subset of $\mathbb{N}$ Since $f(X)$ is the codomain of the bijecttion f'. it follows from prop.7.3**(c)** on p.215 that X is countable. ■

Theorem 7.4. *Let X be a countable set and $A \subseteq X$. Then A is countable.*

PROOF: If X is finite then A is finite by Theorem 7.2 on p.221, and we are done. So assume that X is countably infinite. Then there exists a bijection $\phi : \mathbb{N} \to X$. Let $Y := \mathbb{N}$.

Note that the inverse $\phi^{-1} : X \xrightarrow{\sim} Y$ of ϕ is bijective. Let $B := \{\phi^{-1}(a) : a \in A\}$, and let $f' : A \to B$ be defined by $f'(a) = \phi^{-1}(a)$ for all $a \in A$, i.e., f' is the restriction of ϕ^{-1} to A, with its codomain consisting of all function values of arguments in A. Then f' is bijective according to prop.5.9**(a)** on p.152. It follows from prop.7.3 on p. 215 that $|A| = |B|$.

Since $B \subseteq Y$ and $Y = \mathbb{N}$ we can apply thm.7.1 on p.221. It follows that B and hence A is countable. ■

Corollary 7.2.

(a) *subsets of countable sets are either finite or countably infinite.*
(b) *supersets of uncountable sets are uncountable.*
(c) *Supersets of infinite sets are infinite,*

The proof is left as exercise 7.6 (see p.231). ■

Proposition 7.13 (B/G prop.13.11)**.** *Every infinite set contains a proper subset that is countably infinite.*

The proof is left as exercise 7.7 (see p.231). ■

Proposition 7.14 (B/G prop.13.12)**.**

A set is infinite if and only if it contains a proper subset that is countably infinite.

The proof is left as exercise **??** (see p.**??**). ■

The next proposition is a major stepping stone for proving that countable unions of countable sets are countable.

Proposition 7.15 (B/G Cor.13.16, p.122)**.** $\mathbb{N}^2$ *is countable.*

PROOF: $\boxed{\star}$ Let $f : \mathbb{N} \to \mathbb{N}^2;\ k \mapsto (i,j) = f(k)$ be defined recursively as follows.

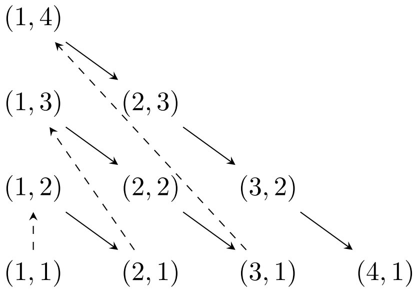
(7.6) $$f(1) \;:=\; (1,1);$$

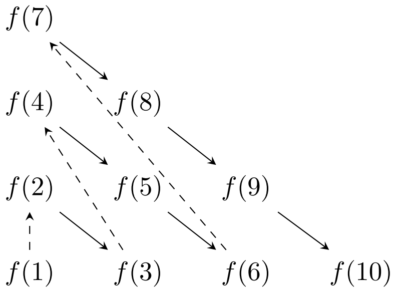
(7.7) $$f(k+1) \;:=\; \begin{cases} (i+1, j-1) \text{ if } f(k) = (i,j) \text{ and } j > 1, & \textbf{(a)} \\ (1,n)) \qquad\quad \text{if } f(k) = (n-1,1). & \textbf{(b)} \end{cases}$$

We will show that f is surjective and injective. For $n \in \mathbb{N}, n \geq 2$, let $D_n := \{(i,j) \in \mathbb{N}^2 : i+j = n\}$. Then The "diagonals" D_n are mutually disjoint, and $\mathbb{N}^2 = \biguplus\limits_{j=2}^{\infty} D_j$.

The following will help to visualize the definition of f. We think of $\mathbb{N}^2$ as a matrix with "infinitely many rows and columns". The left diagram below shows the points of $\mathbb{N}^2$ that belong to the diagonals $D_2, \dots D_5$ in their (x,y)–coordinate form, the one on the right shows them as images of the bijection $f : \mathbb{N} \to \mathbb{N}^2$.

Let $n := i+j$. Equation (7.7)**(a)** specifies that you march southeast on the diagonal D_n if you are not on the bottom row $j = 1$. Equation (7.7)**(b)** specifies that you move from the bottommost point $(n-1,1)$ of D_n to the uppermost point $(1,n)$ of the next diagonal, D_{n+1}.





Assume to the contrary that f is not surjective. Let $A := \{n \in \mathbb{Z}_{\geq 2} : D_n \setminus f(\mathbb{N}) \neq \emptyset\}$, i.e., A contains the indices of all diagonals that have at least one $(i,j) \in \mathbb{N}^2$ that is not a function value. A is not empty because f is not surjective, so it possesses, according to the well–ordering principle, a minimum $n_\star$. Note that $D_2 = \{f(1)\} \subseteq f(\mathbb{N})$, hence $2 \notin A$, hence $n_\star > 2$.

Let $B := \{i \in \mathbb{N} : 1 \leq i < n_\star \text{ and } (i, n_\star - i) \notin f(\mathbb{N})\}$. This is the set of all x–coordinates i of elements of $(i,j) \in D_{n_\star}$ which are not function values of f, and it is not empty because $n_\star \in A$. We apply the well–ordering principle to B and obtain a minimum $1 \leq i_\star < n_\star$.

Case 1: $i_\star \neq 1$. Then $(i_\star - 1, n - i_\star + 1) \in D_{n_\star}$ is a function value $f(k)$ for some $k \in \mathbb{N}$ because $i_\star$ is minimal in B. It follows from (7.7)**(a)** that $f(k+1) = (i_\star, n - i_\star)$. We have reached a contradiction because $i_\star \in B$, hence $(i_\star, n - i_\star)$ is not a function value.

Case 2: $i_\star = 1$. Then $(n_\star - 2, 1) \in D_{n_\star - 1}$ is a function value $f(k)$ for some $k \in \mathbb{N}$ because $n_\star - 1 \notin A$. It follows from (7.7)**(b)** that $f(k+1) = (1, n_\star - 1)$. We have reached a contradiction because $1 \in B$, hence $(1, n_\star - 1)$ is not a function value.

We have proven that f is surjective. We now prove injectivity. Let $k, k' \in \mathbb{N}$ such that $k \neq k'$. We may assume that $k < k'$. Let $i, j, i', j' \in \mathbb{N}$ such that $f(k) = (i,j)$ and $f(k') = (i',j')$. We must prove that $f(k) \neq f(k')$.

It follows from the surjectivity of f and $\mathbb{N}^2 = \biguplus\limits_{j=2}^{\infty} D_j$ that there exists (unique) $n, n' \in \mathbb{N}$ such that $f(k) \in D_n$ and $f(k') \in D_{n'}$. If $n \neq n'$ then $f(k) \neq f(k')$ because $D_n \cap D_{n'}$ and we are done.

So let us assume that $n = n'$. It follows from (7.7)**(a)** that $i' = i + (k' - k)$ and $j' = j - (k' - k)$, thus

$f(k) = (i, j) \neq (i', j') = f(k')$. We have proven injectivity. ■

The proof of prop.7.15 above employs a bijection $f : \mathbb{N} \to \mathbb{N}^2$ which is constructed in a way that can be easily visualized. See the diagrams in the proof of prop.7.15. The drawback is that it was quite complicated to prove that f is in fact bijective. In the following we will construct a different bijection between $\mathbb{N}$ and $\mathbb{N}^2$.

We will first use the uniqueness of prime factorizations to decompose a natural number into a product of factors 2 and an odd number.

Proposition 7.16.

> *Let $n \in \mathbb{N}$. Then*
>
> ***(a)*** *There exist unique $k \in \mathbb{Z}_{\geq 0}$ and $m \in \mathbb{N}$ such that m is odd and $n = 2^k m$.*
>
> ***(b)*** *If $n \neq 1$ then k is the number of times the factor 2 occurs in its prime factorization. Further, either m is the product of all other prime factors, or $m = 1$ if there are no prime factors different from 2.*

PROOF:

If $n = 1$ then the unique factorization $1 = 2^k m$ is obtained with $k = 0$ and $m = 1$. and we have shown both **(a)** and **(b)**.

If $n > 1$ then its unique prime factorization contains zero or more factors of 2. Let k be this number of factors. Let m be the product of all those prime factors of n which are not 2. Then $n = 2^k m$, and m is odd, because otherwise 2 would divide m and hence appear in the prime factorization of m. This proves the existence of the sought after representation, and k and m are precisely as was specified in **(b)**.

Note that we have established on the way that the prime factorization of m does not contain the number 2 as factors, and that the prime factorization of 2^k only contains the number 2.

We now prove uniqueness. Let $k' \in \mathbb{Z}_{\geq 0}$ and $m' \in \mathbb{N}$ such that m' is odd, and such that $n = 2^{k'} m'$. Because m' is odd, its prime factorization does not contain the number 2, and that of $2^{k'}$ only contains the number 2 as factors.

It follows that both m and m' contain exactly the prime factors of n which are not 2, and that both 2^k and $2^{k'}$ contain exactly those which equal 2.

We obtain that $m = m'$ and $k = k'$, and we have established uniqueness. ■

Lemma 7.2. ***Lemma:*** *Let $n \in \mathbb{N}$. Then there exist unique $i, j \in [0, \infty[_{\mathbb{Z}}$ such that $n = 2^i\,(2j + 1)$.*

PROOF: According to prop.7.16 on p.227, there exists a unique pair $(i, m) \in [0, \infty[_{\mathbb{Z}} \times \mathbb{N}$ such that m is odd and

$$n = 2^i \cdot m \tag{7.8}$$

Moreover, it follows from Proposition 6.26 (B/G prop.6.15) on p.196 that there exists $j \in \mathbb{Z}$ such that $m = 2j + 1$. This integer j is unique for the following reason. Let $j' \in \mathbb{Z}$ such that $m = 2j' + 1$. Then $0 = m - m = 2(j - j')$, thus $j = j'$ because there are no zero divisors in $\mathbb{Z}$.

The proof of the lemma is complete if we can prove that $j \geq 0$ and thus $j \in [0, \infty[_{\mathbb{Z}}$. But this is true since $m \in \mathbb{N}$, thus $m \geq 1$, thus $2j + 1 \geq 1$, thus $j \geq 0$. ■

Proposition 7.17.

(a) *The function* $G : \big([0,\infty[_{\mathbb{Z}}\big)^2 \to \mathbb{N};$ $(i,j) \mapsto 2^i\,(2j+1)$ *is a bijection.*

(b) *The function* $F : \mathbb{N}^2 \to \mathbb{N};$ $(i,j) \mapsto 2^{i-1}\,(2j-1)$ *is a bijection.*

PROOF of **(a)**: Note that if $i,j \in [0,\infty[_{\mathbb{Z}}$ then the integers 2^i and $2j+1$ both are positive, hence $2^i(2j+1) \in \mathbb{N}$, hence the assignment $(i,j) \mapsto 2^i(2j+1)$ indeed defines a function with domain $\big([0,\infty[_{\mathbb{Z}}\big)^2$ and codomain $\mathbb{N}$. We must show that g is both injective and surjective.

According to the previous lemma any $n \in \mathbb{N}$ can be written as $n = 2^i\,(2j+1)$ for suitable $i,j \in [0,\infty[_{\mathbb{Z}}$. This proves surjectivity of G.

That lemma also showed that those numbers i and j are unique, thus G is injective.

PROOF of **(b)**: This is an immediate consequence of **(a)** since, if we denote the "even factor" of n

$$\begin{aligned} k = 2^{i-1} \text{ for some } i \in \mathbb{N} \quad &\Leftrightarrow \quad k = 2^i \text{ for some } i \in [0,\infty[_{\mathbb{Z}}\,, \\ m = 2j-1 \text{ for some } j \in \mathbb{N} \quad &\Leftrightarrow \quad m = 2j+1 \text{ for some } j \in [0,\infty[_{\mathbb{Z}}\,. \end{aligned}$$

■

Theorem 7.5 (B/G prop.13.19: Countable unions of countable sets)**.**

The union of countably many countable sets is countable.

PROOF: Let the sets $A_1,\ A_2,\ A_3,\ \dots$ be countable and let $A := \bigcup_{k\in\mathbb{N}} A_k$. We may assume that at least one of those sets A_k is not empty: otherwise their union is empty, hence finite, hence countable, and we are done.

As each of those A_i which is not empty is countable, either A_i 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 will write $a_{(i,j)}$ for $a_i(j)$

We now define the function $f : A \to \mathbb{N}^2,\ \ a \mapsto (i_a, j_a)$ as follows: For each $a \in A$ let $I_a := \{i \in \mathbb{N} : a \in A_i\}$. Since $A := \bigcup_{k\in\mathbb{N}} A_k$, $I_a \neq \emptyset$ and hence has a minimum i_a. Since $a \in A_{i_a}$ and since sets do not contain duplicates of their elements, there is a unique index j_a such that $a = a_{(i_a,j_a)}$.

In other words, we have assigned to each $a \in A$ a unique pair $(i_a, j_a) \in \mathbb{N}^2$ such that $a = a_{(i_a,j_a)}$. This assignment $a \mapsto (i_a, j_a)$ defines a function $f : A \to \mathbb{N}^2$.

If $a, a' \in A$ such that $f(a) = f(a') = (i_a, j_a)$ then both a and a' occupy the same slot j_a in the same set A_{i_a}, hence $a = a'$, thus f is injective. We shrink the codomain of f from $\mathbb{N}^2$ to $f(A)$ and the assignment $a \mapsto (i_a, j_a)$ gives us a bijective function $F : A \xrightarrow{\sim} f(A)$.

$f(A)$ is a subset of the countable set $\mathbb{N}^2$. This proves the theorem because any subset of a countable set is countable (see prop.7.4 on p.225). ■

The following is an easy consequence of the above theorem.

Corollary 7.3. *Let the set X be uncountable and let $A \subseteq X$ be countable. Then the complement $A^{\complement}$ of A is uncountable.*

The proof is left as exercise 7.5 (see p.230). ■

Here are two more corollaries to thm.7.5. Note that the one about the countability of $\mathbb{Z}$ also follows from Theorem 7.3 on p.221.

Corollary 7.4. *The set $\mathbb{Z}$ of all integers is countable.*

PROOF: The set $-\mathbb{N}$ is countable because the function $n \mapsto -n$ is a bijection $\mathbb{N} \xrightarrow{\sim} -\mathbb{N}$, hence

$$Z = \mathbb{N} \cup (-\mathbb{N}) \cup \{0\}$$

is countable as the union of three countable sets. ■

Corollary 7.5. *The rational numbers are countable.*

PROOF: Let $n \in \mathbb{N}$ and $Q_n := \{\frac{m}{n} : m \in \mathbb{Z}\}$. Then $f_n : Q_n \to \mathbb{Z}$, $\frac{m}{n} \mapsto m$ is a bijection because it has as an inverse the function $m \mapsto \frac{m}{n}$. It follows from cor.7.4 that Q_n is countable. By thm.7.5, $\mathbb{Q} = \bigcup \left[Q_n : n \in \mathbb{N}\right]$ is countable as the union of countably many sets. ■

We saw in prop.7.15 that the cartesian product of the two countable factors $\mathbb{N}$ also is countable. The next theorem generalizes this considerably.

Theorem 7.6 (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.

The proof is a triviality for $k = 1$. It is more instructive to choose $k = 2$ for the base case 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.

Hence $X_1 \times X_2 = \bigcup\limits_{x \in X_1} \{x_1\} \times X_2$ is countable according to thm.7.5 on p.228. 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"

$$X_1 \times \cdots \times X_{k+1} = (X_1 \times \cdots \times X_k) \times X_{k+1} \tag{7.9}$$

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 7.6. *Let $n \in \mathbb{N}$. The sets $\mathbb{Q}^n$ and $\mathbb{Z}^n$ are countable.*

PROOF: This follows from the preceding theorem because the sets $\mathbb{Q}$ and $\mathbb{Z}$ are countable. ■

We will examine uncountable sets in ch.10 (Cardinality II: Comparing Uncountable Sets), but we will state a result here concerning a very important example of an uncountable set. The proof of the next theorem is very similar to the proof that the real numbers are uncountable. (See thm.9.12 on p.285.)

Theorem 7.7.

> *Let X be a set which contains at least two elements. Then $X^{\mathbb{N}} = \{(x_n)_{n\in\mathbb{N}} : x_j \in X \ \forall j \in \mathbb{N}\}$ (the set of all sequences with values in X) is uncountable.*

PROOF: Let $a, b \in X$ such that $a \neq b$. We will prove that the subset $A := \{a, b\}^{\mathbb{N}}$ of X is uncountable.

A certainly is not finite since it contains for each $n \in \mathbb{N}$ the sequence $\vec{y}_n = (y^n_j)_{j\in\mathbb{N}}$ which is defined by $y^n_j := a$ if $n \neq j$, $y^n_n := b$. If A were finite then its subset $B := \{\vec{y}_n : n \in \mathbb{N}\}$ also would have to be finite. (See thm.7.4 on p.225.) But B is countably infinite since $n \mapsto \vec{y}_n$ defines a bijection $\mathbb{N} \xrightarrow{\sim} B$. This proves that A is not finite. We are done if we can prove that A also is not countably infinite.

So assume to the contrary that A is countably infinite, i.e., there exist $\vec{x}_1, \vec{x}_2, \dots \in A$ such that $A = \{\vec{x}_n : n \in \mathbb{N}\}$. Note that each $\vec{x}_n$ itself is a sequence $(x^n_j)_{j\in\mathbb{N}}$ in which each member x^n_j is either a or b. We will reach a contradiction by constructing some $\vec{x} \in A$ which is different from $\vec{x}_n$ for each $n \in \mathbb{N}$ since this implies that $\vec{x} \notin A$.

We will obtain such $\vec{x} = (x_j)_{j\in\mathbb{N}}$ by ensuring that each x_j will be different from the diagonal element x^j_j of the infinite grid to the right. Let

$$x_j := \begin{cases} a & \text{if } x^j_j = b, \\ b & \text{otherwise.} \end{cases}$$

$$\begin{array}{c|cccccc} \vec{x_1}: & \colorbox{gray}{x^1_1} & x^1_2 & x^1_3 & x^1_4 & \dots \\ \vec{x_2}: & x^2_1 & \colorbox{gray}{x^2_2} & x^2_3 & x^2_4 & \dots \\ \vec{x_3}: & x^3_1 & 3^2_2 & \colorbox{gray}{x^3_3} & x^3_4 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

Clearly $\vec{x} \in A$ since $A = \{a, b\}^{\mathbb{N}}$ contains any sequence whose members are either a or b. Note that $\vec{x} \neq \vec{x_1}$ since those two sequences differ in their first elements x_1 and x^1_1. Further $\vec{x} \neq \vec{x_2}$ since those two sequences differ in their second elements x_2 and x^2_2. We see that for any $j \in \mathbb{N}$ it is true that $\vec{x} \neq \vec{x_j}$ since those two sequences differ in their j–th elements x_j and x^j_j. It follows from $A = \{\vec{x}_n : n \in \mathbb{N}\}$ that $\vec{x} \notin A$. The assumption that A is countably infinite has allowed us to construct some $\vec{x}$ such that both $\vec{x} \in A$ and $\vec{x} \notin A$. We have reached a contradiction. ■

7.5 Exercises for Ch.7

Exercise 7.1. Prove the following parts of cor.7.1 on p.214 of this document:

- **(a)** If $m < n$ then there exists no surjective function $f : [1, m]_{\mathbb{Z}} \to [1, n]_{\mathbb{Z}}$.
- **(b)** If $m > n$ then there exists no injective function $g : [1, m]_{\mathbb{Z}} \to [1, n]_{\mathbb{Z}}$.
- **(c)** If $m \neq n$ then there exists no bijective function $f : [1, m]_{\mathbb{Z}} \to [1, n]_{\mathbb{Z}}$. □

Exercise 7.2. Prove lemma.7.1 on p.217 of this document: Let X, Ω be sets such that $X \subseteq \Omega$ and $\omega \in X^{\complement}$, and let $\mathfrak{B} := \{A \uplus \{\omega\} : A \in 2^X\}$.

Then the function $F : 2^X \to \mathfrak{B};\ A \mapsto A \uplus \{\omega\}$ is a bijection. □

Exercise 7.3. Prove prop.7.9 on p.220 of this document: Let A be a nonempty, finite subset of $\mathbb{N}$. Then A is bounded. □

Exercise 7.4. Prove prop.7.10 on p.221 of this document: Let $B \subseteq A \subseteq \mathbb{N}$ and assume that A is finite. Then B is finite. □

Exercise 7.5. Prove cor.7.3 on p.228 of this document:

If X is uncountable and $A \subseteq X$ is countable then $A^{\complement}$ is uncountable. □

Exercise 7.6. Prove cor.7.2 on p.225 of this document:

- **(a)** subsets of countable sets are either finite or countably infinite.
- **(b)** supersets of uncountable sets are uncountable.
- **(c)** Supersets of infinite sets are infinite, □

Exercise 7.7. Prove prop.7.13 on p.225 of this document: Every infinite set contains a proper subset that is countably infinite. □

Exercise 7.8. Prove prop.7.14 on p.225 of this document: A set is infinite if and only if it contains a proper subset that is countably infinite. □

8 More on Sets, Relations, Functions and Families

8.1 More on Set Operations

The material in this chapter is a continuation of Chapter 2.6 (Arbitrary Unions and Intersections).

Recall that we had defined unions and intersections of arbitrary collections of sets in Definition 2.29 (Arbitrary unions and intersections) on p.38.

Definition 8.1. 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

$$\bigcup_{i\in\emptyset} A_i \;:=\; \emptyset, \qquad \bigcap_{i\in\emptyset} A_i \;:=\; \Omega. \tag{8.1}$$

Note that this definition is consistent with the fact that

- unions over fewer sets become smaller, so the union over $\emptyset$ should be the smallest set possible, i.e., the empty set,
- intersections over fewer sets become bigger, so the intersection over $\emptyset$ should be the largest set possible, i.e., the universal set. □

We give some more examples of non-finite unions and intersections.

Example 8.1. For any set A we have $A = \bigcup\limits_{a\in A}\{a\}$. According to (8.1) this also is true if $A = \emptyset$. □

The following trivial lemma is useful if you need to prove statements of the form $A \subseteq B$ or $A = B$ for sets A and B. Be sure to understand what it means if you choose $J = \{1, 2\}$ (draw one or two Venn diagrams).

Lemma 8.1 (Inclusion lemma)**.**

Let J be an arbitrary, nonempty index set. Let U, X_j, Y, Z_j, W $(j \in J)$ be sets such that

$$U \subseteq X_j \subseteq Y \subseteq Z_j \subseteq W$$

for all $j \in J$. Then

$$U \subseteq \bigcap_{j\in J} X_j \subseteq Y \subseteq \bigcup_{j\in J} Z_j \subseteq W. \tag{8.2}$$

PROOF: Note that we need at various places in this proof the existence of some $j_0 \in J$, i.e. the assumption that $J \neq \emptyset$ is essential.

(a) Let $x \in U$. Then $x \in X_j$ for all $j \in J$, hence $x \in \bigcap\limits_{j \in J} X_j$. This proves the first inclusion.

(b) Now let $x \in \bigcap\limits_{j \in J} X_j$ and $j_0 \in J$. Then $x \in X_j$ for all $j \in J$; in particular, $x \in X_{j_0}$. It follows from $X_{j_0} \subseteq Y$ that $x \in Y$ and we have shown the second inclusion.

(c) Let $x \in Y$ and $j_0 \in J$. It follows from $Y \subseteq Z_{j_0}$ that $x \in Z_{j_0}$. But then $x \in \{z : z \in Z_j \text{ for some } j \in J\}$, i.e., $x \in \bigcup\limits_{j \in J} Z_j$. This proves the third inclusion.

(d) Finally, assume $x \in \bigcup\limits_{j \in J} Z_j$ It follows from the definitions of unions that there exists $j_0 \in J$ such that $x \in Z_{j_0}$. But then $x \in W$ as W contains Z_{j_0}. It follows that $\bigcup\limits_{j \in J} Z_j \subseteq W$. This finishes the proof of the rightmost inclusion. ■

Definition 8.2 (Disjoint families)**.**

Let J be a nonempty set. We call a family of sets $(A_i)_{i \in J}$ a **mutually disjoint family** if for any two different indices $i, j \in J$ it is true that $A_i \cap A_j = \emptyset$, i.e., if any two sets in that family with different indices are mutually disjoint. □

We recall from Chapter 2.1 (Sets and Basic Set Operations), p.22, Definition 2.11 of a partition: For $\mathfrak{A} \subseteq 2^{\Omega}$, $\mathfrak{A}$ is a partition or a partitioning of Ω if

(a) $A \cap B = \emptyset$ for any two $A, B \in \mathfrak{A}$ such that $A \neq B$, **(b)** $\Omega = \biguplus \Big[A : A \in \mathfrak{A} \Big]$.

We extend this to arbitrary families and hence finite collections and sequences of subsets of Ω:

Definition 8.3 (Partition)**.**

Let J be an arbitrary nonempty set, let $(A_j)_{j \in J}$ be a family of subsets of Ω. We call $(A_j)_{j \in J}$ a **partition** or a **partitioning** of Ω if it is a mutually disjoint family which satisfies $\Omega = \biguplus \Big[A_j : j \in J \Big]$.

In other words,

- $(A_j)_{j \in J}$ is a partition of Ω if and only if $\mathfrak{A} := \{A_j : j \in J\}$ is a partition of Ω.

Note that duplicate nonempty sets cannot occur in a disjoint family of sets because otherwise the condition of mutual disjointness does not hold. □

Example 8.2. 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 nonnegative integers. For $k \in \mathbb{N}$ let $A_k := \{j \in \mathbb{N} : n_{k-1} < j \leq n_k\}$. Then the set $\{A_k : k \in \mathbb{N}\}$ is a partition of $\mathbb{N}$ (**not** of $\mathbb{Z}_{\geq 0}$!) □

Theorem 8.1 (De Morgan's Law)**.** *Let there be a universal set Ω (see (2.8) on p.18). Then the following "duality principle" holds for any indexed family $(A_\alpha)_{\alpha \in I}$ of sets:*

$$\text{(a)}\ \Big(\bigcup_\alpha A_\alpha\Big)^{\complement} = \bigcap_\alpha A_\alpha^{\complement} \qquad \text{(b)}\ \Big(\bigcap_\alpha A_\alpha\Big)^{\complement} = \bigcup_\alpha A_\alpha^{\complement} \tag{8.3}$$

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 unions with intersections and all intersections with unions.

PROOF of De Morgan's law, formula **(a)**:

1) First we prove that $(\bigcup\limits_\alpha A_\alpha)^{\complement} \subseteq \bigcap\limits_\alpha A_\alpha^{\complement}$:

Assume that $x \in (\bigcup\limits_\alpha A_\alpha)^{\complement}$. Then $x \notin \bigcup\limits_\alpha A_\alpha$ which is the same as saying that x does not belong to any of the A_α. That means that x belongs to each $A_\alpha^{\complement}$ and hence also to the intersection $\bigcap\limits_\alpha A_\alpha^{\complement}$.

2) Now we prove that $(\bigcup\limits_\alpha A_\alpha)^{\complement} \supseteq \bigcap\limits_\alpha A_\alpha^{\complement}$:

Let $x \in \bigcap A_\alpha^{\complement}$. Then x belongs to each of the $A_\alpha^{\complement}$ 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 $(\bigcup\limits_\alpha A_\alpha)^{\complement}$. This completes the proof of formula **(a)**.

The proof of formula **(b)** is very similar and given as exercise 8.3 on p.249. ■

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 8.1 (Distributivity of unions and intersections)**.** *Let $(A_i)_{i\in I}$ be an arbitrary family of sets and let B be a set. Then*

$$\bigcup_{i\in I}(B \cap A_i) = B \cap \bigcup_{i \in I} A_i, \tag{8.4}$$

$$\bigcap_{i\in I}(B \cup A_i) = B \cup \bigcap_{i \in I} A_i. \tag{8.5}$$

PROOF: We only prove (8.4). The proof of (8.5) is left as exercise 8.5.

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 $\left(\bigcup_i A_i\right) \cap B$.

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 $x \in \bigcup_i (B \cap A_i)$. ■

Note that the next proposition is about finite unions and can be formulated and proven with what has been taught in chapter 2 (Preliminaries about Sets, Numbers and Functions) on p.13.

Proposition 8.2 (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. Let*

$$\begin{aligned} B_n &:= \bigcup_{j=1}^{n} A_j \;=\; A_1 \cup A_2 \cup \cdots \cup A_n \; (n \in \mathbb{N}),\\ C_1 &:= A_1 \;=\; B_1, \quad C_{n+1} \;:=\; A_{n+1} \setminus B_n \; (n \in \mathbb{N}). \end{aligned}$$

Then,

(a) *The sequence $(B_j)_j$ is increasing: $m < n \Rightarrow B_m \subseteq B_n$.*

(b) *For each $n \in \mathbb{N}$, $\bigcup\limits_{j=1}^{n} A_j \;=\; \bigcup\limits_{j=1}^{n} B_j$.*

(c) *The sets C_j are mutually disjoint and $\bigcup\limits_{j=1}^{n} A_j \;=\; \biguplus\limits_{j=1}^{n} C_j$.*

(d) *The sets C_j $(j \in \mathbb{N})$ form a partition of the set $\bigcup\limits_{j=1}^{\infty} A_j$.*

PROOF of **(a)** and of **(b)**: Left as exercise 8.1 (p.249).

PROOF of **c:** Let $1 \leq j \leq n$. We note that $C_j \subseteq A_j \subseteq B_j \subseteq B_n$ and obtain

$$C_j \cap C_{n+1} \subseteq B_n \cap C_{n+1} \;=\; B_n \cap (A_{n+1} \setminus B_n) \;=\; B_n \cap (A_{n+1} \cap B_n^{\complement}) \;=\; A_{n+1} \cap (B_n \cap B_n^{\complement}) \;=\; \emptyset.$$

We have proved that for any $j, k \in \mathbb{N}$ such that $j < k$ the sets C_j and C_k have empty intersection (we replaced $n+1$ with k) and it follows that the entire sequence of sets C_j is disjoint.

We finally prove that $\bigcup_{j=1}^n A_j \;=\; \bigcup_{j=1}^n B_j \;=\; \biguplus_{j=1}^n C_j$. The first equation follows from **(b)**. To prove the second equation we first show that $\biguplus_{j=1}^n C_j \;\subseteq\; \bigcup_{j=1}^n A_j$. This is immediate from $C_n \subseteq A_n$ for all $n \in \mathbb{N}$.

We finally prove that $\bigcup_{j=1}^n A_j \;\subseteq\; \biguplus_{j=1}^n C_j$. Let $x \in \bigcup_{j=1}^n A_j$. Then $x \in A_j$ for at least one $1 \leq j \leq n$. Let j_0 be the smallest such j. If $j_0 = 1$ then $x \in C_1$ because $C_1 = A_1$, hence $x \in \biguplus_{j=1}^n C_j$ and we are done. Otherwise $x \notin A_j$ for all $1 \leq j < j_0$, hence $x \notin \bigcup_{j=1}^{j_0-1} A_j = B_{j_0-1}$, hence $x \in A_{j_0} \setminus B_{j_0-1}$, i.e., $x \in C_{j_0}$. It follows that $x \in \biguplus_{j=1}^n C_j$.

PROOF of **d:** This is a trivial consequence of **(c)**. ■

8.2 Rings and Algebras of Sets ⋆

Note that this chapter is starred, hence optional.

Definition 8.4 (Rings, algebras, and σ–Algebras of Sets).

A subset $\mathscr{R}$ of 2^Ω (a set of sets!) is called a **ring of sets** if it is closed with respect to the operations "$\cup$" and "$\setminus$", i.e.,

(8.6) $$R_1 \cup R_2 \in \mathscr{R} \text{ and } R_1 \setminus R_2 \in \mathscr{R} \quad \text{whenever } R_1, R_2 \in \mathscr{R}.$$

A subset $\mathscr{A}$ of 2^Ω is called an **algebra of sets** if $\Omega \in \mathscr{A}$ and $\mathscr{A}$ is a ring of sets.
A subset $\mathscr{F}$ of 2^Ω is called a σ**–algebra** if $\mathscr{F}$ is an algebra of sets which satisfies

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \qquad \Rightarrow \qquad \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}$$

σ–algebras are fundamental objects in measure theory and graduate level probability theory. □

Parts **2a** through **2h** of the next proposition have already been encountered in prop.2.4 on p.21 of ch.2.1 (Sets and Basic Set Operations). They have now been tagged with names such as "associativity of $\triangle$" which emphasize the connection to the rings we studied in ch.3 (The Axiomatic Method).

Proposition 8.3.

(1) *Let $\mathscr{R}$ be a ring of sets and $A, B \in \mathscr{R}$. Then $\emptyset \in \mathscr{R}$, $A\triangle B \in \mathscr{R}$, and $A \cap B \in \mathscr{R}$.*
(2) *Let A, B, C, Ω be sets such that $A, B, C \subseteq \Omega$. Then*
(a) $(A\triangle B)\triangle C \;=\; A\triangle(B\triangle C)$ *(associativity of $\triangle$)*
(b) $A\triangle\emptyset \;=\; \emptyset\triangle A \;=\; A$ *(neutral element $\emptyset$ for $\triangle$)*
(c) $A\triangle A \;=\; \emptyset$ *(inverse element $A^{-1} = A$ for $\triangle$)*
(d) $A\triangle B \;=\; B\triangle A$ *(commutativity of $\triangle$)*

Further, we have the following for the intersection operation:

(e) $(A\cap B)\cap C \;=\; A\cap(B\cap C)$ *(associativity of $\cap$)*
(f) $A\cap\Omega \;=\; \Omega\cap A \;=\; A$ *(neutral element Ω for $\cap$)*
(g) $A\cap B \;=\; B\cap A$ *(commutativity of $\cap$)*

And we have the following interrelationship between $\triangle$ and $\cap$:
(h) $A\cap(B\triangle C) \;=\; (A\cap B)\triangle(A\cap C)$ *(distributivity)*

PROOF:

For the proof of **2.a** see the one of prop.2.4. The proofs of the other properties are left as an exercise. ■

Remark 8.1 (Algebras of Sets as Rings)**.**

(1) Prop.8.3**(1)** states that the assignments $(A, B) \mapsto A\triangle B$ and $(A, B) \mapsto A \cap B$ are binary operations on $\mathscr{R}$.
(2) Items **(a) – (d)** of prop.8.3**(2)** assert that $(\mathscr{R}, \triangle)$ is an abelian group with neutral element $\emptyset$ and inverse $A^{-1} = A$.

(3) If $\Omega \in \mathscr{R}$, i.e., $\mathscr{R}$ is an algebra of sets, Items **(e) – (g)** of prop.8.3**(2)** assert that $(\mathscr{R}, \cap)$ is a commutative monoid with unit Ω.

(4) Assume that Ω is not empty. Then the "additive" neutral elenent $\emptyset$ is different from Ω, the "multiplicative" neutral elenent.

(5) **(1) – (4)** plus Proposition 8.3**(2).(h)** imply that, if $\Omega \neq \emptyset$, then $(\mathscr{R}, \triangle, \cap)$ satisfies Definition 3.7 on p.60, i.e., $(\mathscr{R}, \triangle, \cap)$ is a commutative ring with unit.

The above justifies calling $\mathscr{R} = (\mathscr{R}, \triangle, \cap)$ a ring of sets. The name "algebra of sets" for a ring of sets which contains Ω stems from the fact that such systems of subsets of Ω are "boolean algebras". One can view this "algebra of sets = commutative ring with unit" mismatch in names from a different perspective: $(\mathscr{R}, \triangle, \cap)$ is a non–trivial example of what one generally calls a commutative ring in mathematics: an algebraic structure which satisfies all properties of a commutative ring with unit (Definition 3.7 on p.60), except that there need not be a multiplicative unit.

Note that we do not have an integral domain if Ω contains at least two elements ω and ω': Let $A \subseteq \Omega$ such that $\omega \in A$ and $\omega' \in A^{\complement}$. Then $A \neq \emptyset$ and $A^{\complement} \neq \emptyset$ but $A \cap A^{\complement} = \emptyset$, i.e., A and $A^{\complement}$ are a pair of zero divisors in $(\mathscr{R}, \triangle, \cap)$. □

8.3 Cartesian Products of More Than Two Sets

In this chapter we will extend the notion of a Cartesian product to more than two factors. Matter of fact, we will not stop at a finite number of factors and extend that concept to the product of factors X_i where the indices i are the members of an arbitrary index set.

Remark 8.2 (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 8.5 (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 [98]

$$X_1 \times X_2 \times X_3 \times \ldots \times X_N := \{(x_1, x_2, \dots, x_N) : x_j \in X_j \text{ for all } 1 \le j \le N\}$$

Note that the elements of this set are finite sequences in the sense of Definition 7.2 (finite sequences) on p.222.

Two elements $(x_1, x_2, \dots, x_N)$ and $(y_1, y_2, \dots, y_N)$ of $X_1 \times X_2 \times X_3 \times \ldots \times X_N$ are called **equal** if and only if $x_j = y_j$ for all j such that $1 \le j \le N$. 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 8.3 (N–dimensional coordinates)**.** Here is the most important example of a cartesian product of N sets. Let $X_1 = X_2 = \ldots = X_N = \mathbb{R}$. Then

$$\mathbb{R}^N = \{(x_1, x_2, \dots, x_N) : x_j \in \mathbb{R} \text{ for } 1 \le j \le N\}$$

is the set of points in N–dimensional space. You may not be familiar with what those are unless $N = 2$ (see example 5.1 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 note on vectors and coordinates: For $N \le 3$ 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** $(x_1, x_2, \dots, x_N)$ of numbers is called a vector of dimension N. [99] You will read more about this in ch.11 about vectors and vector spaces on page 323.

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 11, p.323 on vectors and vector spaces. □

[98] 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 $N = 3$:

$$X_1 \times (X_2 \times X_3 \times X_4), \quad (X_1 \times X_2) \times (X_3 \times X_4), \quad X_1 \times (X_2 \times X_3 \times X_4),$$

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.

[99] See Definition 7.2 (finite sequences) on p.222.

Example 8.4 (Parallelepipeds). Let $a_1 < b_1, a_2 < b_2, a_3 < b_3$ be real numbers. Then

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \;=\; \{(x,y,z) : a_1 \leq x \leq b_1, a_2 \leq y \leq b_2, a_3 \leq z \leq b_3\}$$

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 $a_j < b_j$ $(j \in \mathbb{N}, j \leq N, a_j, b_j \in \mathbb{R})$. Then

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_N, b_N] \;=\; \{(x_1, x_2, \dots, x_N) : a_j \leq x_j \leq b_j, j \in \mathbb{N}, j \leq N\}$$

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 8.6 (Cartesian Product of a family of sets). $\boxed{\star}$ Let I be an arbitrary, nonempty set (the index set). Let $(X_i)_{i \in I}$ be a family of nonempty sets X_i.

The **cartesian product** of the family $(X_i)_{i \in I}$ is the set

(8.7) $$\prod_{i \in I} X_i \;:=\; \Big(\prod X_i \Big)_{i \in I} \;:=\; \{(x_i)_{i \in I} : x_k \in X_k \; \forall k \in I\}$$

of all familes $(x_i)_{i \in I}$ each of whose members x_j belongs to the corresponding set X_j.
$(x_i)_{i \in I}$, $(y_k)_{k \in I} \in \prod\limits_{i \in I} X_i$ are called **equal** (we write $(x_i)_{i \in I} = (y_k)_{k \in I}$), if $x_j = y_j$ for all $j \in I$.
If all sets X_i are equal to one and the same set X, we also write

(8.8) $$X^I \;:=\; \prod_{i \in I} X \;:=\; \prod_{i \in I} X_i \,.$$

The symbol "$\prod$" is the greek “upper case” letter “Pi” (whose lower case incarnation “π” you are probably more familiar with). □

Remark 8.3. Note that, because I is not empty, $\prod\limits_{i \in I} X_i \; = \emptyset \;\Leftrightarrow\;$ there exists some $i \in I$ such that $X_i = \emptyset$.
Further, two families are equal in the sense of the above definition if and only if they are equal in the sense of Definition 5.21 on p.157. □

Remark 8.4. 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 Definition 5.20 (indexed families) and the subsequent remarks concerining the equivalence of functions and families). In other words,

(8.9) $$Y^X \;=\; \{f : f \text{ is a function with domain } X \text{ and codomain } Y\}. \quad \square$$

8.4 Set Operations involving Direct Images and Preimages

Let X, Y be two nonempty sets and let $f : X \to Y$ be an arbitrary function with domain X and codomain Y. Let $A \subseteq X$ and $B \subseteq Y$. We recall from Definition 5.11 on p.141 that

$$\begin{aligned} f(A) &= \{f(x) : x \in A\} \text{ is the direct image of } A, \\ f^{-1}(B) &= \{x \in X : f(x) \in B\} \text{ is the indirect image or preimage of } B. \end{aligned}$$

We now will examine to which extent direct and indirect images are compatible with unions, intersections, and other basic set operations.

> Unless stated otherwise, X, Y and f are as defined above for the remainder of this chapter: $f : X \to Y$ is a function with domain X and codomain Y.

Proposition 8.4 (f^{-1} is compatible with all basic set ops)**.** *Let J be an arbitrary index set. Let $B \subseteq Y$, $B_j \subseteq Y$ for all j. Then*

$$f^{-1}\Big(\bigcap_{j \in J} B_j\Big) = \bigcap_{j \in J} f^{-1}(B_j) \tag{8.10}$$

$$f^{-1}\Big(\bigcup_{j \in J} B_j\Big) = \bigcup_{j \in J} f^{-1}(B_j) \tag{8.11}$$

$$f^{-1}(B^{\complement}) = \big(f^{-1}(B)\big)^{\complement} \tag{8.12}$$

$$f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2) \tag{8.13}$$

$$f^{-1}(B_1 \Delta B_2) = f^{-1}(B_1) \Delta f^{-1}(B_2) \tag{8.14}$$

PROOF of (8.10): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}\Big(\bigcap_{j \in J} B_j\Big) &\Leftrightarrow f(x) \in \bigcap_{j \in J} B_j \quad \text{(def preimage)} \\ &\Leftrightarrow \forall j\ f(x) \in B_j \quad (\text{def} \cap) \\ &\Leftrightarrow \forall j\ x \in f^{-1}(B_j) \quad \text{(def preimage)} \\ &\Leftrightarrow x \in \bigcap_{j \in J} f^{-1}(B_j) \quad (\text{def} \cap) \end{aligned} \tag{8.15}$$

PROOF of (8.11): Let $x \in X$. Then

$$\begin{aligned} x \in f^{-1}\Big(\bigcup_{j \in J} B_j\Big) &\Leftrightarrow f(x) \in \bigcup_{j \in J} B_j \quad \text{(def preimage)} \\ &\Leftrightarrow \exists j_0 :\ f(x) \in B_{j_0} \quad (\text{def} \cup) \\ &\Leftrightarrow \exists j_0 : x \in f^{-1}(B_{j_0}) \quad \text{(def preimage)} \\ &\Leftrightarrow x \in \bigcup_{j \in J} f^{-1}(B_j) \quad (\text{def} \cup) \end{aligned} \tag{8.16}$$

PROOF of (8.12): Let $x \in X$. Then

$$x \in f^{-1}(B^{\complement}) \Leftrightarrow f(x) \in B^{\complement} \quad \text{(def preimage)}$$
$$\Leftrightarrow f(x) \notin B \quad \text{(def } \complement\text{)}$$
$$\Leftrightarrow x \notin f^{-1}(B) \quad \text{(def preimage)} \tag{8.17}$$
$$\Leftrightarrow x \in f^{-1}(B)^{\complement} \quad \text{(def } \complement\text{)}$$

PROOF of (8.13): Let $x \in X$. Then

$$x \in f^{-1}(B_1 \setminus B_2) \Leftrightarrow x \in f^{-1}(B_1 \cap B_2^{\complement}) \quad \text{(def } \setminus\text{)}$$
$$\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2^{\complement}) \quad \text{(see (8.10))} \tag{8.18}$$
$$\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2)^{\complement} \quad \text{(see (8.12))}$$
$$\Leftrightarrow x \in f^{-1}(B_1) \setminus f^{-1}(B_2) \quad \text{(def } \setminus\text{)}$$

PROOF of (8.14): This follows from $B_1 \Delta B_2 = (B_1 \setminus B_2) \cup (B_2 \setminus B_1)$ and (8.11) and (8.13). ■

Proposition 8.5 (Properties of the direct image)**.**

Let J be an arbitrary index set. Let $A \subseteq X$, $A_j \subseteq X$ for all j. Then

$$f(\bigcap_{j \in J} A_j) \subseteq \bigcap_{j \in J} f(A_j) \tag{8.19}$$

$$f(\bigcup_{j \in J} A_j) = \bigcup_{j \in J} f(A_j) \tag{8.20}$$

PROOF of (8.19): This follows from the monotonicity of the direct image (see 5.15):

$$\bigcap_{j \in J} A_j \subseteq A_i \; \forall i \in J \Rightarrow f(\bigcap_{j \in J} A_j) \subseteq f(A_i) \; \forall i \in J$$
$$\Rightarrow f(\bigcap_{j \in J} A_j) \subseteq \bigcap_{i \in J} f(A_i) \quad \text{(def } \cap\text{)}$$

First proof of (8.20)) - "Expert proof":

$$y \in f(\bigcup_{j \in J} A_j) \Leftrightarrow \exists\, x \in X : \; f(x) = y \text{ and } x \in \bigcup_{j \in J} A_j \quad \text{(def } f(A)\text{)} \tag{8.21}$$
$$\Leftrightarrow \exists\, x \in X \text{ and } j_0 \in J : \; f(x) = y \text{ and } x \in A_{j_0} \quad \text{(def } \cup\text{)} \tag{8.22}$$
$$\Leftrightarrow \exists\, x \in X \text{ and } j_0 \in J : \; f(x) = y \text{ and } f(x) \in f(A_{j_0}) \quad \text{(def } f(A)\text{)} \tag{8.23}$$
$$\Leftrightarrow \exists\, j_0 \in J : \; y \in f(A_{j_0}) \quad \text{(def } f(A)\text{)} \tag{8.24}$$
$$\Leftrightarrow y \in \bigcup_{j \in J} f(A_j) \quad \text{(def } \cup\text{)} \tag{8.25}$$

Alternate proof of (8.20) - 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 "⊆":

(8.26) $$y \in f(\bigcup_{j\in J} A_j) \Rightarrow \exists\, x \in X:\ f(x) = y \text{ and } x \in \bigcup_{j\in J} A_j \quad (\text{def } f(A))$$

(8.27) $$\Rightarrow \exists\, j_0 \in J:\ f(x) = y \text{ and } x \in A_{j_0} \quad (\text{def } \cup)$$

(8.28) $$\Rightarrow y = f(x) \in f(A_{j_0})(\text{def } f(A))$$

(8.29) $$\Rightarrow y \in \bigcup_{j\in J} f(A_j) \quad (\text{def } \cup)$$

B. Proof of "⊇":
This follows from the monotonicity of $A \mapsto f(A)$ (see 5.15):

(8.30) $$A_i \subseteq \bigcup_{j\in J} A_j\ \forall\, i \in J \Rightarrow f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\ \forall\, i \in J$$

(8.31) $$\Rightarrow \bigcup_{i\in J} f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\ \forall\, i \in J \quad (\text{def } \cup) \ \blacksquare$$

The "elementary" proof is barely longer than the first one, but it is so much easier to understand!

Remark 8.5. In general you will not have equality in (8.19). 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 $f(A_1) \cap f(A_2) = [0,\infty]$. Clearly, $\{0\} \subsetneq [0,\infty]$. □

Proposition 8.6 (Direct images and preimages of function composition). *Let X, Y, Z be arbitrary, nonempty sets.*

Let $f : X \to Y$ and $g : Y \to Z$, and let $U \subseteq X$ and $W \subseteq Z$. Then

(8.32) $$(g \circ f)(U) = g\big(f(U)\big) \text{ for all } U \subseteq X\,.$$

(8.33) $$(g \circ f)^{-1}(W) = f^{-1}\big(g^{-1}(W)\big) \text{ for all } W \subseteq Z,\ \text{i.e.,}\quad (g \circ f)^{-1} = f^{-1} \circ g^{-1}\,.$$

PROOF of (8.32): Left as exercise 8.10.

PROOF of (8.33):
a. "⊆": Let $W \subseteq Z$ and $x \in (g \circ f)^{-1}(W)$. Then $g\big(f(x)\big) = (g \circ f)(x) \in W$, hence $f(x) \in g^{-1}(W)$. But then $x \in f^{-1}\big(g^{-1}(W)\big)$. This proves "⊆".
b. "⊇": Let $W \subseteq Z$, $h := g \circ f$, and $x \in f^{-1}\big(g^{-1}(W)\big)$. Then $f(x) \in g^{-1}(W)$, hence $g(f(x)) \in W$, i.e., $h(x) \in W$, hence $x \in h^{-1}(W) = (g \circ f)^{-1}(W)$. This proves "⊇". ■

Proposition 8.7 (Indirect image and fibers of f). *Let X, Y be nonempty sets and let $f : X \to Y$ be a function. We define on the domain X a relation "∼" as follows:*

(8.34) $$x_1 \sim x_2 \Leftrightarrow f(x_1) = f(x_2).$$

(a) *"$\sim$" is an equivalence relation. Its equivalence classes, which we denote by $[x]_f$,* [100] *are*

$$[x]_f \ = \{a \in X : f(a) = f(x)\} \ = \ f^{-1}\{f(x)\}. \quad (x \in X) \tag{8.35}$$

(b) *If $A \subseteq X$ then*

$$f^{-1}(f(A)) \ = \ \bigcup_{a \in A} [a]_f. \tag{8.36}$$

The proof that "$\sim$" is an equivalence relation is left as exercise 8.11.

PROOF of (8.35): 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 $f(a) = f(x) \Leftrightarrow a \in f^{-1}\{f(x)\}$, which is true according to the definition of preimages.

PROOF of (8.36):

Since $f(A) = f(\bigcup_{a \in A}\{a\}) = \bigcup_{a \in A} f\{a\} = \bigcup_{a \in A}\{f(a)\}$ (see 8.20), it follows that

$$f^{-1}(f(A)) \ = f^{-1}(\bigcup_{a \in A}\{f(a)\}) \tag{8.37}$$

$$= \bigcup_{a \in A} f^{-1}\{f(a)\} \quad \text{(see 8.11)} \tag{8.38}$$

$$= \bigcup_{a \in A} [a]_f \quad \text{(see 8.35)} \ \blacksquare \tag{8.39}$$

Corollary 8.1.

$$\text{If } A \subseteq X \quad \text{then} \quad f^{-1}(f(A)) \ \supseteq \ A. \tag{8.40}$$

The proof is left as exercise 8.12 (see p.250). ■

This next example shows how to work with fibers to prove that certain relations are equivalence relations.

Example 8.5. The following are equivalence relations on the set X.

(a) $X = \mathbb{R}$ and $x \sim y \Leftrightarrow |x| = |y|$.
(b) $X = \mathbb{R}_{\neq 0} = \{x \in \mathbb{R} : x \neq 0\}$ and $x \sim y \Leftrightarrow |xy| > 0$.
(c) $X = \mathbb{R}^3$ and $(x,y,z) \sim (u,v,w) \Leftrightarrow z\sin(xy) = w\sin(uv)$.

You can verify this by brute force, but here is an elegant way. Rewrite the equivalence relations as $\alpha \sim \beta \Leftrightarrow F(\alpha) = F(\beta)$ for a suitable function $F(\cdot)$, then apply prop.8.7 (Indirect image and fibers of f).

For the above examples you do this as follows:

(a) $F : X \to \mathbb{R}, \quad x \mapsto |x|$.
(b) $G : X \to \{-1, 1\}, \quad x \mapsto \frac{x}{|x|}$.
(b) $H : X \to \mathbb{R}, \quad (x,y,z) \mapsto z\sin(xy)$. □

[100] $[x]_f$ is called the **fiber over** x of the function f.

Proposition 8.8.

$$\text{(8.41)} \qquad \textit{If } B \subseteq Y \quad \textit{then} \quad f(f^{-1}(B)) \;=\; B \cap f(X).$$

PROOF of "$\subseteq$":

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) $x_0 \in f^{-1}(B) \Rightarrow f(x_0) \in B$ (def. of preimage)

(b) Of course $x_0 \in X$. Hence $f(x_0) \in f(X)$.

(a) and **(b)** together imply that $y = f(x_0) \in B \cap f(X)$.

PROOF of "$\supseteq$":

This part of the proof is left as exercise 8.13 (see p.250). ■

Remark 8.6. Be sure to understand how the assumption $y \in f(X)$ was used. □

Corollary 8.2.

$$\text{(8.42)} \qquad \textit{If } B \subseteq Y \quad \textit{then} \quad f(f^{-1}(B)) \;\subseteq\; B.$$

Trivial as $f(f^{-1}(B)) \;=\; B \cap f(X) \;\subseteq\; B.$ ■

Proposition 8.9.

(a) *Let $A \subseteq X$. If $f : X \to Y$ is injective then $f^{-1}(f(A)) = A$.*

(b) *Let $B \subseteq Y$. If $f : X \to Y$ is surjective then $f\big(f^{-1}(B)\big) = B$.*

(c) *Let $A \subseteq X$ and $B \subseteq Y$. If $f : X \to Y$ is injective and if $B = f(A)$ then $f^{-1}(B) = A$.*

(d) *Let $A \subseteq X$ and $B \subseteq Y$. If $f : X \to Y$ is surjective and if $f^{-1}(B) = A$ then $B = f(A)$.*

(e) *Let $A \subseteq X$ and $B \subseteq Y$. If $f : X \to Y$ is bijective then $B = f(A) \;\Leftrightarrow\; f^{-1}(B) = A$.*

PROOF: Left as exercise 8.14 on p.250. ■

Remark 8.7. It follows from prop.8.9 parts **(a)** and **(b)** together with thm.5.1 (Characterization of inverse functions) on p.145 that if $f : X \to Y$ is a bijection between two nonempty sets X and Y then the direct image function $f : 2^X \to 2^Y; \quad A \mapsto \{f(a) : a \in A\}$ is a bijection between the two power sets of X and Y, and its inverse is the preimage function $f^{-1} : 2^Y \to 2^X; \quad B \mapsto \{x \in X : f(x) \in B\}$.

Proposition 8.10. *Let J be an arbitrary nonempty index set and let $A \subseteq X$, $A_j \subseteq X$ for all j. Let $f : X \to Y$ be bijective. Then the following all are true:*

$$\text{(8.43)} \qquad f(\bigcap_{j \in J} A_j) \;=\; \bigcap_{j \in J} f(A_j)$$

$$\text{(8.44)} \qquad f(\bigcup_{j \in J} A_j) \;=\; \bigcup_{j \in J} f(A_j)$$

$$\text{(8.45)} \qquad f(A^{\complement}) \;=\; f(A)^{\complement}$$

$$\text{(8.46)} \qquad f(A_1 \setminus A_2) \;=\; f(A_1) \setminus f(A_2)$$

$$\text{(8.47)} \qquad f(A_1 \Delta A_2) \;=\; f(A_1) \Delta f(A_2)$$

PROOF: Left as exercise 8.16 on p.250. ■

Note that the remaining content of this chapter has been marked as " $\boxed{\star}$ " (optional)!

Proposition 8.11. $\boxed{\star}$ *Let $f : X \to Y$ be bijective. Let J be an arbitrary nonempty index set and let $(A_j)_{j \in J}$ be a partition of X, i.e., if $i \neq j$ then $A_i \cap A_j = \emptyset$ and $X = \biguplus_j A_j$. Assume further that none of the A_j are enpty. For $j \in J$ let $B_j := f(A_j)$. Then*

- ***(a)*** *$(B_j)_{j \in J}$ is a partition of Y.*
- ***(b)*** *For $j \in J$ we look at the restriction $f\big|_{A_j} : A_j \to Y$ to A_j. Then $f\big|_{A_j}(A_j) = B_j$ and the function*

$$f_j : A_j \to B_j, \qquad x \mapsto f_j(x) := f\big|_{A_j}(x) = f(x)$$

is a bijection.

PROOF: Left as exercise 8.17 on p.250. ■

Corollary 8.3. $\boxed{\star}$ *Let $f : X \to Y$ be bijective. Let $A \subset X, A \neq \emptyset$ (strict inclusion, so $A^{\complement} \neq \emptyset$). Then both*

$$f_A : A \to f(A), \quad x \to f(x) \qquad \text{and} \quad f_{A^{\complement}} : A^{\complement} \to f(A^{\complement}), \quad x \to f(x)$$

are bijections.

PROOF: This follows from prop.8.11, applied to $J = \{1, 2\}, A_1 = A, A_2 = A^{\complement}$. ■

Corollary 8.4. $\boxed{\star}$ *Let $f : X \to Y$ be bijective. Let $a \subset X$ and assume that $X \neq \{a\}$. Then*

$$\tilde{f} : X \setminus \{a\} \to Y \setminus \{f(a)\}, \quad x \to f(x)$$

also is bijective. [101]

PROOF: This follows from 8.4 applied to $A = \{a\}$ and the fact that $f(\{a\}) = \{f(a)\}$. ■

The following two propositions allow you to 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 proposition shows how to preserve bijectivity if two function values need to be switched around.

Proposition 8.12. $\boxed{\star}$ *Let $X, Y \neq \emptyset$, let $f : X \to Y$ be bijective and let $x_1, x_2 \in X$. Let*

$$g(x) := \begin{cases} f(x_2) & \text{if } x = x_1, \\ f(x_1) & \text{if } x = x_2, \\ f(x) & \text{if } x \neq x_1, x_2. \end{cases} \tag{8.48}$$

(In other words, we swap two function arguments). Then $g : X \to Y$ also is bijective.

[101] This is B/G [2] prop.13.2.

PROOF: Left as exercise 8.18 on p.250. ■

A more general version of the above shows how to preserve surjectivity if two function values need to be switched around.

Proposition 8.13. $\boxed{\star}$

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

$$g(x) := \begin{cases} y_2 & \text{if } x \in A_1, \\ y_1 & \text{if } x \in A_2, \\ f(x) & \text{if } x \in B. \end{cases} \tag{8.49}$$

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: Left as exercise 8.19 on p.250. ■

Proposition 8.14. $\boxed{\star}$

Let X, Y be two nonempty sets and let $f : X \to Y$ be surjective. Let $\emptyset \neq B \subsetneq Y$ so that $Y = B \uplus B^{\complement}$ is a partitioning of Y into two nonempty subsets B and $B^{\complement}$. Let $A := \{f \in B\}$. Then the restrictions $f_1 := f\Big|_A : A \to B$ and $f_2 := f\Big|_{A^{\complement}} : A^{\complement} \to B^{\complement}$ of f to A and $A^{\complement}$ are surjections.

PROOF: Left as exercise 8.20 on p.250. ■

8.5 Indicator Functions $\boxed{\star}$

Sometimes it is advantageous to think of the subsets of a universal set Ω as "binary" functions $\Omega \to \{0, 1\}$.

Definition 8.7 (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 defined as

$$1_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{8.50}$$

1_A is called the **indicator function** of the set A. □

Some authors call 1_A the **characteristic function** of A and/or write χ_A or $\mathbb{1}_A$ instead of 1_A.

Recall the following about functions and families: If X and Y are two nonempty sets then Y^X, the "X–fold cartesian product of Y", is the set of all Y–valued families $\left(y_x\right)_{x \in X}$ which are indexed by X. Equivalently Y^X is the set of all functions $f : X \to Y$ with domain X and codomain Y. See Proposition 5.11 (Functions are families and families are functions) on p.160. This will be used in the next proposition which shows that the association of a subset A of Ω with its indicator function 1_A is a bijection.

Proposition 8.15. *Let $\mathscr{F}(\Omega, \{0,1\}) := \{0,1\}^\Omega$ 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.* [102]

***(a)** The mapping*

$$F : 2^\Omega \to \mathscr{F}(\Omega, \{0,1\}), \quad \textit{defined as} \ \ F(A) := 1_A \tag{8.51}$$

which assigns to each subset of Ω its indicator function is injective.

***(b)** Let $f \in \mathscr{F}(\Omega, \{0,1\})$. Further, let $A := \{f = 1\} = f^{-1}(\{1\}) = \{a \in A : f(a) = 1\}$. Then $f = 1_A$.*

(c) $\boxed{\textit{The function } F \textit{ above is bijective.}}$

Its inverse function is

$$G : \ \mathscr{F}(\Omega, \{0,1\}) \to 2^\Omega, \quad \textit{defined as} \ \ G(f) := \{f = 1\}. \tag{8.52}$$

PROOF of **(a)**: This follows from **(c)** which will be proved below.

PROOF of **(b)**: We have

$$\begin{aligned} f(\omega) = 1 \ &\Leftrightarrow \omega \in \{f = 1\} \quad \text{(def. of inverse image)} \\ &\Leftrightarrow \omega \in A \quad \text{(because } A = \{f = 1\}\text{)} \\ &\Leftrightarrow 1_A(\omega) = 1 \quad \text{(def. of indicator function).} \end{aligned}$$

It follows that $f(\omega) = 1$ if and only if $1_A(\omega) = 1$. Since 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., $f = 1_A$. This proves **(b)**.

PROOF of **(c)**: According to theorem 5.1 on p.145 about the characterization of inverse functions **(c)** is proved if we can demonstrate that F and G are inverse to each other. To prove this it suffices to show that

$$G \circ F \ = id_{2^\Omega} \quad \text{and} \quad F \circ G \ = \ id_{\mathscr{F}(\Omega, \{0,1\})}. \tag{8.53}$$

Let $A \in 2^\Omega$, i.e., $A \subseteq \Omega$. 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 $G \circ F = id_{2^\Omega}$. Now let $f \in \mathscr{F}(\Omega, \{0,1\}$ and $\omega \in \Omega$. Then

$$\begin{aligned} \big(F \circ G(f)\big)(\omega) \ &= F(\{f = 1\})(\omega) \ = \ 1_{\{f=1\}}(\omega) \\ &= \begin{cases} 1 & \text{iff } \omega \in \{f = 1\}, \\ 0 & \text{iff } \omega \notin \{f = 1\} \end{cases} = \begin{cases} 1 & \text{iff } f(\omega) = 1, \\ 0 & \text{iff } f(\omega) \neq 1 \end{cases} = \begin{cases} 1 & \text{iff } f(\omega) = 1, \\ 0 & \text{iff } f(\omega) = 0 \end{cases} = \ f(\omega). \end{aligned}$$

The equation next to the last results from the fact that the only possible function values for f are 0 and 1; the equation before that follows from (5.13) (definition of the preimage). It follows from the above chain of equations that $F \circ G(f) \ = \ f \ = \ id_{\mathscr{F}(\Omega, \{0,1\})}(f)$ for all $f \in \mathscr{F}(\Omega, \{0,1\})$, hence $F \circ G \ = \ id_{\mathscr{F}(\Omega, \{0,1\})}$. We have proved (8.53) and hence **(c)**. ■

[102] See remark 8.4 on p.239, ch.8.3 (Cartesian Products of More Than Two Sets).

Let $m, n \in \mathbb{Z}$. We recall from Definition 6.13 (Equivalence Modulo n) (p.197 of ch. 6.10) that $m+n \mod 2$ (the sum mod 2 of m and n) is given by

$$m+n \mod 2 \;=\; \begin{cases} 0 & \Leftrightarrow \quad (m+n)/2 \text{ has remainder } 0 \text{, i.e., } m+n \text{ is even,} \\ 1 & \Leftrightarrow \quad (m+n)/2 \text{ has remainder } 1 \text{, i.e., } m+n \text{ is odd.} \end{cases} \tag{8.54}$$

Proposition 8.16. *Let $m, n, p \in \mathbb{Z}$. Then addition mod 2 is associative, i.e.,*

$$\big(m+n \mod 2\big)+p \mod 2 \;=\; m+\big(n+p \mod 2\big) \mod 2. \tag{8.55}$$

PROOF: This follows from prop.6.34 on p.199 ($\mathbb{Z}_n$ is a commutative ring with unit). [103] ■

Proposition 8.17. *Let A, B, C be subsets of Ω. Then*

$$\mathbb{1}_{A\cup B} \;=\; \max(\mathbb{1}_A, \mathbb{1}_B), \tag{8.56}$$

$$\mathbb{1}_{A\cap B} \;=\; \min(\mathbb{1}_A, \mathbb{1}_B), \tag{8.57}$$

$$\mathbb{1}_{A^{\complement}} \;= 1-\mathbb{1}_A, \tag{8.58}$$

$$\mathbb{1}_{A\triangle B} \;= \mathbb{1}_A + \mathbb{1}_B \mod 2. \tag{8.59}$$

PROOF: The proof of the first three equations is left as an exercise.
PROOF of (8.59): This follows easily from the the fact that

$$(A\triangle B)^{\complement} \;=\; \{\omega\in\Omega : \;[\text{either } \omega\in A\cap B] \text{ or } [\text{neither } \omega \in A \text{ nor } \omega\in B\,]\} \quad ■$$

Prop.8.16 above helps us to prove associativity of symmetric set differences.

Proposition 8.18 (Symmetric set differences $A\triangle B$ are associative)**.** *Let $A, B, C \subseteq \Omega$. Then*

$$(A\triangle B)\triangle C \;= A\triangle(B\triangle C). \tag{8.60}$$

PROOF: This follows easily from (8.59) and and the associativity of $a\oplus b := a+b \mod 2$ as follows. Let $\omega\in\Omega$. Then

$$\begin{aligned} \omega\in(A\triangle B)\triangle C \;&\Leftrightarrow \mathbb{1}_{(A\triangle B)\triangle C}(\omega)=1 \\ &\Leftrightarrow \big(\mathbb{1}_A(\omega)\oplus\mathbb{1}_B(\omega)\big)\oplus\mathbb{1}_C(\omega)=1 \\ &\Leftrightarrow \mathbb{1}_A(\omega)\oplus\big(\mathbb{1}_B(\omega)\oplus\mathbb{1}_C(\omega)\big)=1 \\ &\Leftrightarrow \mathbb{1}_{A\triangle(B\triangle C)}(\omega)=1 \;\Leftrightarrow\; \omega\in A\triangle(B\triangle C). \end{aligned}$$

We obtained the equivalence in the middle from prop.8.16. ■

[103] There also are elementary proofs for this proposition. See exercise 8.23 on p.251.

8.6 Exercises for Ch.8

Exercise 8.1. Prove **(a)** and **(b)** of prop.8.2 (Rewrite unions as disjoint unions) on p.235:

Let $(A_j)_{j\in\mathbb{N}}$ such that $A_j \subseteq \Omega$ for all $j \in \mathbb{N}$. For $n \in \mathbb{N}$ let $B_n := \bigcup\limits_{j=1}^{n} A_j \ = \ A_1 \cup A_2 \cup \cdots \cup A_n$

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 $n \in \mathbb{N}$, $\bigcup\limits_{j=1}^{n} A_j \ = \ \bigcup\limits_{j=1}^{n} B_j$. □

Exercise 8.2. (See example 5.34 on p.157). Let $X := [0,2]$. For $0 \leq x \leq 2$ let $A_x := [x, 2x]$.
(a) What is $\bigcap \Big[A_x : x \in X\Big]$? **(b)** What is $\bigcup \Big[A_x : x \in X\Big]$? □

Exercise 8.3. Prove **(b)** of thm.8.1 (De Morgan's Law):
Let $(A_\alpha)_{\alpha\in I}$ be a family of subsets of a universal set Ω. Then $\big(\bigcap\limits_{\alpha} A_\alpha\big)^{\complement} \ = \ \bigcup\limits_{\alpha} A_\alpha^{\complement}$. □

Exercise 8.4. Supply the missing proofs of prop.8.3 on p.236 of this document. □

Exercise 8.5. Prove the second formula of prop.8.1 (Distributivity of unions and intersections): Let $(A_i)_{i\in I}$ be an arbitrary family of sets and let B be a set. Then

$$\bigcap_{i\in I} (B \cup A_i) \ = B \cup \bigcap_{i\in I} A_i. \ \square$$

Exercise 8.6. Let $(G, \diamond)$ be a group, let $(H_i)_{i\in J}$ be a family of subgroups of G, and let $H := \bigcap\limits_{i\in J} H_i$.
Then H is a subgroup of G. □

Exercise 8.7. Let f be the function $f : [-3,3] \to \mathbb{R}; \quad x \mapsto x^2$.
(a) Is $f \in [-3,3]^{\mathbb{R}}$ or $f \in \mathbb{R}^{[-3,3]}$?
(b) Write f as a family. **Hint:** What is the index set? Domain or codomain? □

Exercise 8.8. Let $f : [-2.\infty[\to \mathbb{R}; \quad x \mapsto x^2$. Compute the following.
(a) $f\big(f^{-1}([-4,4])\big)$, **(b)** $f^{-1}\big(f([0,3])\big)$. □

Exercise 8.9. Prove prop.5.3 on p.143:
(a) $f(\emptyset) \ = \ f^{-1}(\emptyset) \ = \ \emptyset$
(b) $A_1 \ \subseteq \ A_2 \ \subseteq \ X \Rightarrow f(A_1) \ \subseteq \ f(A_2)$
(c) $B_1 \ \subseteq \ B_2 \ \subseteq \ Y \Rightarrow f^{-1}(B_1) \ \subseteq \ f^{-1}(B_2)$
(d) $x \ \in \ X \Rightarrow f(\{x\}) \ = \ \{f(x)\}$
(e) $f(X) \ = \ Y \ \Leftrightarrow$ f is surjective
(f) $f^{-1}(Y) \ = \ X$ always! □

Exercise 8.10. Prove (8.32) of prop.8.6 on p.242: Let X, Y, Z be arbitrary, nonempty sets.
Let $f : X \to Y$ and $g : Y \to Z$. Then $(g \circ f)(U) \ = \ g\big(f(U)\big)$ for all $U \subseteq X$. □

Exercise 8.11. Let X, Y be nonempty sets and let $f : X \to Y$ be a function. We define on the domain X a relation "$\sim$" as follows:

$$x_1 \sim x_2 \Leftrightarrow f(x_1) = f(x_2). \ \square$$

(See prop.8.7 (Indirect image and fibers of f) on p.242). Prove that "$\sim$" is an equivalence relation.

Exercise 8.12. Prove cor.8.1 on p.243 of this document: If $A \subseteq X$ then $f^{-1}(f(A)) \supseteq A$. $\square$

Exercise 8.13. Prove prop.8.8 on p.244 of this document: If $B \subseteq Y$ then $f(f^{-1}(B)) = B \cap f(X)$. $\square$

Exercise 8.14. Prove prop.8.9 on p.244.
Hint: The main tools you need are prop.8.7 on p.242, prop8.8 on p.244, and their corollaries. $\square$

Exercise 8.15. Prove the reverse directions of prop.8.9**(a)** and prop.8.9**(b)** on p.244.

(a) If $f : X \to Y$ satisfies $f^{-1}\big(f(A)\big) = A$ for all $A \subseteq X$ then f is injective.
(b) If $f : X \to Y$ satisfies $f\big(f^{-1}(B)\big) = B$ for all $B \subseteq Y$ then f is surjective.

Exercise 8.16. Prove prop.8.10 on p.244.
Hint: Work with the inverse of f and apply prop.8.4 on p.240. $\square$

Exercise 8.17. Prove prop.8.11 on p.245.
Hint: To prove **(a)**, use prop.8.5 on p.241. $\square$

Exercise 8.18. Prove prop.8.12 on p.245: Let $X, Y \neq \emptyset$, let $f : X \to Y$ be bijective and let $x_1, x_2 \in X$. Let

$$g(x) := \begin{cases} f(x_2) & \text{if } x = x_1, \\ f(x_1) & \text{if } x = x_2, \\ f(x) & \text{if } x \neq x_1, x_2. \end{cases}$$

(In other words, we swap two function arguments). Then $g : X \to Y$ also is bijective. $\square$

Exercise 8.19. Prove prop.8.13 on p.246: 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

$$g(x) := \begin{cases} y_2 & \text{if } x \in A_1, \\ y_1 & \text{if } x \in A_2, \\ f(x) & \text{if } x \in B. \end{cases}$$

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. $\square$

Exercise 8.20. Prove prop.8.14 on p.246: Let X, Y be two nonempty sets and let $f : X \to Y$ be surjective. Let $\emptyset \neq B \subsetneq Y$ so that $Y = B \uplus B^{\complement}$ is a partitioning of Y into two nonempty subsets B and $B^{\complement}$. Let $A := \{f \in B\}$. Then the restrictions $f_1 := f\Big|_A : A \to B$ and $f_2 := f\Big|_{A^{\complement}} : A^{\complement} \to B^{\complement}$ of f to A and to $A^{\complement}$ are surjections. $\square$

Exercise 8.21. Prove prop.8.16 on p.248: Let $m, n, p \in \mathbb{Z}$. Then

$$\big(m+n \mod 2\big) + p \mod 2 \quad = \quad m + \big(n+p \mod 2\big) \mod 2.$$

directly, i.e., without referring to prop.6.34 on p.199 ($\mathbb{Z}_n$ is CRU).
Hint: There are eight possible combinations of zeros and ones for the functions

$$(m,n,p) \to \big(m+n \mod 2\big) + p \mod 2 \qquad \text{and} \qquad (m,n,p) \to m + \big(n+p \mod 2\big) \mod 2.$$

Complete the entries in the table below and show that the entries in the two rightmost columns match. To save space, write $m \oplus n$ for $m+n \mod 2$. To get you started, the row for $m=1, n=0, p=0$ has been already completed.

m	n	p	$m \oplus n$	$n \oplus p$	$(m \oplus n) \oplus p$	$m \oplus (n \oplus p$
0	0	0				
0	0	1				
0	1	0				
0	1	1				
1	0	0	1	0	1	1
1	0	1				
1	1	0				
1	1	1				

□

Exercise 8.22. See [2] B/G project 6.8 on p.58 for the following.
Prove that the following are equivalence relations on $\mathbb{R}^2$.

(a) $(x,y) \sim (u,v) \;\Leftrightarrow\; \sqrt{x^2+y^2} = \sqrt{u^2+v^2}$.
(b) $X = \mathbb{R}_{\neq 0} \;=\; \{x \in \mathbb{R} : x \neq 0\}$ and $x \sim y \;\Leftrightarrow\; |xy| > 0$.
(c) $X = \mathbb{R}^3$ and $(x,y,z) \sim (u,v,w) \;\Leftrightarrow\; z\sin(xy) = w\sin(uv)$.

Hint: See example 8.5 on p.243. □

Exercise 8.23. Let $m, n, p \in \mathbb{Z}$.
Prove that addition mod 2 is associative. (see prop.8.16 on p.248) without referring to prop.6.34. Rather inspect what happens for each of the eight possible combinations of zeros and ones for the functions $(m,n,p) \to \big(m+n \mod 2\big) + p \mod 2$ and $(m,n,p) \to m + \big(n+p \mod 2\big) \mod 2$ □

9 The Real Numbers

9.1 The Ordered Fields of the Real and Rational Numbers

Definition 9.1 (Fields). ★

Let $(F, \oplus, \odot)$ be a commutative ring with unit (see Definition 3.7 on p.60) such that each nonzero element possesses an inverse element with respect to multiplication, i.e., the set $(F \setminus \{0\}, \odot)$ with neutral element 1 is an abelian group. Then we call $(F, \oplus, \odot)$ a **field**. □

Remark 9.1. It follows from thm.3.2 (Uniqueness of the Inverse in Groups) on p.53. that the multiplicative inverse b^{-1} of b is unique. □

Proposition 9.1 (B/G prop.8.6). *Let $(F, \oplus, \odot)$ be a field and $a, b \in F \setminus \{0\}$. Then*

$$(ab)^{-1} \ = \ b^{-1}a^{-1}\,.$$

PROOF: Since $(F \setminus \{0\}, \odot)$ is a group this follows from prop.3.2 on p.53. ■

Proposition 9.2. *Fields are integral domains.*

The proof is left as exercise 9.1 (see p.302). ■

Corollary 9.1 (B/G prop.8.7). *Let $a, b, c \in F$ and $a \neq 0$. If $ab = ac$ then $b = c$.*

PROOF: This follows from cor.3.1 on p.64. ■

We defined for integers m and n their quotient $\frac{n}{m}$ under the condition that $m \mid n$ and $m \neq 0$. We did so by defining $\frac{n}{m}$ as the unique integer j which satisfies $n = j \cdot m$ in case that $n \neq 0$ and by further defining $\frac{0}{m} = 0$. For a field $(F, \oplus, \odot)$ the group property of $(F \setminus \{0\}, \odot)$ gives us an alternative for defining quotients.

Definition 9.2 (Division and Quotients).

Let a, b be elements of a field $(F, \oplus, \odot)$, and let $b \neq 0$. Since b possesses a unique multiplicative inverse b^{-1} (see rem.9.1 on p.252) we can define the function

$$\mathrm{div} : F \times (F \setminus \{0\}) \longrightarrow F; \qquad (a, b) \mapsto a \odot b^{-1}.$$

We call this function the **division** operation on F. It is customary to also write $\frac{a}{b}$ or a/b instead of $a \odot b^{-1}$, and we follow that convention. In particular we may also write $\frac{1}{b}$ instead of b^{-1}. As in the case of the integers we call a the **dividend** or **numerator**, b the **divisor** or **denominator**, and $\frac{a}{b}$ the **quotient** of the expression $\frac{a}{b}$. □

Proposition 9.3. *Let $(F, \oplus, \odot, P)$ be a field and let $a \in R$. If $a \neq 0$ then the function*

$$D : F \to F; \quad x \mapsto a \odot x,$$

is a bijection.

The proof is left as exercise 9.2 (see p.302). ■

The following propositions can be easily shown by rewriting fractions $\frac{u}{v}$ as products uv^{-1} and applying the rules of arithmetic that were proven for integral Domains.

Proposition 9.4 (B/G prop.11.2). *Let $a, b, c, d \in F$ such that $b, d \neq 0$.*

$$\textit{If} \quad \frac{a}{b} = \frac{c}{d} \qquad \textit{then} \quad ad = bc\,.$$

The proof is left as exercise 9.3 (see p.302). ■

Proposition 9.5 (B/G prop.11.3). *Let $a, b, c \in F$ such that $b, c \neq 0$. Then*

$$\frac{ac}{bc} = \frac{a}{b}\,.$$

The proof is left as exercise 9.4 (see p.302). ■

Proposition 9.6 (B/G prop.11.6). *Let $a, b, c, d \in F$ such that $b, d \neq 0$. Then*

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{ad \oplus bc}{bd}\,. \qquad \textit{In particular, } \frac{a}{b} \oplus \frac{(\ominus a)}{b} = 0\,.$$

The proof is left as exercise 9.5 (see p.302). ■

Proposition 9.7. *Let $a, b, c, d \in F$ such that $b, d \neq 0$.*

$$\textit{Then} \quad \frac{a}{b} \odot \frac{c}{d} = \frac{ac}{bd}\,. \qquad \textit{In particular, } \left(\frac{b}{d}\right)^{-1} = \frac{d}{b}\,.$$

The proof is left as exercise 9.6 (see p.302). ■

We introduced in ch.3.4 (Order Relations in Integral Domains) the concept of a positive cone as a means of ordering an integral domain. Since fields are integral domains we can do the same here.

Definition 9.3 (Ordered fields). ★

Let $(F, \oplus, \odot)$ be a field which is ordered by a positive cone P. Then we call $(F, \oplus, \odot, P)$ an **ordered field**. □

Remark 9.2. It is immediate from prop.9.2 on p.252 that ordered fields are ordered integral domains.

Notation 9.1 (Fields and ordered fields).

> Unless this would lead to confusion we will usually write F for a field $(F, \oplus, \odot)$ or for an ordered field $(F, \oplus, \odot, P)$. □

Rules for arithmetic and for inequalities were given for integral domains in ch.3.3 and ch.3.4. We complement those rules with the following propositions which require the existence of the multiplicative inverse x^{-1} for nonzero x. With them we cover all propositions given in B/G ch.8.1 – 8.3.

Proposition 9.8 (B/G prop.8.40).

(a) *Let $a \in F$. Then $a > 0$ if and only if $a^{-1} > 0$, and $a < 0$ if and only if $a^{-1} < 0$.*

(b) *Let $a, b \in F$. If $0 < a < b$ then $0 < \dfrac{1}{b} < \dfrac{1}{a}$.*

The proof is left as exercise 9.7 (see p.302). ■

Corollary 9.2 (B/G prop.11.7). *Let $a, b \in F_{\neq 0}$. Then*

(a) $\dfrac{a}{b} > 0 \Leftrightarrow \dfrac{b}{a} > 0$ *and* $\dfrac{a}{b} < 0 \Leftrightarrow \dfrac{b}{a} < 0$,

(b) $\dfrac{a}{b} > 0 \Leftrightarrow$ *either both $a, b > 0$ or both $a, b < 0$.*

The proof is left as exercise 9.8 (see p.302). ■

The induction axiom resulted in the discrete structure of the integers: If $n \in \mathbb{Z}$ then there are no integers between n and $n+1$. In particular there are none between 0 and 1, and this property of the integers lead to $\mathbb{Z}_{>0} = \mathbb{Z}_{\geq 1}$, i.e., $\mathbb{N} = \mathbb{Z}_{\geq 1}$, and $1 = min(\mathbb{Z}_{>0})$. The existence of a^{-1} for any nonzero $a \in F$ creates an entirely different situation.

Theorem 9.1 (B/G thm.8.43). *Let $a, b \in F$ such that $a < b$. Then*

$$a < \frac{a+b}{2} < b\,.$$

PROOF: The proof is left as exercise 9.9 (see p.302). ■

Theorem 9.2 (B/G thm.8.42).

> *The positive cone P does not have a minimum.*

PROOF: Assume to the contrary that $p_* := \min(P)$ exists. It follows from the previous theorem that

$$0 < \frac{p_*}{2} < p_*\,.$$

Thus $\frac{p_*}{2}$ is an element of P strictly less than $\min(P)$. Contradiction! ■

We have repeatedly worked with Definition 2.14 for the set $\mathbb{R}$ of the real numbers and Definition 2.15 for the set $\mathbb{Q}$ of the rational numbers (see p.25) without giving precise definitions for those mathematical objects. We now are at a point of replacing those informal definitions with mathematically exact ones. Both $\mathbb{Q}$ and $\mathbb{R}$ will turn out to be ordered fields, and this explains the title "The

Ordered Fields of the Real and Rational Numbers" of this chapter. Fields and ordered fields are of extreme importance in the discipline of abstract algebra but for us their only purpose is to allow a unified presentation of many algebraic formulas and inequalities.

Recall for the following that we defined minima, maxima, infima and suprema for arbitrary integral domains in ch.3.5 on p.76.

Axiom 9.1 (Real Numbers)**.** We postulate the existence of a set $\mathbb{R}$ which satisfies the following:

> **(a)** $\mathbb{R}$ is endowed with two binary operations "$+$" (called addition) and "$\cdot$" (called multiplication) and with a positive cone $\mathbb{R}_{>0}$ such that $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$ is an ordered integral domain. As usual we denote the additive unit of this integral domain by 0 and its multiplicative unit by 1.
>
> **(b)** The set $\mathbb{R}_{\neq 0} = \{x \in \mathbb{R} : x \neq 0\}$ is a group with respect to multiplication; thus for each $x \in \mathbb{R}_{\neq 0}$ there exists a unique $x^{-1} \in \mathbb{R}_{\neq 0}$ such that $xx^{-1} = 1$.
>
> **(c)** $\mathbb{R}$ satisfies the **completeness axiom**: Any nonempty subset A of $\mathbb{R}$ which is bounded above possesses a supremum in $\mathbb{R}$ (i.e., $\sup(A) \neq \pm\infty$).
>
> We call this set $\mathbb{R}$ the set of **real numbers**. □

Remark 9.3.

(1) Note that **(a)** and **(b)** together are equivalent to stating that is an ordered field. Thus the real numbers constitute an ordered field which in addition satisfies the completeness axiom.

(2) Definition 3.12 (Intervals in Ordered Integral Domains) on p.70 applies to any ordered integral domain, hence to any ordered field, hence to $\mathbb{R}$. We thus can write $]0, \infty[$ for the positive cone $\mathbb{R}_{>0}$ of the real numbers. □

Note 9.1 (The integers are a subset of the real numbers)**.** We recall that the ordered field $\mathbb{R}$ is an ordered integral domain. (See prop.9.2 on p.252.) We have seen in rem.6.4 on p.178 of ch.6.4 (Embedding the Integers Into an Ordered Integral Domain) that the integers can be embedded into $\mathbb{R}$ in such a way that one can consider them elements of the real numbers.

> We think from this point forward of the set $\mathbb{Z}$ which was defined as a mathematical object in axiom 6.1 on p.167 as a subset of $\mathbb{R}$. In particular, if $n \in \mathbb{Z}$ and $x \in \mathbb{R}$ then we can do comparisons such as $n \leq x$ and construct expressions such as $\sqrt{e^{n/x}}$ by viewing n as an element of $\mathbb{R}$. □

This allows us to define the subset $\mathbb{Q}$ of $\mathbb{R}$ in terms of integers.

Definition 9.4 (Rational numbers)**.**

> We call the set
> $$\mathbb{Q} := \{\, n/d : n \in \mathbb{Z}, d \in \mathbb{N} \,\}$$
> (this is a subset of $\mathbb{R}$!) the set of **rational numbers**.

In other words rational numbers are fractions of integers. □

Theorem 9.3 (The Rational Numbers are an Ordered Field)**.**

> ***(a)*** *The assignments* $(a, b) \mapsto a + b$ *and* $(a, b) \mapsto a \cdot b$ *are binary operations on* $\mathbb{Q}$*, i.e., sums and products of rational numbers are rational numbers.*
>
> ***(b)*** *The triplet* $(\mathbb{Q}, +, \cdot)$ *is an integral domain.*
>
> ***(c)*** *Let* $\mathbb{Q}_{>0} := \mathbb{R}_{>0} \cap \mathbb{Q}$*. Then* $(\mathbb{Q}, +, \cdot, \mathbb{Q}_{>0})$ *is an ordered integral domain which satisfies the following: if* $a, b \in \mathbb{Q}$ *then* $a < b$ *with respect to the ordering induced by* $\mathbb{Q}_{>0}$ *if and only if* $a < b$ *with respect to the ordering induced by* $\mathbb{R}_{>0}$
>
> ***(d)*** $(\mathbb{Q}_{\neq 0}, \cdot)$ *is a (commutative) group.*

PROOF: $\boxed{\star}$

(A) It follows from prop.9.6 on p.253 and prop.9.7 on p.253 that sums and products of two rational numbers are quotients of integers with nonzero denominator, i.e., elements of $\mathbb{Q}$. Part **(a)** follows.

(B) Rational numbers are real numbers since they are quotients of integers and those are real numbers. Thus addition and multiplication of rational numbers is associative, commutative and distributive because this is true for real numbers. Further $0 = \frac{0}{1} \in \mathbb{Q}$ and $1 = \frac{1}{1} \in \mathbb{Q}$. Further, prop.9.6 and prop.9.7 show the following. If $m, n \in \mathbb{Z}$, $n \neq 0$ and $q = \frac{m}{n}$, then $-q = \frac{-m}{n}$. If we assume additionally that $q \neq 0$, i.e., $m \neq 0$, then $q^{-1} = \frac{n}{m}$. Thus the additive and multiplicative inverses of rational numbers not only exist as real numbers, they actually are rational numbers.

In short, $(\mathbb{Q}, \cdot)$ is a commutative monoid, and both $(\mathbb{Q}, +)$ and $(\mathbb{Q} \setminus \{0\}, \cdot)$ are abelian groups.

Since $0 \neq 1$ in $\mathbb{R}$ we have $0 \neq 1$ in $\mathbb{Q}$.

(C) It follows from **B** that $(\mathbb{Q}, +, \cdot)$ satisfies Definition 3.7**(a)–(d)** on p.60 and thus is a commutative ring with unit. Since the integral domain $(\mathbb{R}, +, \cdot)$ does not possess any zero divisors, the same is true for $(\mathbb{Q}, +, \cdot)$. Thus this algebraic object is an integral domain. This proves part **(b)**. We also have seen in **B** that $(\mathbb{Q}, \cdot)$ is a commutative monoid. This proves part **(d)**.

(D) We next prove that the set $\mathbb{Q}_{>0} := \mathbb{R}_{>0} \cap \mathbb{Q}$ is a positive cone for $\mathbb{Q}$. Let $a, b \in \mathbb{Q}_{>0}$. Then $a, b \in \mathbb{R}_{>0}$, hence $a + b, ab \in \mathbb{R}_{>0}$ since $\mathbb{R}_{>0}$ is a positive cone. We have seen in part **A** that $a + b$ and ab are rational, thus they belong to $\mathbb{R}_{>0} \cap \mathbb{Q}$, i.e., $a + b, ab \in \mathbb{Q}_{>0}$. It follows that $\mathbb{Q}_{>0}$ satisfies **(a)** and **(b)** of Definition 3.11 on p.68 of a positive cone.

The additive neutral element 0 does not belong to the positive cone $\mathbb{R}_{>0}$ and thus not to its subset $\mathbb{Q}_{>0}$. Thus Definition 3.11**(c)** is satisfied.

Next we assume that $a \in \mathbb{Q}$ satisfies both $a \notin \mathbb{Q}_{>0}$ and $a \neq 0$. We claim that $a \notin \mathbb{R}_{>0}$. This is true since otherwise we have $a \in \mathbb{R}_{>0} \cap \mathbb{Q}$, i.e., $a \in \mathbb{Q}_{>0}$, and this contradicts our assumption that $a \notin \mathbb{Q}_{>0}$. Since $\mathbb{R}_{>0}$ is a positive cone and neither $a = 0$ nor $a \notin \mathbb{R}_{>0}$ it follows from Definition 3.11**(d)** that $a \in \mathbb{R}_{>0}$. Since a is rational this implies $a \in \mathbb{Q}_{>0}$. Thus Definition 3.11**(d)** is satisfied, and we have proven that $\mathbb{Q}_{>0}$ is a positive cone.

(E) We write "$<_Q$" for the order induced by $\mathbb{Q}_{>0}$ and "$<_R$" for the order induced by $\mathbb{R}_{>0}$, i.e., if $a, b \in \mathbb{Q}$ then

$$a <_R b \quad \Leftrightarrow \quad b - a \in \mathbb{R}_{>0}; \qquad\qquad a <_Q b \quad \Leftrightarrow \quad b - a \in \mathbb{Q}_{>0}.$$

Let $a, b \in \mathbb{Q}$. We now prove that $a <_R b \ \Leftrightarrow \ a <_Q b$.

First assume $a <_R b$. Then $b - a \in \mathbb{R}_{>0}$. Since $b - a \in \mathbb{Q}$, it follows that $b - a \in \mathbb{Q}_{>0}$. Thus $a <_Q b$.

Now assume $a <_Q b$. Then $b - a \in \mathbb{Q}_{>0}$. Since $\mathbb{Q}_{>0} \subseteq \mathbb{R}_{>0}$, it follows that $b - a \in \mathbb{R}_{>0}$. Thus $a <_R b$.

We have shown that part **(c)** follows from **D** and **E**. The proof is complete. ■

Remark 9.4.

(a) Note that $\mathbb{Z} \subseteq \mathbb{Q}$ since any integer n can be written as a fraction $n = \frac{n}{1}$. Further $\mathbb{Q} \subseteq \mathbb{R}$ since $\mathbb{Z} \subseteq \mathbb{R}$ and the quotient $\frac{a}{b}$ of two real numbers a and b where $b \neq 0$ exists as a real number.

(b) It follows from theorem 9.3 that $(\mathbb{Q}, +, \cdot, \mathbb{Q}_{>0})$ is an ordered field, just as is the case for the real numbers. But the rational numbers do not satisfy the completeness axiom. See rem.9.1 on p.258. □

As mentioned before neither the real numbers nor the rational numbers possess the discrete structure of the integers since both ordered fields admit division $\frac{a}{b}$ for nonzero b. We will see in subsequent subchapters of this chapter that both $\mathbb{Q}$ and $\mathbb{R}$ are different in many ways, but here we show an important property that both have in common: The natural numbers which are a subset of both $\mathbb{Q}$ and $\mathbb{R}$ are unbounded in either one of those two sets.

Theorem 9.4 (B/G thm.10.1: $\mathbb{N}$ is unbounded in $\mathbb{R}$)**.**

For any $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ such that $n > x$, i.e., there are no upper bounds for $\mathbb{N}$ in $\mathbb{R}$.

Proof strategy: It is tempting to argue as in prop.6.25 ($\mathbb{N}$ is unbounded in $\mathbb{Z}$) on p.195 which we reproduce here:

Assume to the contrary that there exists an upper bound of $\mathbb{N}$. According to thm.6.9 (extended well-ordering principle) on p.194 $\mathbb{N}$ has a maximum. Let $u^\star := \max(\mathbb{N})$. Then $u^\star + 1$ belongs to $\mathbb{N}$ as the sum of two natural numbers. It follows from $u^\star + 1 > u^\star$ that $u^\star$ is not the largest element of $\mathbb{N}$ and we have reached a contradiction.

The problem is that we could apply the extended well-ordering principle to establish the existence of $u^\star$ only because it holds for nonempty subsets of the set $\mathbb{Z}$ of all integers, and we assumed that there exists an upper bound of $\mathbb{N}$ in that set $\mathbb{Z}$! But what about potential upper bounds of $\mathbb{N}$ which are not integers, maybe not even rational?

To overcome this difficulty we must find an alternate way: We will again do an indirect proof and use the completeness axiom and then work with $\sup(\mathbb{N})$ to obtain a contradiction.

PROOF: Assume to the contrary that there exists an upper bound of $\mathbb{N}$ in $\mathbb{R}$. According to the completeness axiom $u^\star := \sup(\mathbb{N})$ exists as a real number. Since $u^\star$ is the least upper bound of $\mathbb{N}$, $u^\star - \frac{1}{2}$ is not an upper bound of $\mathbb{N}$. Thus there exists $n \in \mathbb{N}$ such that $n > u^\star - \frac{1}{2}$. Thus the natural number $n + 1$ exceeds the upper bound $u^\star$ of $\mathbb{N}$. Contradiction! ■

Corollary 9.3. *There are no upper bounds for $\mathbb{N}$ in $\mathbb{Q}$.*

The proof is left as exercise 9.11 (see p.303). ■

Remark 9.5. One can prove that $\mathbb{N}$ is unbounded in $\mathbb{Q}$ without utilizing the fact that $\mathbb{Q} \subseteq \mathbb{R}$ and thus without referring to the completeness axiom as follows.

Assume to the contrary that $\mathbb{N}$ is bounded in $\mathbb{Q}$, i.e., there exists $u^* \in \mathbb{Q}$ such that

$$k \leq u^* \quad \text{for all } k \in \mathbb{N}\,. \tag{*}$$

Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $u^* = \frac{m}{n}$. It follows from **(*)** and $n \geq 1$ that

$$k \leq kn \leq m \quad \text{for all } k \in \mathbb{N}.$$

But then the integer m is an upper bound for $\mathbb{N}$. This contradicts Proposition 6.25 on p.195.

Remark 9.6 (Contrasting $\mathbb{Z}$ and $\mathbb{R}$)**.**

The Integers:

- **(a)** $\mathbb{Z} = (\mathbb{Z}, +, \cdot)$ is a commutative ring with unit
- **(b)** Cancellation rule (no zero divisors: $\mathbb{Z}$ is an integral domain)
- **(c)** Ordered by the positive cone $P := \mathbb{N}$
- **(d)** Induction axiom: If $A \subseteq \mathbb{N}$ satisfies **(1)** $1 \in A$, **(2)** $\big[n \in A \Rightarrow n+1 \in A\big]$, then $A \supseteq \mathbb{N}$

The Real Numbers:

- **(a)** $\mathbb{R} = (\mathbb{R}, +, \cdot)$ is a commutative ring with unit
- **(b)** $(\mathbb{R}_{\neq 0}, \cdot)$ is an abelian group: each $x \neq 0$ has a multiplicative inverse $\frac{1}{x}$ (implies the cancellation rule, hence $\mathbb{R}$ is an integral domain)
- **(c)** Ordered by the positive cone $P := \mathbb{R}_{>0}$
- **(d)** Completeness axiom: If nonempty $A \subseteq \mathbb{R}$ has upper bounds then $\sup(A)$ exists (as an element of $\mathbb{R}$, i.e. $\sup(A) < \infty$) □

9.2 Minima, Maxima, Infima and Suprema in $\mathbb{R}$ and $\mathbb{Q}$

We had previously discussed minima, maxima, infima and suprema in ordered integral domains. See ch.3.5. We now discuss this subject specifically for the real numbers and the rational numbers.

Remark 9.7.

Let $A \subseteq \mathbb{R}$ be nonempty.

- **(a)** If A is bounded above then it follows from the completeness axiom that its least upper bound $\sup(A) = \min(A_{\mathscr{upp}})$ exists (see axiom 9.1 (Real Numbers) on p.255).
- **(b)** If A is bounded below then it follows from the completeness axiom and cor.3.4 on p.81 that its greatest lower bound $\inf(A) = \max(A_{\mathscr{low}})$ exists.

The above is the core distinction between real numbers and rational numbers. There are bounded sets of rational numbers which do not possess a supremum in $\mathbb{Q}$. □

The following counterexample to this last remark is quite similar to example 3.8**(c)** on p.78 and the reader should review it. We remind the reader that there is no rational number x such that $x^2 = 2$, i.e., $\sqrt{2}$ cannot be expressed as a quotient of two integers. [104]

[104] You will see a strict proof of this assertion in prop.9.28 on p.275.

Example 9.1. For this example we define

$$A_1 := \{x \in \mathbb{R} : x \geq 0 \text{ and } x^2 < 2\},$$
$$A_2 := \{x \in \mathbb{R} : x \geq 0 \text{ and } x^2 \leq 2\},$$
$$A_3 := \{x \in \mathbb{Q} : x \geq 0 \text{ and } x^2 < 2\},$$
$$A_4 := \{x \in \mathbb{Q} : x \geq 0 \text{ and } x^2 \leq 2\}.$$

Note that none of these sets is empty since they all contain the number zero, that each one of them has zero as its minimum (thus also as its infimum), and that each one of them is bounded above: a crude estimate would be 2.

We observe that the sets A_3 and A_4 can be considered as subsets of either the ordered field $(\mathbb{Q}, +, \cdot)$ or the ordered field $(\mathbb{R}, +, \cdot)$ since they are subsets of $\mathbb{Q}$ whereas A_1 and A_2 are subsets of $(\mathbb{R}, +, \cdot)$ but not of $(\mathbb{Q}, +, \cdot)$.

(a) When viewed as subsets of $(\mathbb{R}, +, \cdot)$, all four sets possess a supremum since they are nonempty and bounded. This follows from the completeness axiom. One can show that **(i)** $\sup(A_j)^2 = 2$, i.e., $\sup(A_j) = \sqrt{2}$ for $j = 1, 2, 3, 4$, and that **(ii)** $\sqrt{2} \in A_2$. [105] Thus $\max(A_2)$ exists (and equals $\sup(A_2) = \sqrt{2}$ because max and sup always coincide if the max exists). We claim that $\max(A_3)$ and $\max(A_4)$ do not exist: Let $j = 3, 4$. From $\max(A_j) \in A_j \subseteq \mathbb{Q}$ we obtain $\max(A_j) \in \mathbb{Q}$, i.e., $\sqrt{2} \in \mathbb{Q}$. This contradicts the fact that $\sqrt{2}$ is irrational.
Does $\max(A_1)$ exist? If so then we have

$$\max(A_1) = \sup(A_1) = \sqrt{2}\,, \quad \text{thus} \quad \sqrt{2} \in A_1\,.$$

But this cannot be true since we also have

$$\left(\sqrt{2}\right)^2 = 2, \quad \text{thus} \quad \left(\sqrt{2}\right)^2 \not< 2, \quad \text{thus (by definition of } A_1) \quad \sqrt{2} \notin A_1\,.$$

(e) Matters are very simple for subsets of the integers: It follows from the (extended) well–ordering principle that a nonempty subset of $\mathbb{Z}$ possesses a supremum ⇔ it possesses a maximum ⇔ it is bounded above. Let us see what happens for

$$A_5 := \{x \in \mathbb{Z} : x \geq 0 \text{ and } x^2 < 2\},$$
$$A_6 := \{x \in \mathbb{Z} : x \geq 0 \text{ and } x^2 \leq 2\}.$$

Both are nonempty and bounded subsets of $\mathbb{Z}$. It follows from the extended well–ordering principle that both possess min and max, hence also inf and sup. This is indeed true since $A_5 = A_6 = \{0, 1\}$, thus

$$\min(A_5) = \inf(A_5) = \min(A_6) = \inf(A_6) = 0$$
$$\max(A_5) = \sup(A_5) = \max(A_6) = \sup(A_6) = 1\,. \quad \square$$

[105] See prop.9.25 on p.274 which gives an alternate proof of B/G thm.10.25

Example 9.2 (Example a: Maximum exists)**.** Let $f(x) := 2x$, and $X_1 := \{f(x) : 0 \leq x \leq 1\}$. For each $0 \leq x \leq 1$ we have $f(x) = 2x$, and the biggest possible value is $f(1) = 2$. So the maximum of X_1 exists, and it equals $\max(X_1) = \max\{f(x) : 0 \leq x \leq 1\} = 2$. □

Example 9.3 (Example b: Supremum is finite)**.** Let $f(x) := 2x$, and $X_2 := \{f(x) : 0 \leq x < 1\}$, i.e., we now exclude the right end point 1 at which the maximum value was attained in the previous example. For each $0 \leq x < 1$ we have $f(x) < 2$, so 2 is an upper bound of X_2, hence $\sup(X_2)$ exists and is at most 2. We recall from calculus that the function $f(x) = 2x$ is continuous and hence continuous from the left at $x_0 = 1$. [106] In other words, $f(x)$ will "approach" the value 2 as x approaches $x = 1$ from the left, and it follows that no number less than 2 is an upper bound of X_2, hence $\sup(X_2) = \sup\{f(x) : 0 \leq x < 1\} = 2$.

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: There must be an element $a \in A$ so that $a = m$.

For the supremum it is sufficient that there is a sequence $(a_n)_n$ in A 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 now, because this would require us to delve into the concepts of convergence and contact points. [107] □

Example 9.4 (Example c: Supremum is infinite)**.** Let $f(x) := 2x$, and $X_3 := \{f(x) : x \geq 0\}$. The value $2x$ will exceed all potential upper bounds, and that means that the only reasonable value for $\sup(X_3) = \sup\{f(x) : \ x \geq 0\}$ is $+\infty$.

As in case b above, the max does not exist because there is no $x_0 \in X_3$ such that $f(x_0)$ attains the highest possible value among all $x \in [0, \infty[$. □

Proposition 9.9. *Let $A \subseteq B \subseteq \mathbb{R}$. Then $\inf(A) \geq \inf(B)$ and $\sup(A) \leq \sup(B)$.*

PROOF: The above was proven for ordered integral domains R in prop.3.58 on p.80 under the condition that $\inf(A)$, $\inf(B)$, $\sup(A)$, $\sup(B)$ exist, possibly having value $\pm\infty$. But $\inf(\Gamma)$ exists for any subset Γ of $\mathbb{R}$: If Γ is nonempty and has lower bounds then this follows from the completeness axiom. If $\Gamma = \emptyset$ then $\inf(\Gamma) = \infty$. Otherwise (Γ is not empty and not bounded below) $\inf(\Gamma) = -\infty$. ■

Proposition 9.10 (Supremum and infimum are positively homogeneous)**.**

Let A be a nonempty subset of $\mathbb{R}$ and let $\lambda \in \mathbb{R}_{\geq 0}$. If $\lambda > 0$ or if $\lambda = 0$ and $\sup(A) < \infty$ then

(9.1) *If $\lambda > 0$ or if $\lambda = 0$ and $\sup(A) < \infty$* *then* $\sup(\lambda A) = \lambda \sup(A)$,

(9.2) *If $\lambda > 0$ or if $\lambda = 0$ and $\inf(A) > -\infty$* *then* $\inf(\lambda A) = \lambda \inf(A)$.

PROOF: We only give the proof for the supremum. The proof of (9.2) is similar.

[106] We will get to that in a few pages in def. 9.12 on p.271 (Continuity in $\mathbb{R}$) and rem.9.12 on p.9.12 about one–sided continuity.

[107] We will get to that in ch.**??** on metric spaces.

(9.1) holds for $\lambda = 0$ and $\sup(A) < \infty$ because

$$\sup(0A) \ = \ \sup(\{0\}) \ = \ 0 \ = \ 0 \cdot \sup(A).$$

Thus we may assume $\lambda > 0$. The set $A_{uppb} \ = \ \{u \in \mathbb{R} : u$ is upper bound of $A\}$ then satisfies

(9.3) $$u \in A_{uppb} \ \Leftrightarrow u \geq a \ \forall a \in A \ \Leftrightarrow \ \lambda u \geq \lambda a \ \forall a \in A \ \Leftrightarrow \ \lambda u \in (\lambda A)_{uppb}.$$

Case 1: A is unbounded.
Then λA. also in unbounded.
Thus $\sup(\lambda A) = \sup(A) = \infty$, and hence $\sup(\lambda A) = \lambda \sup(A) = \infty$.
Case 2: A is not empty.
It follows from $\sup(A) \in A_{uppb}$ that $\lambda \sup(A) \in (\lambda A)_{uppb}$,
hence $\lambda \sup(A) \geq \min\big((\lambda A)_{uppb}\big) = \sup(\lambda A)$. It remains to show that $\lambda \sup(A) \leq \sup(\lambda A)$.
We substitute $\frac{v}{\lambda}$ for u in (9.3) and obtain

$$\frac{v}{\lambda} \in A_{uppb} \ \Leftrightarrow \ v \in (\lambda A)_{uppb}\,.$$

It follows from $\sup(\lambda A) \in (\lambda A)_{uppb}$ that

$$\frac{\sup(\lambda A)}{\lambda} \in A_{uppb}, \quad \text{hence} \ \frac{\sup(\lambda A)}{\lambda} \ \geq \ \min\big(A_{uppb}\big) \ = \ \sup(A)\,, \quad \text{hence} \ \ \sup(\lambda A) \ \geq \ \lambda \sup(A)\,.$$

This proves $\lambda \sup(A) \leq \sup(\lambda A)$. ■

Definition 9.5 (bounded functions). $\boxed{\star}$

Given are a nonempty set X and a real-valued function f with domain X.

We call f **bounded above** if the image $f(X) = \{f(x) : x \in X\}$ is bounded above, i.e., if there exists a (possibly very large) number $\gamma_1 > 0$ such that

(9.4) $$f(x) \quad < \quad \gamma_1 \qquad \text{for all arguments } x.$$

We call f **bounded below** if the image $f(X) = \{f(x) : x \in X\}$ is bounded below, i.e., if there exists $\gamma_2 > 0$ such that

(9.5) $$f(x) \quad > \quad -\gamma_2 \qquad \text{for all arguments } x.$$

We call f a **bounded function** if it is both bounded above and below, i.e., if there exists $\gamma > 0$ such that

(9.6) $$|f(x)| \quad < \quad \gamma \qquad \text{for all arguments } x. \ \square$$

We note that f is bounded if and only if its range $f(X)$ is a bounded subset of $\mathbb{R}$. We further 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 9.6 (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–valued function on X. Consider in the following $f(A) = \{f(x) : x \in A\}$, the image of A under f.

The **supremum of** $f(\cdot)$ **on** A is defined as

$$\sup_A f \;:=\; \sup_{x\in A} f(x) \;:=\; \sup f(A) \tag{9.7}$$

The **infimum of** $f(\cdot)$ **on** A is defined as

$$\inf_A f \;:=\; \inf_{x\in A} f(x) \;:=\; \inf f(A). \;\square \tag{9.8}$$

Definition 9.7 (supremum and infimum of families)**.**

The **supremum** and **infimum** of a family of real numbers $(x_i)_{i\in I}$ $(x_i)_{i\in I}$ are defined as

$$\sup\,(x_i) \;:=\; \sup_i\,(x_i) \;:=\; \sup\,(x_i)_i \;:=\; \sup\,(x_i)_{i\in I} \;:=\; \sup_{i\in I} x_i \;:=\; \sup\,\{x_i : i \in I\}. \tag{9.9}$$

$$\inf\,(x_i) \;:=\; \inf_i\,(x_i) \;:=\; \inf\,(x_i)_i \;:=\; \inf\,(x_i)_{i\in I} \;:=\; \inf_{i\in I} x_i \;:=\; \inf\,\{x_i : i \in I\}. \;\square \tag{9.10}$$

The definition above for families extends to sequences.

Definition 9.8 (supremum and infimum of sequences)**.**

Let $I = [k_0, \infty[_{\mathbb{Z}}$ and $x_n \in \mathbb{R}$ for $n \in I$. **Supremum** and **infimum** of $(x_n)_{n\in I}$ are defined as

$$\sup\,(x_n) \;:=\; \sup\,(x_n)_{n\in I} \;:=\; \sup_{n\in I} x_n \;=\; \sup\,\{x_n : n \in I\} \tag{9.11}$$

$$\inf\,(x_n) \;:=\; \inf\,(x_n)_{n\in I} \;:=\; \inf_{n\in I} x_n \;=\; \inf\,\{x_n : n \in I\}. \;\square \tag{9.12}$$

We note that the "duality principle" for min and max, sup and inf (see prop.3.59, prop.3.60 and cor.3.4 on p.80) 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 9.11.

Let X be a nonempty set and $\varphi, \psi : X \to \mathbb{R}$. Let $\emptyset \neq A \subseteq X$. Then

$$\sup\{\varphi(x) + \psi(x) : x \in A\} \;\leq\; \sup\{\varphi(y) : y \in A\} \;+\; \sup\{\psi(z) : z \in A\}, \tag{9.13}$$

$$\inf\{\varphi(x) + \psi(x) : x \in A\} \;\geq\; \inf\{\varphi(y) : y \in A\} \;+\; \inf\{\psi(z) : z \in A\}. \tag{9.14}$$

PROOF:
We only prove (9.13). The proof of (9.14) is similar and left as exercise 9.12 on p.303. [108]

Let $U := \{\varphi(x) + \psi(x) : x \in A\}$, $\;V := \{\varphi(y) : y \in A\}$, $\;W := \{\psi(z) : z \in A\}$. Let $x \in A$.

[108] (9.14) can also be deduced from (9.13) and the fact that $\inf\{\varphi(u) : u \in A\} \;=\; -\sup\{-\varphi(v) : v \in A\}$.

Then $\varphi(x) \leq \sup(V)$ since $\sup(V)$ is an upper bound of V, and $\psi(x) \leq \sup(W)$ since $\sup(W)$ is an upper bound of W, thus $\sup(V) + \sup(W) \geq \varphi(x) + \psi(x)$.

Since this is true for all $x \in A$, we conclude that $\sup(V) + \sup(W)$ is an upper bound of U.

Thus $\sup(V) + \sup(W)$ dominates the least upper bound $\sup(U)$ of U, and this proves (9.13). ■

9.3 Convergence and Continuity in $\mathbb{R}$

You are familiar from calculus with the concepts of convergent sequences and continuous functions whose domain and codomain both are sets of real numbers. We discuss them here in a more rigorous fashion. Convergence and continuity will be generalized in later chapters from $\mathbb{R}$ to so–called metric spaces.

At the start we need to give a definition which makes precise the thought that the sequence (x_n) of real numbers has limit a if eventually [109] all of the x_n will come arbitrarily close to a.

Definition 9.9 (convergence of sequences of real numbers)**.** Let $a \in \mathbb{R}$. We say that a sequence (x_n) of real numbers **converges** [110] to a for $n \to \infty$ if the following is true:

For any $\delta \in]0, \infty[$ (no matter how small), there exists $n_0 \in \mathbb{N}$ such that

$$|a - x_j| < \delta \quad \text{for all } j \geq n_0. \tag{9.15}$$

We write either of

$$a = \lim_{n\to\infty} x_n \qquad \text{or} \qquad x_n \to a \text{ as } n \to \infty \tag{9.16}$$

and we call a the **limit** of the sequence (x_n). □

Remark 9.8. **(a)** The smaller the number δ, the harder it is to enforce the validity of $|a - x_j| < \delta$, the larger n_0 may have to be chosen.

For example, let $x_n := \dfrac{n+1}{n}$, $a := 1$.

If $\delta = 1/10$ then we may choose any $n_0 \geq 11$, since

$$j \geq 11 \Rightarrow |x_j - 1| = \frac{1}{j} \leq \frac{1}{n_0} \leq \frac{1}{11} < \frac{1}{10} = \delta\,.$$

On the other hand, if $\delta = 1/100$ then we need n_0 to be 101 or bigger, since

$$|x_j - 1| \leq \frac{1}{100} \Leftrightarrow \frac{1}{j} > 100 \Leftrightarrow \frac{1}{j} \geq 101\,. \ \square$$

(b) Definition 9.9 can be worded as follows:

- For any $\delta > 0$, $|a - x_j| < \delta|$, **eventually**.

[109] See Definition 7.4 on p.223 and the subsequent Proposition 7.11 for the meaning of "eventually".

[110] We will define convergence of a sequence of items more general than real numbers in ch.12.4 (see Definition 12.10 (convergence of sequences in metric spaces) on p.369).

Definition 9.10 (Open ε-Neighborhood in $\mathbb{R}$). [111]

For $x_0 \in \mathbb{R}$ and $\varepsilon > 0$, let

$$N_\varepsilon(x_0) \ := \,]x_0 - \varepsilon, x_0 + \varepsilon[\, = \ \{x \in \mathbb{R} : |x - x_0| < \varepsilon\}$$

be the set of all elements of $\mathbb{R}$ with a distance to x_0 of strictly less than the number ε (the open interval with center x_0 and radius ε from which the points on the boundary (those with distance equal to ε) are excluded).

(a) We call $N_\varepsilon(x_0)$ the ε–**neighborhood** of x_0. $N_\varepsilon(x_0)$ is often called the **open** ε–**neighborhood** of x_0, to differentiate it from the closed interval $[\, x_0 - \varepsilon, x_0 + \varepsilon \,]$, which is also called the **closed** ε–neighborhood of x_0.

(b) Let $x, y \in \mathbb{R}$ and $\varepsilon > 0$. We say that x and y are ε–**close** if $|x - y| < \varepsilon$. □

Remark 9.9. Clearly two real numbers x and y are ε–close $\Leftrightarrow$ $x \in N_\varepsilon(y)$ $\Leftrightarrow$ $y \in N_\varepsilon(x)$ □

There are two equivalent ways of expressing convergence to $a \in \mathbb{R}$:

(a) No matter how small a δ–neighborhood of a you choose: at most finitely many of the x_n will be located outside that neighborhood.

(b) No matter how small a δ–neighborhood of a you choose: eventually all of the x_n will be found inside that neighborhood.

Example 9.5. Some simple examples for convergence to a real number:

(a) Let $x_n := 1/n$ $(n \in \mathbb{N})$. Then $x_n \to 0$ as $n \to \infty$.

(b) Let $\alpha \in \mathbb{R}$ and $z_n := \alpha^2\pi$ $(n \in \mathbb{N})$. Then the sequence $(z_n)_n$ has limit $\alpha^2\pi$.

(c) More generally let $z_n := x_0$ for some $x_0 \in \mathbb{R}$ $(n \in \mathbb{N})$. Then $\lim\limits_{n\to\infty} z_n = x_0$. □

PROOF of **(a)**: If $\delta > 0$, let n_0 := some integer larger than $1/\delta$. Such a number exists because the natural numbers are not bounded above (see thm.9.4 on p.257). It follows for $n \geq n_0$ that

$$|x_n - 0| \ = \ 1/n \leq 1/n_0 \ < \ \delta.$$

PROOF of **(b)** and **(c)**: **(c)** is left as exercise 9.14 on p.303, and **(b)** follows from **(c)**. ■

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 to $+\infty$ and very small numbers (i.e., very big ones with a minus sign) as being close to $-\infty$.

Definition 9.11 (Limit infinity). $\boxed{\star}$ Given a real number $K > 0$, we define

(9.17a) $$N_K(\infty) := \{x \in \mathbb{R} : x > K\}$$

(9.17b) $$N_K(-\infty) := \{x \in \mathbb{R} : x < -K\}$$

[111] This will generalized to metric spaces in Definition 12.6 on p.366.

We call $N_K(\infty)$ the **K–neighborhood of ∞** and $N_K(-\infty)$ the **K–neighborhood of $-\infty$**. We say that a sequence (x_n) has limit ∞ and we write either of

(9.18) $$x_n \to \infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = \infty$$

if the following is true for any $K \in \mathbb{R}$ (no matter how big): There is an integer n_0 such that all x_j belong to $N_K(\infty)$ for all $j \geq n_0$, i.e., if

for all $K \in \mathbb{N}$ there exists $n_0 \in \mathbb{N}$ such that if $j \geq n_0$ then $x_j > K$.

We say that the sequence (x_n) has limit $-\infty$ and we write either of

(9.19) $$x_n \to -\infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = -\infty$$

if the following is true for any $K \in \mathbb{R}$ (no matter how big): There is an integer n_0 such that all x_j belong to $N_K(-\infty)$ for all $j \geq n_0$. □

(a) There is an equivalent way of stating that the sequence (x_n) has limit ∞: No matter how big a threshold $K > 0$ you choose: eventually all of the x_n will be located above that threshold.
(b) $x_n \to -\infty$ can also be expressed as follows: No matter how big a threshold $K > 0$ you choose: eventually all of the x_n will be located below $-K$.

Remark 9.10.

The majority of mathematicians agrees that there is no "convergence to ∞" or "divergence to ∞". Rather, they say that a sequence has the limit ∞. We follow this convention □

Theorem 9.5 (Limits are uniquely determined)**.**

Let $(x_n)_n$ be a convergent sequence of real numbers. Then its limit is uniquely determined.

PROOF: The proof is left as exercise 9.15 (see p.303). ■

Proposition 9.12 (B/G prop.10.11)**.** *Let $a, b \in \mathbb{R}$. Then $a = b \;\Leftrightarrow\; |a - b| < \varepsilon$ for all $\varepsilon > 0$.*

PROOF: The proof is left as exercise 9.16 (see p.303). ■

Proposition 9.13 (Subsequences of real number sequences with limits)**.**

Let $(x_n)_n$ be a sequence of real numbers with limit $L := \lim\limits_{n\to\infty} x_n$. Let (x_{n_j}) be a subsequence. Then $\lim\limits_{j\to\infty} x_{n_j} = L$.

PROOF: The proof is done separately for $L \in \mathbb{R}$ and for $L = \pm\infty$.

(a) Assume that $(x_n)_n$ is convergent, i.e., $L \in \mathbb{R}$. Let $\varepsilon > 0$. Because the sequence converges, there exists $N \in \mathbb{N}$ such that $|x_j - L| < \varepsilon$ for all $j \geq N$. As $n_j \geq j$ for all j, we conclude that $n_j \geq N$ whenever $j \geq N$, hence $|x_{n_j} - L| < \varepsilon$ for all $j \geq N$. It follows that (x_{n_j}) has limit L.

(b) Assume that $L = \infty$. Let $K \in \mathbb{R}$. Because $\lim\limits_{n\to\infty} x_n = \infty$, there exists $N \in \mathbb{N}$ such that $x_j > K$ for all $j \geq N$. As $n_j \geq j$ for all j and hence $n_j \geq N$ whenever $j \geq N$, we obtain $x_{n_j} > K$ for all $j \geq N$. It follows that (x_{n_j}) has limit ∞.

(c) The case $L = -\infty$ is proved similarly to **(b)**. ■

Note that the above in particular means that subsequences of a convergent sequence converge to the same limit.

Note 9.2 (Notation for limits of monotone sequences)**.**

Let (x_n) be a nondecreasing and y_n a nonincreasing sequence of real numbers.

(a) If $\xi = \lim\limits_{k\to\infty} x_k$ (that limit might be $+\infty$), then we write • $x_n \uparrow \xi \quad (n \to \infty)$

(b) If $\eta = \lim\limits_{j\to\infty} y_j$ (that limit might be $-\infty$), then we write • $y_j \downarrow \eta \quad (j \to \infty)$. □

Proposition 9.14. *[See B/G prop.10.16]*

Let $(x_n)_n$ be a sequence of real numbers such that $\lim\limits_{n\to\infty} x_n$ exists. Let $K \in \mathbb{N}$. For $n \in \mathbb{N}$ let $y_n := x_{n+K}$. Then $(y_n)_n$ has the same limit as $(x_n)_n$.

The proof is left as exercise 9.17 (see p.303). ■

Proposition 9.15 (convergent ⇒ bounded)**.**

Let $(x_n)_n$ be a sequence in $\mathbb{R}$.

- *If the sequence converges, then it is bounded.*

The proof is left as exercise 9.18 (see p.303). ■

The following proposition states that the product of a sequence which converges to zero and a bounded sequence converges to zero.

Proposition 9.16 (bounded times zero–convergent is zero–convergent)**.**

Let $(x_n)_n$ and $(\alpha_n)_n$ be two sequences in $\mathbb{R}$ and let $\alpha \in \mathbb{R}$.

- *If $\lim\limits_{n\to\infty} x_n \ = \ 0$ and if $|\alpha_j| \leq \alpha$ for all $j \in \mathbb{N}$, then*

$$\lim_{j\to\infty} (\alpha_j x_j) \ = \ 0. \tag{9.20}$$

PROOF:

Case 1: $\alpha = 0$. Then $\alpha_j = 0$ and hence $\alpha_j x_j = 0$ for all $j \in \mathbb{N}$. For any $\delta > 0$ let $n_0 = 1$. Then

$$|\alpha_j x_j - 0| \ = \ |\alpha_j x_j| \ = \ 0 < \delta \quad \text{for all } j \in \mathbb{N} \text{ such that } j \geq n_0.$$

This proves convergence $\alpha_j x_j \to 0$.

Case 2: $\alpha \neq 0$, i.e., $|\alpha| > 0$. Let $\delta > 0$. We must show that

$$\text{there is } n_0 \in \mathbb{N} \text{ such that } |\alpha_j x_j| < \delta \text{ for all } j \in \mathbb{N} \text{ such that } j \geq n_0. \tag{9.21}$$

Let $\varepsilon := \delta/|\alpha|$. Then $\varepsilon > 0$ and it follows from $\lim\limits_{j\to\infty} x_j = 0$ that

(9.22) there is $N \in \mathbb{N}$ such that $|x_j| < \varepsilon$ for all $j \in \mathbb{N}$ such that $j \geq N$.

Since also $|\alpha_j \leq \alpha$ for all j we obtain

$$|\alpha_j\, x_j| \ = \ |\alpha_j| \cdot |x_j| \ < \ |\alpha| \cdot \varepsilon \ = \ \delta \quad \text{for all } j \in \mathbb{N} \text{ such that } j \geq N\,.$$

We choose $n_0 := N$ and (9.21) follows. ■

It is very rare that you need to apply Definition 9.9 on p.263 to compute a limit. Rather, the previous proposition and the following set of rules are employed.

Proposition 9.17 (Rules of arithmetic for limits [112] **).**

Let $(x_n)_n$ and $(y_n)_n$ be sequences in $\mathbb{R}$ and $x, y, \alpha \in \mathbb{R}$. Let $\lim\limits_{j\to\infty} x_j = x$ and $\lim\limits_{j\to\infty} y_j = y$. Then

(a) $\lim\limits_{j\to\infty} \alpha = \alpha$,

(b) $\lim\limits_{j\to\infty} (\alpha \cdot x_j) = \alpha \cdot x$, *(constant sequence)*

(c) $\lim\limits_{j\to\infty} (x_j + y_j) = x + y$,

(d) $\lim\limits_{j\to\infty} (x_j \cdot y_j) = x \cdot y$,

(e) *if $x \neq 0$ then* $\lim\limits_{j\to\infty} \dfrac{1}{x_j} = \dfrac{1}{x}$.

PROOF of **(a)**: Exercise 9.13.

PROOF of **(b)**:

Case 1: $\alpha = 0$. Then $\alpha\, x_j$ is the constant sequence $0, 0, \dots$ which converges to $0 = \alpha\, x$ — Done.

Case 2: $\alpha \neq 0$. Let $\delta > 0$. We must show that

(9.23) there is $n_0 \in \mathbb{N}$ such that $|\alpha\, x_j - \alpha\, x| < \delta$ for all $j \in \mathbb{N}$ such that $j \geq n_0$.

Let $\varepsilon := \delta/|\alpha|$. Then $\varepsilon > 0$ and it follows from $\lim\limits_{j\to\infty} x_j = x$ that

(9.24) there is $N \in \mathbb{N}$ such that $|x_j - x| < \varepsilon$ for all $j \in \mathbb{N}$ such that $j \geq N$.

But then $|\alpha\, x_j - \alpha\, x| \ = \ |\alpha| \cdot |x_j - x| \ < \ |\alpha| \cdot \varepsilon \ = \ \delta$ for all $j \in \mathbb{N}$ such that $j \geq N$. We choose $n_0 := N$ and (9.23) is proved.

PROOF of **(c)**:

Let $\delta > 0$. It follows from $\lim\limits_{j\to\infty} x_j = x$ and $\lim\limits_{j\to\infty} y_j = y$ that there exist $N_1, N_2 \in \mathbb{N}$ such that

(9.25) if $j \geq N_1$ then $|x_j - x| < \delta/2$ and if $j \geq N_2$ then $|y_j - y| < \delta/2$.

It follows from the triangle inequality $|A + B| \leq |A| + |B|$ (prop.2.7 on p.45) and from (9.25) that

(9.26) $$|(x_j + y_j) - (x + y)| \ = \ |(x_j - x) + (y_j - y)| \ \leq \ |x_j - x| + |y_j - y)| < \delta/2 + \delta/2 = \delta$$

[112] See [2] B/G (Beck/Geoghegan) prop.10.23

for all $j \geq \max(N_1, N_2)$. Let $n_0 := \max(N_1, N_2)$. It follows from (9.26) that $|(x_j + y_j) - (x+y)| < \delta$ for all $j \geq n_0$. This proves **(c)**.

PROOF of **(d)**:
Let $u_j := (x_j - x)y_j$ and $v_j := x(y_j - y)$ $(j \in \mathbb{N})$. Then

$$x_jy_j - xy \;=\; (x_jy_j - xy_j) + (xy_j - xy) \;=\; (x_j - x)y_j \;+\; x(y_j - y) \;=\; u_j + v_j. \tag{9.27}$$

It follows from parts **(c)** and **(a)** that

$$\begin{aligned}\lim_{j\to\infty}(x_j - x) \;&=\; \lim_{j\to\infty} x_j \;-\; \lim_{j\to\infty} x \;=\; x - x \;=\; 0,\\ \lim_{j\to\infty}(y_j - y) \;&=\; \lim_{j\to\infty} y_j \;-\; \lim_{j\to\infty} y \;=\; y - y \;=\; 0,\end{aligned}$$

i.e., the sequences $x_n - x$ and $y_n - y$ converge to zero.

Moreover The convergent sequences y_n and x (constant sequence!) are bounded by prop.9.15. It now follows from prop.9.16 that $\lim\limits_{j\to\infty} u_j = 0$ and $\lim\limits_{j\to\infty} v_j = 0$. We deduce from (9.27) $x_jy_j = xy + u_j + v_j$ is the sum of three convergent sequences. [113] It follows from part **(c)** that

$$\lim_{j\to\infty} x_ny_n \;=\; \lim_{j\to\infty}(xy) \;+\; \lim_{j\to\infty} u_j \;+\; \lim_{j\to\infty} v_j \;=\; xy + 0 + 0 \;=\; xy.$$

PROOF of **(e)**:
Since $\lim\limits_{j\to\infty} x_n = x$ and $|x| > 0$ there exists $N_1 \in \mathbb{N}$ such that $|x_n - x| \leq |x|/2$ for all $j \geq N_1$. Thus

$$\begin{aligned} |x| = |(x - x_n) + x_n| \;&\leq\; |x - x_n| + |x_n| \;\leq\; \frac{|x|}{2} + |x_n|\\ \Rightarrow \frac{|x|}{2} \;\leq\; |x_n| \;\Rightarrow\; |x|\,|x_n| \;&\geq\; \frac{x^2}{2} \;\Rightarrow\; \frac{1}{|xx_n|} \;\leq\; \frac{2}{x^2}\,. \end{aligned} \tag{9.28}$$

Let $z_n := (x\,x_n)^{-1}$ and $K := \max(2/x^2, |z_1|, |z_2|, \dots, |z_{N_1}|)$. It follows from (9.28) that the sequence $(z_n)_n$ is bounded by K, it follows from part **(a)** that $\lim\limits_{j\to\infty} x = x$, and from part **(c)** that $\lim\limits_{j\to\infty}(x_n - x) = \lim\limits_{j\to\infty}(x_n) - x = 0$, hence $z_n\,(x - x_n) \to 0$ as $n \to \infty$ by prop.9.16. Thus

$$\lim_{j\to\infty}\left(\frac{1}{x_n} - \frac{1}{x}\right) \;=\; \lim_{j\to\infty} \frac{1}{x\,x_n}\cdot(x \;-\; x_n) \;=\; \lim_{j\to\infty} z_n\,(x - x_n) \;=\; 0. \tag{9.29}$$

Let $\delta > 0$. On account of (9.29) there exists $n_0 \in \mathbb{N}$ such that

$$\left|\frac{1}{x_n} - \frac{1}{x}\right| \;=\; \left|\left(\frac{1}{x_n} - \frac{1}{x}\right) - 0\right| \;<\; \delta \;\text{ for all } j \geq n_0. \tag{9.30}$$

This proves convergence of $1/x_n$ to $1/x$. ■

[113] The constant sequence (xy) has limit xy according to part **(a)**

Proposition 9.18.

(a) *Let x_n be a sequence of real numbers that is nondecreasing, i.e., $x_n \leq x_{n+1}$ for all n (see def. 22.1 on p.510), and which is bounded above. Then $\lim\limits_{n\to\infty} x_n$ exists and coincides with* $\sup\{x_n : n \in \mathbb{N}\}$

(b) *If y_n is a sequence of real numbers that is nonincreasing, i.e., $y_n \geq y_{n+1}$ for all n, and which is bounded below. Then $\lim\limits_{n\to\infty} y_n$ exists and coincides with* $\inf\{y_n : n \in \mathbb{N}\}$.

PROOF of **(a)**: Let $x := \sup\{x_n : n \in \mathbb{N}\}$. This is an upper bound of the sequence, hence $x_j \leq x$ for all $j \in \mathbb{N}$. Let $\varepsilon > 0$. x is the smallest upper bound, thus $x - \frac{\varepsilon}{2}$ is not an upper bound, hence there exists $N \in \mathbb{N}$ such that $x - \frac{\varepsilon}{2} \leq x_N$. Because $(x_n)_n$ is nondecreasing, it follows for all $j \geq N$ that $x - \varepsilon \ < \ x - \frac{\varepsilon}{2} \ \leq \ x_N \ \leq \ x_j \ \leq \ x$, hence

$\varepsilon - x > -x_j \geq -x$, hence $\ \varepsilon > x - x_j \geq 0$ for all $j \geq N$, hence $|x_j - x| = x - x_j < \varepsilon$ for all $j \geq N$.

It follows that $\lim\limits_{j\to\infty} x_j = x$, i.e., $x = \sup\limits_{n\in\mathbb{N}} x_n \ = \ \lim\limits_{j\to\infty} x_j$.

The proof of **(b)** is similar to **(a)**. (Alternate proof of **(b)**: apply **(a)** to the sequence $x_n := -y_n$.) ■

Proposition 9.19 (Domination Theorem for Limits)**.**

Let $x_n, y_n \in \mathbb{R}$ be two sequences of real numbers both of which have limits. Assume there is $K \in \mathbb{N}$ such that $x_n \leq y_n$ for all $n \geq K$. Then

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

PROOF: Let $x := \lim\limits_{n\to\infty} x_n$ and $y := \lim\limits_{n\to\infty} y_n$.

Case 1: Neither x nor y is $\pm\infty$.

We will show that $\lim\limits_{n\to\infty} (y_n - x_n) \geq 0$. This suffices to prove the proposition because, according to prop. 9.17.c,

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

We abbreviate $z_n := y_n - x_n$ and $z := \lim\limits_{n\to\infty} z_n$. We assume to the contrary that $z < 0$. Let $\varepsilon := -\frac{1}{2}z$. Then $\varepsilon > 0$. It follows from the definition of limits that there exists $N \in \mathbb{N}$ such that

$$|z_j - z| < \varepsilon, \ \text{ i.e., } |z_j + 2\varepsilon| < \varepsilon, \text{ i.e., } -\varepsilon \ < \ z_j + 2\varepsilon \ < \ \varepsilon, \text{ i.e., } -3\varepsilon \ < \ z_j \ < \ -\varepsilon \ < \ 0 \qquad (\star)$$

holds for all $j \geq N$. Let $n_0 := \max(K, N)$. We obtain for all $j \geq n_0$ from $(\star)$ that $y_j - x_j = z_j < 0$. This contradicts our assumption that $x_j \leq y_j$ for all such j.

Case 2: $x = y = \pm\infty$ or $x = -\infty, y = \infty$:

The proposition is obviously true in those cases.

Case 3: $x = \infty, y = -\infty$:

This is the only case not yet covered. If this is possible then the proposition is false, so our task is to prove that this case cannot occur. We will do so by showing that if $x = \infty$ then $y = \infty$.

Let $\gamma > 0$. Since $x = \lim\limits_n x_n = \infty$ there exists, according to Definition 9.11 (Limit infinity) on p.264, an index n_0 such that $x_j > \gamma$ whenever $j \geq n_0$. Let $N := \max(n_0, K)$. We assumed $x_j \leq y_j$ for all $j \geq K$, thus $y_j \geq x_j > \gamma$ for all $j \geq N$. It follows that $y = \lim\limits_n y_n = \infty$. ■

Corollary 9.4.

Let $x_n, y_n \in \mathbb{R}$ be two sequences of real numbers and $L \in \mathbb{R}$. Assume there is $K \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq K$. Then

$$\lim_{n\to\infty} x_n = L \Leftrightarrow \lim_{n\to\infty} y_n = L, \qquad \lim_{n\to\infty} x_n = \pm\infty \Leftrightarrow \lim_{n\to\infty} y_n = \pm\infty.$$

PROOF:

This is an immediate consequence of the Domination Theorem for Limits. ■

We now give an example that utilizes both convergemce of sequences and (countably) infinitely many unions and intersections.

Proposition 9.20. *For the following note that $[u,v] = \emptyset$ for $u > v$ and $]u,v[= \emptyset$ for $u \geq v$ (see (2.19) on p.27). Let $a, b \in \mathbb{R}$. Then*

$$[a,b] = \bigcap_{n\in\mathbb{N}} \Big]a - \frac{1}{n}, b + \frac{1}{n}\Big[. \tag{9.31}$$

$$]a,b[= \bigcup_{n\in\mathbb{N}} \Big[a + \frac{1}{n}, b - \frac{1}{n}\Big], \tag{9.32}$$

PROOF: ★

(A) PROOF of (9.31)

Case 1: We assume that $a > b$. Then $[a,b] = \emptyset$, hence (9.31) is valid if we can show that there exists $N \in \mathbb{N}$ such that $]a - \frac{1}{N}, b + \frac{1}{N}[$ is empty. We do this as follows. Let $\varepsilon := \frac{a-b}{2}$. Then $\varepsilon > 0$. Since $\lim\limits_{n\to\infty} = 0$ there exists $N \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$ for all $n \geq N$; in particular, $\frac{1}{N} < \varepsilon = \frac{a-b}{2}$. Of course the choice of N depends on x. It follows that

$$\frac{2}{N} < a - b, \quad \text{hence} \quad b + \frac{1}{N} < a - \frac{1}{N}, \quad \text{hence} \quad \Big]a - \frac{1}{N}, b + \frac{1}{N}\Big[= \emptyset.$$

Case 2: We assume that $a = b$. Then $[a,b] = \{a\}$. Clearly $a \in]a - \frac{1}{n}, b + \frac{1}{n}[$ for all n, hence we obtain $\{a\} \subseteq \bigcap_n]a - \frac{1}{n}, b + \frac{1}{n}[$. The proof for case 2 is done if we can show that if $x \neq a$ then there exists $N \in \mathbb{N}$ such that $x \notin]a - \frac{1}{N}, b + \frac{1}{N}[$.

If $x < a$, let $\varepsilon := a - x > 0$. Since $\lim\limits_{n\to\infty} = 0$ there exists $N \in \mathbb{N}$ such that $\frac{1}{N} < \varepsilon = a - x$, i.e., $x < a - \frac{1}{N}$, hence $x \notin]a - \frac{1}{N}, b + \frac{1}{N}[$.

If $x > a$, we can similarly find some$N \in \mathbb{N}$ such that $\frac{1}{N} < x - a$, i.e., $x > a + \frac{1}{N}$, hence $x \notin]a - \frac{1}{N}, b + \frac{1}{N}[$.

Case 3: We assume that $a < b$. If $n \in \mathbb{N}$ then $]a - \frac{1}{n}, b + \frac{1}{n}[\supseteq [a,b]$, hence $[a,b] \subseteq \bigcap_{n\in\mathbb{N}}]a - \frac{1}{n}, b + \frac{1}{n}[$.

We finally show "$\supseteq$". If $x \in \bigcap_{n\in\mathbb{N}}]a - \frac{1}{n}, b + \frac{1}{n}[$ then

$$a - \frac{1}{n} < x < b + \frac{1}{n} \quad \text{for all } n \in \mathbb{N}. \tag{9.33}$$

The proof is done if we can show that this implies both $x \geq a$ and $x \leq b$.

Assume to the contrary that $x < a$. Let $\varepsilon := a - x$. Then $\varepsilon > 0$. Since $\lim\limits_{n\to\infty} = 0$ there exists $N \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$ for all $n \geq N$; in particular, $\frac{1}{N} < \varepsilon$. Of course the choice of N depends on x.

We obtain from $a - x = \varepsilon > \frac{1}{N}$ that $x < a - \frac{1}{N}$. This contradicts (9.33) and we have proved that $x \geq a$. Demonstrating that $x \leq b$ is similar. We have proved "$\supseteq$".

(B) PROOF of (9.32):

This proof is left as exercise 9.19. ■

We now briefly address continuity of functions which map real numbers to real numbers. This subject will be addressed in more detail and in more general settings in ch.13.1.1 on p.402. Let $A \subseteq \mathbb{R}$. Informally speaking, a continuous function $f : A \to \mathbb{R}$ is one whose graph in the xy–plane is a continuous line without any disconnections or gaps. This can be stated in slightly more formal terms by saying that, if the x-values are closely together then the $f(x)$-values must be closely together too.

Here is the formal definition.

Definition 9.12 (Continuity in $\mathbb{R}$ [114] **).** Let $A \subseteq \mathbb{R}$, $x_0 \in A$, and let $f : A \to \mathbb{R}$.

We say that f is **continuous at x_0** and we write

(9.34) $$\lim_{x \to x_0} f(x) = f(x_0)$$

if **any** sequence (x_n) with values in A satisfies the following:

(9.35) $$\text{if } x_n \to x_0 \text{ then } f(x_n) \to f(x_0).$$

In other words, the following must be true for any sequence (x_n) in A:

(9.36) $$\lim_{n\to\infty} x_n = x_0 \quad \Rightarrow \quad \lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x_0).$$

We say that f is **continuous** if f is continuous at x_0 for all $x_0 \in A$. □

Remark 9.11. 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 x_0 and all x_n. □

Remark 9.12 (One–sided Continuity)**.** This example will illustrate the role of the set A in the definition of continuity. If $A = [a, b]$ for two real numbers $a < b$ then continuity of f at a means according to (9.36) that $\lim\limits_{n\to\infty} f(x_n) = f(a)$ for all sequences x_n that approach a but stay on the right of a. Similarly continuity of f at b means that $\lim\limits_{n\to\infty} f(x_n) = f(b)$ for all sequences x_n that approach b but stay on the left of b.

The notation commonly used in those cases is

$$\lim_{x\to a+} f(x) = f(a) \quad \text{and} \quad \lim_{x\to b-} f(x) = f(b)\,.$$

One says in the first case that f is **continuous from the right** at a and in the second case that f is **continuous from the left** at b. □

[114] This definition is about **sequence continuity** of a function. See Definition 13.1 (Sequence continuity) on p.402 where continuity is generalized to metric spaces.

Proposition 9.21. *Let $A \subseteq \mathbb{R}$ and $\gamma \in \mathbb{R}$. The following functions $A \to \mathbb{R}$ are continuous.*

(a) *The constant function $x \mapsto \gamma$,*
(b) *The identity function $id\big|_A : x \mapsto x$.*

PROOF of **(a)**: Let $a, x_n \in A$ such that $\lim\limits_{n\to\infty} x_n = a$. Then $f(x_n) = \gamma$ for all n, and this constant sequence converges to $f(a) = \gamma$. according to example 9.5**(c)**.
PROOF of **(b)**: Let $a, x_n \in A$ such that $\lim\limits_{n\to\infty} x_n = a$. Then $f(x_n) = x_n$ for all n and thus $\lim\limits_{n\to\infty} f(x_n) = \lim\limits_{n\to\infty} x_n = a = f(a)$. ■

Theorem 9.6 (Rules of arithmetic for continuous real–valued functions with domain in $\mathbb{R}$ [115] **).**

Let $A \subseteq \mathbb{R}$ and $\alpha \in \mathbb{R}$. Assume that the functions

$$f(\cdot), g(\cdot), f_1(\cdot), f_2(\cdot), f_3(\cdot), \dots, f_n(\cdot) : A \longrightarrow \mathbb{R}$$

all are continuous at $x_0 \in A$. 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. Specifically, $\alpha f(\cdot) : x \mapsto \alpha \cdot f(x)$ is continuous at x_0. In particular $-f(\cdot) : x \mapsto -f(x) = (-1) \cdot f(x)$ is continuous at x_0.*
(c) *The sum $f + g(\cdot) : x \mapsto f(x) + g(x)$ is continuous at x_0.*
(d) *If $g(x_0) \neq 0$ then the quotient $f/g(\cdot) : x \mapsto f(x)/g(x)$ is continuous at x_0.*
(e) *Any linear combination $\sum\limits_{j=0}^{n} a_j f_j(\cdot) : x \mapsto \sum\limits_{j=0}^{n} a_j f_j(x)$ is continuous in x_0.*

PROOF: This proposition will be generalized in thm.13.3 on p.405 and a full proof will be given there. Here we only give a proof for the product fg to demonstrate how knowledge about the convergence of sequences can be employed to prove statements concerning continuity.
Let $x_n \in A$ for all $n \in \mathbb{N}$ be a sequence such that $\lim\limits_{n\to\infty} x_n = x_0$. It follows from Definition 9.12 (continuity) on p.271 that $\lim\limits_{n\to\infty} f(x_n) = f(x_0)$ and $\lim\limits_{n\to\infty} g(x_n) = g(x_0)$. Thus $\lim\limits_{n\to\infty} f(x_n)g(x_n) = f(x_0)g(x_0)$ by part **(d)** of prop.9.17 (Rules of arithmetic for limits) on p.267. ■

Proposition 9.22. *All polynomials are continuous*

The proof is left as exercise 9.23 on p.304. ■

Proposition 9.23 (The composition of continuous functions is continuous)**.**

Let $A, B \subseteq \mathbb{R}$ be nonempty, $f : A \to \mathbb{R}$ continuous at $x_0 \in A$, and $g : B \to \mathbb{R}$ continuous at $f(x_0)$. Assume further that $f(A) \subseteq B$, i.e., $f(x) \in B$ for all $x \in A$.

Then the composition $g \circ f : X \longrightarrow Y$ *is continuous at x_0.*

The proof is left as exercise 9.20 (see p.304). ■

[115] See Definition 11.7 on p.334 about linear combinations which are referred to in part **(e)**.

Theorem 9.7. *Let $A \subseteq \mathbb{R}$, $x_0 \in A$, and let $f : A \to \mathbb{R}$ be a real–valued function with domain A. Then f is continuous at x_0 if and only if for any $\varepsilon > 0$, no matter how small, there exists $\delta > 0$ such that either one of the following equivalent statements is satisfied:*

$$f\left(N_\delta(x_0) \cap A\right) \subseteq N_\varepsilon\big(f(x_0)\big)\,, \tag{9.37}$$

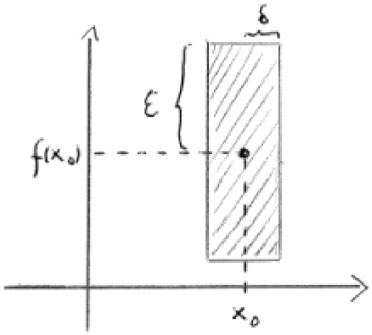
$$f\big(\{x \in A : |x - x_0| < \delta\}\big) \subseteq \{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon\}, \tag{9.38}$$

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon \textit{ for all } x \in A. \tag{9.39}$$

PROOF: A generalized version of this theorem will be proved in a later chapter. [116] ■

Remark 9.13. If we refer to the definition of continuity given in Definition 9.12 (Continuity in $\mathbb{R}$) on p.271 as **sequence continuity** and its equivalent formulation in thm.9.7 as **ε–δ continuity** then this theorem states that both formulations are equivalent, and it is thus OK to refer to either property simply as continuity.

Figure 9.1: ε-δ continuity: Any $\big(x, f(x)\big)$ such that $x_0 - \delta < x < x_0 + \delta$ must be contained in the shaded rectangle $]x_0 - \delta, x_0 + \delta[\times]f(x_0) - \varepsilon, f(x_0) + \varepsilon[$.



The following generalization of thm.9.7 allows us to to restrict our focus to small $\varepsilon > 0$ and begin the proof that a function is continuous at some point x_0 with a statement similar to this one:

"Let $\varepsilon > 0$. We may assume without restricting generality that $\varepsilon < 1$."

Being able to do so is sometimes convenient.

Proposition 9.24. *Let $A \subseteq \mathbb{R}$, $x_0 \in A$, and let $f : A \to \mathbb{R}$ be a real–valued function with domain A. Then f is continuous at x_0 if and only if there exists $\varepsilon^\star > 0$ which satisfies the following:*
for any $\varepsilon \in]0, \varepsilon^\star[$ there exists $\delta > 0$ such that either one of the following equivalent statements is satisfied:

(a) $f\big(\{x \in A : |x - x_0| < \delta\}\big) \subseteq \{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon\}$,

(b) $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$ *for all* $x \in A$.

The proof is left as exercise 9.25 (see p.304). ■

[116] See thm.13.1 on p.403.

9.4 Rational and Irrational Numbers

Proposition 9.25 (B/G thm.10.25)**.** *Let* $A := \{a \in \mathbb{R}_{>0} : a^2 < 2\}$.
Then $r := \sup(A)$ *exists and* $r^2 = 2$.

PROOF: We write $A_{\mathit{upp}\mathscr{b}}$ for the set of upper bounds of A (see Definition 3.16 on p.77).

First, we prove the existence of r.

It follows from $\frac{14}{10} > 0$ and $\left(\frac{14}{10}\right)^2 < 2$ that $A \neq \emptyset$. It follows from $\frac{15}{10} > 0$ and $\left(\frac{15}{10}\right)^2 > 2$ and $a^2 < 2$ for all $a \in A$ that $a^2 < \left(\frac{15}{10}\right)^2$.

Hence, from prop.3.49 (Generalization of B/G prop.10.5), we obtain $a < \frac{15}{10}$ for all $a \in \mathbb{R}$ and we conclude that $\frac{15}{10}$ is an upper bound of A.

This not only proves that $r = \sup(A)$ exists, but also that $\frac{14}{10} < r \leq \frac{15}{10}$

Let $x_n := r - \frac{1}{n}$ and $y_n := r + \frac{1}{n}$ $(n \in \mathbb{N})$.

Second, we prove that $x_n^2 \leq 2$ for all n.

It follows from $r > \frac{14}{10}$ and $n \geq 1$ that $x_n > 0$. It further follows from $x_n < r$ that $x_n \notin A_{\mathit{upp}\mathscr{b}}$, i.e., $x_n < a$ for some $a \in A$, hence $x_n^2 < a^2 < 2$.

Third, we prove that $y_n^2 \geq 2$ for all n.

It follows from $r + \frac{1}{n} > r$ that y_n is an upper bound of A which is different from r, its least upper bound. We show that $y_n \in A$ as follows: If $y_n \in A$, then $y_n = r + \frac{1}{n} \leq u$ for any upper bound u of A, hence $r + \frac{1}{n} \leq r$. We have reached a contradiction. So we have $y_n \notin A$ and hence, as $y_n > 0$, $y_n^2 \geq 2$.

We conclude the proof as follows. It follows from the rules of arithmetic for limits (prop.9.17) that

$$\lim_{n\to\infty} y_n^2 - \lim_{n\to\infty} x_n^2 = \lim_{n\to\infty} \frac{4r}{n} = 4r \cdot \lim_{n\to\infty} \frac{1}{n},$$

and the latter expression is zero because $\lim\limits_{n\to\infty} \frac{1}{n} = 0$ (see example 9.5.a on p.264). So y_n^2 and x_n^2 have the same limit. We established in parts 2 and 3 of the proof that $x_n^2 \leq 2 \leq y_n$. It follows from prop.9.19 (Domination Theorem for Limits) on p.269 that

$$\lim_{n\to\infty} x_n^2 \leq \lim_{n\to\infty} 2 = 2 \leq \lim_{n\to\infty} y_n^2 = \lim_{n\to\infty} x_n^2,$$

hence $\lim\limits_{n\to\infty} x_n^2 = 2$. We use again the rules of arithmetic for limits and $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0$ to conclude

$$\lim_{n\to\infty} x_n = r + \lim_{n\to\infty} \frac{1}{n} = r, \quad \text{hence,} \quad 2 = \lim_{n\to\infty} x_n^2 = \left(\lim_{n\to\infty} x_n\right)^2 = r^2. \ \blacksquare$$

Definition 9.13 (Lowest terms)**.** We repeat here the following from Definition 2.15 on page 25 of Chapter 2.

Let $q := \frac{d}{n}$ $(d, n \in \mathbb{Z}, d \neq 0)$ be a rational number. We say that d and n are a representation of q in **lowest terms** or that q is written in lowest terms if

a. d and n have no common factors,

b. $n \in \mathbb{N}$. □

We have the following simple proposition.

Proposition 9.26. *Let $q = \frac{m}{n}$ $(m, n \in \mathbb{Z}, n \neq 0)$ be a nonzero rational number. Then q is written in lowest terms if and only if $n \in \mathbb{N}$ and m and n are relatively prime.*

PROOF:

This is immediate from Proposition 6.39 on p.203 which states that two integers are relatively prime if and only if they share no common (prime) factors. ■

Proposition 9.27 (B/G prop.11.5)**.**

Let $m, n, s, t \in \mathbb{Z}$ be such that m and n do not have any common factors.

$$\textit{If}\ \ \frac{m}{n} = \frac{s}{t}\,, \qquad \textit{then}\ \ m\ \textit{divides}\ \ s\ \ \textit{and}\ \ n\ \textit{divides}\ \ t.$$

The proof is left as exercise 9.26 (see p.304). ■

Proposition 9.28 (B/G prop.11.10)**.** *The real number $\sqrt{2}$ is irrational.*

PROOF: The proof given here is from [2] Beck/Geoghegan: The Art of Proof.
Assume to the contrary that $\sqrt{2} \in \mathbb{Q}$. Then there exists a

$$\text{lowest term representation} \quad \sqrt{2} = \frac{m}{n}$$

with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Matter of fact, $m \geq 0$ since $\sqrt{2} \geq 0$. We have

$$\sqrt{2} = \frac{m}{n} \Rightarrow \frac{2n}{m} = \frac{m}{n} \overset{prop.9.27}{\Longrightarrow} n \mid m \Rightarrow \frac{m}{n} \in \mathbb{Z}\,, \quad \text{i.e., } \sqrt{2} \in \mathbb{Z}\,.$$

thus $\sqrt{2} = \frac{m}{n}$ is an integer.
Since it is true for $x, y \in [0, \infty[$ that $x < y \Leftrightarrow x^2 < y^2$ (see Proposition 3.49 on p.75) and since $1^2 = 1$, $(\sqrt{2})^2 = 2$, $2^2 = 4$, it follows from $0 < 1 < 2 < 4$ that

$$0 < 1 < \sqrt{2} < 2\,.$$

This contradicts the fact that there are no integers between 1 and 2. See Proposition 6.2 on p.171. ■

Definition 9.14 (Perfect Squares)**.**

Let $n \in \mathbb{Z}$ We call n a **perfect square** if there exists $k \in \mathbb{Z}$ such that $n = k^2$. In other words, the set of all perfect squares is the set $0, 1, 4, 9, \ldots$. □

Theorem 9.8 (B/G thm.11.12)**.**

Let $n \in \mathbb{Z}_{\geq 0}$. If n is not a perfect square then $\sqrt{n}$ is irrational.

PROOF: Left as an exercise. ■

Remark 9.14. The above theorem states that we have a dichotomy:

If n is a nonnegative integer then its square root is either an integer or irrational.

Note that the above yields another proof that $\sqrt{2}$ is irrational: 2 is not a perfect square. □

Proposition 9.29 (B/G prop.11.13)**.** *Let m and n be nonzero integers. Then $\frac{m}{n}\sqrt{2}$ is irrational.*

PROOF: Left as an exercise. ■

Theorem 9.9 (B/G ch.11: n–th root)**.** *Let n be an integer ≥ 2 and $x \in \mathbb{R}_{>0}$. Then there exists $r \in \mathbb{R}_{>0}$ such that $r^n = x$ and r is uniquely determined.*

PROOF: Left as an exercise. Adopt the proof of prop.9.25 or of B/G thm.10.25. ■

Definition 9.15 (n–th root)**.**

Let n be an integer ≥ 2 and $x \in \mathbb{R}_{>0}$. We write $\sqrt[n]{x}$ for the uniquely defined $r \in \mathbb{R}_{>0}$ such that $r^n = x$, and we extend this definition to $n = 1$ by defining $\sqrt[1]{x} := x$. We call $\sqrt[n]{x}$ the **n–th root** of x. □

Proposition 9.30 (B/G prop.11.16)**.** *Let $n \in \mathbb{Z}_{\geq 2}$. Then $\sqrt[n]{2}$ is irrational.*

PROOF: Assume to the contrary that $\sqrt[n]{2} = \frac{m}{k}$ for suitable $k, n \in \mathbb{Z}$. We may assume that k and m are in lowest terms, i.e., their prime number factorizations

$$m = p_1 p_2 \cdots p_i \text{ and } k = q_1 q_2 \cdots q_j$$

have no factors in common. The same is also true for

$$m^{n-1} = p_1^{n-1} p_2^{n-1} \cdots p_i^{n-1} \text{ and } k^{n-1} = q_1^{n-1} q_2^{n-1} \cdots q_j^{n-1},$$

hence the right hand side of $\dfrac{2k}{m} = \dfrac{m^{n-1}}{k^{n-1}}$ also is in lowest terms. It follows from B/G prop.11.5 that $k^{n-1} | m$. Further, $k | k^{n-1}$, hence $k | m$, and we conclude that $\sqrt[n]{2} = \frac{m}{k}$ is an integer. Note for the following that $r > 0$, i.e., $\frac{m}{k} > 0$.

Case 1: $\frac{m}{k}$ is zero or 1, i.e., $\sqrt[n]{2} \leq 1$, hence $2 = (\sqrt[n]{2})^n \leq 1$. We have reached a contradiction.

Case 2: $\frac{m}{k} > 1$. Then $\frac{m}{k} \geq 2$, hence

$$2 = (\sqrt[n]{2})^n = \left(\frac{m}{k}\right)^n \geq 2^n > 2\,.$$

Again we have reached a contradiction. ■

Proposition 9.31 (B/G prop.11.17)**.**

Let $x, y \in \mathbb{R}$ such that $x < y$. Then there exists irrational z such that $x < z < y$.

The proof is left as exercise 9.28 (see p.304). ■

Proposition 9.32 (B/G cor.11.18)**.** *There is no smallest positive irrational number.*

PROOF: Left as exercise 9.27 on p.304 ■

9.5 Geometric Series

The following chapter provides some basic facts about geometric series. More advanced material on series of real numbers is deferred to ch.13.2.2 (Infinite Series) on p.420 because it needs the concept of Cauchy sequences. [117]

Definition 9.16 (Real–valued Sequences and Series). $\boxed{\star}$

A sequence (a_j) is called a **real–valued sequence** if each a_j is a real number.

For any such sequence, we can build another sequence (s_n) as follows:

(9.40) $$s_1 := a_1; \quad s_2 := a_1 + a_2; \quad s_3 := a_1 + a_2 + a_3; \cdots \quad s_n := \sum_{k=1}^{n} a_k$$

We write this more compactly as

(9.41) $$a_1 + a_2 + a_3 + \cdots \;=\; \sum a_k,$$

and we call any such object which represents a sequence of partial sums a **series**. Loosely speaking, a series is a sum of infinitely many terms. We call (s_n) the sequence of **partial sums** associated with the series $\sum a_k$.

Let $s \in \mathbb{R}$. We say that the **series converges** to s and we write

(9.42) $$\sum_{k=1}^{\infty} a_k \;=\; s$$

if this is true for the associated sequence of partial sums (9.40), i.e., if $\lim\limits_{n\to\infty} s_n = s$. We then also say that the **series has limit** s.

We say that the **series has limit** $\pm\infty$ if $\lim\limits_{n\to\infty} s_n = \pm\infty$. In this case we write

(9.43) $$\sum_{k=1}^{\infty} a_k \;= \pm\infty.$$

We adopt for series the convention we did in rem.9.10 on p.265 for sequences: A series with limit $-\infty$ or ∞ never ever converges or diverges to $\pm\infty$. Instead we simply say that $\sum a_k$ diverges. □

Proposition 9.33 (Limits of Geometric Series).

(a) *Let* $|q| < 1$*. Then* $\lim\limits_{j\to\infty} q^n = 0$.

(b) (9.44) $$\sum_{j=0}^{n} q^j \;=\; \frac{1-q^{n+1}}{1-q}\,,$$

(c) (9.45) $$\sum_{j=0}^{\infty} q^j \;=\; \frac{1}{1-q}\,.$$

[117] See ch.12.10 (Completeness in Metric Spaces) on p.390

PROOF: The proof of **(b)** is a repetition of prop.6.21 on p.183, and **(c)** follows from **(a)** and **(b)**.
Proof of **(a)**: Clearly **(a)** is true for $q = 0$.
If $0 < q < 1$ then $q^{n+1} = q \cdot q^n < q^n$, i.e., the sequence q^n is strictly decreasing and bounded below by zero. It follows from prop.9.18**(b)** on p.269 that $a := \lim\limits_{n\to\infty} q^n$ exists and $0 \leq a \leq q < 1$. Moreover from prop.9.14 on p.266 and prop.9.17**(b)** on p.267,

$$a \;=\; \lim_{n\to\infty} q^n \;=\; \lim_{n\to\infty} q^{n+1} \;=\; \lim_{n\to\infty} (q \cdot q^n) \;=\; q \cdot \lim_{n\to\infty} q^n \;=\; qa,$$

i.e., $a(1-q) = 0$. Since $\mathbb{R}$ has no zero divisors and $q < 1$ this is only possible if $a = 0$, i.e., $\lim\limits_{n\to\infty} q^n = 0$.
Now assume $-1 < q < 0$. Let $\varepsilon > 0$. We just have seen that $\lim\limits_{n\to\infty} |q|^n = 0$, i.e., there exists $N \in \mathbb{N}$ such that $|\,|q|^n - 0\,| \;<\; \varepsilon$ for all $n \in \mathbb{N}$ such that $n \geq N$. But then

$$|q^n - 0| \;=\; |q^n| = \;|q|^n \;=\; |q|^n - 0 \;=\; |\,|q|^n - 0\,| \;<\; \varepsilon$$

for all $n \in \mathbb{N}$ such that $n \geq N$, i.e., $\lim\limits_{n\to\infty} q^n = 0$. ■

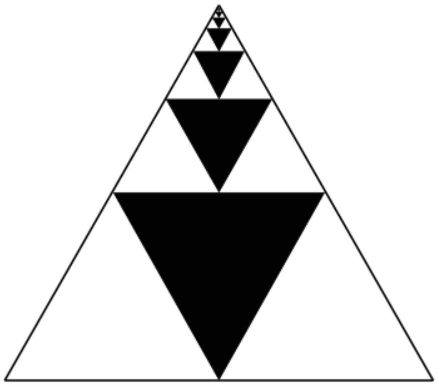
Example 9.6 (Cover page of the B/G book)**.** Examine the geometric series with $q = \dfrac{1}{4}$:

$$\sum_{j=0}^{\infty}(1/4)^j \;= \frac{4}{3} \;=\; 1 \;+\; \sum_{j=1}^{\infty}(1/4)^j, \quad \text{i.e.,} \sum_{j=1}^{\infty}(1/4)^j; \;=\; \frac{1}{3}$$

It has the following geometric meaning:

(a) each subsequent iteration has half the height (similar triangles), hence the triangles of each iteration have $1/4$ the area of the previous one. That means that if $(1/4)^j$ is the area of each of the triangles in iteration j then $(1/4) \cdot (1/4)^j$ is the area of each of the triangles in iteration $j+1$.

(b) in each horizontal slice the shaded triangle has $1/3$ the area of the entire slice. That means that the limit of the sum of the shaded triangles is $1/3$ of the total area.



The picture to the left illustrates the above. You can find it on the cover page of [2] B/G: Art of Proof. □

9.6 Decimal Expansions of Real and Rational Numbers

When we gave an informal definition of the real numbers in Definition 2.13 on p.24 of ch.2 we introduced them as decimal numerals, i.e., strings composed of an integer followed by a decimal point followed by a sequence of decimal digits. In this chapter, which closely follows [2] B/G: Art of Proof, we make precise the connection between such decimal numerals and the real numbers as we have defined them in this chapter in axiom 9.1 on p.255, i.e., as an ordered integral domain $(\mathbb{R}, +, \cdot, \mathbb{R}_{>0})$ which satisfies the completeness axiom.

Notation 9.2 (Decimal digits)**.**

Note that $[0,9]_{\mathbb{Z}}$ is according to notations 2.1 on p.28 (and also according to Definition 3.12 on p.70 equal to the set $\{0,1,2,3,4,5,6,7,8,9\}$ of decimal digits. □

Definition 9.17 (Decimal Expansion)**.**

Let $x \in \mathbb{R}_{\geq 0}$, $d_0 \in \mathbb{Z}_{\geq 0}$, and $(d_j)_{j \in \mathbb{N}}$ a sequence of decimal digits d_j such that

$$x = d_0 + \sum_{j=1}^{\infty} d_j 10^{-j} = \sum_{j=0}^{\infty} d_j 10^{-j}. \tag{9.46}$$

Then we call both the word $d_0.d_1d_2d_3\dots$ (of infinite length) and also the corresponding sequence $(d_0, d_1, d_2, \dots) = (d_j)_{j=0}^{\infty}$ a **decimal expansion** of the nonnegative real number x.

Note above and in the following definitions the "decimal period" between d_0 and d_1!

We do not distinguish between $\sum\limits_{j=0}^{\infty} d_j 10^{-j}$, $d_0.d_1d_2d_3\dots$, and $(d_j)_{j=0}^{\infty}$ and think of those expression as different notations for the same real number.

We extend the above definition to $x \in \mathbb{R}_{<0}$ as follows. If $-x$ has a decimal expansion $-x = d_0 + \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ then we call the word $-d_0.d_1d_2d_3\dots$ and also the corresponding sequence $(-d_0, d_1, d_2, \dots) = -d_0, (d_j)_{j=1}^{\infty}$ a decimal expansion of x.

We may omit leading zeros of the integer d_0 and trailing zeros of the digits $d_1, d_2, \dots$. We further may omit the decimal point together with all digits d_j to the right of that decimal point if $d_j = 0$ for all $j \in \mathbb{N}$. In other words, if $x = \sum\limits_{j=0}^{\infty} d_j 10^{-j}$ and if $d_j = 0$ for all $j \in \mathbb{N}$ then we may write either of d_0, $d_0.$, or $d_0.0$ for x. □

Remark 9.15. We have seen in ch.6.13 (The Base–β Representation of the Integers) the following. If we set $\beta := 10$ then the "integer part" d_0 of x can be written as a sum $d_0 = \sum\limits_{i=0}^{\mu} d'_i 10^i$ for appropriate $\mu \in \mathbb{Z}_{\geq 0}$ and decimal digits $d'_i \in [0,9]_{\mathbb{Z}}$. It would seem to make a lot of sense to to incorporate that base–10 expansion of the integer part d_0 into (9.46), and thus write

$$x \;=\; \sum_{j=\mu'}^{\infty} d'_j 10^{-j}.$$

for some suitable, possibly negative, $\mu' \in \mathbb{Z}$ and $d'_j \in [0,9]_{\mathbb{Z}}$ $(j \geq \mu')$, but we will stay with (9.46) and in this way remain consistent with [2] B/G (Beck/Geoghegan), ch.12.2 (Decimals) and also with Wikipedia: https://en.wikipedia.org/wiki/Decimal_representation. □

Remark 9.16.

Just as we did in ch.6.13 (The Base–β Representation of the Integers) on p.205, we could have generalized decimal representations (9.46) to representations

$$x \;=\; d_0 \,+\, \sum_{j=1}^{\infty} d_j \beta^{-j}\,,$$

i.e., we could have replaced the number 10 with $\beta \in [2,\infty[_{\mathbb{N}}$, and we could have replaced the decimal digits with base β digits $d_j \in [0,\beta-1]_{\mathbb{N}}$. We will not attempt to strive for such generality here. □

Next, we prove the existence of a decimal representation for any nonnegative real number and examine under what circumstances such a representation is not unique. The foundation for the existence and uniqueness proofs of decimal representations are the formulas of prop.9.33 for geometric series. We will see that now. Note that much of the following proposition is part of B/G prop.12.4 and prop.12.5 on p.116.

Proposition 9.34 (Geometric series for decimals). *Let $n \in \mathbb{N}$ and $d_j \in [0,9]_{\mathbb{Z}}$ for $j \geq n$. Then,*

(a) $0 \;\leq\; 9 \displaystyle\sum_{j=n}^{\infty} 10^{-j} \;=\; \frac{1}{10^{n-1}}$,

(b) $\displaystyle\sum_{j=n}^{\infty} d_j 10^{-j} \;\leq\; \frac{1}{10^{n-1}}$,

(c) $\displaystyle\sum_{j=n}^{\infty} d_j 10^{-j} \;=\; \frac{1}{10^{n-1}} \quad\Leftrightarrow\quad d_j = 9$ *for all* $j \geq n$.

PROOF of **(a)** and **(b)**: Left as exercise 9.29 (see p.305).

PROOF of **(c)**: The contrapositive of "⇒" is the following statement: If at least one digit d_j is not 9 then $\sum\limits_{j=n}^{\infty} d_j 10^{-j} \;\neq\; \frac{1}{10^{n-1}}$. To prove it we introduce the following notation. For $k \geq n$ let

$$s_k \;=\; \sum_{j=n}^{k} d_j 10^{-j}, \quad s \;=\; \sum_{j=n}^{\infty} d_j 10^{-j}, \quad s'_k \;=\; 9\sum_{j=n}^{k} 10^{-j}, \quad s' \;=\; 9\sum_{j=n}^{\infty} 10^{-j}.$$

If there is $j_0 \geq n$ such that $d_{j_0} < 9$ then $s'_k - s_k \geq 10^{-j_0}$ for all $k \geq j_0$. We use the Domination Theorem for Limits (prop.9.19 on p.269) and the fact that the difference of the limits is the limit of the differences, and we obtain

$$s' - s \ = \ \lim_{k\to\infty}(s'_k - s_k) \ \geq \ 10^{-j_0}.$$

This proves "⇒". The "⇐" direction is immediate from part **(a)**. ■

Lemma 9.1. *Let* $x \ = \ m + \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ *(*$m \in \mathbb{Z}_{\geq 0}, d_j \in [0,9]_{\mathbb{Z}}$ *for all* $j \in \mathbb{N}$*) be a decimal expansion of a nonnegative real number* x*. Then*

(a) $m \ \leq \ x \ \leq \ m+1,$

(b) $x = m \ \ \Leftrightarrow \ \ d_j = 0$ *for all* $j \in \mathbb{N}$*,*

(c) $x = m+1 \ \ \Leftrightarrow \ \ d_j = 9$ *for all* $j \in \mathbb{N}$*.*

The proof is left as exercise 9.30 (see p.305). ■

The above lemma is a generalization of the following proposition.

Proposition 9.35 ((B/G prop.12.7))**.**

Let $m \in \mathbb{Z}$ *and* $d_j \in [0,9]_{\mathbb{Z}}$ *for all* $j \in \mathbb{N}$ *such that* $m \ + \ \sum\limits_{j=1}^{\infty} \dfrac{d_j}{10^j} \ = 1$*. Then either* $m = 1$ *and* $d_j = 0$ *for all* $j \in \mathbb{N}$ *or* $m = 0$ *and* $d_j = 9$ *for all* $j \in \mathbb{N}$*. In other words,* 1.00000... *and* 0.999999... *are the only two decimal representations of the number* 1*.*

PROOF: Apply lemma 9.1**(c)** with $m = 0$. ■

Theorem 9.10 (Existence of Decimal Expansions (B/G thm.12.6))**.** *Every real number has a decimal expansion.*

PROOF: We give here the idea behind the proof. For an exact proof see ch.12 of B/G [2].

Let $x \in \mathbb{R}$. It suffices the theorem for nonnegative x: If $x < 0$ then we obtain a decimal expansion for x by taking the one for $-x$ and preceding it with a minus sign.

We will find $d_0 \in \mathbb{Z}_{\geq 0}$ and $d_1, d_2, \dots \in [0,9]_{\mathbb{Z}}$ such that $x = d_0 + \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ as follows.

Step 1: The intervals $[k, k+1[$ $(n \in [0,\infty[_{\mathbb{Z}})$ are a partitioning of $\mathbb{R}_{\geq 0}$, thus there exists a unique nonnegative integer k such that $k \leq x < k+1$. Let $d_0 = k$. If $x = d_0$ we are done, otherwise we continue.

Step 2: The intervals $[d_0 + \frac{k}{10}, d_0 + \frac{k+1}{10}[$ $(k = 0, 1, \dots 9)$ are a partitioning of $[d_0, d_0+1[$, thus there exists a unique digit k such that $d_0 + \frac{k}{10} \leq x < d_0 + \frac{k+1}{10}$. Let $d_1 = k$. If $x = d_0 + \frac{d_1}{10}$ we are done, otherwise we continue.

Step n: The intervals $\left[\sum\limits_{j=0}^{n-2} d_j 10^{-j} + \frac{k}{10^{n-1}}, \sum\limits_{j=0}^{n-2} d_j 10^{-j} + \frac{k+1}{10^{n-1}}\right[$ $(k = 0, 1, \dots 9)$ are a partitioning of $\left[\sum\limits_{j=0}^{n-2} d_j 10^{-j}, \sum\limits_{j=0}^{n-2} d_j 10^{-j} + \frac{1}{10^{n-2}}\right[$, thus there exists a unique digit k such that

$$\sum_{j=0}^{n-2} d_j 10^{-j} + \frac{k}{10^{n-1}} \ \leq \ x \ < \ \sum_{j=0}^{n-2} d_j 10^{-j} + \frac{k+1}{10^{n-1}}.$$

Let $d_{n-1} = k$. If $x = \sum\limits_{j=0}^{n-2} d_j 10^{-j} + \frac{k}{10^{n-1}}$ we are done, otherwise we continue ... ■

Theorem 9.11 (Uniqueness of Decimal Expansions (B/G thm.12.8)).

Let $x \in \mathbb{R}_{\geq 0}$ have two different decimal representations

$$x \;=\; d_0 \,+\, \sum_{j=1}^{\infty} \frac{d_j}{10^j} \;=\; e_0 \,+\, \sum_{j=1}^{\infty} \frac{e_j}{10^j}\,, \tag{9.47}$$

where $d_0, e_0 \in [0, \infty[_{\mathbb{Z}}$ and $d_j, e_j \in [0, 9]_{\mathbb{Z}}$ for all $j \in \mathbb{N}$. Further, let K be the smallest subscript such that $d_K \neq e_K$. Then we have the following:

If $d_K < e_K$, then • $e_K = d_K + 1$, • $e_j = 0$ and $d_j = 9$ for all $j > K$.

PROOF: It follows from (9.47) and $d_j = e_j$ for $0 \leq j < K$ that

$$10^K \sum_{j=K}^{\infty} \frac{d_j}{10^j} = 10^K \sum_{j=K}^{\infty} \frac{e_j}{10^j}, \text{ i.e., } d_K + \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} \;=\; e_K + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j}, \tag{9.48}$$

$$\text{hence } \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} = (e_K - d_K) + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j}. \tag{9.49}$$

$$\text{Thus } \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} \leq \sum_{j=1}^{\infty} \frac{9}{10^j} \;=\; 1 \;\leq\; (e_K - d_K) + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j} \;=\; \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j}. \tag{9.50}$$

Here the first equality of (9.50) follows from prop.9.35 (with $m = 0$). We obtain the second inequality of (9.50) from $(e_K - d_K) \geq 1$ and $e_{K+j} \geq 0$. The last equality of (9.50) follows from (9.49).

(9.50) can only be valid if all inequalities are equalities. In particular we obtain

$$1 \;= (e_K - d_K) + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j} \;=\; \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j}.$$

According to prop.9.35 it follows, together with $e_K - d_K \;\geq 1$, that $e_K - d_K \;= 1$, $e_{K+j} \;= 0$ for all $j \geq 1$, and $d_{K+j} = 9$ for all $j \geq 1$.

This proves the theorem. ■

Remark 9.17. Note that if $x \in \mathbb{R}$ has a finite decimal expansion $x \;=\; m + \sum\limits_{j=1}^{K} \frac{d_j}{10^j}$ $(K \in \mathbb{N})$, i.e., one in which only finitely many digits d_j are not zero, then x is rational as the finite sum of rational terms m and $d_j 10^{-j}$. □

Corollary 9.5. *If a real number has different decimal expansions then it is rational.*

The proof is left as exercise 9.31 (see p.305). ■

As an application of decimal representations we prove a more powerful version of thm.9.1 on p.254 which states, when applied to $\mathbb{R}$ (rather than $\mathbb{Q}$) that for any two real numbers x and y there exists $z \in \mathbb{R}$ such that $x < z < y$. It turns out that we may choose z as a rational number.

Proposition 9.36 (B/G prop.11.8)**.**

Let $x, y \in \mathbb{R}$ be such that $x < y$. Then there exists $q \in \mathbb{Q}$ be such that $x < q < y$.

PROOF:
Case 1: $x \geq 0$, and hence $y > 0$.
It follows from thm.9.1 on p.254 that there exists $z \in \mathbb{R}$ be such that $x < z < y$. Since z is not negative it has a decimal representation $z = \sum\limits_{j=0}^{\infty} d_j 10^{-j}$ such that $d_0 \in \mathbb{Z}_{\geq 0}$ and d_j is a decimal digit for each $j > 0$. For $n \in \mathbb{N}$

$$\text{let } z_n := z - \sum_{j=0}^{n} d_j 10^{-j}. \text{ Then } z - z_n = \sum_{j=n+1}^{\infty} d_j 10^{-j} \leq 9 \cdot \sum_{j=n+1}^{\infty} 10^{-j} = 10^{-n}.$$

Thus $z - z_n$ converges to zero. Since $z > x$ it follows that there exists $n_0 \in \mathbb{N}$ such that $z - z_n = |z - z_n| < z - x$, and hence $z_n > x$ for all $n \geq n_0$. In particular $x < z_{n_0} < z < y$. Since z_{n_0} is rational as a (finite) sum of rational numbers $d_j 10^{-j}$, $q := z_{n_0}$ is a rational number which satisfies $x < q < y$.
Case 2: $x < 0$ and $y > 0$. We choose $q := 0$.
Case 3: $y \leq 0$, and hence $x < 0$.
According to the already proven case 1 there exists $q' \in \mathbb{Q}$ such that $-y < q' < -x$. Let $q := -q'$. Then $q \in \mathbb{Q}$, and q satisfies $x < q < y$. ■

The following is copied from B/G ch.12 for convenience:

Definition 9.18 (Repeating Decimals)**.**

A nonnegative decimal

$$x = m.d_1 d_2 \ldots = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad (d_j \in \{0, 1, 2 \ldots, 9\})$$

is **repeating** if there are natural numbers N and p such that

$$d_{N+n+kp} = d_{N+n} \ \forall\, 0 \leq n < p,\ k \in \mathbb{N}. \ \square$$

Note that the above definition INCLUDES THE CASE $p = 1$ and $d_N = 0$ (finite expansion!).

Proposition 9.37 (B/G Prop.12.11, p.119)**.** *Every repeating decimal represents a rational number.*

PROOF: $\boxed{\star}$ Let $x = (m, d_1, d_2, \ldots) = m + \sum\limits_{j=1}^{\infty}$ be a repeating decimal, i.e., there exist $N, p \in \mathbb{N}$ such that for all $0 \leq n < p$ and for all $k \in \mathbb{N}$,

$$d_{N+n+kp} = d_{N+n}.$$

Summation of the terms that span the first K periods yields

$$\sum_{j=N}^{N+Kp-1} d_j 10^{-j} = \sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+kp+n} 10^{-(N+kp+n)} = \sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)} \tag{9.51}$$

The last equation results from the periodicity of length p. We take limits $K \to \infty$ on both sides above [118] and we obtain

$$\begin{aligned}\sum_{j=N}^{\infty} d_j 10^{-j} &= \lim_{K\to\infty} \sum_{j=N}^{N+Kp-1} d_j 10^{-j}\\ &= \lim_{K\to\infty} \sum_{k=0}^{K-1}\sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)}\\ &= \sum_{k=0}^{\infty}\sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)}\,.\end{aligned}$$

Then

$$\begin{aligned}x &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{j=N}^{\infty} d_j 10^{-j}\\ &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{k=0}^{\infty}\sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n+kp}}\\ &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n}} \sum_{k=0}^{\infty} \frac{1}{(10^p)^k}\\ &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n}} \sum_{k=0}^{\infty} (10^{-p})^k\\ &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n}} \frac{1}{1-10^{-p}}.\end{aligned}$$

Hence x is a finite sum of rational numbers and therefore rational. ■

Example 9.7. We illustrate (9.51) with the following example. Let $x = 0.1234\overline{567}$. Then the period 567 has length $p = 3$ and period 1 starts at $N = 5$, period 2 starts at $N = 8$, period 3 starts at $N = 11$, period 4 starts at $N = 14$, period 5 starts at $N = 17, \dots$. Summation of the first four period $(K = 4)$ thus starts at $j = N = 5$ and ends at $j = 17 - 1 = 16$. Formula (9.51) becomes

$$\sum_{j=N}^{N+Kp-1} d_j 10^{-j} = \sum_{j=5}^{5+4\cdot 3-1} d_j 10^{-j} = \sum_{j=5}^{16} d_j 10^{-j}$$

and the second expression becomes

$$\sum_{k=0}^{K-1}\sum_{n=0}^{p-1} d_{N+kp+n} 10^{-(N+kp+n)} = \sum_{k=0}^{3}\sum_{n=0}^{3-1} d_{5+3k+n} 10^{-(5+3k+n)} = \sum_{k=0}^{3}\sum_{n=0}^{2} d_{5+3k+n} 10^{-(5+3k+n)}$$

[118] There is a catch in the first equation: Let $s_i := \sum_{j=N}^{i} d_j 10^{-j} (i \geq N)$ and $y_K := s_{N+Kp-1} = \sum_{j=N}^{N+Kp-1} d_j 10^{-j}$ $(K \in \mathbb{N})$. Then $(y_K)_K$ is a subsequence of $(s_i)_i$. How do we know that both $(y_K)_K$ and $(s_i)_i$ have the same limit? This is shown in prop.9.13 (Subsequences of sequences with limits) on p.265.

In that last expression, $k = 0$ covers the digits from 5 to 7, $k = 1$ covers the digits from 8 to 10, $k = 2$ covers the digits from 11 to 13 and $k = 3$ covers the digits from 14 to 16, so the net effect is that of $\sum_{j=5}^{16} d_j 10^{-j}$: summation of the terms between $j = 5$ and $j = 16$. □

Remark 9.18. The fact that fractions are repeating decimals (including the period $\bar{0}$) (see B/G thm. 12.13) is illustrated by long division of $2 \div 7$. The sequence of remainders is

$$2 - 6 - 4 - 5 - 1 - 3 - 2...$$

and once we have the same remainder, we are in an endless loop.

Note 9.3 (Decimal expansions of real numbers)**.**

> Let $x \in \mathbb{R}$.
> **(a)** x has at most two different decimal expansions.
> **(b)** If x has two expansions then one is all zeros except for finitely many digits and the other is all nines except for finitely many digits.
> **(c)** If x has more than one expansion then x is rational.
> **(d)** x is a repeating decimal if and only if $x \in \mathbb{Q}$. □

9.7 Countable and Uncountable Subsets of the Real Numbers

We have seen that the rational numbers are countably infinite (prop.7.5 on p.229), just as the natural numbers and the integers, even though they are "dense" on the real numbers line in the following sense. Between any two fractions $x < y$ there exists a fraction z such that $x < z < y$, e.g., their arithmetic mean $\frac{x+y}{2}$. In that way $\mathbb{Q}$ and $\mathbb{R}$ are alike, but the real numbers are not comparable in size to the rational numbers (and not to the integers as well), as the next theorem will show.

Theorem 9.12. *The real numbers are uncountable.*

This proof is a more elaborate version of the one given in B/G [2] (thm.13.22, p.125).

Let $A := \{ \sum\limits_{j=1}^{\infty} d_j 10^{-j} : d_j = 3 \text{ or } d_j = 4 \text{ for all } j \in \mathbb{N} \}$, i.e., $A \subseteq \mathbb{R}$ is the set of all decimals $0.d_1 d_2 d_3 \dots$ for which each digit d_n is either 3 or 4.

Let $x, x' \in A$, i.e., $x = \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ and $x' = \sum\limits_{j=1}^{\infty} d'_j 10^{-j}$ for suitable digits d_j and d'_j, all of which are either 3 or 4. Since no digits are 8 or 9 it follows from the uniqueness theorem for decimal expansions (thm.9.11 on p.282) that $x = x'$ implies $d_j = d'_j$ for all $j \in \mathbb{N}$. This can be expressed as follows: Let $S := \{3, 4\}^{\mathbb{N}} = \{(d_j)_{j \in \mathbb{N}} : d_j = 3 \text{ or } d_j = 4 \text{ for all } j \in \mathbb{N}\}$. Then $F : S \to A; \ \ (d_j)_{j \in \mathbb{N}} \mapsto \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ is injective. But F also is surjective: If $x \in A$ then $x = \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ for suitable d_j which are either 3 or 4. Thus $(d_j)_{j \in \mathbb{N}} \in S$ and $F\big((d_j)_j\big) = x$. This proves surjectivity of F.

Since F is bijective we conclude that $|A| = |S|$. It follows from thm.7.7 on p.230 that S is uncountable, thus A is uncountable, thus its superset $\mathbb{R}$ is uncountable. ■

Remark 9.19. The above proof should look familiar: It is based on the same principle as that of thm.7.7 on p.230. There a and b take the role of the digits 3 and 4. In fact, we can obtain the uncountablility of $\mathbb{R}$ from that theorem as follows: It follows from thm.9.11 (uniqueness of decimal expansions) on p.282 That if $x, y \in [0,1[$ have decimal expansions $x = \sum\limits_{j=1}^{\infty} \frac{d_j}{10^j}$ and $y = \sum\limits_{j=1}^{\infty} \frac{e_j}{10^j}$ such that each one of the digits d_j, e_j is either 3 or 4 then $x = y \Leftrightarrow d_j = e_j$ for all $j \in \mathbb{N}$. In other words, if X is the set of all real numbers between 0 and 1 whose decimal expansions have digits which are exclusively 3 or 4 then the assignment $\sum\limits_{j=1}^{\infty} \frac{d_j}{10^j} \mapsto (d_j)_{j \in \mathbb{N}}$ defines a bijection $X \xrightarrow{\sim} \{3,4\}^{\mathbb{N}}$. The set $\{3,4\}^{\mathbb{N}}$ is uncountable by virtue of thm.9.11, thus X is uncountable, thus its superset $\mathbb{R}$ is uncountable.

Real numbers can be partitioned into rational and irrational numbers. One can also partition them into so called algebraic numbers and transcendental numbers.

Definition 9.19 (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 9.38 (B/G Prop.13.21)**.**

The set of all algebraic numbers is countable.

The proof is left as exercise 9.32 (see p.305). ■

Proposition 9.39.

Let $k, m, n \in \mathbb{N}$. Then $\sqrt[k]{\frac{m}{n}}$ is algebraic.

PROOF: Let $p(x) := nx^k - m$. Then

$$p\big(\sqrt[k]{m/n}\big) = n\big(\sqrt[k]{m/n}\big)^k - m = \frac{nm}{n} - m = 0. \blacksquare$$

Proposition 9.40.

Let $r \in \mathbb{Q}$. Then r is algebraic.

The proof is left as exercise 9.33 (see p.305). ■

Note that if $m, n > 0$ then the above is a special case of prop.9.39 (let $k := 1$).

Here are some trivial consequences of the fact that $\mathbb{R}$ is uncountable (see thm. 9.12, p.9.12 and B/G Thm.13.22).

Proposition 9.41.

The set of all transcendental numbers and that of all irrational numbers are uncountable.

PROOF: the uncountable real numbers are the disjoint union of the countable rational numbers with the irrational numbers, and they also are the disjoint union of the countable algebraic numbers with the transcendentals. The assertion follows from cor.7.3. ■

9.8 Limit Inferior and Limit Superior

Definition 9.20 (Tail sets of a sequence)**.**

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Let

$$T_n \;:=\; \{x_j : j \in \mathbb{N} \text{ and } j \geq n\} \;=\; \{x_n, x_{n+1}, x_{n+2}, x_{n+3}, \dots\} \tag{9.52}$$

be what remains in the sequence after we discard the first $n-1$ elements. We call $(T_n)_{n\in\mathbb{N}}$ the n–th **tail set** of the sequence $(x_k)_k$. □

Remark 9.20. 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 $T_m \supseteq T_n$.

c. It follows from (b) and prop.9.9 on p.260 and prop.9.18 on p.269 that

$$\begin{aligned} \beta_n \;&:=\; \sup(T_n) \;\text{ is nonincreasing, hence } \lim_{n\to\infty} \beta_n = \inf_n \beta_n;\\ \alpha_n \;&:=\; \inf(T_n) \;\text{ is nondecreasing, hence } \lim_{n\to\infty} \alpha_n = \sup_n \alpha_n. \end{aligned}$$

These limits can also be expressed as follows.

$$\begin{aligned} \lim_{n\to\infty} \big(\sup\{x_j : j \in \mathbb{N}, j \geq n\}\big) \;&=\; \lim_{n\to\infty} \big(\sup(T_n)\big) \;=\; \inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big),\\ \lim_{n\to\infty} \big(\inf\{x_j : j \in \mathbb{N}, j \geq n\}\big) \;&=\; \lim_{n\to\infty} \big(\inf(T_n)\big) \;=\; \sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big). \end{aligned} \tag{9.53}$$

An expression like $\sup\{x_j : j \in \mathbb{N}, j \geq n\}$ can be written more compactly as $\sup\limits_{j\in\mathbb{N}, j\geq n} \{x_j\}$. Moreover, when dealing with sequences (x_n), it is understood in most cases that $n \in \mathbb{N}$ or $n \in \mathbb{Z}_{\geq 0}$ and the last expression simplifies to $\sup\limits_{j\geq n}\{x_j\}$. This can also be written as $\sup\limits_{j\geq n}(x_j)$ or $\sup\limits_{j\geq n} x_j$.

In other words, (9.53) becomes

$$\begin{aligned} \inf_{n\in\mathbb{N}} \big(\sup_{j\geq n} x_j\big) \;&=\; \inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big) \;=\; \lim_{n\to\infty} \big(\sup(T_n)\big) \;=\; \lim_{n\to\infty} \big(\sup_{j\geq n} x_j\big),\\ \sup_{n\in\mathbb{N}} \big(\inf_{j\geq n} x_j\big) \;&=\; \sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big) \;=\; \lim_{n\to\infty} \big(\inf(T_n)\big) \;=\; \lim_{n\to\infty} \big(\inf_{j\geq n} x_j\big). \;\square \end{aligned} \tag{9.54}$$

The above leads us to the following definition:

Definition 9.21. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ with tail sets $T_n = \{x_j : j \in \mathbb{N}, j \geq n\}$.

Assume that T_n is bounded above for some $n \in \mathbb{N}$ (and hence for all $n \in \mathbb{N}$). We call

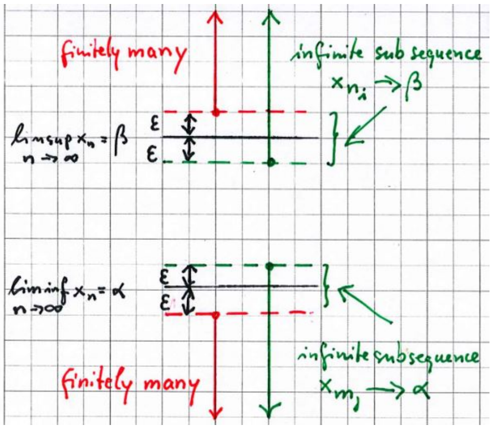
$$\limsup_{n\to\infty} x_j \;:=\; \lim_{n\to\infty}\big(\sup_{j\geq n} x_j\big) \;=\; \inf_{n\in\mathbb{N}}\big(\sup_{j\geq n} x_j\big) \;=\; \inf_{n\in\mathbb{N}}\big(\sup(T_n)\big)$$

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 (and hence for all $n \in \mathbb{N}$). We call

$$\liminf_{n\to\infty} x_j \;:=\; \lim_{n\to\infty}\big(\inf_{j\geq n} x_j\big) \;=\; \sup_{n\in\mathbb{N}}\big(\inf_{j\geq n} x_j\big) \;=\; \sup_{n\in\mathbb{N}}\big(\inf(T_n)\big)$$

the **lim inf** or **limit inferior** of the sequence (x_n).
If, for each n, T_n is not bounded below then we say $\liminf\limits_{n\to\infty} x_j = -\infty$. □



(a) Given $\varepsilon > 0$:

- There is $N_1 = N_1(\varepsilon)$: $k \geq N_1 \;\Rightarrow\; x_k - \beta \;\leq\; \frac{\varepsilon}{2}$;
- There is $N_2 = N_2(\varepsilon)$: $k \geq N_2 \;\Rightarrow\; \alpha - x_k \;\leq\; \frac{\varepsilon}{2}$;
- $k \geq \max(N_1, N_2) \;\Rightarrow\;$ both $|x_k - \beta| \leq \frac{\varepsilon}{2}$, $|x_k - \alpha| \leq \frac{\varepsilon}{2}$;
- Hence, only finitely many x_k above $\beta + \varepsilon$ or below $\alpha - \varepsilon$;

(b) For **any** j there is $n_j \geq j$ such that $x_{n_j} > \beta_{n_j} - \frac{1}{j} \geq \beta - \frac{1}{j}$

- Hence, with **(a)**, eventually all x_{n_j} satisfy $|x_{n_j} - \beta| < \varepsilon$
- Hence, there is a subsequence x_{n_j} such that $\lim\limits_{j\to\infty} x_{n_j} = \beta$.

Theorem 9.13 (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 $n_1 < n_2 < \dots \in \mathbb{N}$ 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 $n_1 < n_2 < \dots \in \mathbb{N}$ 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:

Step 1: Let $\varepsilon > 0$. It follows from $\beta_n = \sup(T_n) = \sup\{x_j : j \geq n\}$ and $\beta_n \downarrow \beta = \limsup_n x_n$ that $\beta_n < \beta + \varepsilon$ for all $n \geq N$ 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 $n \geq N$, 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 **a2** of thm.9.13.

Step 2: We create subsequences $(x_{n_j})_j$ and $(\beta_{n_j})_j$ such that

$$\beta_{n_j} \geq x_{n_j} > \beta_{n_j} - 1/j \tag{9.55}$$

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 $\beta_1 \geq x_k > \beta_1 - 1$. We set $n_1 := k$.

Having constructed $n_1 < n_2 < \cdots < n_k$ such that $\beta_{n_j} \geq x_{n_j} > \beta_{n_j} - 1/j$ for all $j \leq k$ 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

$$x_{n_k+i} > \beta_{n_k+1} - \frac{1}{k+1}. \tag{9.56}$$

Let $n_{k+1} := n_k + i$. The sequence β_n nonincreasing (i.e., decreasing) and it follows from

$$n_{k+1} = n_k + i \geq n_k + 1$$

that $\beta_{n_{k+1}} \leq \beta_{n_k+1}$. But then (9.56) implies that

$$x_{n_{k+1}} > \beta_{n_{k+1}} - \frac{1}{k+1}.$$

We note that $x_{n_{k+1}} \leq \beta_{n_{k+1}}$ 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 (9.56) we have

$$\beta_{n_{k+1}} \geq x_{n_{k+1}} > \beta_{n_k+1} - \frac{1}{k+1}. \tag{9.57}$$

It follows that $x_{n_{k+1}}$ satisfies (9.55) and the the proof of step 2 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

$$\lim_k \beta_{n_k} = \beta, \ \lim_k \beta_{n_k} - \frac{1}{k} = \lim_k \beta_{n_k} - \lim_k \frac{1}{k} = \beta - 0 = \beta,$$

and x_{n_j} is "sandwiched" between two sequences which both converge to the same limit β.

Step 4. No subsequence of (x_n) can converge to a number u bigger than β: Let

$$\varepsilon := \frac{1}{2}(u - \beta).$$

It follows from step 1 that all but finitely many x_j satisfy $x_j \leq \beta + \varepsilon$, hence $x_j \leq u - \varepsilon$.

We conclude that $|u - x_j| \geq \varepsilon$ for $j \geq N$ and no subsequence converging to u can be extracted. This proves **a1** of thm.9.13.

Step 5. We still must prove the missing half of thm.9.13.**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 the subsequence constructed in step 2. It follows from (9.55) and $\beta_{n_j} \geq \beta$ and $1/j < \varepsilon$ that $x_{n_j} > \beta - \varepsilon$. This proves the missing half of thm.9.13.**a2**.

Uniqueness of β: Let $v > \beta$ and $\varepsilon := (v - \beta)/3$. 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.9.13.**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.9.13.**a2**. We have proved that $\limsup_n x_n$ is uniquely determined by the inequalities of thm.9.13.**a2** and we have shown both **a1** and **a2** of thm.9.13.

Parts **b1** and **b2** of thm.9.13 follow now easily from applying parts **a1** and **a2** to the sequence $y_n := -x_n$. ■

Theorem 9.14 (Characterization of limits via limsup and liminf). *Let $(x_n)_{n \in \mathbb{N}}$ be a bounded sequence in $\mathbb{R}$.*

The sequence (x_n) converges to a real number if and only if liminf and limsup for that sequence coincide. Moreover, if such is the case then

$$\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n. \tag{9.58}$$

PROOF of "$\Rightarrow$": Let $L := \lim\limits_{n\to\infty} x_n$. It follows from prop.9.13 (Characterization of limsup and liminf) parts **a1** and **b1** on p.288 that subsequences can be found which converge to $\liminf\limits_{n\to\infty} x_n$ and to $\limsup\limits_{n\to\infty} x_n$. We also know from prop.9.13 (Subsequences of sequences with limits) on p.265 that any convergent subsequence has limit L. This proves $\liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n = L$.

PROOF of "$\Leftarrow$": [119] Let $L := \liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$. Let $\varepsilon > 0$, $J_1 := \{j \in \mathbb{N} : x_j < L - \frac{\varepsilon}{2}\}$, $J_2 := \{j \in \mathbb{N} : x_j > L + \frac{\varepsilon}{2}\}$, $J := J_1 \cup j_2$.

It follows from thm.9.13 (Characterization of limsup and liminf) parts **a2** and **b2** that J_1, J_2, and hence J only contain at most finitely many indices.

If $J = \emptyset$ let $n_0 := 1$, else let $n_0 := \max(J) + 1$.

Then $L - \frac{\varepsilon}{2} \leq x_j \leq L + \frac{\varepsilon}{2}$, hence $|x_j - L| < \varepsilon$ for all $j \geq n_0$. It follows that $L = \lim\limits_{n\to\infty} x_n$. ■

Proposition 9.42. *Let $x_n, x'_n \in \mathbb{R}$ be two sequences of real numbers. Assume there is $K \in \mathbb{N}$ such that $x_n \leq x'_n$ for all $n \geq K$. Then*

$$\liminf_{n\to\infty} x_n \leq \liminf_{n\to\infty} x'_n \qquad \textit{and} \qquad \limsup_{n\to\infty} x_n \leq \limsup_{n\to\infty} x'_n.$$

PROOF:

[119] Here is a shorter proof of "$\Leftarrow$". The drawback: It makes use of prop.9.43 and prop.9.44. Let $L := \liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$. Let $\varepsilon > 0$. We know from (9.61), p.293 and (9.64), p.295 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 $L = \lim\limits_{n\to\infty} x_n$. ■

We only prove prove the limsup inequality because once we have that, we apply it to the sequences $(-x_n)_n$ and $(-x'_n)_n$ which satisfy $-x'_n \leq -x_n$ for all $n \geq K$. We obtain

$$-\liminf_{n\to\infty} x'_n \;=\; \limsup_{n\to\infty}(-x'_n) \leq \limsup_{n\to\infty}(-x_n) \;=\; -\liminf_{n\to\infty} x_n,$$

hence $\liminf\limits_{n\to\infty} x_n \leq \liminf\limits_{n\to\infty} x'_n$. and this proves the liminf inequality of the proposition.

Case 1: Both sequences are bounded.

Let $u := \limsup_n x_n$ and $u' := \limsup_n x'_n$. We assume to the contrary that $u > u'$. Then $\varepsilon := \frac{u-u'}{2} > 0$.

According to cor.9.8 on p.294 there are infinitely many $x_{n_1}, x_{n_2}, \dots$ such that $x_{n_j} > u - \varepsilon$. At most finitely of those n_j can be less than K. We discard those and there still are infinitely many $n_j \geq K$ such that $x_{n_j} > u - \varepsilon$.

As $x'_i \geq x_i$ for all $i \geq K$, it follows that there are infinitely many n_j such that

$$x'_{n_j} \geq x_{n_j} > u - \varepsilon = u' + \varepsilon.$$

We employ cor.9.8 a second time. It also states that there are at most finitely many x'_{n_j} such that $x'_{n_j} \geq u' + \varepsilon$. We have reached a contradiction.

Case 2: Not both sequences are bounded above.

If both are bounded below, $\liminf_n x_n \leq \liminf_n x'_n$ is obtained just as in case 1, otherwise this is covered in case 3. We now observe what happens to the limits superior.

Case 2a: x_n is not bounded above.

Then neither is x'_n, hence all tailsets for both sequences have sup = ∞, hence $\limsup_n x_n = \limsup_n x'_n = \infty$.

Case 2b: x'_n is not bounded above.

Then all tailsets for x'_n have sup = ∞, hence $\limsup_n x'_n = \infty$, hence $\limsup_n x_n \leq \limsup_n x'_n$.

Case 3: Not both sequences are bounded below.

If both are bounded above, $\limsup_n x_n \leq \limsup_n x'_n$ is obtained just as in case 1, otherwise this is covered in case 2. We now observe what happens to the limits inferior.

Case 2a: x'_n is not bounded below.

Then neither is x_n and we that all tailsets for both sequences have inf = $-\infty$, hence $\liminf_n x_n = \liminf_n x'_n = \infty$.

Case 2b: x_n is not bounded above.

Then all tailsets for x_n have inf = $-\infty$, hence $\liminf_n x_n = -\infty$, hence $\liminf_n x_n \leq \liminf_n x'_n$.

■

Here is the first corollary to prop.9.42.

Corollary 9.6. *Let $x_n, y_n \in \mathbb{R}$ be two sequences of real numbers. Assume there is $K \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq K$. Then*

$$\limsup_{n\to\infty} x_n = \limsup_{n\to\infty} y_n \quad \textit{and} \quad \liminf_{n\to\infty} x_n = \liminf_{n\to\infty} y_n.$$

PROOF: Immediate from prop.9.42. ■

Here is the second corollary to prop.9.42.

Corollary 9.7. *Let $x_n \geq 0$ such that $\limsup\limits_{n\to\infty} x_n = 0$. Then $(x_n)_n$ converges to zero.*

The proof is left as exercise 9.36 (see p.305). ■

Remark 9.21. Prop.9.42 leads to an easy alternate proof of prop.9.19 (Domination Theorem for Limits) on p.269 which states that if two sequences $x_n, y_n \in \mathbb{R}$ satisfy $x_n \leq y_n$ eventually then $\lim\limits_{n\to\infty} x_n \leq \lim\limits_{n\to\infty} y_n$.

PROOF: Immediate from prop.9.42. ■

Example 9.8. Let $A \subseteq \mathbb{R}$.

$$x_n := \begin{cases} 1 + (-1)^n/n & \text{if } n \in I^* := \{1,2,5,6,9,10,13,14,\dots\} \\ -1 + (-1)^n/n & \text{if } n \in I_* := \{3,4,7,8,11,12,15,16,\dots\}. \end{cases} \tag{9.59}$$

We can write the elements of T^* as a sequence $m_1 < m_2 < \dots$ and the elements of T_* as a sequence $n_1 < n_2 < \dots$.[120] Let $j \in \mathbb{N}$. Clearly there exist indices $k, l \in \mathbb{N}$ which increase with j such that $x_{m_j} = 1 + \frac{1}{k}$ and $x_{n_j} = -1 - \frac{1}{l}$. Thus $x_{m_j} \downarrow 1$ and $x_{n_j} \uparrow -1$ as $j \to \infty$.
Since the index sets I^* and I_* partition $\mathbb{N}$, i.e., $\mathbb{N} = I^* \biguplus I_*$, there are no subsequences of (x_n) which can have a limit other than ± 1. It is immediate from parts **a1** and **b1** of thm.9.13 on p.288 that $\liminf\limits_{n\to\infty} x_n = -1$ and $\limsup\limits_{n\to\infty} x_n = 1$.
This example also illustrates part **a2** of thm.9.13: Since $x_{n_j} < 0$ for all j and $\lim\limits_{j\to\infty} x_{m_j} = 1$ implies that eventually all x_{m_j} will be ε–close to 1 there can be at most finitely many indices n such that

$$x_n > \limsup_{n\to\infty} x_n + \varepsilon.$$

Moreover the ε–closeness to 1 of eventually all x_{m_j} implies that $x_{m_j} > \limsup\limits_{n\to\infty} x_n - \varepsilon$ for infinitely many indices m_j, hence

$$x_n > \limsup_{n\to\infty} x_n - \varepsilon$$

for infinitely many $n \in \mathbb{N}$. □

The remainder of this chapter on liminf and limsup is **optional material**.

[120] We can do this recursively as follows:

$$m_1 = 1, m_2 = m_1 + 2 + (-1)^1, m_3 = m_2 + 2 + (-1)^2, \dots, m_{j+1} = m_j + 2 + (-1)^j, \dots,$$

and

$$n_1 = 3, n_2 = n_1 + 2 + (-1)^1, n_3 = n_2 + 2 + (-1)^2, \dots, n_{j+1} = n_j + 2 + (-1)^j, \dots.$$

(Alternatively: $n_j = m_j + 2$.)

Proposition 9.43. $\boxed{\star}$ *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded above with tail sets T_n.*

(A) *Let*

$$\begin{aligned}
\mathscr{U} &:= \{y \in \mathbb{R} : T_n \cap [y,\infty[\neq \emptyset \textit{ for all } n \in \mathbb{N}\},\\
\mathscr{U}_1 &:= \{y \in \mathbb{R} : \textit{ for all } n \in \mathbb{N} \textit{ there exists } k \in \mathbb{Z}_{\geq 0} \textit{ such that } x_{n+k} \geq y\},\\
\mathscr{U}_2 &:= \{y \in \mathbb{R} : \ \exists \textit{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \textit{ such that } x_{n_j} \geq y \textit{ for all } j \in \mathbb{N}\},\\
\mathscr{U}_3 &:= \{y \in \mathbb{R} : x_n \geq y \textit{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{9.60}$$

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.*** $$u - \varepsilon \in \mathscr{U} \quad \textit{and} \quad u + \varepsilon \notin \mathscr{U} \quad \textit{for all } \varepsilon > 0. \tag{9.61}$$
- ***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:

$$x \in T_n \ \Leftrightarrow \ x = x_j \text{ for some } j \geq n \ \Leftrightarrow \ x = x_{n+k} \text{ for some } k \in \mathbb{Z}_{\geq 0}$$

from which it follows that $x \in T_n \cap [y,\infty[\ \Leftrightarrow \ x = x_{n+k} \geq y$ for some $k \geq 0$.

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 $T_1 \cap [y,\infty[\neq \emptyset$ there is some $x \in T_1$ such that $y \leq x < \infty$, i.e., $x \geq y$. Because $x \in T_1 = \{x_1, x_2, \dots\}$ we have $x = x_{n_1}$ for some integer $n_1 \geq 1$; we have proved the existence of n_1.

Induction assumption: Assume that $n_1 < n_2 < \dots < n_{j_0}$ have already been picked.

Induction step: As $y \in \mathscr{U}_1$ there is $k \in \mathbb{Z}_{\geq 0}$ such that $x_{(n_{j_0}+1)+k} \geq y$. We set $n_{j_0+1} := n_{j_0} + 1 + k$. 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 $n_1 < n_2 < n_3 < \dots \in \mathbb{N}$ 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 $x_j \geq y$. 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 $j \geq n$

Pick any one of those, say j'. Then $x_{j'} \in T_n$ and $x_{j'} \geq y$. 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 $[y_2,\infty[\subseteq [y_1,\infty[$ and $T_n \cap [y_2,\infty[\neq \emptyset$ for all $n \in \mathbb{N}$ that $T_n \cap [\,y_1,\infty[\neq \emptyset$ for all $n \in \mathbb{N}$, i.e., $y_1 \in \mathscr{U}$.

We conclude that $\mathscr{U}$ is 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 - (9.61) 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 - (9.61) part 2**, $u + \varepsilon \notin \mathscr{U}$:

This is trivial as $u + \varepsilon > u = \sup(\mathscr{U})$ implies that $y \leq u < u + \varepsilon$ 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

$$u \ - \ 1/j \ \leq x_{n_j} \ \leq \ u \ + \ 1/j. \tag{9.62}$$

Base case: We have proved as part of **C.1** that $x_n \geq u + 1$ 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 $x_n \geq u - 1$. Infinitely many of those n must exceed K. We pick one of them and that will be n_1. Clearly, n_1 satisfies (9.62) and this proves the base case.

Induction step: Let us now assume that $n_1 < n_2 < \cdots < n_k$ satisfying (9.62) have been constructed. $x_n \geq u + 1/(k+1)$ 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 $x_n \geq u - 1/(k+1)$. 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 (9.62) 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 (9.62) that $|x_{n_j} - u| \ \leq 1/j \ < \ 1/N < \varepsilon$ for all $j \geq n$ 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 (9.61), $u + \varepsilon \notin \mathscr{U}$. 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 $|w - x_j| \geq \varepsilon$ for all $j \geq n$ and that rules out the possibility of finding n_j such that $\lim\limits_{j\to\infty} x_{n_j} = w$. ■

Corollary 9.8. $\boxed{\star}$ *As in prop.9.43, 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.9.43. ■

When we form the sequence $y_n = -x_n$ then the roles of upper bounds and lower bounds, max and min, inf and sup are reversed. Example: x is an upper bound for $\{x_j : j \geq n\}$ if and only if $-x$ is a lower bound for $\{y_j : j \geq n\}$.

The following "dual" version of prop. 9.43 is a direct consequence of the duality of upper/lower bounds, min/max, inf/sup. See prop.3.59, prop.3.60 and cor.3.4 on p.80.

Proposition 9.44. $\boxed{\star}$ *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded below with tail sets T_n.*

(A) *Let*

$$\begin{aligned}
\mathscr{L} &:= \{y \in \mathbb{R} : T_n \cap]-\infty, y] \neq \emptyset \textit{ for all } n \in \mathbb{N}\},\\
\mathscr{L}_1 &:= \{y \in \mathbb{R} : \textit{ for all } n \in \mathbb{N} \textit{ there exists } k \in \mathbb{Z}_{\geq 0} \textit{ such that } x_{n+k} \leq y\},\\
\mathscr{L}_2 &:= \{y \in \mathbb{R} : \ \exists \textit{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \textit{ such that } x_{n_j} \leq y \textit{ for all } j \in \mathbb{N}\},\\
\mathscr{L}_3 &:= \{y \in \mathbb{R} : x_n \leq y \textit{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{9.63}$$

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.** (9.64)* $$l + \varepsilon \in \mathscr{L} \quad \textit{and} \quad l - \varepsilon \notin \mathscr{L}$$

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.9.43. ■

Proposition 9.45. $\boxed{\star}$ *Let* (x_n) *be a bounded sequence of real numbers. As in prop. 9.43 and prop 9.44, let*

$$\begin{aligned}
u &= \sup(\mathscr{U}) &= \sup\{y \in \mathbb{R} : T_n \cap [y, \infty[\neq \emptyset \textit{ for all } n \in \mathbb{N}\},\\
l &= \inf(\mathscr{L}) &= \inf\{y \in \mathbb{R} : T_n \cap]-\infty, y] \neq \emptyset \textit{ for all } n \in \mathbb{N}\},
\end{aligned} \tag{9.65}$$

Then $\quad u = \limsup\limits_{n\to\infty} x_j \quad$ *and* $\quad l = \liminf\limits_{n\to\infty} x_j.$

Proof that $u = \limsup\limits_{n\to\infty} x_j$: Let

$$\beta_n := \sup_{j \geq n} x_j, \quad \beta := \inf_n \beta_n = \limsup_{n\to\infty} x_n. \tag{9.66}$$

We will prove that β has the properties listed in prop.9.43.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

$$b \in \mathscr{U} \text{ for } b < \beta \quad \text{and} \quad a \notin \mathscr{U} \text{ for } a > \beta. \tag{9.67}$$

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 $j \geq n_0$. 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 $j \geq n_0$.
From that we conclude that $T_n \cap [a, \infty[= \emptyset$ for all $j \geq n_0$. 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 $\beta \leq \beta_n$ 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 (9.67). 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 $(-x_n)$. ■

The material presented above will allow a greatly shortened proof of thm.9.13 (Characterization of limsup and liminf) on p.288.

Remark 9.22 (Simplified proof of thm.9.13 (Characterization of limsup and liminf)). $\boxed{\star}$

We know from prop.9.45 on p.295 that $\limsup\limits_{n\to\infty} x_n$ is the unique number u described in part **C** of prop.9.43, p.293:
$u - \varepsilon \in \mathscr{U}$ and $u + \varepsilon \notin \mathscr{U}$ 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.9.45) means that there are infinitely many n such that $x_n \geq u - \varepsilon$ and $u + \varepsilon \notin \mathscr{U} = \mathscr{U}_3$ means that there are at most finitely many n such that $x_n \geq u + \varepsilon$. This proves **a1** and **a2**.
We also know from prop.9.45 that $\liminf\limits_{n\to\infty} x_n$ is the unique number l described in part **C** of prop.9.44, p.294: $l + \varepsilon \in \mathscr{L}$ 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.9.45) means that there are infinitely many n such that $x_n \leq l + \varepsilon$ and $l - \varepsilon \notin \mathscr{L} = \mathscr{L}_3$ means that there are at most finitely many n such that $x_n \leq l - \varepsilon$. This proves **b1** and **b2**. ■ □

9.9 Sequences of Sets and Indicator functions and their liminf and limsup $\boxed{\star}$

Let Ω be a nonempty 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.

Example 9.9. The following are examples of sequences of real–valued functions.

(a) $f_n : [0,1] \to \mathbb{R};\ x \mapsto x^n$ is a sequence of real–valued functions.
(b) Let $\Omega := \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. be the unit circle in the Euclidean plane. Then $\varphi_n : \Omega \to \mathbb{R};\ \varphi_n(x,y) := \sqrt{x^2 + y^2}$ is a sequence of real–valued functions.
(c) Let $f := \mathbb{R} \to \mathbb{R}$ be a (fixed, but arbitrary) function which is infinitely often differentiable at all its arguments, i.e., $D_n f(x_0) := f^{(n)}(x_0) = \frac{d^n f}{dx^n}\big|_{x=x_0}$ exists for all $x_0 \in \mathbb{R}$ and all $n \in \mathbb{N}$. Then $h_n : \mathbb{R} \to \mathbb{R};\ x \mapsto D_n f(x)$ is a sequence of real–valued functions. □

Definition 9.22. $\boxed{\star}$ [limsup and liminf of a sequence of real–valued functions]

Let Ω be a nonempty 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 $\omega \in \Omega$. We define

$$\liminf_{n\to\infty} f_n \;:\; \Omega \to \mathbb{R} \quad \text{as follows: } \omega \mapsto \liminf_{n\to\infty} f_n(\omega), \tag{9.68}$$

$$\limsup_{n\to\infty} f_n \;:\; \Omega \to \mathbb{R} \quad \text{as follows: } \omega \mapsto \limsup_{n\to\infty} f_n(\omega). \;\square \tag{9.69}$$

It is possible to extend the above definition to situations in which the sets $\{f_n(\omega) : n \in \mathbb{N}\}$ are not bounded. [121]

Remark 9.23. We recall from thm.9.13 (Characterization of limsup and liminf) on p.288 that

$$\liminf_{n\to\infty} f_n(\omega) \;= \inf\{\alpha \in \mathbb{R} : \lim_{j\to\infty} f_{n_j}(\omega) \;=\; \alpha \text{ for some subsequence } n_1 < n_2 < \dots\}, \tag{9.71}$$

$$\limsup_{n\to\infty} f_n(\omega) \;= \sup\{\beta \in \mathbb{R} : \lim_{j\to\infty} f_{n_j}(\omega) \;=\; \beta \text{ for some subsequence } n_1 < n_2 < \dots\}. \;\square \tag{9.72}$$

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.8.15 on p.247 that any such function is the indicator function 1_A of the set

$$A \;:=\; \{f_n = 1\} \;=\; f_n^{-1}(\{1\}) \;=\; \{\omega \in \Omega : f_n(\omega) = 1\} \;\subseteq\; \Omega.$$

Proposition 9.46 (liminf and limsup of $\{0,1\}$–functions)**.** *Let $\Omega \neq \emptyset$ and $f_n : \Omega \to \{0,1\}$. Let $\omega \in \Omega$. Then both $\liminf_n f_n(\omega)$ and $\limsup_n f_n(\omega)$ can only be equal to zero or one. Further,*

$$\liminf_{n\to\infty} f_n(\omega) \;= 1 \;\Leftrightarrow\; f_n(\omega) \;= 1 \textit{ eventually}, \tag{9.73}$$

$$\limsup_{n\to\infty} f_n(\omega) \;= 1 \;\Leftrightarrow\; f_n(\omega) \;= 1 \textit{ for infinitely many } n \in \mathbb{N}. \tag{9.74}$$

PROOF: It follows from (9.71), (9.72) and $0 \leq f_n(\omega) \leq 1$ that $0 \leq \liminf_n f_n(\omega) \leq \limsup_n f_n(\omega) \leq 1$. We conclude from (9.71) 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) (9.73) but also that either $\liminf_n f_n(\omega) \;= 1$ or $\liminf_n f_n(\omega) \;= 0$

[121] In more advanced texts you will find the following

Definition 9.23 (Extended real–valued functions). $\boxed{\star}$ 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–valued 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 their arguments to be $\pm\infty$. Example 1: if $F(x) = \infty$ and $F(y) = \infty$, what is $F(x) - F(y)$?) Example 2: The following rule is applied to products of extended real numbers:

$$0 \cdot \pm\infty \;=\; \pm\infty \cdot 0 \;=\; 0 \tag{9.70}$$

We conclude from (9.72) 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 all except at most finitely many n. This proves not only (both directions(!) of) (9.74) but also that either $\limsup_n f_n(\omega) \ = 1$ or $\limsup_n f_n(\omega) \ = 0$. ■

Next, we look at indicator functions 1_{A_n} of a sequence of sets $A_n \subseteq \Omega$.

Definition 9.24. $\boxed{\star}$

Let $A_n \subseteq \Omega \, (n \in \mathbb{N})$. We define

(9.75) $$A_\star := \bigcup_{n\in\mathbb{N}} \bigcap_{j\geq n} A_j, \qquad A^\star := \bigcap_{n\in\mathbb{N}} \bigcup_{j\geq n} A_j. \ \square$$

Proposition 9.47.

Let $\omega \in \Omega$. Then

(9.76) $\omega \in A_\star \ \Leftrightarrow \omega \in A_n$ *eventually, i.e., $\omega \in A_n$ for all except at most finitely many $n \in \mathbb{N}$.*

(9.77) $\omega \in A^\star \ \Leftrightarrow \omega \in A_n$ *for infinitely many $n \in \mathbb{N}$,*

(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 $1 \leq n_1 < n_2 < \dots$ 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 $k \leq n_k$ (think!) and it follows from $\omega \notin A_{n_k}$ and $A_{n_k} \supseteq \bigcap\limits_{j\geq n_k} A_j \supseteq \bigcap\limits_{j\geq k} A_j$ that there is no $k \in \mathbb{N}$ such that $\omega \in \bigcap\limits_{j\geq k} A_j$.

But then $\omega \notin \bigcup\limits_{k} \bigcap\limits_{j\geq k} A_j = A_\star$ and we are done with the proof of **(a)**.

(b) Proof that $\omega \in A_n$ for all except at most finitely many $n \in \mathbb{N}$ implies $\omega \in A_\star$:
By assumption there exists some $N = N(\omega) \in \mathbb{N}$ such that $\omega \in A_n$ for all $n \geq N$.

It follows that $\omega \in \bigcap\limits_{n\geq N} A_n \subseteq \bigcup\limits_{m\in\mathbb{N}} \bigcap\limits_{n\geq m} A_n \ = A_\star$ 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 $1 \leq n_1 < n_2 < \dots$ such that $\omega \in A_{n_j}$ for all $j \in \mathbb{N}$.

We observe that $\omega \in \bigcup\limits_{j\geq n} A_j$ for all $n \in \mathbb{N}$. As $\omega \in \bigcup_{j\geq 1} A_j$ there exists $n_1 \geq 1$ 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 $1 \leq j \leq k$ then we find n_{k+1} as follows: As $\omega \in \bigcup_{j\geq n_k+1} A_j$ there exists $n_{k+1} \geq n_k + 1$ 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 $\Gamma_n := \bigcup_{j\geq n} A_j$.

Let $1 \le n_1 < n_2 < \dots$ such that $\omega \in A_{n_j}$ for all $j \in \mathbb{N}$. Let $k \in \mathbb{N}$.
Then $n_k \ge k$, 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 9.48 (liminf and limsup of indicator functions)**.**

Let $A_n \subseteq \Omega\,(n \in \mathbb{N})$ and let $A_\star, A^\star$ be the sets defined in (9.75). Then

$$1_{A_\star} = \liminf_{n\to\infty} 1_{A_n} \quad \textit{and} \quad 1_{A^\star} = \limsup_{n\to\infty} 1_{A_n} \tag{9.78}$$

PROOF: Let $\omega \in \Omega$. Then

$$\begin{aligned} 1_{A_\star}(\omega) = 1 \;\Leftrightarrow\; \omega \in A_\star \;&\Leftrightarrow \omega \in A_n \text{ for all except at most finitely many } n \in \mathbb{N} \\ &\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for all except at most finitely many } n \in \mathbb{N} \\ &\Leftrightarrow \liminf_n 1_{A_n}(\omega) = 1 \end{aligned} \tag{9.79}$$

The second equivalence follows from prop.9.47 and the last equivalence follows from prop.9.46 and this proves the first equation. Similarly we have

$$\begin{aligned} 1_{A^\star}(\omega) = 1 \;\Leftrightarrow\; \omega \in A^\star \;&\Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N} \\ &\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for infinitely many } n \in \mathbb{N} \\ &\Leftrightarrow \limsup_n 1_{A_n}(\omega) = 1 \end{aligned} \tag{9.80}$$

Again the second equivalence follows from prop.9.47 and the last equivalence follows from prop.9.46. ■

This last proposition is the reason for the following definition.

Definition 9.25. $\boxed{\star}$ [limsup and liminf of a sequence of sets]

Let Ω be a nonempty set and let $A_n \subseteq \Omega$ $(n \in \mathbb{N})$. We define

$$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{j \ge n} A_j, \tag{9.81}$$

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{j \ge n} A_j. \tag{9.82}$$

We call $\liminf_{n\to\infty} A_n$ the **limit inferior** and $\limsup_{n\to\infty} A_n$ the **limit superior** of the sequence A_n.

We note that $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$ if and only if the functions $\liminf_{n\to\infty} 1_{A_n}$ and $\limsup_{n\to\infty} 1_{A_n}$ coincide (prop. 9.48) which is true if and only if the sequence $1_{A_n}(\omega)$ has a limit for all $\omega \in \Omega$ (thm.9.14 on p.290). In this case we define

$$\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n \tag{9.83}$$

and we call this set the **limit** of the sequence A_n. □

Note 9.4 (Notation for limits of monotone sequences of sets). [122]

Let (A_n) be a nondecreasing 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 nonincreasing sequence of sets, i.e., $B_1 \supseteq B_2 \supseteq \dots$ and let $B := \bigcap_n B_n$. We write suggestively

$$A_n \uparrow A \quad (n \to \infty), \qquad B_n \downarrow B \quad (n \to \infty). \ \square$$

Example 9.10. Let $A_n \subseteq \Omega$.

(9.84) **(a)** If $A_n \uparrow$ then $\displaystyle \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n.$

(9.85) **(b)** If $A_n \downarrow$ then $\displaystyle \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n. \ \square$

Note 9.5 (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 $T_n := \{x_j : j \geq n\}$.

⚠

(a) Do not confuse $\displaystyle \liminf_{n\to\infty} x_n = \sup_n \big(\inf(T_n)\big)$ with $\displaystyle \liminf_{n\to\infty} T_n = \bigcup_n \big(\bigcap_{k\geq n} T_k\big)$.

(b) Do not confuse $\displaystyle \limsup_{n\to\infty} x_n = \inf_n \big(\sup(T_n)\big)$ with $\displaystyle \limsup_{n\to\infty} T_n = \bigcap_n \big(\bigcup_{k\geq n} T_k\big)$.

Those two concepts are very different: $\liminf\limits_n x_n$ $(\limsup\limits_n x_n)$ 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 9.10 and the fact the the sequence of tailsets T_n is always nonincreasing. $\square$

We conclude this subchapter with a remark about the usefulness of the liminf and limsup of a sequence of sets.

Remark 9.24. In probability theory one models events, i.e., pre–images that correspond to collections of random outcomes, as sets. For example, if we have "random variables"

$$X_K : \Omega \longrightarrow [1,6]_{\mathbb{Z}}\,; \quad X_k(\omega) = k\text{–th throw of a die} \quad (\omega \text{ indicates randomness})$$

$$\text{then} \quad A_k := \{X_k = 5 \textbf{ or } X_k = 6\} = X_k^{-1}\big(\{5,6\}\big)$$

is the event that the $k-th$ throw resulted in a 5 or a 6.

Proposition 9.47 on p.298 then implies, in connection with Definition 9.25, the following.

[122] See note 9.2 on p.266.

If $(A_n)_n$ is a sequence of events then

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \text{the event that } A_n \text{ happens eventually}\,,\\ \limsup_{n\to\infty} A_n &= \text{the event that } A_n \text{ happens infinitely often}\,.\end{aligned}$$

Considering that "eventually" means that there are at most finitely many exceptions,

$$\liminf_{n\to\infty} A_n = \text{the event that } A_n \text{ does } \textbf{NOT} \text{ happen at most finitely many times}\,.$$

In the context of repeatedly rolling a die,

$$\begin{aligned}\liminf_{n\to\infty}\{X_n = 5 \textbf{ or } X_n = 6\} &= \text{the event that a 5 or 6 will be rolled eventually}\,,\\ \limsup_{n\to\infty}\{X_n = 5 \textbf{ or } X_n = 6\} &= \text{the event that a 5 or 6 will be rolled infinitely often}\,.\end{aligned}$$

An example for the usefulness of is Kolmogorov's zero–one law: Let $P(B)$ denote the probability that the event B happens. If the events A_n happen independently of each other then

$$\begin{aligned}&\text{either } P\big(\liminf_{n\to\infty} A_n\big) = 0 \qquad \text{or } P\big(\liminf_{n\to\infty} A_n\big) = 1\,,\\ &\text{either } P\big(\limsup_{n\to\infty} A_n\big) = 0 \qquad \text{or } P\big(\limsup_{n\to\infty} A_n\big) = 1\,.\ \blacksquare\end{aligned}$$

9.10 Sequences that Enumerate Parts of $\mathbb{Q}$ $\boxed{\star}$

We will briefly study the following remarkable property of the real numbers: One can find a single sequence $q_n \in \mathbb{R}$ of real numbers whose members come arbitrarily close to every real number.

We informally defined the real numbers in ch.2.3 (Numbers) on p.24 as the set of all decimals, i.e., all numbers x which can be written as

$$x = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad \text{where } d_j \text{ is a digit, i.e., } d_j = 0,1,2,\dots,9, \tag{9.86}$$

$$\text{i.e., } x = \lim_{k\to\infty} x_k \quad \text{where } x_k = m + \sum_{j=1}^{k} d_j 10^{-j}. \tag{9.87}$$

Each x_k is a (finite) sum of fractions, hence $x_k \in \mathbb{Q}$.

We proved in cor.7.5 on p.229 that $\mathbb{Q}$ and hence all of its subsets are countable: 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 $A := \mathbb{Q}$ as follows.

Let $x \in \mathbb{R}$ have the representation (9.87). Then $x_k \in \mathbb{Q}$ for each $k \in \mathbb{N}$, hence 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 (9.87) that $q_{n_k} \to x$ as $k \to \infty$. In other words, we have proved the following

Theorem 9.15 (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 $x = \lim\limits_{k\to\infty} q_{n_k}$ $\blacksquare$.

Remark 9.25.

(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.9.15: 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 $A := \{x \in \mathbb{R} : x^2 \leq 2\} \ = [\,-\sqrt{2}, \sqrt{2}\,]$ and let $A_{\mathbb{Q}} := A \cap \mathbb{Q} = \{q \in \mathbb{Q} : q^2 \leq 2\}$. A is of such a shape that for any $x \in A$ the partial sums $x_k \ = \ m \ + \ \sum\limits_{j=1}^{k} d_j 10^{-j}$ which converge to x belong to $A_{\mathbb{Q}}$. (Why? Especially, why also for $x = \pm\sqrt{2}$?) □

9.11 Exercises for Ch.9

9.11.1 Exercises for Ch.9.1 (The Ordered Fields of the Real and Rational Numbers)

Exercise 9.1. Prove prop.9.2 on p.252 of this document: Fields are integral domains. □

Exercise 9.2. Prove prop.9.3 on p.253 of this document: yadayada □

Exercise 9.3. Prove prop.9.4 on p.253 of this document:
Let $a, b, c, d \in F$ such that $b, d \neq 0$. If $\frac{a}{b} = \frac{c}{d}$ then ad = bc. □

Exercise 9.4. Prove prop.9.5 on p.253: If $a, b, c \in F$ such that $b, c \neq 0$. then $\frac{ac}{bc} = \frac{a}{b}$. □

Exercise 9.5. Prove prop.9.6 on p.253: If $a, b, c, d \in F$ such that $b, d \neq 0$ then $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$. □

Exercise 9.6. Prove prop.9.7 on p.253:
If $a, b, c, d \in F$ such that $b, d \neq 0$. then $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$. In particular, $(\frac{b}{d})^{-1} = \frac{d}{b}$. □

Exercise 9.7. Prove prop.9.8 on p.254 of this document:

(a) Let $a \in F$. Then $a > 0$ if and only if $a^{-1} > 0$.

(b) Let $a, b \in F$. If $0 < a < b$ then $0 < \frac{1}{y} < \frac{1}{x}$. □

Exercise 9.8. Prove prop.9.2 on p.254 of this document: Let $a, b \in F_{\neq 0}$. Then

(a) $\frac{a}{b} > 0$ if and only if $\frac{b}{a} > 0$ and $\frac{a}{b} < 0$ if and only if $\frac{b}{a} > 0$,

(b) $\frac{a}{b} > 0$ if and only if either both $a, b > 0$ or both $a, b < 0$. □

Exercise 9.9. Prove thm.9.1 on p.254 of this document:
Let $a, b \in R$ such that $a < b$. Then there exists $c \in R$ such that $a < c < b$. □

Exercise 9.10. Prove thm.9.3 on p.256 of this document:

(a) The assignments $(a, b) \mapsto a + b$ and $(a, b) \mapsto a \cdot b$ are binary operations on $\mathbb{Q}$, i.e., sums and products of rational numbers are rational numbers.

(b) The triplet $(\mathbb{Q}, +, \cdot)$ is an integral domain.

(c) Let $\mathbb{Q}_{>0} := \mathbb{R}_{>0} \cap \mathbb{Q}$. Then $(\mathbb{Q}, +, \cdot, \mathbb{Q}_{>0})$ is an ordered integral domain which satisfies the following: if $a, b \in \mathbb{Q}$ then $a < b$ with respect to the ordering induced by $\mathbb{Q}_{>0}$ if and only if $a < b$ with respect to the ordering induced by $\mathbb{R}_{>0}$

(d) $(\mathbb{Q}_{\neq 0}, \cdot)$ is a (commutative) group.

Hint: We suggest you break down the proof as follows.

(a) The assignments $(a,b) \mapsto a+b$ and $(a,b) \mapsto a \cdot b$ are binary operations on $\mathbb{Q}$, i.e., sums and products of rational numbers are rational numbers.

(b) It follows from $\mathbb{Q} \subseteq \mathbb{Q} \subseteq \mathbb{R}$ (HOW?) that both "+" and "$\cdot$" are associative and commutative, that they satisfy distributivity (see Definition 3.7**(c)** on p.60) and that 0 and 1 are the neutral elements for "+" and "$\cdot$" respectively and the are rational numbers.

(c) Supply the missing pieces which prove that $(\mathbb{Q}, +, \cdot)$.

(d) Use what you know about $\mathbb{R}_{>0}$ to first show that $\mathbb{Q}_{>0}$ satisfies Definition 3.11**(a)** – **(c)** of a positive cone (see p.68).

(e) Use what you know about $\mathbb{R}_{>0}$ to show that $\mathbb{Q}_{>0}$ satisfies Definition 3.11**(d)** of a positive cone (not as obvious as proving properties Definition 3.11**(a)** – **(c)**).

(f) Prove that if $a, b \in \mathbb{Q}$ then $a < b$ with respect to the ordering induced by $\mathbb{Q}_{>0}$ if and only if $a < b$ with respect to the ordering induced by $\mathbb{R}_{>0}$.

(g) Prove that any rational nonnegative number possesses a multiplicative inverse. □

Exercise 9.11. Prove cor.9.3 on p.257 of this document: There are no upper bounds for $\mathbb{N}$ in $\mathbb{Q}$. □

9.11.2 Exercises for Ch.9.2 (Minima, Maxima, Infima and Suprema)

Exercise 9.12. Prove (9.14) of prop.9.11 on p.262 of this document: Let X be a nonempty set and $\varphi, \psi : X \to \mathbb{R}$. Let $A \subseteq X$. Then $\inf\{\varphi(x) + \psi(x) : x \in A\} \geq \inf\{\varphi(y) : y \in A\} + \inf\{\psi(z) : z \in A\}$. □

9.11.3 Exercises for Convergence

Exercise 9.13. Prove example 9.5 part **(c)**: Let $z_n := x_0$ for some $x_0 \in \mathbb{R}$ $(n \in \mathbb{N})$. Then $\lim\limits_{n\to\infty} z_n = x_0$. If that is too abstract, try to prove the special case **(b)** first. □

Exercise 9.14. Prove example 9.5**(c)** on p.264 of this document: Let $z_n := x_0$ for some $x_0 \in \mathbb{R}$ $(n \in \mathbb{N})$. Then $\lim\limits_{n\to\infty} z_n = x_0$. □

Exercise 9.15. Prove thm.9.5 on p.265 of this document: Let $(x_n)_n$ be a convergent sequence of real numbers. Then its limit is uniquely determined. □

Exercise 9.16. Prove prop.9.12 on p.265 of this document: Let $a, b \in R$. Then $a = b \Leftrightarrow |a - b| < \varepsilon$ for all $\varepsilon > 0$. □

Exercise 9.17. Prove prop.9.14 on p.266: Let $(x_n)_n$ be a sequence of real numbers such that $\lim\limits_{n\to\infty} x_n$ exists. Let $K \in \mathbb{N}$. For $n \in \mathbb{N}$ let $y_n := x_{n+K}$. Then $(y_n)_n$ has the same limit.
Hint: Use prop.9.13. □

Exercise 9.18. Prove prop.9.15 on p.266 of this document: Let $(x_n)_n$ be a sequence in $\mathbb{R}$ with limit $a \in \mathbb{R}$. Then this sequence is bounded. □

Exercise 9.19. Prove (9.32) of prop.9.20 on p.270: Let $a, b \in \mathbb{R}$. Then $]a, b[= \bigcup\limits_{n\in\mathbb{N}} [a + 1/n, b - 1/n]$. Adapt the proof of (9.31) but note that this one is simpler. There are only two cases to worry about: $a \geq b$ (very easy!) vs $a < b$. □

9.11.4 Exercises for Continuity

Exercise 9.20. Prove prop.9.23 on p.272 of this document: Let $A, B \subseteq \mathbb{R}$ be nonempty, $f : A \to \mathbb{R}$ continuous at $x_0 \in A$ and $g : B \to \mathbb{R}$ continuous at $f(x_0)$. Assume further that $f(A) \subseteq B$, i.e., $f(x) \in B$ for all $x \in A$. Then the composition $g \circ f : X \to Y$ is continuous at x_0. □

Exercise 9.21. Let $A \in \mathbb{R}$ be an interval with endpoints $a < b$, i.e., A is either of $]a, b[$, $]a, b\,]$, $[\,a, b[$, $[\,a, b\,]$. Let $a < c < b$ and $A_1 := \{x \in A : x \leq c\}$ and $A_2 := \{x \in A : x \geq c\}$.

Let $f_1 : A_1 \to \mathbb{R}$ and $f_2 : A_2 \to \mathbb{R}$ be two continuous, real–valued functions such that $f_1(c) = f_2(c)$. Prove that the "spliced" function

$$f(x) := \begin{cases} f_1(x) & \text{for } x \in A_1, \\ f_2(x) & \text{for } x \in A_2 \end{cases} \tag{9.88}$$

is continuous on A. □

Exercise 9.22. Let $a, b, c, d \in \mathbb{R}$ such that $a < b$ and $c < d$. Let $f :]a, b[\to]c, d[$ be bijective and strictly monotone, i.e., strictly increasing or decreasing. Prove that both f and f^{-1} are continuous.

Hint: Use thm.9.7 on p.273.

Exercise 9.23. Prove prop.9.22 (All polynomials are continuous) on p.272. □

Exercise 9.24. Let $n \in \mathbb{N}$ and let $p(x) := \sum\limits_{j=0}^{n} a_j x^j$ $(a_j \in \mathbb{R}, a_n \neq 0)$ be a polynomial. Let $A \subseteq \mathbb{R}$ be unbounded. Prove that the direct image $p(A)$ is unbounded. **Hint:** Show that there is a sequence $x_n \in A$ such that $x_n \to \infty$ or $x_n \to -\infty$. Examine $y_k := \frac{p(x_k)}{a_n x_k^n}$ and use prop.9.17 (rules of arithmetic for limits). □

Exercise 9.25. Prove prop.9.24 on p.273 of this document:

Let $A \subseteq \mathbb{R}$, $x_0 \in A$, and let $f : A \to \mathbb{R}$ be a real–valued function with domain A. Then f is continuous at x_0 if and only if there exists $\varepsilon^\star > 0$ which satisfies the following:

for any $\varepsilon \in]0, \varepsilon^\star]$ there exists $\delta > 0$ such that either one of the following equivalent statements is satisfied:

$$\begin{array}{ll} \textbf{(a)} & f\big(\{x \in A : |x - x_0| < \delta\}\big) \subseteq \{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon\}, \\ \textbf{(b)} & |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon \text{ for all } x \in A. \end{array}$$

Hint: If $\varepsilon > \varepsilon^\star$, what can you say about the sets $\{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon^\star\}$ and $\{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon\}$ □

9.11.5 Exercises for Ch.9.4 (Rational and Irrational Numbers)

Exercise 9.26. Prove prop.9.27 on p.275 of this document: Let $m, n, s, t \in \mathbb{Z}$ be such that m and n do not have any common factors. If $\frac{m}{n} = \frac{s}{t}$ then m divides s and n divides t. □

Exercise 9.27. Prove cor.9.32 on p.276 of this document: There is no smallest positive irrational number. □

Exercise 9.28. Prove prop.9.31 on p.276 of this document: If $x, y \in \mathbb{R}$ such that $x < y$ then there exists irrational z such that $x < z < y$. □

9.11.6 Exercises for Geom. series and Decimal Expansions

Exercise 9.29. Prove **(a)** and **(b)** of prop.9.34 on p.280 of this document:
Let $n \in \mathbb{Z}_{\geq 0}$ and $d_j \in [0,9]_{\mathbb{Z}}$ for $j \geq n$. Then

$$\textbf{(a)} \quad 0 \leq 9 \sum_{j=n}^{\infty} 10^{-j} = \frac{1}{10^{n-1}}, \qquad \textbf{(b)} \quad \sum_{j=n}^{\infty} d_j 10^{-j} \leq \frac{1}{10^{n-1}}. \ \square$$

Exercise 9.30. Prove lemma 9.1 on p.281 of this document: Let $x = m + \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ ($m \in \mathbb{Z}_{\geq 0}, d_j \in [0,9]_{\mathbb{Z}}$ for all $j \in \mathbb{N}$) be a decimal expansion of a nonnegative real number x. Then

(a) $m \leq x \leq m+1$,
(b) $x = m \Leftrightarrow d_j = 0$ for all $j \in \mathbb{N}$,
(c) $x = m+1 \Leftrightarrow d_j = 9$ for all $j \in \mathbb{N}$. □

Exercise 9.31. Prove cor.9.5 on p.282 of this document:
If $x \in \mathbb{R}$ has different decimal expansions then $x \in \mathbb{Q}$. □

9.11.7 Exercises for Ch.9.7 (Countable and Uncountable Subsets of the Real Numbers)

Exercise 9.32. Prove prop.9.38 on p.286 of this document:
The set of all algebraic numbers is countable.

Hint: Show that the sets $P_n := \{\text{polynomials } p(x) = \sum\limits_{j=0}^{n} a_j x^j : a_j \in \mathbb{Z} \text{ and } -n \leq a_j \leq n\}$ are finite. Everything else will be easy. □

Exercise 9.33. Prove prop.9.40 on p.286 of this document: Let $r \in \mathbb{Q}$. Then r is algebraic. □

9.11.8 Exercises for Ch.9.8 (Limit Inferior and Limit Superior)

Exercise 9.34. Let $a, b \in \mathbb{R}$ such that $a < b$ and let $(x_n)_n$ be a sequence such that $x_j \in \{a, b\}$ for all j. Prove the following:

(a) If $x_j = a$ eventually then $\limsup\limits_{j\to\infty} x_j = a$, else $\limsup\limits_{j\to\infty} = b$.
(b) If $x_j = b$ eventually then $\liminf\limits_{j\to\infty} x_j = b$, else $\liminf\limits_{j\to\infty} = a$.

Exercise 9.35. Prove cor.9.6 on p.291:
Let $x_n, y_n \in \mathbb{R}$ be two sequences of real numbers. Assume there is $K \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq K$. Then

$$\limsup_{n\to\infty} x_n = \limsup_{n\to\infty} y_n \quad \text{and} \quad \liminf_{n\to\infty} x_n = \liminf_{n\to\infty} y_n. \ \square$$

Exercise 9.36. Prove cor.9.7 on p.292 of this document:
Let $x_n \in [0, \infty[$ such that $\limsup\limits_{n\to\infty} x_n = 0$. Then $(x_n)_n$ converges to zero. □

Exercise 9.37. Let $x_n := (-1)^n$ for $n \in \mathbb{N}$. Prove that $\liminf_n x_n = -1$ and $\limsup_n x_n = 1$ by working with the tailsets of that sequence. Do not use anything after Definition 9.20 (Tail sets of a sequence) on p.287! **Hint:** What is α_n and β_n?

Very similar to the previous exercise, but slightly more difficult:

Exercise 9.38. Let $x_n := (-1)^(n + \frac{1}{n})$ for $n \in \mathbb{N}$. Prove that $\liminf_n x_n \; = -1$ and $\limsup_n x_n \; = 1$ by working with the tailsets of that sequence. Do not use anything after Definition 9.20 (Tail sets of a sequence) on p.287! **Hint:** Compute α_n and β_n for $n = 1, 2, 3, 4, 5, 6$ to see the pattern: Look separately at odd and even indices to prove that both $|\beta_n - 1|$ and $|\alpha_n + 1|$ are either $\frac{1}{n}$ or $\frac{1}{n+1}$. What follows for the convergence behavior of α_n and β_n?

Exercise 9.39. Prove the assertions of example 9.10 on p.300. □

Exercise 9.40. It was mentioned in rem.9.21 on p.292 that use of prop.9.42 leads to an easy, alternate, proof of prop.9.19 (Domination Theorem for Limits):

If two sequences $x_n, y_n \in \mathbb{R}$ satisfy $x_n \leq y_n$ eventually then $\lim\limits_{n \to \infty} x_n \leq \lim\limits_{n \to \infty} y_n$.

Do this alternate proof! □

10 Cardinality II: Comparing Uncountable Sets

If we want to compare sets based on their sizes then we have a good idea how to go about it, at least as far as finite sets (excluding the empty set) and countable sets are concerned. We can biject them to a subset of the natural numbers, and then compare their images in $\mathbb{N}$: If X_1 and X_2 are finite sets such that we have bijective functions $X_1 \xrightarrow{f_1} [n_1]$ and $X_2 \xrightarrow{f_2} [n_2]$ such that $n_1 \leq n_2$, and if C is countably infinite with a bijection $C \xrightarrow{c} \mathbb{N}$, then we have

$$0 = |\emptyset| < |X_1| = n_1 \leq |X_2| = n_2 < |C| = \infty. \tag{10.1}$$

At times, when we compare sets, we want to know more about them than just a single number (if we think of infinity as a number). Note that we have a corresponding chain of injections

$$\emptyset \rightarrowtail X_1 \rightarrowtail X_2 \rightarrowtail C. \tag{10.2}$$

Would it be feasible to use injective functions as a means to compare sets as far as their size is concerned? Of course we lose information if we boil down the information about two sets to whether or not there exists an injection from one of them to the other, but for many purposes it turns out to be fruitful to know whether such is the case, and still not worry about any more detail. For example, all countably infinite and uncountable sets have the same size ∞, but it turns out that there are many degrees of uncountability if one considers a set X no biger than a set Y if there exists an injection $X \rightarrowtail Y$.

The above leads us to the definition of cardinality.

10.1 The Cardinality of a Set

Definition 10.1 (Cardinality Comparisons)**.** Given are two arbitrary sets X and Y. We say that

(a) X, Y **have same cardinality**, and we write $\mathbf{card}(X) = \mathbf{card}(Y)$, if either both $X, Y \neq \emptyset$ and there is a bijection $f : X \xrightarrow{\sim} Y$, or if both X and Y are empty. Otherwise we write $\mathbf{card}(X) \neq \mathbf{card}(Y)$

(b) the **cardinality of** X **is less than or equal to the cardinality of** Y, and we write $\mathbf{card}(X) \leq \mathbf{card}(Y)$, if there is an injective mapping $f : X \to Y$ or if X is empty.

(c) the **cardinality of** X **is less than the cardinality of** Y, and we write $\mathbf{card}(X) < \mathbf{card}(Y)$, if both $\text{card}(X) \leq \text{card}(Y)$ and $\text{card}(Y) \neq \text{card}(X)$, i.e., if either $X = \emptyset$ and $Y \neq \emptyset$, or there is an injective mapping but not a bijection $f : X \to Y$.

(d) the **cardinality of** X **is greater than or equal to the cardinality of** Y, and we write $\mathbf{card}(X) \geq \mathbf{card}(Y)$, if $\text{card}(Y) \leq \text{card}(X)$.

(e) the **cardinality of** X **is greater than the cardinality of** Y, and we write $\mathbf{card}(X) > \mathbf{card}(Y)$, if $\text{card}(Y) < \text{card}(X)$. □

Note the following concerning the above definition.

(a) It does not specify how $\mathbf{card}(X)$ itself is defined. This will be done in Definition 10.2 on p.308.

(b) It covers the cases $X = \emptyset$ and/or $Y = \emptyset$. We have $\mathbf{card}(\emptyset) \leq \mathbf{card}(Y)$ for any set Y, $\mathbf{card}(\emptyset) \neq \mathbf{card}(Y)$ if Y is not empty, and $\mathbf{card}(X) = \mathbf{card}(Y)$ if $X = Y = \emptyset$. □

Example 10.1. Let A, B be two sets such that $A \subseteq B$. Then $\text{card}(A) \leq \text{card}(B)$.

PROOF:

Case 1: $A = \emptyset$. It then is true by definition that $\text{card}(\emptyset) \leq \text{card}(B)$ for any set B.

Case 2: $A \neq \emptyset$. It follows from $B \supseteq A$ that $B \neq \emptyset$. Further the mapping $x \mapsto x$ is injective. This proves $\text{card}(A) \leq \text{card}(B)$. ■

Theorem 10.1 (B/G thm.13.31)**.**

*Let X be a set. Then **card**$(X) <$ **card**(2^X).*
In other words, X can be injected into 2^X, but it is not possible to find bijective $f : X \xrightarrow{\sim} 2^X$.

Proof: The function $x \mapsto \{x\}$ is an injection from X into 2^X, hence **card**$(X) \leq$ **card**(2^X).

It remains to show that there is no bijective function $f : X \xrightarrow{\sim} 2^X$. We will show that it is not even possible to find surjective f with domain X and codomain 2^X.

We assume to the contrary that such f exists. Let

$$\Gamma := \{x \in X : x \notin f(x)\}.$$

Obviously $\Gamma \subseteq X$, i.e., $\Gamma \in 2^X$. f is surjective, hence there exists $x_0 \in X$ such that $f(x_0) = \Gamma$.

Case 1: Assume $x_0 \in \Gamma$. Then $x_0 \notin f(x_0)$, i.e., $x_0 \notin \Gamma$. We have a contradiction.

Case 2: Assume $x_0 \notin \Gamma$. Then $x_0 \in f(x_0)$, i.e., $x_0 \in \Gamma$. Again, we have a contradiction.

We conclude that there is no surjective $f : X \to 2^X$.

Proposition 10.1. *Let X, Y be two sets such that card(X) = card(Y). Then card(2^X) = card(2^Y).*

PROOF: The proof is left as exercise 10.2. ■

10.2 Cardinality as a Partial Ordering

We assume in this subchapter that all sets are subsets of a universal set Ω. Having such a universal set allows us to declare on its power set 2^Ω equivalence relations. If we work with specific sets, e.g. the set $\mathbb{R}$ of all real numbers, we assume implicitly that those sets are contained in Ω.

We defined in Definition 10.1 on p.307 the meaning of **card**$(X) =$ **card**(Y) and **card**$(X) \leq$ **card**(Y) for two sets X and Y but we never defined the expression **card**(X) per se. This will be done now.

Definition 10.2 (Cardinality as an Equivalence Class)**.** $\boxed{\star}$ Let $X, Y \subseteq \Omega$.

We call X and Y equivalent and we write $X \sim Y$, if and only if **card**$(X) =$ **card**(Y), i.e., either both X and Y are empty, or both are not empty and there is a bijection $f : X \xrightarrow{\sim} Y$.

The proposition following this definition shows that "$\sim$" is indeed an equivalence relation on 2^Ω. This justifies to define for a set $X \subseteq \Omega$ its **cardinality** as follows:

(10.3) $\quad \textbf{card}(X) := [X] \quad$ (the equivalence class of X w.r.t "$\sim$") .

In other words,

(10.4) $\quad \textbf{card}(\emptyset) := \{\emptyset\},$

(10.5) $\quad \textbf{card}(X) := \{Y \subseteq \Omega : \exists \text{ bijection } X \to Y\}$ if $X \neq \emptyset$. □

Proposition 10.2. *$X \sim Y$ as defined above is an equivalence relation on 2^Ω.*

PROOF:

Proof strategy: How about this? The equals relation is reflexive symmetric and transitive. Since $X \sim Y$, if and only if $\textbf{card}(X) = \textbf{card}(Y)$, $X \sim Y$ inherits those properties from the equals relation $\textbf{card}(X) = \textbf{card}(Y)$.

Here is the problem. When we defined $\textbf{card}(X) = \textbf{card}(Y)$ in Definition 10.1 on p.307 we did so without giving any meaning to the expressions card(X) and card(Y). Rather, we defined this expression, and hence $X \sim Y$, to mean the following:

(10.6) $\quad X \sim Y \quad \Leftrightarrow \quad$ either $\quad X = Y = \emptyset \quad$ or $\quad \left[\, X, Y \neq \emptyset \text{ and there exists a bijection } X \xrightarrow{\sim} Y \,\right]$

For this reason a correct proof of this proposition must refer to (10.6).

Let $X, Y, Z \subseteq \Omega$.

Case 1. $X = \emptyset$.

Reflexivity: Clearly, $X = X = \emptyset$, hence $X \sim X$.

Symmetry: If $X \sim Y$ then it follows from (10.6). that $Y = \emptyset$. Thus $X = Y = \emptyset$, thus $Y \sim X$.

Transitivity: Assume that $X \sim Y$ and $Y \sim Z$. Since $X = \emptyset$ and $X \sim Y$ it follows from (10.6) that $Y = \emptyset$. Since $Y = \emptyset$ and $Y \sim Z$ it follows from (10.6) that $Z = \emptyset$. Thus $X = Z = \emptyset$, thus $X \sim Z$.

Case 2. $X \neq \emptyset$.

Reflexivity: $id_X : x \mapsto x$ is a bijection $X \to X$, hence $X \sim X$.

Symmetry: If $X \sim Y$ then it follows from (10.6) and $X \neq \emptyset$ that $Y \neq \emptyset$ and that there exists a bijection $f : X \xrightarrow{\sim} Y$. The inverse $f^{-1} : Y \to X$ then is a bijection $Y \xrightarrow{\sim} X$. It follows that $Y \sim X$

Transitivity: Assume that $X \sim Y$ and $Y \sim Z$. Since $X \neq \emptyset$ and $X \sim Y$ it follows from (10.6) that $Y \neq \emptyset$ and that there exists a bijection $f : X \xrightarrow{\sim} Y$. Since $Y \neq \emptyset$ and $Y \sim Z$ it follows from (10.6) that $Z \neq \emptyset$ and that there exists a bijection $g : Y \xrightarrow{\sim} Z$. It follows from prop.5.5**(c)** that the composition $g \circ f$ of the two bijective functions f and g is a bijection $X \xrightarrow{\sim} Z$. ■

Next we collect some material to prove the Cantor–Schröder–Bernstein Theorem. This theorem allows us to prove antisymmetry of the relation

$$\textbf{card}(X) \leq \textbf{card}(Y), \quad \text{defined on the set } \mathscr{A} := \{\textbf{card}(X) : X \subseteq \Omega\}.$$

Proposition 10.3. *Let X', X'', Y', Y'' be nonempty sets such that $X' \cap X'' = \emptyset$ and $Y' \cap Y'' = \emptyset$ Let $f' : X' \to Y'$ and $f'' : X'' \to Y''$. Then the function*

$$f : X' \biguplus X'' \longrightarrow Y' \biguplus Y'' ; \qquad x \mapsto \begin{cases} f'(x) & \text{if } x \in X', \\ f''(x) & \text{if } x \in X'', \end{cases}$$

satisfies the following:

(a) *If f' and f'' are injective then f is injective.*
(b) *If f' and f'' are surjective then f is surjective.*
(c) *If f' and f'' are bijective then f is bijective.*

The proof is left as exercise 10.3 (see p.322). ■

The following theorem from Tarski and the proof of the subsequent Cantor–Schröder–Bernstein Theorem have been found in the online article

https://chiasme.wordpress.com/2013/11/20/a-short-proof-of-cantor-bernstein-schroeder-theorem/

(A short proof of Cantor–Bernstein–Schröder Theorem). The suggestion to prove Cantor–Schröder–Bernstein with help of Tarski's Theorem was given to the author by David Biddle.

Theorem 10.2 (Tarski's Fixed Point Theorem)**.**

Let Ω be a set and let $\varphi : 2^\Omega \longrightarrow 2^\Omega$ be nondecreasing with respect to "$\subseteq$", i.e.,

$$A, B \subseteq \Omega \text{ and } A \subseteq B \quad \Rightarrow \quad \varphi(A) \subseteq \varphi(B)\,.$$

Then φ has a ***fixed point****, i.e., there exists an argument $A_0 \in 2^\Omega$ such that $\varphi(A_0) = A_0$.*

PROOF: Let

$$\mathfrak{F} \;:=\; \{A \in 2^\Omega : A \subseteq \varphi(A)\}\,, \qquad A_0 := \bigcup [A : A \in \mathfrak{F}]\,.$$

We will show that A_0 is a fixed point for φ. First we prove

(A) $$A_0 \;\subseteq\; \varphi(A_0)\,.$$

To see this we observe that

$$A \in \mathfrak{F} \;\Rightarrow\; A \subseteq \varphi(A)\,. \quad \text{Since } \varphi \text{ is nondecreasing and } A \subseteq A_0, \quad \varphi(A) \subseteq \varphi(A_0)\,.$$
$$\text{Thus } A \subseteq \varphi(A_0) \quad \text{for all } A \in \mathfrak{F}\,, \quad \text{Thus } A_0 \;=\; \bigcup [A : A \in \mathfrak{F}] \;\subseteq\; \varphi(A_0)\,.$$

We have shown **(A)**. It remains to prove

(B) $$\varphi(A_0) \;\subseteq\; A_0\,.$$

We just proved that $A_0 \subseteq \varphi(A_0)$. Since φ is nondecreasing, $\varphi(A_0) \subseteq \varphi\big(\varphi(A_0)\big)$.

$$\text{Thus } \varphi(A_0) \in \mathfrak{F}, \;\; \text{thus } \varphi(A_0) \subseteq \bigcup [A : A \in \mathfrak{F}]\,, \;\; \text{i.e., } \varphi(A_0) \subseteq A_0\,.$$

We have shown that A_0 is a fixed point for φ. ■

Example 10.2. Can one replace in Tarski's fixed point theorem the assumption φ nondecreasing with φ nonincreasing? In other words, if $\varphi : 2^\Omega \longrightarrow 2^\Omega$ satisfies

$$A, B \subseteq \Omega \text{ and } A \subseteq B \quad \Rightarrow \quad \varphi(A) \supseteq \varphi(B)\,,$$

will there be $A_0 \subseteq \Omega$ such that $\varphi(A_0) = A_0$?

Here is a counterexample. Assume that Ω contains at least two arguments. Let $a_* \in \Omega$ and define

$$\varphi(A) \;:=\; \begin{cases} \emptyset & \text{if } a_* \in A\,, \\ \Omega & \text{if } a_* \in A\,, \end{cases} \qquad (A \subseteq \Omega).$$

The table below shows both that φ is nonincreasing and that there is no fixed point.

Let $B \subseteq A \subseteq \Omega$. We look at all possible cases, depending on whether or not $a_* \in B$, $a_* \in A$. Note that $a_* \in B$ and $a_* \in A$ cannot happen, since $B \subseteq A$.

$a_* \in B$	$a_* \in A$	$\varphi(B) = \emptyset$	$\varphi(A) = \emptyset$
$a_* \notin B$	$a_* \notin A$	$\varphi(B) = \Omega$	$\varphi(A) = \Omega$
$a_* \notin B$	$a_* \in A$	$\varphi(B) = \Omega$	$\varphi(A) = \emptyset$ □

Theorem 10.3 (Cantor–Schröder–Bernstein's Theorem [123]).

Let X and Y be nonempty sets. Let there be injective functions

$$f : X \rightarrowtail Y \qquad \text{and} \qquad g : Y \rightarrowtail X\,.$$

Then there exists a bijection $X \xrightarrow{\sim} Y$.

PROOF: The proof given here is based on the Tarski Fixed Point Theorem. Let

$$\varphi : 2^X \longrightarrow 2^X\,; \qquad A \mapsto g\big(Y \setminus f(X \setminus A)\big)\,.$$

Since $A \mapsto X \setminus A$ is nonincreasing and the direct image function $U \mapsto f(U)$ is nondecreasing, $A \mapsto f(X \setminus A)$ is nonincreasing, thus $A \mapsto Y \setminus f(X \setminus A)$ is nondecreasing.

Since the direct image function $V \mapsto g(V)$ is nondecreasing, $\varphi : A \mapsto g\big(Y \setminus f(X \setminus A)\big)$ is nondecreasing. It follows from Theorem 10.2 (Tarski's Fixed Point Theorem) on p.310 that there exists $A_0 \in X$ such that $\varphi(A_0) = A_0$.

Let $X_* \subseteq X$ and $Y_* \subseteq Y$. Since f and g are injective, it follows from Proposition 5.9 on p.152 that the restrictions

$$f_* : X_* \longrightarrow f(X_*)\,; \quad x \mapsto f(x) \qquad \text{and} \qquad g_* : Y_* \longrightarrow g(Y_*)\,; \quad y \mapsto g(y)$$

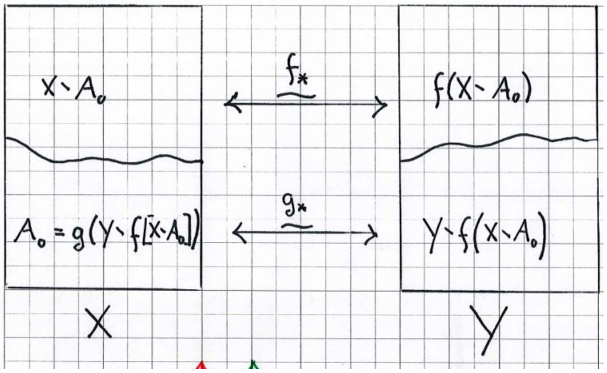
are bijections. We apply this to the sets $X_* := X \setminus A_0$ and $Y_* := Y \setminus f(X \setminus A_0)$:

$$g_* : Y \setminus f(X \setminus A_0) \xrightarrow{\sim} g\big(Y \setminus f(X \setminus A_0)\big) \qquad \text{and} \qquad f_* : X \setminus A_0 \xrightarrow{\sim} f(X \setminus A_0)$$

thus are bijections. Since $A_0 = \varphi(A_0) = g\big(Y \setminus f(X \setminus A_0)\big)$, we have bijections

$$\textbf{(A)} \qquad (g_*)^{-1} : A_0 \xrightarrow{\sim} Y \setminus f(X \setminus A_0) \qquad \text{and} \qquad f_* : X \setminus A_0 \xrightarrow{\sim} f(X \setminus A_0).$$

[123] Named after the German mathematicians Friedrich Wilhelm Karl Ernst Schröder (1841 – 1902), Georg Ferdinand Ludwig Philipp Cantor (1845 – 1918), Felix Bernstein (1878 – 1956)



The picture to the right illustrates those bijections.

Next, we observe that any two sets U and V satisfy $U \cup V = (U \setminus V) \biguplus U$. Hence,

- if $V \subseteq U$ then $U = U \cup V = (U \setminus V) \biguplus U$.

It follows that we have unions of disjoint sets

- $X = (X \setminus A_0) \biguplus A_0$
- $Y = \big(Y \setminus f(X \setminus A_0)\big) \biguplus f(X \setminus A_0)$.

By Proposition 10.3 on p.310, the bijections $(g_*)^{-1}$ and f_* in **(A)** can be combined into a bijection

$$h : A_0 \biguplus X(\setminus A_0) \xrightarrow{\sim} (Y \setminus f(X \setminus A_0)) \biguplus f(X \setminus A_0) \qquad \text{i.e.,} \qquad X \xrightarrow{\sim} Y\,. \ \blacksquare$$

Corollary 10.1.

The relation $\textbf{card}(X) \leq \textbf{card}(Y)$ *partially orders the set* $\mathscr{A} := \{\textbf{card}(X) : X \subseteq \Omega\}$.

PROOF: We must show reflexivity, antisymmetry, and transitivity.

Case 1: None of the sets involved is empty.

Reflexivity is obvious, antisymmetry follows from Cantor-Schröder-Bernstein and transitivity follows from prop.5.5**(a)**: The composition of two injective functions is injective.

Case 2: At least one of the sets involved is empty:

We use the fact that if $\textbf{card}(A) \leq \textbf{card}(B)$ and $B = \emptyset$ then $A = \emptyset$.

As an example we prove antisymmetry: If $X = \emptyset$ and $\textbf{card}(X) \leq \textbf{card}(Y)$ and $\textbf{card}(Y) \leq \textbf{card}(X)$ then the second "$\leq$" implies $Y = \emptyset$, i.e., $X = Y$. On the other hand, If $Y = \emptyset$ and $\textbf{card}(X) \leq \textbf{card}(Y)$ and $\textbf{card}(Y) \leq \textbf{card}(X)$ then the first "$\leq$" implies $X = \emptyset$, i.e., $X = Y$. ■

Theorem 10.4.

Let $X, Y \subseteq \Omega$. *Then*

$$\textbf{card}(X) \leq \textbf{card}(Y) \qquad \textit{or} \qquad \textbf{card}(Y) \leq \textbf{card}(X)$$

In other words, "$\leq$" is a total ordering [124] *on the set of all cardinalities for subsets of* Ω.

PROOF: The proof will be given in thm.15.2, p.454, of ch.15 (Applications of Zorn's Lemma). ■

As an application of the Cantor–Schröder–Bernstein theorem we will prove that one can biject any two intervals of real numbers, no matter whether one or both of them are open, closed, or half–open.

Theorem 10.5.

Let $a, b \in \mathbb{R}$ *such that* $a < b$. *Let* A *be one of* $]a, b[,\ ]a, b\,],\ [\,a, b[,\ [\,a, b\,]$.

$$\textit{Then} \quad \textit{card}(A) = \textit{card}(\mathbb{R})\,.$$

[124] See Definition 5.5 (Linear orderings) on p.131.

PROOF:

(a) $F : \mathbb{R} \to \,]-1,1[\,;\ \ x \mapsto \frac{x}{|x|+1}$ has the function $G(y) \ := \begin{cases} \frac{y}{1-y} & \text{if } x \geq 0, \\ \frac{y}{1+y} & \text{if } x < 0. \end{cases}$ as an inverse. The proof of this is tedious but elementary if one observes that $y \geq 0$ if and only if $x \geq 0$. This makes it easy to solve $y = \frac{x}{|x|+1}$ for x. The details are left to the reader. It follows that card($]-1,1[$) = card($\mathbb{R}$).

(b) Let $c > 0$. The function $x \mapsto cx$ is a bijection from $]-1,1[$ to $]-c,c[$ because it has the function $y \mapsto \frac{y}{c}$ as an inverse. It follows from part **(a)** that card($]-c,c[$) = card($\mathbb{R}$).

(c) If $\lambda \in \mathbb{R}$ then $x \mapsto x + \lambda$ bijects A to $A + \lambda$ (Inverse: $y \mapsto y - \lambda$). Let $c := \frac{b-a}{2}$ and $B := A - \frac{a+b}{2}$. Then B is the interval with endpoints $-c$ and c where either endpoint of B is included/excluded if and only if such is the case for the corresponding endpoint of A. Note that card(B) = card(A) because B is the image of A under the bijection $x \mapsto x - \frac{b-a}{2}$.

(d) It follows from $]-c,c[\ \subseteq\ B$ and part **(b)** that card($\mathbb{R}$) $=$ card($]-c,c[$) $\leq$ card(B), and it then follows from $B \subseteq \mathbb{R}$ that card($\mathbb{R}$) $\leq$ card(B) $\leq$ card($\mathbb{R}$).

It is a consequence of the Cantor–Schröder–Bernstein Theorem which is formulated and proved later on (ch.10.2 [125] , thm.10.3 on p.311) that there exists a bijection between B and $\mathbb{R}$. We saw in part **(c)** that card(B) = card(A). This proves card(A) = card($\mathbb{R}$). ■

We have previously seen that $\mathbb{R}$ is uncountable by proving that the subset of all real numbers $x = \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ such that $d_j = 3$ or $d_j = 4$ can be bijected to the uncountable set $\{3,4\}^{\mathbb{N}}$. See Theorem 9.12 on p.285. The next theorem, which is another application of the Cantor–Schröder–Bernstein Theorem, uses this fact to prove that the set of real numbers and the power set of $\mathbb{N}$ can be bijected.

Theorem 10.6.

$$\text{(10.7)} \qquad \mathit{card}(\mathbb{R}) \ = \ \mathit{card}(2^{\mathbb{N}})\,.$$

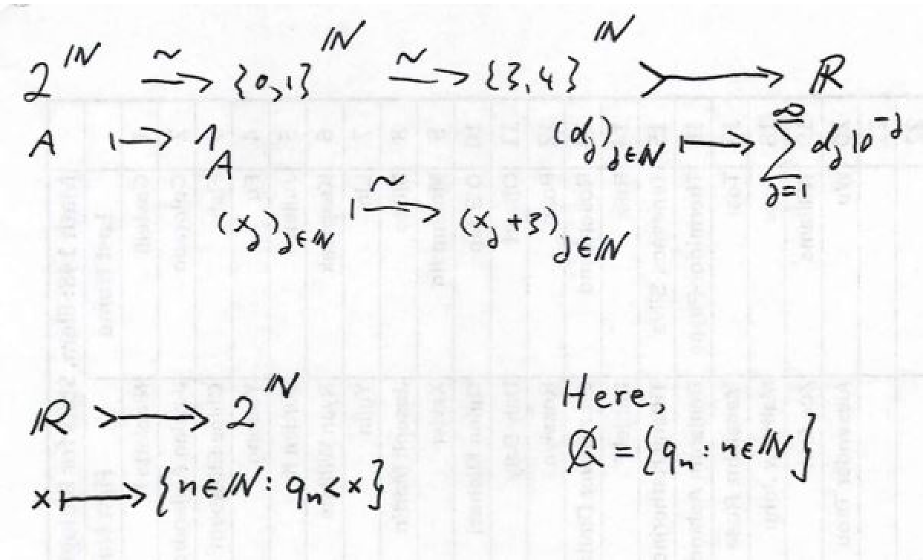
PROOF: First, we show that there exists an injection $f : 2^{\mathbb{N}} \to \mathbb{R}$.

- **(a)** Proposition 8.15 on p.247 shows that there is a bijection $f_1 : 2^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$.
- **(b)** Clearly, the function defined by $0 \mapsto 3$ and $1 \mapsto 4$ is a bijection $f_2 : \{0,1\}^{\mathbb{N}} \to \{3,4\}^{\mathbb{N}}$.
- **(c)** Moreover, the proof of Theorem 9.12 on p.285 shows that $\{3,4\}^{\mathbb{N}}$ can be bijected to the subset $\{x = 0.d_1d_2d_3\cdots : d_j = 3 \text{ or } d_j = 4\}$ of $\mathbb{R}$. Thus, we have an injection $f_3 : \{3,4\}^{\mathbb{N}} \to \mathbb{R}$.
- **(d)** It follows that $f : f_3 \circ f_2 \circ f_1$ is an injection $2^{\mathbb{N}} \to \mathbb{R}$.

Next, we show that there exists an injection $g : \mathbb{R} \to 2^{\mathbb{N}}$.

- **(e)** Since $\mathbb{Q}$ is countable, there is a sequence $q_n \in \mathbb{Q}$ such that $\mathbb{Q} = \{q_n : n \in \mathbb{N}\}$.
- **(f)** $x \mapsto \{n \in \mathbb{N} : q_n < x\}$ defines a function $g : \mathbb{R} \to 2^{\mathbb{N}}$. We claim that g is injective:
  Let $x, x' \in \mathbb{R}$ such that $x \neq x'$. We may assume that $x < x'$. According to Proposition 9.36 on p.283, there exists $n_* \in \mathbb{N}$ such that $x < q_{n_*} < x'$. Since $n_* \notin g(x_1)$ and $n_* \in g(x_2)$, it follows that $g(x_1) \neq g(x_2)$. This proves injectivity of g

[125] "Cardinality as a Partial Ordering". The proof of the Cantor–Schröder–Bernstein Theorem is by no means trivial.



$$2^{\mathbb{N}} \xrightarrow{\sim} \{0,1\}^{\mathbb{N}} \xrightarrow{\sim} \{3,4\}^{\mathbb{N}} \rightarrowtail \mathbb{R}$$

$$A \mapsto 1_A \qquad (x_j)_{j\in\mathbb{N}} \overset{\sim}{\longmapsto} (x_j+3)_{j\in\mathbb{N}} \qquad (\alpha_j)_{j\in\mathbb{N}} \longmapsto \sum_{j=1}^{\infty} \alpha_j 10^{-j}$$

$$\mathbb{R} \rightarrowtail 2^{\mathbb{N}}$$

$$x \longmapsto \{n\in\mathbb{N} : q_n < x\}$$

Here, $\mathbb{Q} = \{q_n : n\in\mathbb{N}\}$

It follows from **(d)** and **(g)** that there are injections $f : 2^{\mathbb{N}} \to \mathbb{R}$ and $g : \mathbb{R} \to 2^{\mathbb{N}}$. We apply Cantor–Schröder–Bernstein and conclude that $\mathbb{R}$ and $2^{\mathbb{N}}$ can be bijected. Thus, $\operatorname{card}(\mathbb{R}) = \operatorname{card}(2^{\mathbb{N}})$. ■

10.3 Alternate Proofs of the Cantor–Schröder–Bernstein Theorem ⋆

The author has seen different proofs of the Cantor–Schröder–Bernstein Theorem. They all are a lot more complicated than the one given in Chapter , They are given here since they are interesting exercises with respect to working with sets.

I - First alternate proof:

The following proof of the Cantor–Schröder–Bernstein Theorem and the material that precedes it closely follow Chapter 11 of [6] Chartrand, G., Polimeni, A. and Zhang, Ping: Mathematical Proofs: A Transition to Advanced Mathematics.

Definition 10.3. Let $\emptyset \neq Y \subseteq X$ and $f : X \to Y$. Then we define for each $n \in [0,\infty[_{\mathbb{Z}}$,

$$f^n : X \longrightarrow Y\,; \qquad f^0(x) := x\,, \qquad f^n(x) := f\big(f^{n-1}(x)\big) \text{ if } n \in \mathbb{N}\,.$$

We call f^n the **n–th iterate of** f. □

Remark 10.1. It follows from $Y \subseteq X$ and the monotonicity of the direct image function that

$$X \supseteq Y \supseteq f(X)\,, \text{ thus } f(X) \supseteq f(f(X)) = f^2(X)\,, \text{ thus } f^2(X) \supseteq f(f^2(X)) = f^3(X)\,, \dots$$

It follows that Y is big enough to function as codomain for each iterate $f^1, f^2, \dots$ since

(10.8) $$Y \supseteq f(X) \supseteq f^m(X) \supseteq f^n(X)\,, \qquad \text{for all integers } 1 \leq m < n\,.$$ □

Example 10.3. Let $f : [0,1/2] \to [0,1/3]; x \mapsto x^2$.

Note that $[0,1/3] \subseteq [0,1/2]$ Then $f^0(x) = f(x) = x^2$ and $f^n(x) = x^{2n}$.

Also note that $0 \leq x \leq 1/2$ implies $0 \leq x^{2n} \leq 1/(2^{2n}) < 1/3$ and thus $[0,1/3]$ is sufficiently large as codomain for each f^n (except $n = 0$). □

Lemma 10.1. *Let $\emptyset \neq B \subseteq A$ such that there exists an injection $f : A \rightarrowtail B$. Then there exists a bijection between A and B.*

PROOF: If $B = A$ then the identity id_A bijects A to B and the proof is finished. We thus assume $B \subsetneq A$, hence $A \setminus B \neq \emptyset$. Since nothing needs to be shown if f is surjective we also assume that

(A) $$f(A) \neq B\,, \qquad \text{i.e.,} \quad B \setminus f(A) \neq \emptyset\,.$$

(B) $$\text{Let} \quad B' := \{f^n(x) : x \in A \setminus B,\, n \in \mathbb{N}\}$$

We obtain from **(A)** that $B \setminus f(A) \neq \emptyset$, hence $B' \neq \emptyset$.
It follows from (10.8) on p.314 that $f^n(x) \in f(A)$ for all $x \in A$ and $n \in \mathbb{N}$, hence

(C) $$B' \subseteq f(A)\,, \quad \text{for all } n \in \mathbb{N}\,.$$

(D) $$\text{Let} \quad C := B' \cup (A \setminus B) \qquad \text{and} \qquad D := B \setminus B'\,.$$

It follows from $f^0(x) = x$, hence $f^0(A \setminus B) = A \setminus B$, that

(E) $$C = \{f^n(x) : x \in A \setminus B,\, n \in [0, \infty[_{\mathbb{Z}}\}\,.$$

We want to define functions which have C and D as their domains and/or codomains, and this requires $C \neq \emptyset$ and $D \neq \emptyset$. Clearly, $C \neq \emptyset$ because $\emptyset \neq A \setminus B \subseteq C$.
To see that $D \neq \emptyset$, note that **(A)** and **(C)** yield $B \setminus f(A) \neq \emptyset$ and $B' \subseteq f(A)$. Thus

$$\emptyset \neq B \setminus f(A) \subseteq B \setminus B' = D\,.$$

Next we show that $f(C) \subseteq B'$, i.e., if $x \in C$ then $f(x) \in B'$. We separately consider $x \in A \setminus B$ and $x \in B'$.
(i) $x \in A \setminus B$: Since $f(x) = f^1(x)$, we obtain from **(B)** that $f(x) \in B'$.
(ii) $x \in B'$: It follows from **(B)** that there exists $n \in \mathbb{N}$ and $\tilde{x} \in A \setminus B$ such that $x = f^n(\tilde{x})$, thus, again by **(B)**, $f(x) = f^{n+1}(\tilde{x}) \in B'$.
Since $f(C) \subseteq B'$ we can downsize the codomain of f to B' if we restrict f to C. In other words, the following is a valid definition of a function.

(F) $$f_C : C \longrightarrow B' \qquad f_C(x) := f(x)\,.$$

Next we prove that f_C is bijective, i.e., f_C is surjective and injective.
Let $y \in B'$. It follows from **(B)** that there exists $n \in \mathbb{N}$ and $x \in A \setminus B$ such that $y = f^n(x)$.
(i) $n = 1$: Then $y = f(x)$. Since $x \in A \setminus B$ and $A \setminus B \subseteq C$ and thus $f_C(x) = f(x)$, we found $x \in C$ such that $f_C(x) = y$.
(ii) $n > 1$, i.e., $n = k + 1$ for some $k \in \mathbb{N}$: Let $\tilde{x} = f^k(x)$. Then $\tilde{x} \in B'$ by **(B)**. Since $B' \subseteq C$ and $f_C(c) = f(c)$ for all $c \in C$, we found $\tilde{x} \in C$ such that

$$f_C(\tilde{x}) = f(\tilde{x}) = f\big(f^{k+1}(x)\big) = f\big(f^n(x)\big) = y\,.$$

(i) and **(ii)** imply that f_C is surjective. Moreover, f_C is injective as a restriction of the injection f to a smaller domain. Thus f_C is bijective. Next we show

(**G**) $$B' \cap D = \emptyset \qquad \text{and} \qquad C \cap D = \emptyset\,.$$

It is trivial that $B' \cap D = \emptyset$ since $D = B \setminus B'$ by **(D)** shows that D and B' have no elements in common. This also helps to see that $C \cap D = \emptyset$:

$$D \cap C = (B \setminus B') \cap \big((A \setminus B) \cup B'\big) = B \cap (A \setminus B') \cap \big((A \setminus B) \cup B'\big) \subseteq B \cap (A \setminus B').$$

Assume $x \in D$. In other words, $x \in B \setminus B'$. Then **(i)**: $x \in B$, thus $x \notin A \setminus B$. Also, **(ii)**: $x \notin B'$. Thus **(i)** and **(ii)** together imply that $x \notin (A \setminus B) \cup B'$, i.e., $x \notin C$

We have shown that $x \in D \Rightarrow x \notin C$ and thus $C \cap D = \emptyset$. We have proved **(G)**. We recall that f_C bijects C to B'. Obviously the identity function $id_D : x \mapsto x$ bijects D to itself. By Proposition 10.3 on p.310, the function

(**H**) $$h : C \biguplus D \longrightarrow B' \uplus D\,; \qquad x \mapsto \begin{cases} f_C(x) & \text{if } x \in C, \\ x & \text{if } x \in D, \end{cases}$$

bijects $C \biguplus D$ to $B' \biguplus D$

Next we show that $C \biguplus D = A$. It follows from **(E)** that

(**I**) $$C \supseteq \{f^0(x) : x \in A \setminus B\}\,.C = \{x : x \in A \setminus B\} = A \setminus B\,.$$

Moreover, since $C \supseteq B'$,

(**J**) $$C \cup D = C \cup (B \setminus B') \supseteq C \cup (B \setminus C) \supseteq C \cup B\,.$$

Since $C \supseteq A \setminus B$ by **(I)** and $C \cup D \supseteq C \cup B$ by **(J)**,

$$C \cup D = C \cup B \supseteq (A \setminus B) \cup B \supseteq A\,.$$

But all sets occurring above are subsets of A. It follows that $C \biguplus D = A$.

Next we show that $B' \biguplus D = B$. It follows from **(C)** that $B' \subseteq f(A)$. Since B is the codomain of f we further have $f(A) \subseteq B$. Thus $B' \subseteq B$, thus $B' \cup (B \setminus B') = B$, thus

$$B' \cup D = B' \cup (B \setminus B') = B\,.$$

It follows that the function h defined in **(H)** is a bijection

$$h : A \xrightarrow{\sim} B\,.$$

We have proved the lemma. ■

With the help of Lemma 10.1 the proof of the Cantor–Schröder–Bernstein Theorem is a simple affair.

ALTERNATE PROOF I of Theorem 10.3:

Since g is injective, the function

$$g_* : Y \longrightarrow g(Y)\,; \qquad y \mapsto g(y)$$

is bijective.

Since the function $\varphi := g_* \circ f : X \longrightarrow g(Y)$ is injective as the composition of two injective functions

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g_* \circ f"'] & Y \arrow[d, "g_*"] \\ & g(Y) \end{tikzcd}$$

and since $g(Y) \subseteq X$ we can apply Lemma 10.1 with X instead of A and $g(Y)$ instead of B. It follows that there exists a bijection

$$h : X \xrightarrow{\sim} g(Y)\,.$$

The function g_* is bijective, hence it has an inverse $g_*^{-1} : g(Y) \xrightarrow{\sim} Y$. Thus $g_*^{-1} \circ h : X \xrightarrow{\sim} Y$ is bijective as the composition of two bijections. ■

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[dr, "g_*^{-1} \circ h"'] & g(Y) \arrow[d, "g_*^{-1}"] \\ & Y \end{tikzcd}$$

II - Second alternate proof:

The next proof of the Cantor–Schröder–Bernstein Theorem and the material that precedes it closely follow the presentation in [11] Haaser/Sullivan: Real Analysis.

Lemma 10.2. *Let A_1, A_2, A_3 be nonempty sets such that $A_1 \supseteq A_2 \supseteq A_3$. and such that there exists a bijection $f : A_1 \xrightarrow{\sim} A_3$. Then there exists a bijection $g : A_1 \xrightarrow{\sim} A_2$.*

PROOF: We define a sequence of sets A_n for $n \geq 4$ as $A_n := f(A_{n-2})$. Note that if follows from the surjectivity of f that $f(A_1) = A_3$ and thus we obtain

$$A_n \;=\; f(A_{n-2}) \quad \text{for} \quad n \geq 3.$$

Next we define sets B_k as follows:

$$B_0 \;:=\; \bigcap_{n=1}^{\infty} A_n; \qquad B_k := A_k \setminus A_{k+1} \quad \text{for} \quad n \in \mathbb{N}.$$

We will prove the following:

(a) The sequence $(A_n)_{n=1}^{\infty}$ is nonincreasing, i.e., $k < n \;\Rightarrow\; A_k \supseteq A_n$.
(b) The sets B_k are mutually disjoint: $i, j \in [0, \infty[\,\Rightarrow\; A_i \cap A_j = \emptyset$.
(c) **c1.** $A_1 \;=\; \bigcup\limits_{k=0}^{\infty} B_k$; **c2.** $A_2 \;=\; \bigcup \big[B_k : k \geq 0, k \neq 1\big]$.
(d) $k \in \mathbb{N} \;\Rightarrow\; f(B_k) = B_{k+2}$.
(e) The function g defined as $g(x) \;:=\; \begin{cases} f(x) & \text{if } x \in \biguplus\big[B_{2k-1} : k \geq 1\big], \\ x & \text{if } x \in \biguplus\big[B_{2k} : k \geq 0\big] \end{cases}$ is a bijection $A_1 \xrightarrow{\sim} A_2$.

The function g defined in **(e)** above is the bijection which this lemma claims to exists.

PROOF of **(a)**:

We reformulate **(a)** as follows: Let $n \in \mathbb{N}$. For all $i < j \leq n$ it is true that $A_i \supseteq A_j$. The proof is done by strong induction on n.

Base cases: $n \leq 3$ The above is true for $n = 1, 2, 3$ since we assumed $A_1 \supseteq A_2 \supseteq A_3$.

Induction assumption: Since the base cases cover $n+1 \leq 3$ we may assume that $n+1 \geq 4$, i.e., $n \geq 3$:

We have some $n \geq 3$ such that if $1 \leq i < j \leq n$ then $A_i \supseteq A_j$. (IA)

We must prove that if $1 \leq i, j \leq n+1$ then $A_i \supseteq A_j$. $(\star)$.

There is nothing to prove if $i = j$. We may also assume that $j = n+1$ since otherwise $1 \leq i < j \leq n$ and the induction assumption allows us to conclude that $A_i \supseteq A_j$. There is no need for the variable j since it has been fixed to $n+1$.

We moreover may assume that $i \geq 3$: If we can prove $(\star)$ for $3 \leq i < n+1$ then it is true that $A_3 \supseteq A_{n+1}$. Since $A_1 \supseteq A_2 \supseteq A_3$. $A_i \supseteq A_{n+1}$ will also be true for $i = 1$ and $i = 2$.

Thus it suffices to prove that if $3 \leq i < n+1$ then $A_i \supseteq A_{n+1}$. $(\star)$.

That is trivial: Since $1 \leq i-2 < n-1$ the induction assumption yields $A_{i-2} \supseteq A_{n-1}$. Since the direct image function is monotone (see 5.16 on p.143) it follows that $f(A_{i-2}) \supseteq f(A_{n-1})$, i.e., $A_i \supseteq A_{n+1}$.

PROOF of **(b)**:

Let $1 \leq i < j$. Then

$$B_i \cap B_j \ = \ (A_i \setminus A_{i+1}) \cap (A_j \setminus A_{j+1}) \ \subseteq \ (A_i \setminus A_{i+1}) \cap A_j \ \subseteq \ (A_i \setminus A_{i+1}) \cap A_{i+1} \ = \ \emptyset.$$

Here the last "$\subseteq$" follows from the fact that A_n is nonincreasing and $j \geq i+1$, and the last equation follows from the definition of "$\setminus$".

PROOF of **c1**: For the proof of "$\supseteq$" we note that $A_1 \supseteq A_k \supseteq A_k \setminus A_{k+1} = B_k$ for all $k \in \mathbb{N}$, and that trivially $A_1 \supseteq B_0 = \bigcap\limits_{k=1}^{\infty} A_k$.

For the reverse inclusion let $x \in A_1$ and $J := \{j \in \mathbb{N} : x \in A_j\}$. There are two cases: **Case 1**: J is unbounded. Then $J = \mathbb{N}$ since the sets A_j are nonincreasing. Thus $x \in A_j$ for all $j \geq 0$, thus $x \in \bigcap\limits_{n=1}^{\infty} A_n$, i.e., $x \in B_0$, thus $x \in \bigcup\limits_{k=0}^{\infty} B_k$. **Case 2**: J is bounded. Note that $J \neq \emptyset$ since $1 \in J$, thus $j^* := \max(J)$ exists according to the extended well-ordering principle. It follows from $x \in A_{j^*}$ and $x \notin A_{j^*+1}$ that $x \in B_{j^*} = A_{j^*} \setminus A_{j^*+1}$, thus $x \in \bigcup\limits_{k=0}^{\infty} B_k$.

PROOF of **c2**: For the proof of "$\supseteq$" we note that $A_2 \supseteq A_k \supseteq A_k \setminus A_{k+1} = B_k$ for all $k \geq 2$, and that trivially $A_2 \supseteq B_0 = \bigcap\limits_{k=1}^{\infty} A_k$.

For the reverse inclusion let $x \in A_2$ and $J := \{j \in \mathbb{Z} : j \geq 2 \text{ and } x \in A_j\}$. **Case 1**: J is unbounded. Then $J = [2, \infty[_{\mathbb{Z}}$ since the sets A_j are nonincreasing. Thus $x \in A_j$ for all $j \geq 2$, thus $x \in \bigcap\limits_{n=2}^{\infty} A_n$ which equals $\bigcap\limits_{n=1}^{\infty} A_n$ since $A_1 \supseteq \bigcap\limits_{n=2}^{\infty} A_n$, i.e., $x \in B_0$.It follows hat $x \in \bigcup \big[B_k : k \geq 0, k \neq 1\big]$. **Case 2**: J is bounded. Then $j^* := \max(J)$ exists according to the extended well-ordering principle. It follows from $x \in A_{j^*}$ and $x \notin A_{j^*+1}$, thus $x \in B_{j^*} = A_{j^*} \setminus A_{j^*+1}$, thus $x \in \bigcup \big[B_k : k \geq 0, k \neq 1\big]$.

PROOF of **(d)**: Since the bijective f is compatible with all set operation we have $f(U \setminus V) = f(U) \setminus f(V)$ for any two sets U, V (see (8.46) on p.244). It follows for any $k \in \mathbb{N}$ that

$$f(B_k) \ = \ f(A_k \setminus A_{k+1}) \ = \ f(A_k) \setminus f(A_{k+1}) \ = \ A_{k+2} \setminus A_{k+3} \ = \ B_{k+2}.$$

PROOF of **(e)**: Let $O := \biguplus \big[B_{2k-1} : k \geq 1\big]$ and $E := \biguplus \big[B_{2k} : k \geq 0\big]$. We note that it is appropriate to write $\uplus$ instead of $\cup$ since we proved in **(b)** that the sets B_k are mutually disjoint, and that this then implies that the assignment $x \mapsto g(x)$ is unambiguous: either $x \in O$ and $g(x) = f(x)$ or $x \in E$ and $g(x) = x$. Further we obtain from **c1** that the domain $O \uplus E = \bigcup\limits_{k=0}^{\infty} B_k$ of f equals A_1. Since f satisfies $f\Big(\bigcup\limits_i U_i\Big) = \bigcup\limits_i f(U_i)$ for any family $(U_i)_i$ such that each U_i belongs to the domain A_1 of f (see prop.8.5 (Properties of the direct image) on p.241)) it follows that

$$g(A_1) \;=\; g(O \biguplus E) \;=\; g(O) \cup g(E) = f(O) \cup E \;=\; f\Big(\biguplus_{k\geq 1} B_{2k-1}\Big) \;\cup\; \biguplus_{k\geq 0} B_{2k} : k \geq 0.$$

We proved in **(d)** that $f(B_k) = f(B_{k+2})$ for all $k \in \mathbb{N}$. Thus

$$g(A_1) \;=\; \biguplus_{k\geq 1} f(B_{2k-1}) \;\cup\; \biguplus_{k\geq 0} B_{2k} \;=\; \biguplus_{k\geq 1} B_{2k+1} \;\uplus\; \biguplus_{k\geq 0} B_{2k} \;=\; B_0 \uplus \biguplus \big[B_j : k \geq 2\big] \;=\; A_2.$$

We have proven that the function $g : A_1 \to A_2$ is surjective. All that remains to prove the lemma, i.e., that there exists a bijection $A_1 \xrightarrow{\sim} A_2$ is to show that g is injective.

So let $x, x' \in A_1$ such that $x \neq x'$. It follows from **(b)** and **c1** that there exist unique indices $j, k \geq 0$ such that $x \in B_j$ and $x' \in B_k$. Since both mappings $x \mapsto f(x)$ and $x \mapsto x$ are injective $g(x) \neq g(x')$ in the case that $j = k$.

We may thus assume that $j \neq k$. If both j and k are odd then it follows from the injectivity of f that $g(x) = f(x) \neq f(x') = g(x')$ and we are done. If both j and k are even then $g(x) = x \neq x' = g(x')$. Again we are done.

Finally assume that j is odd and k is even. Then $j + 2$ is odd and thus must be different from the even index k, thus B_{j+2} and B_k are disjoint. Since $g(x) \in g(B_j) = f(B_j) = B_{j+2}$ and $g(x') = x' \in B_k$ and those two sets have empty intersection it follows that $g(x) \neq g(x')$. We have proven injectivity and thus bijectivity of $g : A_1 \xrightarrow{\sim} A_2$. ■

ALTERNATE PROOF II of Theorem 10.3:

Let $\;B_1 \;:=\; f'(X), \;\; A_2 \;:=\; g'(Y), \;\; A_3 \;:=\; g'(B_1).$

Then $\;B_1 \subseteq Y\;$ and $\;A_2 \subseteq X,\;$ thus $\;A_3 \;=\; g'(B_1) \;\subseteq\; g'(Y) \;=\; A_2 \;\subseteq\; X.$

Further, $\;g'\big(f'(X)\big) \;=\; g'(B_1) \;=\; A_3.$

Since the function $g' \circ f'$ is injective as the composition of two injective functions this proves that $g' \circ f' : X \xrightarrow{\sim} A_3$ is bijective.

It follows from lemma 10.2 above that there exists a bijection

$$f : X \xrightarrow{\sim} A_2 \;=\; g'(Y)\,.$$

Since $g' : Y \to A$ is injective we obtain from this function a bijection

$$g : Y \xrightarrow{\sim} A_2$$

by simply downsizing the codomain to $g'(Y)$ But then the function

$$g^{-1} \circ f \quad \text{is a bijection} \quad X \xrightarrow{\sim} Y$$

as the composition of two bijective functions. ■

III - Third alternate proof:

The follwing last proof of the Cantor–Schröder–Bernstein's Theorem is a more detailed version of the one found in the chapter Further Topics F: Cardinal Number and Ordinal Number of [2] B/G (Beck/Geoghegan).

ALTERNATE PROOF III of Theorem 10.3:

We have no interest in any particulars of the sets X and Y. We only are interested in establishing the existence of a bijection $h : X \to Y$. We hence may assume that X and Y are mutually disjoint, replacing X with $\{1\} \times X$ and Y with $\{2\} \times Y$ if necessary (see remark 5.16 on p.150).

Let $f : X \xrightarrow{\sim} f'(X)$ and $g : Y \xrightarrow{\sim} g'(Y)$ be the bijective functions obtained from the injections f' and g' by restricting their codomains to the images of their domains. We note that the subsets $f'(X) \subseteq Y$ and $g'(Y) \subseteq X$ also are disjoint and that $f(X) = f'(X),\ g(Y) = g'(Y)$. Let

$$\sigma : f(X) \biguplus g(Y) \to X \biguplus Y; \qquad z \mapsto \begin{cases} f^{-1}(z) & \text{if } z \in f(X), \\ g^{-1}(z) & \text{if } z \in g(Y), \end{cases} \tag{10.9}$$

i.e., $\sigma\big|_{f(X)} = f^{-1}$ and $\sigma\big|_{g(Y)} = g^{-1}$. Note that if $y \in f(X)$ then $\sigma(y) \in X$; if $x \in g(Y)$ then $\sigma(x) \in Y$. We can create iterates

$$\sigma^2(z) = \sigma\big(\sigma(z)\big),\ \ \sigma^3(z) = \sigma\big(\sigma^2(z)\big),\ \ \dots, \sigma^{n+1}(z) = \sigma\big(\sigma^n(z)\big),\ \ \dots,$$

just as long as $\sigma^n(z) \in f(X) \biguplus g(Y)$. We further define σ^0 for all $z \in f(X) \biguplus g(Y)$ as $\sigma^0(z) := z$. We associate with each $z \in X \biguplus Y$ a "score" $N(z) \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ as follows.

- **(a)** If $\sigma^k(z) \in f(X) \biguplus g(Y)$ for all $k \in \mathbb{N}$ then $N(z) := \infty$.
- **(b)** If $\sigma^k(z) \notin f(X) \biguplus g(Y)$ for some $k \in \mathbb{N}$ then $N(z) := \min\{\, j \geq 0 : \sigma^j(z) \notin f(X) \biguplus g(Y)\, \}$.

Note that **(b)** implies the following: If $z = \sigma^0(z) \notin f(X) \biguplus g(Y)$ then $N(z) = 0$.

Depending on whether we start out with $x \in g(Y)$ or $y \in f(X)$, we obtain the following finite or infinite sequences:

$$\begin{aligned} &\text{if } x \in g(Y): \quad x \overset{\sigma}{\to} f^{-1}(x) \overset{\sigma}{\to} g^{-1}(f^{-1}(x)) \overset{\sigma}{\to} f^{-1}(g^{-1}(f^{-1}(x))) \overset{\sigma}{\to} \dots, \\ &\text{if } y \in f(X): \quad y \overset{\sigma}{\to} g^{-1}(y) \overset{\sigma}{\to} f^{-1}(g^{-1}(y)) \overset{\sigma}{\to} g^{-1}(f^{-1}(g^{-1}(y))) \overset{\sigma}{\to} \dots. \end{aligned}$$

If $N(z) < \infty$ then the sequence will terminate after $N(z)$ iterations. Let

$$\begin{aligned} X_E &:= \{x \in X : N(x) \text{ is even}\}, X_O := \{x \in X : N(x) \text{ is odd}\}, X_\infty := \{x \in X : N(x) = \infty\}, \\ Y_E &:= \{y \in Y : N(y) \text{ is even}\}, Y_O := \{y \in Y : N(y) \text{ is odd}\}, Y_\infty := \{y \in Y : N(y) = \infty\}. \end{aligned}$$

The above defines partitions $X = X_E \biguplus X_O \biguplus X_\infty$ and $Y = Y_E \biguplus Y_O \biguplus Y_\infty$ of X and Y.

Each of the functions f, g, σ changes the score of its argument from odd to even and from even to odd. Hence

$$\begin{aligned} &f(X_E) \subseteq Y_O,\ f^{-1}(Y_O) \subseteq X_E, \qquad f(X_O) \subseteq Y_E,\ f^{-1}(Y_E) \subseteq X_O, \\ &g(Y_E) \subseteq X_O,\ g^{-1}(X_O) \subseteq Y_E, \qquad g(Y_O) \subseteq X_E,\ g^{-1}(X_E) \subseteq Y_O, \\ &f(X_\infty) \subseteq Y_\infty,\ f^{-1}(Y_\infty) \subseteq X_\infty, \qquad g(Y_\infty) \subseteq X_\infty,\ g^{-1}(X_\infty) \subseteq Y_\infty, \end{aligned} \tag{10.10}$$

We define a bijection $h : X \overset{\sim}{\to} Y$ as follows:

$$h : X \ \to \ Y; \qquad x \ \mapsto \begin{cases} \sigma(x) = g^{-1}(x) & \text{if } x \in X_O \biguplus X_\infty, \\ f(x) & \text{if } x \in X_E. \end{cases}$$

Note that $g^{-1}(x)$ is defined for all $x \in X_O \biguplus X_\infty$ because we then have $N(x) > 0$.

We show that h is injective: Let $x_1, x_2 \in X$ such that $x_1 \neq x_2$. There are four cases.

Case 1: Both $x_1, x_2 \in X_O \biguplus X_\infty$. Then $h(x_1) = g^{-1}(x_1) \neq g^{-1}(x_2) = h(x_2)$ because the bijectivity of g implies that of g^{-1}. In particular, g^{-1} is injective.

Case 2: Both $x_1, x_2 \in X_E$. Then $h(x_1) = f(x_1) \neq f(x_2) = h(x_2)$ because f is injective.

Case 3: $x_1 \in X_O \biguplus X_\infty$ and $x_2 \in X_E$. It follows from (10.10) that $h(x_1) = g^{-1}(x_1) \in Y_E \uplus Y_\infty$ and that $h(x_2) = f(x_2) \in Y_O$. Because $Y_E \uplus Y_\infty$ and Y_O have no elements in common, it follows that $h(x_1) \neq h(x_2)$. We have proved that h is injective.

Case 4: $x_2 \in X_O \biguplus X_\infty$ and $x_1 \in X_E$. Injectivity of h follows from case 3 because we can switch the roles of x_1 and x_2.

We finally show that h is surjective: Let $y \in Y$. There are two cases.

Case (i): $y \in Y_E \biguplus Y_\infty$. It follows from (10.10) that $g(y) \in X_O \biguplus X_\infty$, hence

$$h(g(y)) \ = \ \sigma(g(y)) \ = \ g^{-1}(g(y)) \ = \ y.$$

Here the second equation follows from (10.9). We have found an item in X which is mapped by h to y. and this proves that h is surjective.

Case (ii): $y \in Y_O$. It follows from (10.10) that $f^{-1}(y) \in X_E$, hence $h(f^{-1}(y)) = f(f^{-1}(y)) = y$. Again we have found an item in X which is mapped by h to y. We have proved that h is surjective also in this case. ■

10.4 Exercises for Ch.10

Exercise 10.1. Let Ω be a set and let $\varphi : 2^\Omega \to 2^\Omega$ satisfy $\ A, B \subseteq \Omega$ and $A \subseteq B \ \Rightarrow \ \varphi(A) \subseteq \varphi(B)$.

$$\text{Let} \quad \mathfrak{F} \ := \ \{A \in 2^\Omega : A \subseteq \varphi(A)\}, \quad A_0 := \bigcup [A : A \in \mathfrak{F}].$$

The proof of Tarski's fixed point theorem (Theorem 10.2 on p.310) shows that A_0 is a fixed point for φ, i.e., $\varphi(A_0) = A_0$. Modify this proof to show the following:

$$\text{Let} \quad \mathscr{E} \ := \ \{B \in 2^\Omega : \varphi(B) \subseteq B\}, \quad B_0 := \bigcap [B : B \in \mathscr{E}].$$

Then B_0 also is a fixed point for φ.

Exercise 10.2. Prove prop.10.1 on p.308:

Let X, Y be two sets such that card(X) = card(Y). Then card(2^X) = card(2^Y). □

Exercise 10.3. Prove prop.10.3 on p.310 of this document: Let X, X', Y, Y' be nonempty sets such that $X \cap X' = \emptyset$ and $Y \cap Y' = \emptyset$ Let $f : X \rightarrowtail Y$ and $f' : X' \rightarrowtail Y'$ injective functions. Then

$$h : X \biguplus X' \longrightarrow Y \uplus Y' \, ; \qquad x \ \mapsto \begin{cases} f(x) & \text{if } x \in X, \\ f'(x) & \text{if } x \in X', \end{cases}$$

is an injection.

Note that part of the proof is showing that the relation $\{(x,y) \in (X \biguplus X') \times (Y \uplus Y') : y = h(x)\}$ indeed defines the graph of a function. □

11 Vectors and Vector spaces

11.1 $\mathbb{R}^n$: Euclidean Space

Most if not all of the material of this chapter with the exception of ch.11.2.2 (normed Vector Spaces) on p.341 is familiar to anyone who took a linear algebra course or, in case of two or three dimensional space, to those who took a course in multivariable calculus.

11.1.1 n–Dimensional Vectors

The following definition of an n–dimensional vector is a special case of a vector, something which will be defined as an element of an abstract **"vector space"** . [126]

Definition 11.1 (n–dimensional vectors). ⋆

Let $n \in \mathbb{N}$. An **n–dimensional vector** is a finite, ordered collection $\vec{v} = (x_1, x_2, \dots, x_n)$ of real numbers $x_1, x_2, \dots, x_n$, n is called the **dimension** of the vector $\vec{v}$. □

Thus, $\vec{v}$ is an element of $\mathbb{R}^n$. [127]

Here are some examples of vectors:

Example 11.1 (Two–dimensional vectors). The two–dimensional vector $\vec{v}$ 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 11.2 (Three–dimensional vectors). $\vec{v_t} = \big(3 - t, 15, \sqrt{5t^2 + \frac{22}{7}}\big) \in \mathbb{R}^3$ 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., $\vec{v}_{-2} = \big(5, 15, \sqrt{20 + \frac{22}{7}}\big)$. note that

$$F : \mathbb{R} \to \mathbb{R}^3 \qquad t \mapsto F(t) = \vec{v_t}$$

defines a function from $\mathbb{R}$ into $\mathbb{R}^3$ in the sense of definition (5.7) on p.134. Each argument s has assigned to it one and only one argument $\vec{v_s} = \big(3 - s, 15, \sqrt{5s^2 + \frac{22}{7}}\big) \in \mathbb{R}^3$.

Or, is it rather that we have three functions

$$\begin{aligned} &x(\cdot) : \mathbb{R} \to \mathbb{R} && t \to x(t) = 3 - t, \\ &y(\cdot) : \mathbb{R} \to \mathbb{R} && t \to y(t) = 15, \\ &z(\cdot) : \mathbb{R} \to \mathbb{R} && t \to z(t) = \sqrt{5t^2 + \tfrac{22}{7}}, \end{aligned}$$

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 $\vec{v_t}$. □

[126] The definition of an abstract vector space will be given in Definition 11.5 on p.329.

[127] See Definition 8.5 (Cartesian Product of three or more sets) on p.237) concerning $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n\text{times}}.$.

Example 11.3 (One–dimensional vectors)**.** 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 $a \in \mathbb{R}$.

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 $\vec{w_2} = 5.7a$. $\square$

Example 11.4 (Vectors as functions)**.** An n–dimensional vector $\vec{x} = (x_1, x_2, x_3, \cdots, x_n)$ can be interpreted as a real–valued function

$$\begin{aligned} &f_{\vec{x}}(\cdot) : \{1, 2, 3, \cdots, n\} \longrightarrow \mathbb{R} \qquad m \mapsto x_m, \ \text{ i.e.,} \\ &f_{\vec{x}}(1) = x_1,\ f_{\vec{x}}(2) = x_2,\ \cdots,\ f_{\vec{x}}(n) = x_n, \end{aligned} \tag{11.1}$$

This can be done in reverse. Any real–valued function $f(\cdot) : \{1, 2, 3, \cdots, n\} \longrightarrow \mathbb{R}$ can be associated with the vector $\vec{v}_{f(\cdot)}$ that lists the function values $f(j)$:

$$\vec{v}_{f(\cdot)} := (f(1), f(2), f(3), \cdots, f(n)) \in \mathbb{R}^n. \ \square \tag{11.2}$$

Definition 11.2 (Transposed matrix)**.** $\boxed{\star}$

Let A be a matrix with m rows and n columns. We will write $A = \big((a_{ij})\big)$ to express that a_{ij} denotes the "cell" at the intersection of row i and column j. ($i \in [1, m]_{\mathbb{Z}}$ and $j \in [1, n]_{\mathbb{Z}}$).

$$A \;=\; \begin{bmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}, & a_{m2}, & \dots, & a_{mn}(t) \end{bmatrix}.$$

If A is a matrix with m rows and n columns, and if a_{ij} denotes the "cell" at the intersection of row i and column j, then we denote by $A^{\top}$ the "flipped" matrix which has row i of A as its i–th column, and column j of A as its j–th row.

In other words, if $A = \big((a_{ij})\big)$ and if $A^{\top} = \big((a^*_{k\ell})\big)$ then $a^*_{ij} = a_{ji}$ for all $i \in [1, m]_{\mathbb{Z}}$ and $j \in [1, n]_{\mathbb{Z}}$. We call $A^{\top}$ the **transpose** or **transposed matrix** of A.

$$A^{\top} \;=\; \begin{bmatrix} a_{11}, & a_{21}, & \dots, & a_{m1} \\ a_{12}, & a_{22}, & \dots, & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n}, & a_{2n}, & \dots, & a_{mn}(t) \end{bmatrix}.$$

We usually do not work with matrix multiplication and do not care whether we think of a vector as a column vector or a row vector. For theoretical reasons, most books on linear algebra define vectors $\vec{x} \in \mathbb{R}^n$ as column vectors, as seen on the right.

$$\vec{x} \;=\; \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Since this is very annoying to typeset column vectors, we will often write instead

$$\vec{x} \;=\; (x_1, \dots, x_n)^{\top}$$

and indicate by means of the transpose symbol $(\cdot)^{\top}$ that $\vec{x}$ is a column vector. $\square$

11.1.2 Addition and Scalar Multiplication for n–Dimensional Vectors

Definition 11.3 (Addition and scalar multiplication in $\mathbb{R}^n$)**.** $\boxed{\star}$ 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.
We define the **sum** $\vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ as the vector $\vec{z}$ with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_n = x_n + y_n; \tag{11.3}$$

We define the **scalar product** $\alpha\vec{x}$ of α and $\vec{x}$ as the vector $\vec{w}$ with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_n = \alpha x_n. \ \square \tag{11.4}$$

Figure 11.1 below describes vector addition.

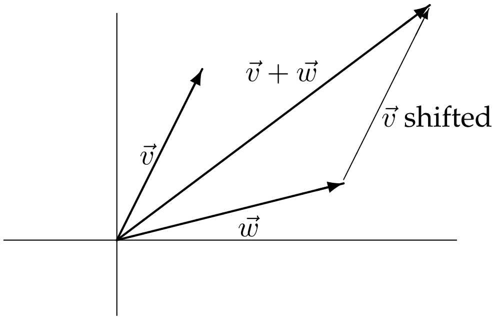


Figure 11.1: Adding two vectors.

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.

11.1.3 Length of n–Dimensional Vectors and the Euclidean Norm

It is customary to write $\|\vec{v}\|_2$ for the length, often also called the **Euclidean norm**, of the vector $\vec{v}$.

Example 11.5 (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 $\|\sqrt{2}\|_2 = |\sqrt{2}| \approx 1.414$. $\square$

Example 11.6 (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 11.2). How long is that arrow?

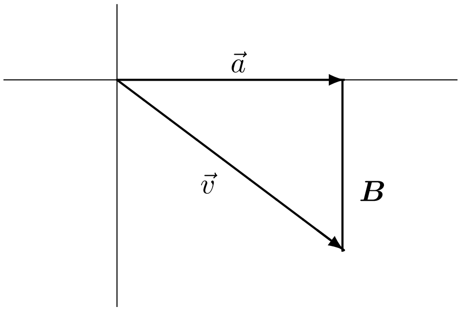


Figure 11.2: Length of a 2–dimensional vectors.

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,

$$\|(x,y)\|_2 = \sqrt{x^2+y^2} \;\; \square \tag{11.5}$$

Example 11.7 (Length of three–dimensional vectors)**.** This is not so different from the two–dimensional case. 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$(\boldsymbol{B}) = |z| = 12$.

We connect the origin $(0,0,0)$ with the point $(4,-3,0)$ in the xy–plane (the endpoint of $\boldsymbol{B}$).

We can apply what we know 2–dimensional vectors because this arrow is contained in the xy–plane. Matter of fact, we have a genuine two–dimensional vector $\vec{a} = (4,-3)$ because the line starts at the origin. Observe that $\vec{a}$ has the same values 4 and -3 for its x– and y–coordinates as the original vector $\vec{v}$. [128] 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

$$\|\vec{v}\|_2^2 = z^2 + \|(x,y)\|_2^2 \;=\; 144 + 25 = 169, \quad \text{hence} \quad \|\vec{v}\|_2 = 13.$$

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. We conclude that it is true in general that $\|(x,y,z)\|^2 = x^2+y^2+z^2$, hence

$$\|(x,y,z)\| = \sqrt{x^2+y^2+z^2} \;\square \tag{11.6}$$

The previous examples show how to extend the concept of "length" to vector spaces of any finite dimension:

Definition 11.4 (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:

$$\|\vec{v}\|_2 \;=\; \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2} \;=\; \sqrt{\sum_{j=1}^n {x_j}^2}. \;\square \tag{11.7}$$

The above definition is important enough to write the special cases for $n = 1,2,3$ where $\|\vec{v}\|_2$ coincides with the length of $\vec{v}$:

$$\begin{aligned} \text{1–dimensional}: &\quad \|(x)\|_2 \;=\; \sqrt{x^2} \;=\; |x| \\ \text{2–dimensional}: &\quad \|(x,y)\|_2 \;=\; \sqrt{x^2+y^2} \\ \text{3–dimensional}: &\quad \|(x,y,z)\|_2 \;=\; \sqrt{x^2+y^2+z^2} \end{aligned} \tag{11.8}$$

Proposition 11.1 (Properties of the Euclidean norm)**.**

Let $n \in \mathbb{N}$. Then the Euclidean norm has the following properties, when viewed as a function

$$\|\cdot\|_2 : \mathbb{R}^n \to \mathbb{R}; \qquad \vec{v} = (x_1, x_2, \dots, x_n) \mapsto \|\vec{v}\|_2 = \sqrt{\sum_{j=1}^n {x_j}^2} \;:$$

[128] You will learn in the chapter on vector spaces that the vector $\vec{a} = (4,-3)$ is the projection on the xy–coordinates $\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2; \quad (x,y,z) \mapsto (x,y)$ of the vector $\vec{v} = (4,-3,12)$. See Example 11.19) on p.337.

(11.9a) $\|\vec{v}\|_2 \geq 0 \quad \forall \vec{v} \in \mathbb{R}^n \quad$ *and* $\quad \|\vec{v}\|_2 = 0 \;\Leftrightarrow\; \vec{v} = 0$ ***(positive definiteness)***
(11.9b) $\|\alpha\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}$ ***(absolute homogeneity)***
(11.9c) $\|\vec{v}+\vec{w}\|_2 \leq \|\vec{v}\|_2 + \|\vec{w}\|_2 \quad \forall\, \vec{v}, \vec{w} \in \mathbb{R}^n$ ***(triangle inequality)***

PROOF:
(a) It is certainly true that $\|\vec{v}\|_2 \geq 0$ for any n–dimensional vector $\vec{v}$ because it is defined as $+\sqrt{K}$ where the quantity K is, as a sum of squares, nonnegative. 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 $\vec{v} = (x_1, x_2, \ldots, x_n)$ be a vector in $\mathbb{R}^n$ such that $\|\vec{v}\|_2 = 0$. This means that $\sqrt{\sum\limits_{j=1}^{n} {x_j}^2} = 0$ which is only possible if everyone of the nonnegative x_j is zero. In other words, $\vec{v}$ must be the zero vector 0.
(b) Let $\vec{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} \|\alpha\vec{v}\|_2 &= \sqrt{\sum_{j=1}^{n} (\alpha x_j)^2} = \sqrt{\sum_{j=1}^{n} \alpha^2 {x_j}^2} = \sqrt{\alpha^2 \sum_{j=1}^{n} {x_j}^2} = \sqrt{\alpha^2}\sqrt{\sum_{j=1}^{n} {x_j}^2} \\ &= \sqrt{\alpha^2}\|\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \end{aligned}$$

because it is true that $\sqrt{\alpha^2} = |\alpha|$ for any real number α (see assumption 2.1 on p.28).
(c) The proof will only be given for $n = 1, 2, 3$.
***n* = 1** : (11.9.c) simply is the triangle inequality for real numbers (see (2.5) on 28) and we are done.
***n* = 2, 3** : Look back at the picture about addition of vectors in the plane or in space (see p.325). 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 $\|\vec{v}+\vec{w}\|_2 \leq \|\vec{v}\|_2 + \|\vec{w}\|_2$ 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 (11.9.c). An exact proof will be given for arbitrary $n \in \mathbb{N}$ as a consequence of the so–called Cauchy–Schwartz inequality (cor.11.1). The inequality itself is stated and proved in prop.11.12 on p.343 in the section which discusses inner products (dot products) on vector spaces. ■

11.2 General Vector Spaces

11.2.1 Vector spaces: Definition and Examples

Part of this follows [4] Brin, Matthew and Marchesi, Gerald: Linear Algebra, a text for Math 304, Spring 2016.

Mathematicians are very fond of looking at 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–valued functions

Well, that was disingenuous. We saw that real numbers and one–dimensional vectors are sort of the same (see 11.3 on p.324). We also saw that n–dimensional vectors can be thought of as real–valued functions with domain $X = \{1, 2, 3, \cdots, n\}$. (see 11.4 on p.324). 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 (11.1.2) on p.325). 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 11.8 (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:

$$100a = 100(-3, 1/5) = (-300, 20). \ \square$$

In the last example we avoided using the notation "$\vec{x}$" with the cute little arrows on top for vectors. The reason is that this notation is not all that popular in math, even for n–dimensional vectors, and definitely not for 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 [12]). This definition is quite lengthy because a set needs to satisfy many rules to be a vector space.

Definition 11.5 (Vector spaces (linear spaces))**.** $\boxed{\star}$ A nonempty set V is called a **vector space** or **linear space** and we call its elements **vectors** if V satisfies the following:

(A) There exists a binary operation $+ : V \times V \to V;\ (x, y) \mapsto x + y$ on V such that $(V, +)$ is an abelian group (see def. 3.2 on p.52). We call $x + y$ the **sum** of x and y. Note that $(V, +)$ being an abelian group means that the following properties hold for "+":

1. $x + y = y + x$ for all $x, y \in V$ (**commutativity**);

2. $(x+y)+z \;=\; x+(y+z)$ for all $x,y,z \in V$ (**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) There exists a function $\cdot : \mathbb{R} \times V \to V;\ (\alpha, x) \mapsto \alpha \cdot x$, i.e., any real number α and vector x uniquely determine a vector $\alpha \cdot x$. It is customary to simply write αx for $\alpha \cdot x$. This vector is called the **scalar product** of α and x, and 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. $\alpha(x+y) \;=\; \alpha x + \alpha y$; □

Remark 11.1. $\boxed{\star}$ We state for the reader's convenience the above definition of a vector space V one more time in a more easily remembered form.

(a) V is nonempty and comes with two assignments:
$+ : V \times V \to V;\ (x,y) \mapsto x+y$, the sum of x and y,
$\cdot : \mathbb{R} \times V \to V;\ (\alpha, x) \mapsto \alpha \cdot x$, (also written αx), the scalar product of α and x.

(c) $(V,+)$ is an abelian group. We write 0 (null vector) for its neutral element, $-x$ for the inverse of a vector x, and $x-y$ for $x+(-y)$.

(d) $\alpha(\beta x) \;=\; (\alpha\beta)x$ for all $\alpha, \beta \in \mathbb{R}$ and $x \in V$.

(e) $1 \cdot x \;=\; x$ for all $x \in V$. (1 is the real number 1).

(f) Two distributive laws:
$(\alpha+\beta)x \;=\; \alpha x + \beta x$,
$\alpha(x+y) \;=\; \alpha x + \alpha y$. □

Definition 11.6 (Subspaces of vector spaces). $\boxed{\star}$

Let V be a vector space and let $A \subseteq V$ be a nonempty subset of V such that

- 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.
The set $\{0\}$ which only contains the null vector 0 of V is called the **nullspace**. □

Remark 11.2 (Closure properties of linear subspaces)**.**

(a) Note that if $\alpha = 0$ then $\alpha x = 0$. it follows that the null vector belongs to any subspace.
(b) We ruled out the case $A = \emptyset$ but did not require that A be a strict subset of V ((2.3) on p.16). In other words, the entire vector space V is a subspace of itself.
(c) It is trivial to verify that the nullspace $\{0\}$ is a subspace. □

Proposition 11.2 (Subspaces are vector spaces)**.**

A subspace of a vector space is a vector space, i.e., it satisfies all requirements of definition (11.5).

PROOF: None of the equations 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. Such can never be the case. Here is 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 $\lambda x + \lambda y$ holds because it holds for x and y as elements of V. ■

Remark 11.3 (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 11.9 (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 11.10 (Vector space $\mathbb{R}^n$)**.** The sets $\mathbb{R}^n$ of n–dimensional vectors become vector spaces if addition and scalar multiplication are defined as in (11.3) on p.325. □

The following example should be thought of as the **definition** of the very important function spaces $\mathscr{F}(X,\mathbb{R})$, $\mathscr{B}(X,\mathbb{R})$, $\mathscr{C}(X,\mathbb{R})$.

Example 11.11 (Vector spaces of real–valued functions)**.** Let X be an arbitrary, nonempty set. Then

$$\mathscr{F}(X,\mathbb{R}) \;:= \{f(\cdot) : f(\cdot) \text{ is a real–valued function on } X\} \tag{11.10}$$

denotes the set of all real–valued functions with domain X [129] and

$$\mathscr{B}(X,\mathbb{R}) \;:=\; \{g(\cdot) : g(\cdot) \text{ is a bounded real–valued function on } X\}$$

denotes the subset of all bounded real–valued functions with domain X.
Let $A \subseteq \mathbb{R}$. Then

$$\mathscr{C}(A,\mathbb{R}) \;:= \{\psi(\cdot) : \psi(\cdot) \text{ is a continuous real–valued function on } A\}$$

denotes the set of all real–valued continuous functions with domain A. [130]
We list separately the case $X = [a,b]$ where $a,b \in \mathbb{R}$ such that $a < b$. Then

$$\mathscr{C}([a,b],\mathbb{R}) \;:=\; \{h(\cdot) : h(\cdot) \text{ is a continuous real–valued function for } a \le x \le b\}$$

denotes the set of all contnuous real–valued functions with domain $[a,b]$. Note that, for continuous functions, we had to restrict our choice of domain to subsets of real numbers because there is no notion of continuity for functions on abstract domains (and codomains).

If you define addition and scalar multiplication as in (5.17) on p.153, then each of these sets of real–valued functions becomes a vector space for the following reasons:

I: You can verify properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ because then you just deal with ordinary real numbers.

II: The sum of two bounded functions and the product of a bounded function with a scalar is a bounded function. In other words, "$+$" associates with any two elements $f,g \in \mathscr{B}(X,\mathbb{R})$ a third item $f+g \in \mathscr{B}(X,\mathbb{R})$ and "$\cdot$" associates with any $f \in \mathscr{B}(X,\mathbb{R})$ and $\alpha \in \mathbb{R}$ a third item $\alpha \cdot f \in \mathscr{B}(X,\mathbb{R})$.

III: Likewise, the sum of two continuous functions and the product of a continuous function with a scalar is a continuous function. As for bounded functions, "$+$" associates with any two elements $f,g \in \mathscr{C}([a,b],\mathbb{R})$ a third item $f+g \in \mathscr{C}([a,b],\mathbb{R})$ and "$\cdot$" associates with any $f \in \mathscr{C}([a,b],\mathbb{R})$ and $\alpha \in \mathbb{R}$ an item $\alpha \cdot f \in \mathscr{C}([a,b],\mathbb{R})$.

It follows from the above that all three function sets are vector spaces and also that **1)** $\mathscr{B}(X,\mathbb{R})$ is a subspace of $\mathscr{F}(X,\mathbb{R})$, **2)** $\mathscr{C}(X,\mathbb{R})$ is a subspace of $\mathscr{F}(X,\mathbb{R})$.

We will see in ch.14 (Compactness) on p.434 that continuous functions defined on a closed interval are bounded. It follows that

$$\mathscr{C}([a,b],\mathbb{R}) \;\subseteq\; \mathscr{B}([a,b],\mathbb{R}) \;\subseteq\; \mathscr{F}([a,b],\mathbb{R}).$$

We deduce from this that **3)** $\mathscr{C}([a,b],\mathbb{R})$ also is a subspace of $\mathscr{B}([a,b],\mathbb{R})$.

It should be noted that, for example, continuous function need **not** be bounded on **open** intervals $]a,b[$, as the example $f(x) = \frac{1}{x}$ demonstrates for $a = 0$ and $b = 1$.

We summarize the above:

[129] Note that $\mathscr{F}(X,\mathbb{R}) = \mathbb{R}^X$ (see remark 8.4, p.239 which follows Definition 8.6 of the Cartesian Product of a family of sets.)

[130] Continuity for such functions was discussed in ch.9.3 on p.263.

$$\begin{aligned}
\mathscr{F}(X,\mathbb{R}) &= \{f(\cdot) : f(\cdot) \text{ is a real–valued function on } X\}\\
\mathscr{B}(X,\mathbb{R}) &= \{g(\cdot) : g(\cdot) \text{ is a bounded real–valued function on } X\}\\
\mathscr{C}([a,b],\mathbb{R}) &= \{h(\cdot) : h(\cdot) \text{ is a continuous real–valued function for } a \leq x \leq b\}
\end{aligned}$$

- We have subspace relationships $\mathscr{B}(X,\mathbb{R}) \subseteq \mathscr{F}(X,\mathbb{R})$
- We have subspace relationships $\mathscr{C}([a,b],\mathbb{R}) \subseteq \mathscr{B}([a,b],\mathbb{R}) \subseteq \mathscr{F}([a,b],\mathbb{R})$ □

Here are some more examples.

Example 11.12 (Subspace $\{(x,y) : x = y\}$)**.** The set $V := \{(x,x) : x \in \mathbb{R}\}$ 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. It follows that the subset L of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (see (11.6) on p.330). □

A proof for the following is omitted even though it is not difficult:

Example 11.13 (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 11.14 (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:

$$\vec{x} := (x_1, x_2, \dots, x_m, \underbrace{0, 0, \dots, 0}_{n-m \text{ times}}) \quad \square$$

Example 11.15 (All finite–dimensional vectors)**.** Let

$$\mathfrak{S} \;:=\; \bigcup_{n \in \mathbb{N}} \mathbb{R}^n \;=\; \mathbb{R}^1 \cup \mathbb{R}^2 \cup \ldots \cap \mathbb{R}^n \cup \ldots$$

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 define the sum $\vec{x} + \vec{y}$ as the vector

$$\vec{z} := (x_1 + y_1, x_2 + y_2, \dots, x_m + y_m, y_{m+1}, y_{m+2}, \dots, y_n) \quad \square$$

Example 11.16 (All sequences of real numbers)**.** Let $\mathbb{R}^{\mathbb{N}} = \prod\limits_{j \in \mathbb{N}} \mathbb{R}$ (see (8.7) on p.239). Is this the same set as $\mathfrak{S}$ from the previous example? The answer is No for the following reason: Each element $x \in \mathfrak{S}$ is of some finite dimension, say n, meaning 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 other words, $\mathbb{R}^{\mathbb{N}}$ 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 11.7 (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

$$\sum_{j=0}^{n} \alpha_j x_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_n x_n \tag{11.11}$$

a **linear combination** of the vectors x_j . The multipliers $\alpha_1, \alpha_2, \dots$ are called **scalars**. □

In other words, linear combinations are sums of scalar multiples of vectors. The expression in (11.11) always is an element of V, no matter how big $n \in \mathbb{N}$ was chosen:

Proposition 11.3 (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 $\sum\limits_{j=0}^{n} \alpha_j x_j$ also belongs to V. Note that this is also true for subspaces, because those are vector spaces, too.

PROOF: Trivial. ■

Proposition 11.4.

Let V be a vector space and let $(W_i)_{i \in I}$ be a family of subspaces of V. Let $W := \bigcap \big[W_i : i \in I \big]$. 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 11.8 (Linear span)**.** ⋆

Let V be a vector space and $A \subseteq V$. Then the set

$$span(A) := \{ \sum_{j=1}^{k} \alpha_j x_j : k \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in A \; (1 \leq j \leq k) \,\}. \tag{11.12}$$

of all linear combinations of vectors in A is called the **span** or **linear span** of A. □

Proposition 11.5.

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

$$x_{1,1},\ x_{1,2}, \dots,\ x_{1,n_1},\ x_{2,1},\ x_{2,2}, \dots,\ x_{2,n_2}, \dots,\ x_{k,1},\ x_{k,2}, \dots,\ x_{k,n_k}. \;\blacksquare$$

Theorem 11.1.

Let V be a vector space and $A \subseteq V$.

Let $\mathscr{V} := \{W \subseteq V : W \supseteq A$ and W is a subspace of $V\}$. Then $span(A) = \bigcap \left[W : W \in \mathscr{V}\right]$.

PROOF: Clearly, $span(A) \supseteq A$ It follows from prop.11.5 that $span(A) \in \mathfrak{V}$, hence $span(A) \supseteq \bigcap \left[W : W \in \mathfrak{V}\right]$.

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 $span(A) \subseteq \bigcap \left[W : W \in \mathfrak{V}\right]$. ■

Remark 11.4 (Linear span(A) = subspace generated by A)**.** Let V be a vector space and $A \subseteq V$. Theorem 11.1 justifies to call $span(A)$ the **subspace generated by** A. □

Definition 11.9 (linear mappings)**.** $\boxed{\star}$

Let V_1, V_2 be two vector spaces. Let the function $f(\cdot) : V_1 \to V_2$ satisfy

$$f(x+y) = f(x) + f(y) \quad \forall x, y \in V_1 \qquad \textbf{additivity} \tag{11.13a}$$

$$f(\alpha x) = \alpha f(x) \quad \forall x \in V_1,\ \forall \alpha \in \mathbb{R} \qquad \textbf{homogeneity} \tag{11.13b}$$

Then we call $f(\cdot)$ a **linear function** or **linear mapping**. □

Note 11.1 (Note on homogeneity)**.** We encountered “absolute homogeneity” when examining the properties of the Euclidean norm ((11.9) on p.328). That is not the same concept as homogeneity for linear functions because you had to take the absolute value $|\alpha|$ instead of α. □

Remark 11.5 (Linear mappings are compatible with linear combinations)**.** We saw in the last proposition that vector spaces are closed with respect to linear combinations. Linear functions and linear combinations work harmoniously in the following sense:

(A): The image of the sum is the sum of the images,

(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.

In other words, linear mappings preserve or are structure compatible with linear combinations. Matter of fact, **(A)** asserts that f is a homomorphism $(V_1,+) \to (V_2,+)$ from the group $(V_1,+)$ to the group $(V_2,+)$. See Definition 3.6 (Homomorphisms and isomorphisms) on p.59 and the preceding remarks on structure compatibility. □

The proof of item C in the previous remark is given in the next proposition.

Proposition 11.6 (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, i.e.,

$$f(\sum_{j=0}^{n} \lambda_j x_j) \;=\; \sum_{j=0}^{n} \lambda_j f(x_j). \tag{11.14}$$

PROOF by induction on n: We first note that $f(\lambda x) = \lambda f(x)$ because linear mappings preserve scalar multiples. This proves the base case $n = 1$. Because linear mappings also preserve the addition of any two vectors, the proposition holds for $n = 2$. Our induction assumption is

$$f(\sum_{j=0}^{k} \lambda_j x_j) \;=\; \sum_{j=0}^{k} \lambda_j f(x_j) \quad \text{for all } 1 \leq k < n.$$

We use it in the second equation ($k = 2$) and the third equation ($k = n-1$) 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 11.17 (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 11.18 (Projections on any coordinate)**.** More generally, let $n \in \mathbb{N}$ and $1 \leq j \leq n$.

$$\pi_j(\cdot) : \mathbb{R}^n \to \mathbb{R} \qquad (x_1, x_2, \dots, x_n) \mapsto x_j$$

is called the **projection** on the jth coordinate or the j**th coordinate function**.
A specific example for $n = 2$: Let $\vec{v} := (3.5, -2) \in \mathbb{R}^2$. Then $\pi_1(\vec{v}) = 3.5$ and $\pi_2(\vec{v}) = -2$. □

Example 11.19 (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 $n - m$ coordinates. Mathematicians express this as follows:
Let $m, n, i_1, i_2, \dots, i_m \in \mathbb{N}$ such that $m < n$ and $1 \leq i_1 < i_2 < \cdots < i_m \leq n$. The map

$$\pi_{i_1, i_2, \dots, i_m}(\cdot) : \mathbb{R}^n \to \mathbb{R}^m \qquad (x_1, x_2, \dots, x_n) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_m}) \tag{11.15}$$

is called the **projection** on the coordinates $i_1, i_2, \dots, i_m$.[131] □

Example 11.20. Let $x_0 \in A$. The mapping

$$\varepsilon_{x_0} : \mathscr{F}(A, \mathbb{R}) \to \mathbb{R}; \qquad f(\cdot) \mapsto f(x_0) \tag{11.16}$$

which assigns to any real–valued function on A its value at the specific point x_0 is linear because if $h(\cdot) = \sum\limits_{j=0}^{n} a_j f_j(\cdot)$ then

$$\varepsilon_{x_0}(\sum_{j=0}^{n} a_j f_j) = \varepsilon_{x_0}(h) = h(x_0) = \sum_{j=0}^{n} a_j f_j(x_0) = \sum_{j=0}^{n} a_j \varepsilon_{x_0}(f_j).$$

$\varepsilon_{x_0}(\cdot)$ is called the **abstract integral** with respect to point mass at x_0. □

Lemma 11.1 ($F \circ span = span \circ F$)**.** *[4] 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*

$$F(span(A)) = span(F(A)). \tag{11.17}$$

Proof: See Brin/Marchesi Linear Algebra, general lemma 4.1.7. ■

Definition 11.10 (Linear dependence and independence)**.** $\boxed{\star}$

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 $(1 \leq j \leq k)$ • $\sum\limits_{j=1}^{k} \alpha_j x_j = 0.$

(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 $\sum\limits_{j=1}^{k} \alpha_j x_j = 0$ then $\alpha_j = 0$, for all $1 \leq j \leq k$. □

[131] 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 (11.5) on p.325).

Definition 11.11 (Basis of a vector space). ⋆

Let V be a vector space and $B \subseteq V$. B is called a **basis** of V if both
- B is linearly independent
- $span(B) = V$. □

For the Kronecker delta which appears in the next item, see Definition 2.21 on p.29.

Definition 11.12 (Standard basis of $\mathbb{R}^n$). ⋆

Let $n \in \mathbb{N}$. For $i \in [1,n]_{\mathbb{Z}}$, let $\vec{e}^{(i)} := \left(\delta_{i1}, \delta_{i2}, \dots, \delta_{in}\right)^\top$.
Here, δ_{ij} denotes the Kronecker delta: $\delta_{ii} = 1$ for all i and $\delta_{ij} = 0$ for $i \neq j$. Thus,

$$\vec{e}^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \vec{e}^{(n)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Let $B := \{\vec{e}^{(i)} : i \in [1,n]_{\mathbb{Z}}\}$. Then B is a basis of $\mathbb{R}^n$ which we call the **standard basis**, also the **canonical basis**, of $\mathbb{R}^n$. □

Remark 11.6. If $\vec{x} = \left(x_1, \dots, x_n\right)^\top$, then

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i \vec{e}^{(i)}.$$

From this it is immediate that
- The left hand side of the above is the zero vector if and only if $x_1 = x_2 = \cdots = x_n = 0$. Thus, the standard basis is a set of linearly independent vectors.
- Every $\vec{x} \in \mathbb{R}^n$ is a linear combination of $\vec{e}^{(1)}, \dots, \vec{e}^{(n)}$. Thus, the linear span of B is $\mathbb{R}^n$.

It follows that the standard basis is, in fact, a basis of $\mathbb{R}^n$. □

Lemma 11.2.

Let V be a vector space and $A \subseteq V$. Assume that A is linearly independent but not a basis. Let $y \in span(A)^\complement$. Then $A \cup \{y\}$ is linearly independent.

Proof: Let $A' := A \cup \{y\}$ and let $x_1, x_2, \dots x_k$ be distinct elements of A' and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$) such that

$$\sum_{j=1}^{k} \alpha_j x_j = 0 \tag{11.18}$$

We must show that each α_j is zero.

Case 1: $y \neq x_j$ for all j:
Then $\{x_1, \dots, x_k\} \subseteq A' \setminus \{y\} = A$. It follows from the linear independence of A that each α_j is zero.

Case 2: $y = x_{j_0}$ for some $1 \leq j_0 \leq k$:
We first show that $\alpha_{j_0} = 0$. This is true because otherwise

$$x_{j_0} = \sum_{j \neq j_0} \frac{-\alpha_j}{\alpha_{j_0}} x_j \tag{11.19}$$

would be a linear combination of elements of A, contrary to the assumption that $x_{j_0} = y \in span(A)^{\complement}$. We deduce from (11.18) and $\alpha_{j_0} = 0$ that

$$\sum_{j \neq j_0} \alpha_j x_j = 0. \tag{11.20}$$

It follows from $\{x_j : j \neq j_0\} \subseteq A' \setminus \{y\} = A$ and the linear independence of A that $\alpha_j = 0$ for all $j \neq j_0$. ■

Theorem 11.2.

Let V be a vector space with a finite basis $B = \{b_1, \dots, b_k\}$. then any other basis of V has the same size k.

PROOF:
See, e.g., [4] Brin/Marchesi Linear Algebra. ■

This last theorem gives rise to the following definition.

Definition 11.13 (Dimension of vector spaces). ⋆

- Let V be a vector space with a finite basis $B = \{b_1, \dots, b_k\}$. We call k the **dimension** of V and we write $\dim(V) = k$.
- If V does not possess a finite basis then we say that V has infinite dimension and we write $\dim(V) = \infty$. □

The following proposition gives an example of an infinite linearly independent set.

Proposition 11.7.

For $a \in \mathbb{R}$ define $f_a(\cdot) \in \mathscr{B}(\mathbb{R}, \mathbb{R})$ as follows.

$$f_a(x) := \begin{cases} 0 & \text{if } x \neq a, \\ 1 & \text{if } x = a. \end{cases}$$

Then $\mathscr{A} := \{f_a : a \in \mathbb{R}\}$ is a linearly independent subset of $\mathscr{B}(\mathbb{R}, \mathbb{R})$.

PROOF:
We write $0(\cdot)$ for the zero function on $\mathbb{R}$. Let $n \in \mathbb{N}, a_1, \dots a_n \in \mathbb{R}$, and $\alpha_1, \dots \alpha_n \in \mathbb{R}$ such that

$$f := \sum_{j=1}^{k} \alpha_j f_{a_j} = 0(\cdot), \text{ i.e., } f(x) = \sum_{j=1}^{k} \alpha_j f_{a_j}(x) = 0 \text{ for all } x \in \mathbb{R}.$$

We must show that then $\alpha_i = 0$ for all integers $1 \leq i \leq k$. We have $0 = f(a_i) = \sum_{j=1}^{k} f_{a_j}(a_i) = \alpha_i$.

■

Proposition 11.8.

Let V be a vector space and let U be a (linear) subspace of V. Let $x_0 \in V$. Let $\tilde{U} := \{u + \lambda x_0 : u \in U \text{ and } \lambda \in \mathbb{R}\}$. Then $\tilde{U} = span(U \cup \{x_0\})$.

PROOF:

PROOF of $\subseteq$): Let $x \in \tilde{U}$, i.e., $x = u + \lambda x_0$ for some $u \in U$ and $\lambda \in \mathbb{R}$. Clearly x is a linear combination of $u \in U$ and x_0, hence $x \in span(U \cup \{x_0\})$.

PROOF of $\supseteq$): Let $x \in span(U \cup \{x_0\})$. By the definition of spans there exists $k \in \mathbb{N}$, $u_1, \dots, u_k \in U$ and $\alpha, \alpha_1, \dots \alpha_k \in \mathbb{R}$ such that $x = \sum_{j=1}^{k} \alpha_j u_j + \alpha x_0$. Let $u := \sum_{j=1}^{k} \alpha_j u_j$. Then $x = u + \alpha x_0$, hence $x \in \tilde{U}$. ■

Proposition 11.9.

Let V and V' be two vector spaces and let U be a proper (linear) subspace of V, i.e., $U \subsetneq V$. Let $x_0 \in U^{\complement}$, $y_0 \in V'$. Let $f := U \to V'$ be a linear function from U into V'. Let $\alpha \in \mathbb{R}$. Then

$$g : U \biguplus \{x_0\} \to V'; \quad g(x) := \begin{cases} f(x) & \text{if } x \in U, \\ y_0 & \text{if } x = x_0, \end{cases} \tag{11.21}$$

uniquely extends to a linear function $\tilde{f} : span(U \biguplus \{x_0\}) \to V'$ as follows:

$$\tilde{f}(x + \alpha\, x_0) := f(x) + \alpha\, y_0 \quad \text{for } x \in U, \alpha \in \mathbb{R}\,. \tag{11.22}$$

PROOF:

Let $\tilde{U} := span(U \biguplus \{x_0\})$. It follows from prop.11.8 on p.340 that any $x \in \tilde{U}$ is of the form $x = u + \alpha x_0$ for some suitable $u \in U$ and $\alpha \in \mathbb{R}$.

It follows that the function $\tilde{f}$ defined in (11.22) is in fact defined on all of $\tilde{U}$. Clearly $\tilde{f}$ coincides with g on $U \biguplus \{x_0\}$, hence $\tilde{f}$ extends g from $U \biguplus \{x_0\}$ to $\tilde{U}$.

Proof of linearity of $\tilde{f}$:

Let x_1 and $x_2 \in \tilde{U}$, i.e., there exist $u_1, u_2 \in U$ and $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $x_1 = u_1 + \alpha_1 x_0$ and $x_2 = u_2 + \alpha_2 x_0$. Let $\lambda \in \mathbb{R}$. To prove linearity of $\tilde{f}$ we must show that $\tilde{f}(x_1 + \lambda x_2) = \tilde{f}(x_1) + \lambda \tilde{f}(x_2)$.

$$\begin{aligned} &\tilde{f}(x_1 + \lambda x_2) = \tilde{f}\big((u_1 + \alpha_1 x_0) + \lambda(u_2 + \alpha_2 x_0)\big) = \tilde{f}\big((u_1 + \lambda u_2) + (\alpha_1 x_0 + \lambda \alpha_2 x_0)\big) \\ &= \tilde{f}\big((u_1 + \lambda u_2) + (\alpha_1 + \lambda \alpha_2) x_0\big) = f(u_1 + \lambda u_2) + (\alpha_1 + \lambda \alpha_2) y_0 \\ &= \big(f(u_1) + \lambda f(u_2)\big) + (\alpha_1 y_0 + \lambda \alpha_2 y_0) = \big(f(u_1) + \alpha_1 y_0\big) + \lambda\big(f(u_2) + \alpha_2 y_0\big) = \tilde{f}(x_1) + \lambda \tilde{f}(x_2). \end{aligned}$$

The linearity of f on U was used in the fifth equation. Everything else is utilizing (11.22) and grouping terms differently. This finishes the proof of linearity of $\tilde{f}$ on $\tilde{U}$.

It remains to show the uniqueness of $\tilde{f}$. So let $h : \tilde{U} \to V'$ linear such that $h(x) = \tilde{f}(x)$ for all $x \in U \biguplus \{x_0\}$. We must prove that $h(x) = \tilde{f}(x)$ for all $x \in \tilde{U}$. Let $x \in \tilde{U}$, i.e., $x = u + \alpha x_0$ for some

$u \in U$ and $\alpha \in \mathbb{R}$. Then

$$h(x) \ = h(u + \alpha x_0) \ = \ h(u) + \alpha h(x_0) \ = \ \tilde{f}(u) + \alpha \tilde{f}(x_0) \ = \ \tilde{f}(u + \alpha x_0) \ = \ \tilde{f}(x).$$

The third equality results from $h\Big|_{U \uplus \{x_0\}} = \tilde{f}\Big|_{U \uplus \{x_0\}}$, the second and fourth equalities from the linearity of h and $\tilde{f}$. This proves uniqueness of $\tilde{f}$. ■

11.2.2 Normed Vector Spaces

Definition 11.4 on p.327 in ch.11.1.3 (Length of n–Dimensional Vectors and the Euclidean Norm) gave the definition of the Euclidean norm $\|\vec{x}\|_2 = \sqrt{\sum_{j=1}^{n} x_j^2}$ in $\mathbb{R}_n$. We saw that in dimensions $n = 1, 2, 3$ that $\|\vec{x}\|_2$ equals the length of the vector $\vec{x}$ and that prop.11.1 on p. 327 "proved" informally for $n = 1, 2, 3$ that $\| \cdot \|_2$ satisfies the following three properties:

(a) positive definiteness,
(b) absolute homogeneity,
(c) triangle inequality.

In this chapter we define the norm $\|x\|$ of a vector x in an abstract vector space as a function which satisfies the above three properties, and hence generalizes the concept of the length of a vector in n–dimensional space to more general vector spaces. Before we give give this definition, we first introduce the concept of an inner product $x \bullet y$ of two vectors x and y. We will see that some of the most important norms, the Euclidean norm among them, can be derived from inner products.

The following definition of inner products and proof of the Cauchy–Schwartz inequality were taken from "Calculus of Vector Functions" (Williamson/Crowell/Trotter [17]).

Definition 11.14 (Inner product [132]).

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:

(11.23a) $x \bullet x \geq 0 \quad \forall x \in V \quad \text{and} \quad x \bullet x = 0 \ \Leftrightarrow \ x = 0$ **positive definiteness**

(11.23b) $x \bullet y = y \bullet x \quad \forall x, y \in V$ **symmetry**

(11.23c) $(x + y) \bullet z = x \bullet z + y \bullet z \quad \forall \, x, y, z \in V$ **additivity**

(11.23d) $(\lambda x) \bullet y = \lambda (x \bullet y) \quad \forall \, x, y \in V \quad \forall \, \lambda \in \mathbb{R}$ **homogeneity**

We call such a function an **inner product**. □

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 11.15 (Bilinearity). ★

[132] also called **dot product**, e.g., in [4] Brin/Marchesi Linear Algebra, ch.6, Orthogonality.

Let V be a vector space with a function

$$B : V \times V \to \mathbb{R}; \qquad (x,y) \mapsto B(x,y).$$

$B(\cdot,\cdot)$ is called **bilinear** if it is linear in each argument, i.e., the mappings

$$\begin{aligned} B_1 :&V \to \mathbb{R}; \quad x \mapsto B(x,y)\\ B_2 :&V \to \mathbb{R}; \quad y \mapsto B(x,y) \end{aligned}$$

are both linear. □

Proposition 11.10 (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

(11.24a) $$(a+b)\bullet(x+y) = a\bullet x \ + \ b\bullet x \ + \ a\bullet y \ + \ b\bullet y$$

(11.24b) $$(x+y)\bullet(x+y) = x\bullet x \ + \ 2(x\bullet y) \ + \ y\bullet y$$

(11.24c) $$(x-y)\bullet(x-y) = x\bullet x \ - \ 2(x\bullet y) \ + \ y\bullet y$$

PROOF of **(b)** and **(c)**: Left as an exercise.

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.

The proof of **(b)** and **(c)** is left as exercise 11.2 (see p.357). ■

The following is the most important example of an inner product.

Proposition 11.11 (Inner product on $\mathbb{R}^n$)**).**

Let $n \in \mathbb{N}$. Then the real–valued function

(11.25) $$(\vec{x},\vec{y}) \ \mapsto \ x_1y_1 + x_2y_2 + \ldots + x_ny_n \ = \sum_{j=1}^{n} x_jy_j\,,$$

where $\vec{x} = (x_1,\ldots x_n)$ *and* $\vec{y} = (y_1,\ldots y_n)$, *is an inner product on* $\mathbb{R}^n \times \mathbb{R}^n$.

PROOF:

(a) For $\vec{x} = \vec{y}$ we obtain $\vec{x}\bullet\vec{x} \ = \displaystyle\sum_{j=1}^{n} x_j^2$ and positive definiteness of the inner product follows from $\displaystyle\sum_{j=1}^{n} x_j^2 = 0 \ \Leftrightarrow \ x_j^2 = 0\,\forall j \ \Leftrightarrow \ x_j = 0\,\forall j.$

(b) Symmetry is clear because $x_jy_j = y_jx_j$.

(c) Let $\vec{z} = (z_1,\ldots z_n)$. Additivity follows from the fact that $(x_j+y_j)z_j = x_jz_j + y_jz_j$.

(d) Homogeneity follows from the fact that $(\lambda x_j)y_j = \lambda(x_jy_j)$. ■

Proposition 11.12 (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,

$$(x \bullet y)^2 \;\leq\; (x \bullet x)\,(y \bullet y).$$

PROOF:
Step 1: We assume first that $x \bullet x = y \bullet y = 1$. Then

$$\begin{aligned} 0 \;&\leq (x-y) \bullet (x-y) \\ &= x \bullet x - 2x \bullet y + y \bullet y = 2 - 2(x \bullet y) \end{aligned}$$

where the first equality follows from proposition (11.10) on p.342. Thus $2(x \bullet y) \;\leq\; 2$, i.e.,

$$x \bullet y \;\leq\; 1. \tag{11.26}$$

Since this inequality holds for any vectors x, y such that $x \bullet x = y \bullet y = 1$ and since absolute homogeneity (11.23d) implies $(-x) \bullet (-x) = (-1)^2 x \bullet x = 1$ we may replace x with $-x$ and obtain

$$-(x \bullet y) \;=\; (-x) \bullet y \;\leq\; 1. \tag{11.27}$$

It follows from (11.26) and (11.27) that $|x \bullet y| \;\leq\; 1$, thus $(x \bullet y)^2 \;\leq\; 1$, i.e., $(x \bullet y)^2 \;\leq\; (x \bullet x)\,(y \bullet y)$. The Cauchy–Schwartz inequality is thus true under the assumption $x \bullet x = y \bullet y = 1$.

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 nonnegative numbers $x \bullet x$ and $y \bullet y$, nonnegative.

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 $u \bullet v \;\leq\; 1$. It follows that

$$(x \bullet y)/(\sqrt{x \bullet x}\sqrt{y \bullet y}) \;=\; (x/\sqrt{x \bullet x}) \bullet (y/\sqrt{y \bullet y}) \;\leq\; 1$$

We multiply both sides with $\sqrt{x \bullet x}\sqrt{y \bullet y}$,

$$x \bullet y \;\leq\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We replace x by $-x$ and obtain

$$-(x \bullet y) \;\leq\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

Because $|x \bullet y|$ is either of $-(x \bullet y)$ or $(x \bullet y)$, it follows from the last two inequalities that

$$|x \bullet y| \;\leq\; \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We square this and obtain

$$(x \bullet y)^2 \leq (x \bullet x)\,(y \bullet y)$$

and the Cauchy–Schwartz inequality is proven. ■

Definition 11.16 (sup–norm of bounded real–valued functions)**.**

Let X be an arbitrary, nonempty set. Let $f : X \to \mathbb{R}$ be a bounded real–valued function on X, i.e., there exists $K \geq 0$ such that $|f(x)| \leq K$ for all $x \in X$. Let

(11.28) $$\|f\|_\infty := \sup\{|f(x)| : x \in X\}$$

We call $\|f\|_\infty$ the **supremum norm** or **sup–norm** of the function f. □

Proposition 11.13 (Properties of the sup norm)**.**

Let X be an arbitrary, nonempty set. Let

$$\mathscr{B}(X,\mathbb{R}) := \{h(\cdot) : h(\cdot) \text{ is a bounded real–valued function on } X\}$$

(see example 11.11 on p. 332). Then the sup–norm

$$\|\cdot\|_\infty : \mathscr{B}(X,\mathbb{R}) \to \mathbb{R}_+, \qquad h \mapsto \|h\|_\infty = \sup\{|h(x)| : x \in X\}$$

satisfies the following:

(11.29a) $\|f\|_\infty \geq 0 \;\forall\, f \in \mathscr{B}(X,\mathbb{R})$ *and* $\|f\|_\infty = 0 \;\Leftrightarrow\; f(\cdot) = 0$ ***positive definiteness***

(11.29b) $\|\alpha f(\cdot)\|_\infty = |\alpha| \cdot \|f(\cdot)\|_\infty \;\forall\, f \in \mathscr{B}(X,\mathbb{R}), \forall\, \alpha \in \mathbb{R}$ ***absolute homogeneity***

(11.29c) $\|f(\cdot) + g(\cdot)\|_\infty \leq \|f(\cdot)\|_\infty + \|g(\cdot)\|_\infty \;\forall\, f, g \in \mathscr{B}(X,\mathbb{R})$ ***triangle inequality***

PROOF: The proof is left as exercise 11.1 on p.357. ■

Note 11.2. We previously discussed the Euclidean norm

(11.30) $$\|\vec{x}\|_2 = \sqrt{\sum_{j=1}^{n} {x_j}^2}$$

for n–dimensional vectors $\vec{x} = (x_1, x_2, \dots, x_n)$. You saw in (11.1) on p.327 that it satisfies positive definiteness, absolute homogeneity and the triangle inequality, just like the sup–norm. [133] Those are properties which you associate with the length or size of an object. A very rich mathematical theory can be developed for a generalized definition of length which is based just on those properties. □

As mentioned before, mathematicians like to define new objects that are characterized by a certain 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–valued functions. Accordingly we give a special name to a function defined on a vector space which satisfies positive definiteness, homogeneity and the triangle inequality.

[133] Actually, the proof that $\|\cdot\|_2$ satisfies the triangle inequality was given only for dimensions $1, 2, 3$. It will be proved in this chapter that it is true for all dimensions n. See cor.11.1 (Inner products define norms) on p.347.

Definition 11.17 (Normed vector spaces)**.** Let V be a vector space with a real–valued function

$$\|\cdot\| : \quad V \to \mathbb{R} \qquad x \mapsto \|x\|$$

which satisfies

(11.31a)	$\|x\| \geq 0 \quad \forall x \in V \quad$ and $\quad \|x\| = 0 \;\Leftrightarrow\; x = 0$	**positive definiteness**
(11.31b)	$\|\alpha x\| = \lvert\alpha\rvert \cdot \|x\| \quad \forall x \in V, \forall \alpha \in \mathbb{R}$	**absolute homogeneity**
(11.31c)	$\|x+y\| \leq \|x\| + \|y\| \quad \forall\, x, y \in V$	**triangle inequality**

We call $\|\cdot\|$ a **norm** on V and we call V a **normed vector space**.

We write $(V, \|\cdot\|)$ instead of V when we wish to emphasize what norm on V we are discussing. □

Example 11.21 (Vector space of polynomials with sup–norm)**.** Let $A \subseteq \mathbb{R}$. It follows from (5.10) and (9.22) that the set $\mathscr{P} := \{p(\cdot) : p(\cdot)$ is a polynomial on $A\}$ of all polynomials on an arbitrary nonempty subset A of the real numbers is a subspace of the vector space $\mathscr{C}(A, \mathbb{R})$. (see example (13.1) on p.406.

If A is bounded then any polynomial p on A is bounded, hence its sup–norm

$$\|p\|_\infty \;=\; \sup\{\lvert p(x)\rvert : x \in A\}$$

is finite, and $(\mathscr{P}, \|\cdot\|_\infty)$ is a normed vector space.

If A is not bounded, then $\|p\|_\infty$ is not finite for all $p \in \mathscr{P}$. Matter of fact, it can be shown that, if A is not bounded, then the only polynomials for which $\|p(\cdot)\|_\infty < \infty$ are the constant functions on A. □

Remark 11.7. Let $(V, \|\cdot\|)$ be a normed vector space and let $\gamma > 0$.
Let $p : V \to \mathbb{R}$ be defined as $p(x) := \gamma\|x\|$. Then p also is a norm.

PROOF: The proof is left as exercise 11.3. ■

Definition 11.18 (p–norms for $\mathbb{R}^n$)**.** ⋆

Let $p \geq 1$. It will be proved in prop.11.16 on p.350 that the function

$$\vec{x} \mapsto \|\vec{x}\|_p \;:=\; \Big(\sum_{j=1}^{n} \lvert x_j\rvert^p\Big)^{1/p} \tag{11.32}$$

is a norm on $\mathbb{R}^n$). This norm is called the **p–norm** on $\mathbb{R}^n$). The Euclidean norm is a p–norm; it is the 2-norm on $\mathbb{R}^n$). □

Remark 11.8. We have seen that a vector space can be endowed with more than one norm.

- **(a)** We have seen in Remark11.7 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)** The p–norms define a collection of different norms for $\mathbb{R}^n$. □

The following theorem shows that an inner product can be associated in a natural fashion with a norm.

Theorem 11.3 (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

$$\| \cdot \|_{\bullet} : x \ \mapsto \ \|x\| \ = \ \sqrt{(x \bullet x)}$$

defines a norm on V

PROOF:
Positive definiteness : follows immediately from that of the inner product.
Absolute homogeneity : Let $x \in V$ and $\lambda \in \mathbb{R}$. Then

$$\|\lambda x\|_{\bullet} \ = \sqrt{(\lambda x) \bullet (\lambda x)} \ = \ \sqrt{\lambda\lambda(x \bullet x))} \ = \ |\lambda|\sqrt{x \bullet x} \ = \ |\lambda|\|x\|_{\bullet}.$$

Triangle inequality : Let $x, y \in V$. Then

$$\begin{aligned} \|x+y\|_{\bullet}{}^2 \ &= (x+y) \bullet (x+y) \\ &= x \bullet x \ + \ 2(x \bullet y) \ + \ y \bullet y \\ &\leq x \bullet x \ + \ 2|x \bullet y| \ + \ y \bullet y \\ &\leq x \bullet x \ + \ 2\sqrt{x \bullet x}\sqrt{y \bullet y} \ + \ y \bullet y \\ &= \|x\|_{\bullet}{}^2 \ + \ 2\|x\|_{\bullet}\,\|y\|_{\bullet} \ + \ \|y\|_{\bullet}{}^2 \\ &= \big(\|x\|_{\bullet} \ + \ \|y\|_{\bullet}\big)^2. \end{aligned}$$

The second equation uses bilinearity and symmetry of the inner product. The first inequality expresses the simple fact that $\alpha \leq |\alpha|$ for any number α. The second inequality uses Cauchy–Schwartz. The next equality just substitutes the definition $\|x\|_{\bullet} = \sqrt{(x \bullet x)}$ 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 $a = \|x\|_{\bullet}$ and $b = \|y\|_{\bullet}$.

We take square roots in the above inequality $\|x+y\|_{\bullet}{}^2 \ \leq \ \big(\|x\|_{\bullet} \ + \ \|y\|_{\bullet}\big)^2$ and obtain $\|x+y\|_{\bullet} \leq \|x\|_{\bullet} \ + \ \|y\|_{\bullet}$, the triangle inequality we set out to prove. ■

Definition 11.19 (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

$$\| \cdot \|_{\bullet} : x \ \mapsto \ \|x\|_{\bullet} \ := \ \sqrt{(x \bullet x)} \tag{11.33}$$

is called the **norm associated with the inner product** $\bullet(\cdot,\cdot)$. □

It was stated in prop.11.1 on p. 327 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 11.1.

The Euclidean norm in $\mathbb{R}^n$*:* $\|(x_1, x_2, \dots, x_n)\|_2 = \sqrt{\sum_{j=1}^{n} {x_j}^2}$ *(see def.11.4 on p.327) is a norm.*

PROOF: This follows from the fact that

$$\vec{x} \bullet \vec{y} = \sum_{j=1}^{n} x_j y_j \qquad \text{where } \vec{x} = (x_1, \dots, x_n) \text{ and } \vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$$

defines an inner product on $\mathbb{R}^n \times \mathbb{R}^n$ (see prop.11.11) for which $\|(x_1, x_2, \dots, x_n)\|_2$ is the associated norm. ■

We now look at an inner product on the vector space $\mathscr{C}([a,b], \mathbb{R})$ of all continuous real–valued functions on the interval $[a,b]$ which was defined in example 11.11 (Vector spaces of real–valued functions) on p.332. We use the terminology of [15] Stewart, J: Single Variable Calculus) for the following.

Definition 11.20. $\boxed{\star}$ Let $a, b \in \mathbb{R},\ a < b$ and assume that $f, g : [a,b] \to \mathbb{R}$ are integrable functions. (See example 5.21 on p.140.)

(a) We call the definite integral $\int_a^b f(x)dx$ the **net area** between the graph of f, the x–axis, and the vertical lines through $(a, 0)$ $(y = a)$ and $(b, 0)$ $(y = b)$. The above integral treats areas above the x–axis as positive and below the x–axis as negative, i.e., the net area is the difference between the areas above the x–axis and those below the x–axis.

(b) We call $\int_a^b |f(x)|dx$ the **area** between the graph of f, the x–axis, and the vertical lines $y = a$ and $y = b$. Note that $f(x)$ has been replaced by its absolute value $|f(x)|$. In contrast to the net area, areas below the x–axis are also counted positive. □

(c) We call $\int_a^b f(x) - g(x)dx$ the **net area** between the graphs of f and g and the vertical lines $y = a$ and $y = b$. We call $\int_a^b |f(x) - g(x)|dx$ the **area** between the graphs of f and g and the vertical lines $y = a$ and $y = b$. □

Example 11.22. Let $f : [-1, 1];\quad x \mapsto 4x^3$. The antiderivative (see example 5.21 on p.140) of $f(\cdot)$ is $x \mapsto x^4$ and we compute net area and area as follows:

(a) Net area = $\int_{(-1)}^{1} 4x^3 dx = x^4\Big|_{-1}^{1} = 1 - 1 = 0;$

(b) Area

$$= \int_{(-1)}^{1} 4|x^3|dx = \int_{(-1)}^{0} (-4x^3)dx + \int_{0}^{1} 4x^3 dx$$
$$= -x^4\Big|_{-1}^{0} + -x^4\Big|_{0}^{1} = \big(0 - (-1)\big) + \big(1 - 0\big) = 2. \quad □$$

Let $a, b \in \mathbb{R}$ such that $a < b$. We remember from example 5.21 on p.140 that continuous functions are integrable. This allows us to compare for $f \in \mathscr{C}([a,b],\mathbb{R})$ the expressions

$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}, \quad \int_a^b |f(x)|dx, \text{ and } \int_a^b \big(f(x)\big)^2 dx. \tag{11.34}$$

All three expressions give in a sense the size of f. The sup–norm measures it as the biggest possible displacement from zero, the integral over the absolute value measures the area between the graphs of the functions $x \mapsto f(x)$ and $x \mapsto 0$, and the last expression does the same with the square of f. In many respects the use of areas is superior to using the biggest difference to zero.

Squaring $f(\cdot)$ rather than using its absolute value has some mathematical advantages. One of them is that the function $(f,g) \mapsto \int_a^b f(x)g(x)dx$ defines an inner product on $\mathscr{C}([a,b],\mathbb{R})$ whose associated norm is $f \mapsto \int_a^b \big(f(x)\big)^2 dx$. We will discuss that now. In preparation we prove the following proposition.

Proposition 11.14.

Let $a, b \in \mathbb{R}$ such that $a < b$. and let $f : [a,b] \to \mathbb{R}$ be continuous. Then,

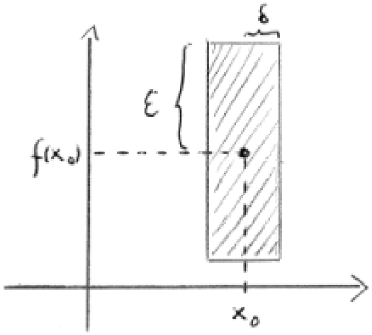
$$\int_a^b f(x)dx = 0 \quad \textit{if and only if} \quad f(x) = 0 \ \textit{for all } x \in \,]a,b[. \ \square$$

PROOF: Assume that there is $a < x_0 < b$ such that $f(x_0) \neq 0$, i.e., $f(x_0) > 0$. Let $\varepsilon := \frac{f(x_0)}{2}$. As f is continuous at x_0 there exists according to thm.9.7 on p.273 some $\delta > 0$ such that

$$|f(x_0) - f(x)| < \varepsilon, \text{ hence } f(x) > f(x_0) - \varepsilon = \frac{f(x_0)}{2} = \varepsilon \text{ for all } x_0 - \delta < x < x_0 + \delta. \tag{11.35}$$

Continuity at x_0:

If $|x - x_0| < \delta$ then $|f(x_0) - f(x)| < \varepsilon$:
The graph of f stays within the rectangle with corners $(x_0 \pm \delta, f(x_0) \pm \varepsilon)$.



Let $g : [a,b] \to \mathbb{R}$ be defined as follows.

$$g(x,y) = \begin{cases} \varepsilon & \text{if } x_0 - \delta < x < x_0 + \delta \\ 0 & \text{else.} \end{cases}$$

It follows from (11.35) that $f \geq g$, hence $\int_a^b f(x)dx \geq \int_a^b g(x)dx = (2\delta)\varepsilon > 0$. ■

Proposition 11.15.

Let $a, b \in \mathbb{R}$ such that $a < b$. Then the mapping

$$(f,g) \mapsto f \bullet g := \int_a^b f(x)g(x)dx \tag{11.36}$$

defines an inner product on $f \in \mathscr{C}([a,b],\mathbb{R})$. □

PROOF: We must prove positive definiteness, symmetry, and linearity in the left argument. In the following let $f, g, h \in \mathscr{C}([a,b],\mathbb{R})$ and $\lambda \in \mathbb{R}$.

(a) Positive definiteness: It follows from $f^2(x) \geq 0$ that $f \bullet f = \int_a^b f^2(x)dx \geq 0$. Clearly, if 0 denotes as usual the zero function $x \mapsto 0$ then $0 \bullet 0 = 0$. It remains to be shown that if $\int_a^b f^2(x)dx \geq 0$ then $f = 0$. This follows from prop.11.14.

(b) Symmetry:

$$f \bullet g = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = g \bullet f.$$

(c) Additivity and homogeneity: This can be deduced from the well–known formulas

$$\int_a^b f(x) + g(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx \quad \text{and} \quad \int_a^b \lambda g(x)dx = \lambda \int_a^b g(x)dx.$$

as follows:

$$(f+g) \bullet h = \int_a^b \big(f(x) + g(x)\big)h(x)dx = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx = f \bullet h + g \bullet h,$$
$$(\lambda f) \bullet g = \int_a^b \lambda f(x)g(x)dx = \lambda \int_a^b f(x)g(x)dx = \lambda(f \bullet g). \blacksquare$$

According to Definition 11.19 (norm for an inner product) and thm.11.3 (inner products define norms) we now define the norm associated with $f \bullet g = \int_a^b f(x)g(x)dx$.

Definition 11.21 (L_2–Norm for continuous functions)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$. Let $f \bullet g$ be the the following inner product on the space $\mathscr{C}([a,b],\mathbb{R})$ of all continuous functions $[a,b] \to \mathbb{R}$:

$$f \bullet g := \int_a^b f(x)g(x)dx. \tag{11.37}$$

The L^2**–norm**. of f is the norm associated with that inner product:

$$\|\cdot\|_{L^2} : f \mapsto \|f\|_\bullet = \sqrt{\int_a^b f^2(x)dx}\,. \tag{11.38}$$

□

We saw in Definition 11.18 that the Euclidean norm is the p–norm $\|\vec{x}\|_p = \left(\sum\limits_{j=1}^n |x_j|^p\right)^{1/p}$ for the special case $p = 2$. There is an analogue for the L^2 norm.

Definition 11.22 (L^p–norms for $\mathscr{C}([a,b],\mathbb{R})$). ⋆

Let $a,b \in \mathbb{R}$ such that $a < b$ and $p \geq 1$. It will be shown in prop.11.17 (The L^p–norm is a norm) on p.350 that

$$f \mapsto \|f\|_{L^p} := \left(\int_a^b |f(x)|^p dx\right)^{1/p} \tag{11.39}$$

is a norm on $\mathscr{C}([a,b],\mathbb{R})$. This norm is called the L^p**–norm** of f. □

11.2.3 The Inequalities of Young, Hoelder, and Minkowski ⋆

Note that this chapter is starred, hence optional.

Proposition 11.16 (The p–norm in $\mathbb{R}^n$ is a norm).

Let $p \in [1,\infty[$. Then the p-norm $\vec{x} \mapsto \|\vec{x}\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$ *is a norm in $\mathbb{R}^n$.*

PROOF:
(a). Positive definiteness:
Clearly, $\sum_{j=1}^n |x_j|^p \geq 0$ because each term $|x_j|^p$ is nonnegative, hence $\|\vec{x}\|_p = \sqrt{\sum_{j=1}^n |x_j|^p} \geq 0$. Note that $\|\vec{x}\|_p = 0$ is only possible if $|x_j|^p = 0$ for all indices j, because, if $x_{j_0} \neq 0$ for some j_0 then $|x_{j_0}|^p > 0$, hence $\left(\|\vec{x}\|_p\right)^{1/p} \geq |x_{j_0}|^p > 0$.
(b). Absolute homogeneity:
If $\lambda \in \mathbb{R}$ then

$$\|(\lambda\vec{x})\|_p = \left(\sum_{j=1}^n (|\lambda|\,|x_j|)^p\right)^{1/p} = \left(|\lambda|^p \sum_{j=1}^n |x_j|^p\right)^{1/p} = |\lambda|\left(\sum_{j=1}^n |x_j|^p\right)^{1/p} = |\lambda|\|\vec{x}\|_p.$$

(c). Triangle inequality for $p = 1$:
It follows from $|x_j + y_j| \leq |x_j| + |y_j|$ for all j that

$$\|\vec{x}+\vec{y}\|_1 = \sum_{j=1}^n |x_j + y_j| \leq \sum_{j=1}^n |x_j| + \sum_{j=1}^n |y_j| = \|\vec{x}\|_1 + \|\vec{y}\|_1.$$

(d). Triangle inequality for $p > 1$:
This is Minkowski's inequality for $(\mathbb{R}^n, \|\cdot\|_p)$ (thm.11.7 below). That $\|\cdot\|_2$ satisfies the triangle inequality (i.e., $p = 2$) also follows independently from cor.11.1 on p.347. ■

Proposition 11.17 (The L^p–norm is a norm).

Let $p \in [1,\infty[$ and let $a,b \in \mathbb{R}$ such that $a < b$.
Then the L^p-norm $f \mapsto \|f\|_{L^p} = \left(\int_a^b |f(x)|^p\right)^{1/p}$ *is a norm in $\mathscr{C}([a,b],\mathbb{R})$.*

PROOF:

(a). Positive definiteness:
Follows from prop.11.14 on p.348 and the fact that $x \mapsto |f(x)|^p$ is a nonnegative and continuous function.

(b). Absolute homogeneity:
If $\lambda \in \mathbb{R}$ then

$$\begin{aligned}\|(\lambda f)\|_{L^p} &= \Big(\int_a^b (|\lambda|\,|f(x)|)^p\,dx\Big)^{1/p}\\ &= \Big(|\lambda|^p \int_a^b |f(x)|^p\,dx\Big)^{1/p} = |\lambda|\Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} = |\lambda|\|f\|_{L^p}.\end{aligned}$$

(c). Triangle inequality for $p = 1$:
It follows from $|f(x) + g(x)| \le |f(x)| + |g(x)|$ for all x that

$$\begin{aligned}\|f+g\|_{L^1} &= \int_a^b |f(x)+g(x)|\,dx \le \int_a^b (|f(x)| + |g(x)|)\,dx\\ &= \int_a^b |f(x)|\,dx + \int_a^b |g(x)|\,dx = \|f\|_{L^1} + \|g\|_{L^1}.\end{aligned}$$

(d). Triangle inequality for $p > 1$:
This is Minkowski's inequality for L^p–norms (thm.11.5 below). That $\|\cdot\|_{L^2}$ satisfies the triangle inequality (i.e., $p = 2$) also follows independently from cor.11.1 on p.347. ■

We were referring to Minkowksi's inequalities for $(\mathbb{R}^n, \|\cdot\|_p)$ and L^p–norms when proving the triangle inequality for those norms. We now build the machinery that will allows us to prove those inequalities.

Proposition 11.18 (Young's Inequality)**.**

Let $a, b > 0$ and let $p, q > 1$ be ***conjugate indices****, i.e.,*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{11.40}$$

Then ***Young's inequality*** *holds:*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}. \tag{11.41}$$

PROOF:

Step 1: We show that $q - 1 = \frac{1}{p-1}$:

$$\begin{aligned}&\frac{1}{p} + \frac{1}{q} = 1 \Rightarrow q + p = pq \Rightarrow q(1-p) = -p\\ &\Rightarrow q = \frac{p}{p-1} \Rightarrow q - 1 = \frac{p-(p-1)}{p-1} = \frac{1}{p-1}.\end{aligned} \tag{11.42}$$

Step 2: The functions

$$\varphi : [0,\infty[\, \to\, [0,\infty[;\quad x \mapsto x^{p-1} \quad \text{and} \quad \psi : [0,\infty[\, \to\, [0,\infty[;\quad y \mapsto y^{q-1}$$

are inverse to each other because we have

$$\psi\big(\varphi(x)\big) \;=\; \psi\big(x^{p-1}\big) \;=\; \big(x^{p-1}\big)^{q-1} \;\overset{(\star)}{=}\; \big(x^{p-1}\big)^{1/(p-1)} \;=\; x$$

($(\star)$ follows from step 1). We further have

$$\varphi\big(\psi(y)\big) \;=\; \varphi\big(y^{q-1}\big) \;=\; \big(y^{q-1}\big)^{p-1} \;\overset{(\star\star)}{=}\; \big(y^{q-1}\big)^{1/(q-1)} \;=\; y$$

($(\star\star)$ again follows from step 1). Note that those two functions are continuous (actually, differentiable) and strictly increasing because $\varphi'(t) = (p-1)t^{(p-2)} > 0$ and $\psi'(t) = (q-1)t^{(q-2)} > 0$ for all $t \geq 0$. We further have $\varphi(0) = 0 = \psi(0)$.

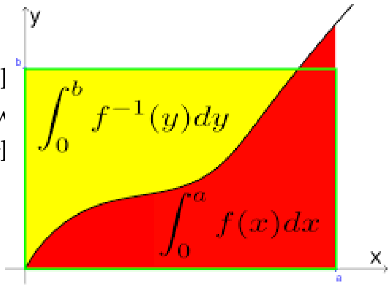
Step 3: Let $f : [0,\infty[\, \to\, [0,\infty[$ be a continuous and strictly increasing (hence invertible) function such that $f(0) = 0$. Then the following is true for any two real numbers $a, b > 0$:

$$ab \;\leq\; \int_0^a f(x)dx \;+\; \int_0^b f^{-1}(y)dy. \tag{11.43}$$

To prove this, we distinguish three cases. Either $b < f(a)$ or $b > f(a)$ or $b = f(a)$.

The picture to the right shows what happens if $b < f(a)$: Tl rectangle ab is covered by the areas determined by the tw integrals, but not all of the area of $\int_0^a f(x)dx$ is covered by tl rectangle.

Source: https://brilliant.org/wiki/youngs-inequality/



If $b > f(a)$ then the situation is similar, except that now not all of the area of $\int_0^b f^{-1}(y)dy$ is covered by the rectangle ab. Finally, if $b = f(a)$, the area covered by the two integrals matches the rectangle.

Step 4: We now apply the above to the function $y = f(x) = x^{p-1}$.
The inverse function is $x = f^{-1}(y) = y^{1/(p-1)} = y^{q-1}$ (see (11.42)). We integrate and obtain

$$\int_0^a f(x)dx \;=\; \int_0^a x^{p-1} \;=\; \frac{x^p}{p}\bigg|_0^a \;=\; \frac{a^p}{p}, \qquad \int_0^a f{-}1(y)dy \;=\; \int_0^a y^{q-1} \;=\; \frac{y^q}{q}\bigg|_0^b \;=\; \frac{b^q}{q}.$$

Young's inequality (11.41) now follows from (11.43). ■

Theorem 11.4 (Hoelder's inequality for L^p–norms).

Let $a, b \in \mathbb{R}$ such that $a < b$. Let $p, q > 1$ be conjugate indices, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{11.44}$$

*Then **Hoelder's inequality** is true:*

(11.45)
$$\|fg\|_{L^1} \leq \|f\|_{L^p}\|g\|_{L^q}, \quad i.e., \int_a^b |f(x)g(x)|dx \leq \Big(\int_a^b |f(x)|^p dx\Big)^{1/p} \Big(\int_a^b |g(x)|^q dx\Big)^{1/q}.$$

PROOF: We note that the composite function $x \mapsto |f(x)|^p$ is continuous, hence integrable, as the composite of the three continuous functions $x \mapsto f(x)$, $y \mapsto |y|$, and $z \mapsto z^p$.

Note that $\|f\|_{L^p} = 0$ is only possible if $|f(x)|^p = 0$, i.e., $f(x) = 0$ for all x (see prop.11.14 on p.348). Likewise, $\|g\|_{L^q} = 0$ implies $g(x) = 0$ for all x. In either case, $\int_a^b f(x)g(x) = 0$ and (11.45) is trivially satisfied. So we may assume that both $\|f\|_{L^p} > 0$ and $\|g\|_{L^q} > 0$ For some fixed $x \in [a, b]$ let

$$A := \|f\|_{L^p}, \quad a_x := \frac{|f(x)|}{A}, \qquad B := \|g\|_{L^q}, \quad b_x := \frac{|g(x)|}{B}.$$

It follows from Young's inequality (11.41) that $a_x b_x \leq \frac{a_x^p}{p} + \frac{b_x^q}{q}$. We integrate both sides of that inequality $\int_a^b \cdots dx$ and obtain from the monotonicity of the integral (see example 5.21 on p.140) that

$$\int_a^b a_x b_x dx \leq \int_a^b \Big(\frac{a_x^p}{p} + \frac{b_x^q}{q}\Big) dx. \tag{11.46}$$

i.e.,

$$\begin{aligned} \frac{1}{AB}\int_a^b |f(x)g(x)|\,dx &\leq \int_a^b (\frac{|f(x)|^p}{pA^p} + \frac{|g(x)|^q}{qB^q})\,dx \\ &= \frac{1}{pA^p}\int_a^b |f(x)|^p\,dx + \frac{1}{qB^q}\int_a^b |g(x)|^q\,dx. \end{aligned} \tag{11.47}$$

We use

$$\int_a^b |f(x)|^p dx = (\|f\|_{L^p})^p, \quad \int_a^b |g(x)|^q dx = (\|g\|_{L^q})^q \tag{11.48}$$

in (11.47) and obtain

$$\frac{1}{AB}\int_a^b |f(x)g(x)|dx \leq \frac{A^p}{pA^p} + \frac{B^q}{qB^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

It follows from the definition of A and B that

$$\int_a^b |f(x)g(x)|dx \leq AB = \|f\|_{L^p}\,\|g\|_{L^q}. \blacksquare$$

Theorem 11.5 (Minkowski's inequality for L^p–norms)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$ and let $p \in [1, \infty[$. Then ***Minkowski's inequality*** *is true:*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}, \ \ i.e., \tag{11.49}$$

$$\Big(\int_a^b |f(x) + g(x)|^p \, dx\Big)^{1/p} \leq \Big(\int_a^b |f(x)|^p \, dx\Big)^{1/p} + \Big(\int_a^b |g(x)|^p \, dx\Big)^{1/p}. \tag{11.50}$$

PROOF: This follows for $p = 1$ from part **(c)** of the proof of prop.11.17. We may assume that $p > 1$. Let q be the conjugate index to p, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1, \text{ hence } (p-1)q = p \tag{11.51}$$

(see (11.42)). Let $a \leq x \leq b$. Then

$$|f(x) + g(x)|^p = |f(x) + g(x)| \, |f(x) + g(x)|^{p-1} \leq |f(x)| \, |f(x) + g(x)|^{p-1} + |g(x)| \, |f(x) + g(x)|^{p-1}.$$

The last inequality follows from $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ and $|f(x) + g(x)|^{p-1} \geq 0$. We integrate and obtain

$$\int_a^b |f(x) + g(x)|^p \, dx \leq \int_a^b |f(x)| \, |f(x) + g(x)|^{p-1} \, dx + \int_a^b |g(x)| \, |f(x) + g(x)|^{p-1} \, dx. \tag{11.52}$$

We apply Hoelder's inequality to the first of the two integrals on the right–hand side of (11.52) and obtain

$$\begin{aligned} \int_a^b \big(|f(x)|\big) \big(|f(x) + g(x)|^{p-1}\big) \, dx &\leq \Big(\int_a^b \big(|f(x)|\big)^p \, dx\Big)^{1/p} \Big(\int_a^b \big(|f(x) + g(x)|^{p-1}\big)^q \, dx\Big)^{1/q} \\ &= \Big(\int_a^b |f(x)|^p \, dx\Big)^{1/p} \Big(\int_a^b |f(x) + g(x)|^{(p-1)q} \, dx\Big)^{1/q} \\ &= \Big(\int_a^b |f(x)|^p \, dx\Big)^{1/p} \Big(\int_a^b |f(x) + g(x)|^p \, dx\Big)^{1/q}. \end{aligned} \tag{11.53}$$

The last equality results from $(p-1)q = p$ (see (11.51)). Similarly, we obtain from the second integral on the right–hand side of (11.51) the following:

$$\int_a^b \big(|g(x)|\big) \big(|f(x) + g(x)|^{p-1}\big) \, dx \leq \Big(\int_a^b |g(x)|^p \, dx\Big)^{1/p} \Big(\int_a^b |f(x) + g(x)|^p \, dx\Big)^{1/q}. \tag{11.54}$$

We apply (11.53) and (11.54) to (11.52) and obtain

$$\begin{aligned} \int_a^b |f(x) + g(x)|^p \, dx &\leq \Big(\int_a^b |f(x)|^p \, dx\Big)^{1/p} \Big(\int_a^b |f(x) + g(x)|^p \, dx\Big)^{1/q} \\ &+ \Big(\int_a^b |g(x)|^p \, dx\Big)^{1/p} \Big(\int_a^b |f(x) + g(x)|^p \, dx\Big)^{1/q}. \end{aligned} \tag{11.55}$$

Minkowski's inequality (11.50) is trivially satisfied if $\int_a^b |f(x)+g(x)|^p dx = 0$, so we may assume that $\int_a^b |f(x)+g(x)|^p dx > 0$. This allows us to divide each term in (11.55) by $\big(\int_a^b |f(x)+g(x)|^p dx\big)^{1/q}$. We obtain

$$(11.56)\qquad \Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1-1/q} \le \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} + \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p}.$$

Note that $1-\frac{1}{q} = \frac{1}{p}$ because $\frac{1}{q}+\frac{1}{p} = 1$, and (11.56) reads

$$\Big(\int_a^b |f(x)+g(x)|^p\,dx\Big)^{1/p} \le \Big(\int_a^b |f(x)|^p\,dx\Big)^{1/p} + \Big(\int_a^b |g(x)|^p\,dx\Big)^{1/p}. \ \blacksquare$$

Theorem 11.6 (Hoelder's inequality for the p–norms)**.**

Let $n \in \mathbb{N}$
and $\vec{x} = (x_1,\dots x_N), \vec{y} = (y_1,\dots y_N) \in \mathbb{R}^n$. Let $p,q > 1$ be conjugate indices, i.e.,

$$(11.57)\qquad \frac{1}{p}+\frac{1}{q} = 1.$$

*Then **Hoelder's inequality** in $\mathbb{R}^n$ is true:*

$$(11.58)\qquad \sum_{j=1}^n |x_jy_j| \le \|\vec{x}\|_p\|\vec{y}\|_q,\ \ \text{i.e.,} \sum_{j=1}^n |x_jy_j| \le \Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p}\Big(\sum_{j=1}^n |y_j|^q\Big)^{1/q}.$$

PROOF: Let $\vec{x},\vec{y} \in \mathbb{R}^n$. If $\vec{x} = 0$ or $\vec{y} = 0$ then $\sum_{j=1}^n |x_jy_j| = 0$ and (11.58) is trivially satisfied. We hence may assume that both $\vec{x} \neq 0$ and $\vec{y} \neq 0$.

It follows from part **(a)** of the proof of prop.11.16 (positive definiteness of $\|\cdot\|_p$ for all p) on p.350 that $\|\vec{x}\|_p > 0$ and $\|\vec{y}\|_q > 0$ For some fixed index $1 \le j \le n$ let

$$A := \|\vec{x}\|_p,\ \ a_j := \frac{|x_j|}{A}, \qquad B := \|\vec{y}\|_q,\ \ b_x := \frac{|y_j|}{B}.$$

It follows from Young's inequality (11.41) that

$$a_jb_j \le \frac{a_j^p}{p} + \frac{b_j^q}{q}.$$

We take sums $\sum_{j=1}^n \cdots$ of both sides of that inequality and obtain from the monotonicity of summation

$$(11.59)\qquad \sum_{j=1}^n a_jb_j \le \sum_{j=1}^n \Big(\frac{(a_j)^p}{p} + \frac{(b_j)^q}{q}\Big),$$

i.e.,

$$(11.60)\qquad \frac{1}{AB}\sum_{j=1}^n |x_jy_j| \le \sum_{j=1}^n \Big(\frac{|a_j|^p}{pA^p} + \frac{|b_j|^q}{qB^q}\Big) = \frac{1}{pA^p}\sum_{j=1}^n |a_j|^p + \frac{1}{qB^q}\sum_{j=1}^n |b_j|^q.$$

But

$$\sum_{j=1}^{n} |x_j|^p \;=\; (\|f\|_p)^p, \quad \sum_{j=1}^{n} |y_j|^q \;=\; (\|g\|_q)^q. \tag{11.61}$$

It follows from (11.60) that

$$\frac{1}{AB}\sum_{j=1}^{n} |x_j y_j| \;\leq\; \frac{A^p}{pA^p} \;+\; \frac{B^q}{qB^q} \;=\; \frac{1}{p} \;+\; \frac{1}{q} \;=\; 1,$$

and we deduce from the definition of A and B that

$$\sum_{j=1}^{n} |x_j y_j| \;\leq\; AB \;=\; \|\vec{x}\|_p \, \|\vec{x}\|_q. \;\blacksquare$$

Theorem 11.7 (Minkowski's inequality for $(\mathbb{R}^n, \|\cdot\|_p)$).

*Let $n \in \mathbb{N}$ and $\vec{x} = (x_1, \dots x_N)$, $\vec{y} = (y_1, \dots y_N) \in \mathbb{R}^n$. Let $p \in [1, \infty[$. Then **Minkowski's inequality for** $(\mathbb{R}^n, \|\cdot\|_p)$ is true:*

$$\|\vec{x} + \vec{y}\|_p \;\leq \|\vec{x}\|_p \;+\; \|\vec{y}\|_p, \;\; i.e., \tag{11.62}$$

$$\Big(\sum_j |x_j + y_j|^p\Big)^{1/p} \;\leq\; \Big(\sum_j |f(x)|^p\Big)^{1/p} \;+\; \Big(\sum_j |g(x)|^p\Big)^{1/p}. \tag{11.63}$$

PROOF: This follows for $p = 1$ from part **(c)** of the proof of prop.11.17. We hence may assume that $p > 1$. Let q be the conjugate index to p, i.e.,

$$\frac{1}{p} \;+\; \frac{1}{q} \;=\; 1, \text{ hence } (p-1)q \;=\; p \tag{11.64}$$

(see (11.42)). Let $a \leq x \leq b$. Then

$$|x_j + y_j|^p \;= |x_j + y_j| \, |x_j + y_j|^{p-1} \;\leq\; |x_j| \, |x_j + y_j|^{p-1} \;+\; |y_j| \, |x_j + y_j|^{p-1}.$$

The last inequality follows from $|x_j + y_j| \leq |x_j| + |y_j|$ and $|x_j + y_j|^{p-1} \geq 0$. We sum and obtain

$$\sum_j |x_j + y_j|^p \;\leq \sum_j |x_j| \, |x_j + y_j|^{p-1} \;+\; \sum_j |y_j| \, |x_j + y_j|^{p-1}. \tag{11.65}$$

Hoelder's inequality applied to the first of the two integrals on the right–hand side of (11.65) yields

$$\begin{aligned} \sum_j \big(|x_j|\big) \, \big(|x_j + y_j|^{p-1}\big) \;&\leq \Big(\sum_j \big(|x_j|\big)^p\Big)^{1/p} \Big(\sum_j \big(|x_j + y_j|^{p-1}\big)^q\Big)^{1/q} \\ &= \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^{(p-1)q}\Big)^{1/q} \\ &= \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \end{aligned} \tag{11.66}$$

The last equality results from $(p-1)q = p$ (see (11.64)). Similarly, we obtain from the second integral on the right–hand side of (11.65) the following:

$$\sum_j \left(|y_j|\right) \left(|x_j + y_j|^{p-1}\right) \leq \Big(\sum_j |y_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \tag{11.67}$$

We apply (11.66) and (11.67) to (11.65) and obtain

$$\begin{aligned} \sum_j |x_j + y_j|^p &\leq \Big(\sum_j |x_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q} \\ &\quad + \Big(\sum_j |y_j|^p\Big)^{1/p} \Big(\sum_j |x_j + y_j|^p\Big)^{1/q}. \end{aligned} \tag{11.68}$$

Minkowski's inequality (11.63) is trivially satisfied if $\sum_j |x_j + y_j|^p = 0$, so we may assume that $\sum_j |x_j + y_j|^p > 0$. We divide each term in (11.68) by $\big(\sum_j |x_j + y_j|^p\big)^{1/q}$ and obtain

$$\Big(\sum_j |x_j + y_j|^p\Big)^{1-1/q} \leq \Big(\sum_j |x_j|^p\Big)^{1/p} + \Big(\sum_j |y_j|^p\Big)^{1/p}. \tag{11.69}$$

Note that $1 - \frac{1}{q} = \frac{1}{p}$ because $\frac{1}{q} + \frac{1}{p} = 1$, and (11.69) reads

$$\Big(\sum_j |x_j + y_j|^p\Big)^{1/p} \leq \Big(\sum_j |x_j|^p\Big)^{1/p} + \Big(\sum_j |y_j|^p\Big)^{1/p}. \blacksquare$$

In chapter **??** on the topology of metric spaces (p. **??**) 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 nonempty set.

11.3 Exercises for Ch.11

Exercise 11.1. Prove prop.11.13 on p.344: Let X be an arbitrary, nonempty set. Then the function $\|\cdot\|_\infty : \mathscr{B}(X,\mathbb{R}) \to \mathbb{R}_+, \quad h \to \|h\|_\infty = \sup\{|h(x)| : x \in X\}$ defines a norm. □

Exercise 11.2. Prove parts **(a)** and **(b)** of prop.11.10 (Algebraic properties of the inner product) on p.342:

Let V be a vector space with inner product $\bullet(\cdot,\cdot)$. Let $a, b, x, y \in V$. Then

(a) $(a+b) \bullet (x+y) = a \bullet x \,+\, b \bullet x \,+\, a \bullet y \,+\, b \bullet y$

(b) $(x+y) \bullet (x+y) = x \bullet x \,+\, 2(x \bullet y) \,+\, y \bullet y$

(c) $(x-y) \bullet (x-y) = x \bullet x \,-\, 2(x \bullet y) \,+\, y \bullet y$ □

Exercise 11.3. Prove Remark 11.7 on p.345:

Let $(V, \|\cdot\|)$ be a normed vector space and let $\gamma > 0$. Let $p : V \to \mathbb{R}$ be defined as $p(x) := \gamma\|x\|$. Then p also is a norm. □

Exercise 11.4. Prove that the p–norm (see Definition 11.18 on p.345) is a norm on $\mathbb{R}^n$ for the special case $p = 1$:

$$\|\vec{x}\|_1 = \sum_{j=1}^{n} |x_j| \;\square$$

11.4 Blank Page after Ch.11

This page is intentionally left blank.

12 Metric Spaces and Topological Spaces – Part I

There is a branch of Mathematics, called topology, which deals with the concept of closeness. The definition of the limit 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 also can be phrased in terms of closeness: Continuous functions 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 metric spaces (X, d): sets X that are equipped with a distance function $(x, y) \mapsto d(x, y)$. Even this limited context will significantly generalize the material of ch.9.3 (Convergence and Continuity in $\mathbb{R}$) and ch.11.2.2 (Normed Vector Spaces).

12.1 Definition and Examples of Metric Spaces

A metric is a real–valued 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 nonnegative, 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 11.1.3 on p.325 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

$$\text{dist}(\vec{v}, \vec{w}) \ = \ \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2}$$

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

$$\text{dist}(\vec{v}, \vec{w}) \ = \ \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + (w_3 - v_3)^2}.$$

In the next chapter we will generalize the concept of distance to more general objects.

Definition 12.1 (Metric spaces)**.** Let X be an arbitrary, nonempty set.
A **metric** on X is a real–valued 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:

(12.1a) $d(x,y) \geq 0 \quad \forall x,y \in X \quad \text{and} \quad d(x,y) = 0 \;\Leftrightarrow\; x = y$ **positive definiteness**

(12.1b) $d(x,y) = d(y,x) \quad \forall x,y \in X$ **symmetry**

(12.1c) $d(x,z) \leq d(x,y) + d(y,z) \quad \forall\, x,y,z \in X$ **triangle inequality**

Let $x,y \in X$ and $\varepsilon > 0$. We say that x and y are ε–**close** if $d(x,y) < \varepsilon$. 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 12.1 (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., $\vec{v} = -\vec{w}$. We only care about size and not about direction.

"Triangle inequality": If you directly drive from x to z then this will take less fuel than if you make a stopover at an intermediary y. □

Remark 12.2. Do not make the mistake and think of X as a set of numbers or vectors! For example, we might deal with

$$X \;:=\; \{\text{ all students who are currently taking this class }\}.$$

We can define the distance of any two students s_1 and s_2 as

$$d(s_1,s_2) = \begin{cases} 0 & \text{for } s_1 = s_2, \\ 1 & \text{for } s_1 \neq s_2. \end{cases}$$

We will learn later in this subchapter that the above function is called the discrete metric on X and satisfies indeed the definition of a metric. [134] □

The triangle inequality generalizes to more than two terms.

[134] see Definition 12.3 on p.362 and prop.12.2 directly thereafter.

Proposition 12.1.

Let (X, d) be a metric space. Let $n \in \mathbb{N}$ and $x_1, x_2, \dots, x_n \in X$. Then

$$d(x_1, x_n) \ \leq \ \sum_{j=1}^{n-1} d(x_j, x_{j+1}) \ = \ d(x_1, x_2) \ + \ d(x_2, x_3) \ + \ d(x_{n-1}, x_n). \tag{12.2}$$

The proof is left as exercise 12.1 on p.398. ■

Before we give some examples of metric spaces, here is a theorem that tells you that a vector space with a norm (see Definition 11.17 on p.345), becomes a metric space as follows:

Theorem 12.1 (Norms define metric spaces).

Let $(V, \| \cdot \|)$ be a normed vector space. Then the function

$$d_{\|\cdot\|}(\cdot,\cdot) : V \times V \to \mathbb{R}_{\geq 0}; \qquad (x, y) \mapsto d_{\|\cdot\|}(x, y) := \|y - x\| \tag{12.3}$$

defines a metric space $(V, d_{\|\cdot\|})$.

PROOF: The proof is left as exercise 12.2 on p.398. ■

Definition 12.2 (Metric induced by a norm [135]).

We say that the metric $d_{\|\cdot\|}(\cdot,\cdot)$ defined by (12.3) is **induced by the norm** $\| \cdot \|$. and that $d_{\|\cdot\|}(\cdot,\cdot)$ is **derived from the norm** $\| \cdot \|$, or that $d_{\|\cdot\|}(\cdot,\cdot)$ is **associated with the norm** $\| \cdot \|$. □

Here are some examples of metric spaces.

Example 12.1 (($\mathbb{R}$ with $d_{|\cdot|}(a, b) = |b - a|$)**.** According to thm.12.1 $(\mathbb{R}, d_{|\cdot|})$ 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 $d_{|\cdot|}(x, y) \ = \ |x - y| \ = \ |-(y - x)| \ = \ |y - x| \ = \ d_{|\cdot|}(y, x)$.

The triangle inequality for a metric follows from $|a + b| \leq |a| + |b|$ (see prop.2.5 on p.28):

$$\begin{aligned} d_{|\cdot|}(x, z) \ &= |x - z| \ = \ |(x - y) - (z - y)| \\ &\leq |x - y| + |z - y| \ = \ d_{|\cdot|}(x, y) + d_{|\cdot|}(z, y) \ = \ d_{|\cdot|}(x, y) + d_{|\cdot|}(y, z). \ \square \end{aligned}$$

Example 12.2 (bounded real–valued functions with $d_{\|\cdot\|_\infty} f, g) = $ sup–norm of $g(\cdot) - f(\cdot)$)**.**

$$d_{\|\cdot\|_\infty}(f, g) \ = \ \|g - f\|_\infty \ = \ \sup\{|g(x) - f(x)| : x \in X\} \tag{12.4}$$

is a metric on the set $\mathscr{B}(X, \mathbb{R})$ of all bounded real–valued functions on X. This follows from thm.12.1 and prop.11.13 on p. 344, according to which $(\mathscr{B}(X, \mathbb{R}), \| \cdot \|_\infty)$ is a normed vector space. □

[135] Compare this to Definition 11.19 (Norm for an inner product) on p.346.

Example 12.3 (continuous real–valued functions on [a,b]with $d_{\|\cdot\|_{L^2}}(f,g) = \|g-f\|_{L^2}$)**.** We will see in ch.12.2 on p.362 that $\|g-f\|_\infty$ is a good measure for the difference of the functions f and g and that an often even better measure is that of the area difference between their graphs which is given by the netric

$$d_{\|\cdot\|_{L^2}}(f,g) \ = \ \|g-f\|_{L^2} \ = \ \sqrt{\int_a^b \big(g(x)-f(x)\big)^2 dx}. \tag{12.5}$$

(See Definition 11.21 on p.349). □

Example 12.4 ($\mathbb{R}^n$ with the Euclidean metric)**.**

$$d_{\|\cdot\|_2}(\vec{x},\vec{y}) \ = \ \sqrt{(y_1-x_1)^2+(y_2-x_2)^2+\ldots+(y_n-x_n)^2} \ = \ \sqrt{\sum_{j=1}^n (y_j-x_j)^2}$$

This follows from the fact that the Euclidean norm is a norm on the vector space $\mathbb{R}^n$ (see cor.11.1 on p.347.) □

Just in case you think that all metrics are derived from norms, here is an extremely important counterexample.

Definition 12.3 (Discrete metric)**.**

Let X be nonempty. Then the function

$$d(x,y) = \begin{cases} 0 & \text{for } x=y \\ 1 & \text{for } x \neq y \end{cases}$$

on $X \times X$ is called the **discrete metric** on X. □

The above definition makes sense because of the following proposition.

Proposition 12.2. *The discrete metric satisfies the properties of a metric.*

PROOF: Obviously the function is nonnegative and it is zero if and only if $x = y$. Symmetry is obvious too.

The triangle inequality $d(x,z) \leq d(x,y) + d(y,z)$ is certainly true 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

$$d(x,z) = 1 \leq d(x,y) + d(y,z),$$

and this proves the triangle inequality. ■

12.2 Measuring the Distance of Real–Valued Functions

How do we compare two functions? Let us make our lives easier: How do we compare two real–valued 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 12.4 (Maximal displacement distance between real–valued functions [136] **).**

Let X be an arbitrary, nonempty set and let $f(\cdot), g(\cdot) : X \to \mathbb{R}$ be two real–valued functions on X. We define the **maximal displacement distance** , also called the **sup–norm distance** or $\|\cdot\|_\infty$ **distance** , between $f(\cdot)$ and $g(\cdot)$ as

$$d_\infty(f,g) \ := \ \|f(\cdot) - g(\cdot)\|_\infty \ = \ \sup\{|f(x) - g(x)| : x \in X\}, \tag{12.6}$$

i.e., as the metric induced by the sup–norm on the set $\mathscr{B}(X, \mathbb{R})$ of all bounded real–valued function on X. □

Remark 12.3. We will see in prop.13.7 on p.417 of ch.13.2.1 on convergence of function sequences that the sup–norm induced metric is suitable to measure what will be called "uniform convergence" of real–valued functions. As a metric, the distance measure of two functions f, g satisfies positive definiteness, symmetry and the 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 $h(\cdot)$. □

Remark 12.4. Figure 12.1 illustrates the last definition. 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.

[136] See example 12.2 on p.361.

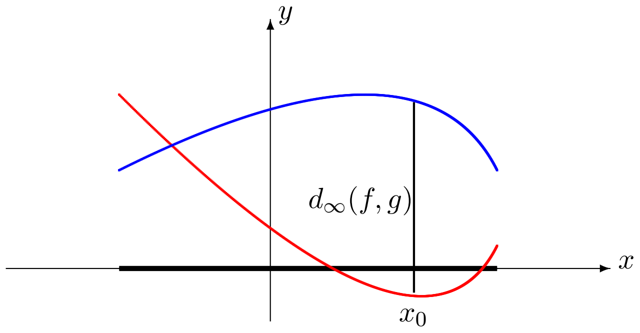


Figure 12.1: Distance of two real–valued functions.

Figure 12.2 allows you to visualize for a given $\delta > 0$ and $f : X \to \mathbb{R}$ the "δ–neighborhood" of $f(\cdot)$ defined as

$$N_\delta(f) \;:=\; \{g : X \to \mathbb{R} : d(_\infty f, g) < \delta\} \;=\; \{g(\cdot) : X \to \mathbb{R} : \sup_{x \in X} |f(x) - g(x)| < \delta\}, \tag{12.7}$$

i.e., the set of all functions $g(\cdot)$ with distance less than δ from $f(\cdot)$.

You draw the graph of $f(\cdot) + \delta$ (the graph of $f(\cdot)$ shifted up by the amount of δ) and the graph of $f(\cdot) - \delta$ (the graph of $f(\cdot)$ shifted down 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 $f(\cdot)$.

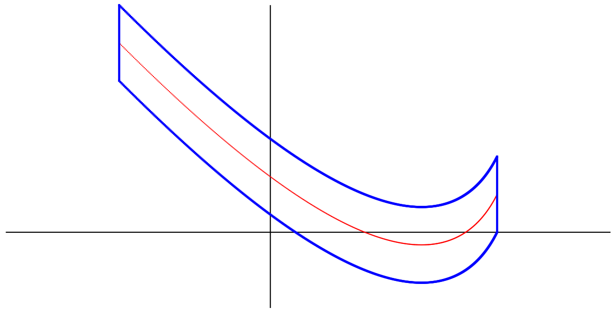


Figure 12.2: δ–neighborhood of a real–valued function.

In other words, assuming that the domain A is a single, connected chunk and not a collection of several 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. □

Definition 12.5 (Mean distances between real–valued functions)**.** Let $a, b \in \mathbb{R}$ such that $a < b$ and let $f(\cdot), g(\cdot) : X \to \mathbb{R}$ be two continuous real–valued functions on X. We define the **mean square distance** between $f(\cdot)$ and $g(\cdot)$ on $[a, b]$ as

$$d_{L^2}(f,g) \;:=\; d_{\|\cdot\|_{L^2}(f,g)} \;=\; \|g - f\|_{L^2} \;=\; \left(\int_a^b (g(x) - f(x)^2 dx\right)^{1/2}, \tag{12.8}$$

i.e., as the metric induced by the L^2–norm on the set $\mathscr{C}_{\mathscr{B}}$([a,b], $\mathbb{R}$) of all continuous and bounded real–valued function on $[a, b]$.

We further define the **mean distance** between $f(\cdot)$ and $g(\cdot)$ on $[a, b]$ as

$$d_{L^1}(f,g) \;:=\; d_{\|\cdot\|_{L^1}(f,g)} \;=\; \|g - f\|_{L^1} \;=\; \int_a^b |g(x) - f(x|dx, \tag{12.9}$$

i.e., as the metric induced by the L^1–norm on the set $\mathscr{C}_{\mathscr{B}}$([a,b], $\mathbb{R}$). □

Remark 12.5. We saw in Definition 11.20, example 11.22, and Definition 11.21 on pp.347 that both

$$d_{L^1}(f,g) \;:= d_{\|\cdot\|_{L^1}(f,g)} \;=\; \|g - f\|_{L^1} \;=\; \int_a^b |g(x) - f(x)|dx, \tag{12.10}$$

$$d_{L^2}(f,g) \;:= d_{\|\cdot\|_{L^2}(f,g)} \;=\; \|g - f\|_{L^2} \;=\; \left(\int_a^b (g(x) - f(x)^2 dx\right)^{1/2}, \tag{12.11}$$

are often better suitable than the distance derived from the sup–norm to measure the distance of two functions. One of the drawbacks from an instructor's perspective is that there is no picture like figure 12.2 to visualize the set of all functions with an L^1–distance or L^2–distance from a given function.

12.3 Neighborhoods and Open Sets

(A) Given a point $x_0 \in \mathbb{R}$ (a real number) and $\varepsilon > 0$, we can look at

$$\begin{aligned} N_\varepsilon(x_0) = (x_0 - \varepsilon, x_0 + \varepsilon) \;&= \{x \in \mathbb{R} : x_0 - \varepsilon < x < x_0 + \varepsilon\} \\ &= \{x \in \mathbb{R} : d(x, x_0) = |x - x_0| < \varepsilon\} \end{aligned} \tag{12.12}$$

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 (12.1) on p.361).

(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

$$\begin{aligned} N_\varepsilon(\vec{x}_0) \;&= \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\} \\ &= \{\vec{x} \in \mathbb{R}^2 : d_{\|\cdot\|_2}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{x}_0\|_2 < \varepsilon\} \end{aligned} \tag{12.13}$$

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

$$\begin{aligned} N_\varepsilon(\vec{x}_0) &= \{(x,y,z) \in \mathbb{R}^3 : (\vec{x} - \vec{x}_0)^2 + (\vec{y} - \vec{y}_0)^2 + (\vec{z} - \vec{z}_0)^2 < \varepsilon^2\} \\ &= \{\vec{x} \in \mathbb{R}^3 : d_{\|\cdot\|_2}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{x}_0\|_2 < \varepsilon\} \end{aligned} \tag{12.14}$$

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 $(V, \|\cdot\|)$ and a vector $x_0 \in V$, we can look at

$$N_\varepsilon(x_0) \;=\; \{x \in V : \|x - x_0\| < \varepsilon\} \tag{12.15}$$

which is the set of all vectors in V 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).

(E) Given a bounded real–valued function $f \in \mathscr{B}(X, \mathbb{R})$, we can look at the sets $N_\varepsilon(f)$ $(\varepsilon > 0)$ defined in (12.7) on p.364, i.e., the set of all functions $g(\cdot)$ with distance less than ε from $f(\cdot)$.

(F) Given is a closed interval $[a, b]$ $(a, b \in \mathbb{R})$. For a continuous (hence bounded) real–valued function $f \in \mathscr{B}([a, b], \mathbb{R})$, we can look at the sets

$$N_\varepsilon(f) \;=\; \{g \in \mathscr{B}([a, b], \mathbb{R}) : \|g - f\|_{L^2} < \varepsilon\}, \tag{12.16}$$

i.e., the set of all functions $g(\cdot)$ such that $\sqrt{\int_a^b \big(g(x) - f(x)\big)^2 dx} < \varepsilon$ (see Definition 11.21 on p.349)

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 12.6 (ε-Neighborhood)**.**

Given a metric space (X, d), $x_0 \in X$ and $\varepsilon > 0$, let

$$N_\varepsilon(x_0) \;=\; \{x \in X : d(x, x_0) < \varepsilon\} \tag{12.17}$$

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. 12.4 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 12.7 (Interior points in metric spaces)**.**

Given is a metric space (X, d).
An element $a \in A \subseteq X$ is called an **inner point** or **interior point** of A if we can find some $\varepsilon > 0$ (no mattter how small), so that $N_\varepsilon(a) \subseteq A$. □

Definition 12.8 (Open sets in metric spaces).

Given is a metric space (X, d).
A set all of whose members are interior points is called an **open set**. □

Proposition 12.3.

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

$$d(z, x) \ \leq \ d(z, y) + d(y, x) \ < \delta + \varepsilon.$$

In other words, each element z of $N_\delta(y)$ is $\delta + \varepsilon$–close to x. Hence $N_\delta(y) \subseteq N_{\delta+\varepsilon}(x)$. ■

Proposition 12.4. *$N_\varepsilon(x_0)$ is an open set*

PROOF: It is worth while to examine this proof [137] 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,

$$\varepsilon - d(a, x_0) \ = 2\delta \tag{12.18}$$

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?

$$d(b, x_0) \leq d(b, a) + d(a, x_0) < \delta + (\varepsilon - 2\delta) = \varepsilon - \delta < \varepsilon$$

In the above chain, the first inequality is a consequence of the triangle inequality. The second one reflects the fact that $b \in N_\delta(a)$ and uses (12.18).

We have proved that for any $b \in N_\delta(a)$ it is true that $b \in N_\varepsilon(x_0)$ hence $N_\delta(a) \subseteq N_\varepsilon(x_0)$.

This proves that a is an interior point of $N_\varepsilon(x_0)$. But a is an arbitrary point in $N_\varepsilon(x_0)$. It follows that $N_\varepsilon(x_0)$ is open. ■

Proposition 12.5 (Open intervals are open in $(\mathbb{R}, d_{|\cdot|})$).

Let $a, b \in \mathbb{R}$ such that $a < b$. Then the open interval $]a, b[$ is an open set in $(\mathbb{R}, d_{|\cdot|})$.

PROOF: The proof is left as exercise 12.3 on p.398. ■

Definition 12.9 (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. □

[137] A shorter proof can be given if the previous proposition is used.

Remark 12.6.

(a) You will see very often 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 $x \in X$, one can find an open neighborhood U_x of x such that $U_x \subseteq A_x$. Thus, very often **the open neighborhoods are the important ones**. Accordingly, there are many theorems where it is assumed that some given neighborhood is open.

(b) The empty set is not a neighborhood of any $x \in X$, since the condition $x \in \emptyset$ is never satisfied. □

Proposition 12.6 (Metric Spaces are Hausdorff Spaces)**.**

Let (X, d) be a metric space and let x, y be two different elements of X. Then there exist neighborhoods N_x of x and N_y of y such that $N_x \cap N_y = \emptyset$.

PROOF:

Let $\varepsilon := \frac{1}{2}d(x,y)$ and let $x' \in N_\varepsilon(x)$. We must show that $x' \notin N_\varepsilon(y)$, i.e., $d(x',y) \geq \varepsilon$. Assume to the contrary that $d(x',y) < \varepsilon$. It follows from $x' \in N_\varepsilon(x)$ that $d(x,x') < \varepsilon$. Thus

$$d(x,y) \leq d(x,x') + d(x',y) < \varepsilon + \varepsilon = d(x,y).$$

We have reached a contradiction. ■

Remark 12.7. ★ We will encounter in Chapter.12.5 objects more general than metric spaces, the so called topological spaces, which allow to define neighborhoods of a point. There are such spaces for which this proposition is not true. See Remark 12.13 on p.374 about the "indiscrete topology".

A topological space with the property that any two of its elements can be "separated" by disjoint neighborhoods is called a **Hausdorff space** [138] or also a **T2 space**. □

Theorem 12.2 (Metric spaces are topological spaces)**.**

The following is true about open sets of a metric space (X, d):

(12.19a) *An arbitrary union* $\bigcup_{i \in I} U_i$ *of open sets U_i is open.*

(12.19b) *A finite intersection* $U_1 \cap U_2 \cap \ldots \cap U_n$ $(n \in \mathbb{N})$ *of open sets is open.*

(12.19c) *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.

[138] so named after the German mathematician Felix Hausdorff

PROOF of b: Let $x \in U := U_1 \cap U_2 \cap \ldots \cap U_n$. Then $x \in U_j$ for all $1 \leq j \leq n$ 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 [139]

$$N_\varepsilon(x) \subseteq N_{\varepsilon_j}(x) \subseteq U_j \ (1 \leq j \leq n), \quad \text{hence} \quad N_\varepsilon(x) \subseteq \bigcap_{j=1}^{n} U_j.$$

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.4 (Logic). ■

This last theorem provides the underpinnings for the definition of abstract topological spaces, a subject which will be touched upon in ch.12.5 on p.372.

12.4 Convergence

You have already encountered the precise definition of the convergence of sequences of real numbers in ch.9.3. It is only a small step to generalize this concept to all metric spaces and therefore also to all normed vector spaces.

Definition 12.10 (Convergence of Sequences in Metric Spaces)**.**

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 the x_n will eventually come arbitrarily close to a in the following sense:

Let δ be a (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 $j \geq n_0$.

This can also be expressed as follows:

(12.20) For all $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that $d(a, x_j) < \delta$ for all $j \geq n_0$.

Here is an yet another way of expressing convergence of $(x_n)_n$ to a:

- No matter how small a neighborhood of a is given, all members x_n will eventually be inside that neighborhood.

We write either of

(12.21) $$a = \lim_{n\to\infty} x_n \qquad \text{or} \qquad x_n \to a$$

and we call a the **limit** of the sequence (x_n) □

[139] 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.

Theorem 12.3 (Limits in metric spaces are uniquely determined)**.**

> *Let (X, d) be a metric space and 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

$$d(x_n, L_1) < \varepsilon \, \forall n \geq N_1 \text{ and } d(x_n, L_2) < \varepsilon \, \forall n \geq N_2.$$

It follows that, for $n \geq \max(N_1, N_2)$, [140]

$$d(L_1, L_2) \;\leq\; d(L_1, x_n) + d(x_n, L_2) \;<\; 2\varepsilon \;=\; d(L_1, L_2)$$

and we have reached a contradiction. ■

Proposition 12.7.

> *Let (X, d) be a metric space and $L, x_n \in X$ $(n \in \mathbb{N})$. Let $\delta_n \in \mathbb{R}_{>0}$ such that $\delta_n \to 0$ as $n \to \infty$. Assume further that $x_n \in N_{\delta_n}(L)$ for all $n \in \mathbb{N}$. Then $\lim\limits_{n\to\infty} x_n \;=\; L$.*

PROOF:

Let $\varepsilon > 0$. It follows from $\lim\limits_{k\to\infty} \delta_k = 0$ that there exists $n_0 \in \mathbb{N}$ such that $\delta_k < \varepsilon$ for all $k \geq n_0$. Because $x_k \in N_{\delta_k}(L)$ implies $d(x_k, L) < \delta_k$, we conclude that $d(x_k, L) \; < \varepsilon$ for all $k \geq n_0$, and hence that $\lim\limits_{k\to\infty} x_n = L$. ■

For the special case $\delta_n = \frac{1}{n}$ we obtain

Corollary 12.1.

> *Let (X, d) be a metric space and $L, x_n \in X$ $(n \in \mathbb{N})$ such that $d(x_n, L) \leq \frac{1}{n}$ for all $n \in \mathbb{N}$. Then $\lim\limits_{n\to\infty} x_n \;=\; L$.*

PROOF: Obvious from prop.12.7. ■

We proved for constant sequences of real numbers that they are convergent. This is also true for sequences with values in metric spaces.

Proposition 12.8.

> *Let (X, d) be a metric space, $L \in X$ and $x_n = L$ for all $n \in \mathbb{N}$. Then $\lim\limits_{n\to\infty} x_n \;=\; L$.*

PROOF:

This follows from cor.12.1 above since $x_n = L \;\Rightarrow\; d(x_n, L) = 0 \;\Rightarrow\; d(x_n, L) \leq \frac{1}{n}$ for all $n \in \mathbb{N}$ but we should be able to prove this directly from the definition of convergence. This is how:

[140] You could have used $N_1 + N_2$ instead. Do you see why?

Let $\delta > 0$ and $n_0 = 1$. Then $d(x_n, L) = 0 < \delta$ for all $n \geq 1$, i.e., (12.10) on p.369 is satisfied. ■

The following proposition shows that the limit behavior of a sequence is a property of its tail, i.e., it does not depend on the first finitely many indices.

Proposition 12.9. [141]

Let x_n, y_n be two sequences in a metric space (X,d). Assume there is $K \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq K$. Let $L \in X$ Then

$$\lim_{n\to\infty} x_n = L \Leftrightarrow \lim_{n\to\infty} y_n = L.$$

PROOF: The proof is left as exercise 12.9 on p.399. ■

Compare the next proposition to Proposition 9.13 on p.265.

Proposition 12.10 (Subsequences of sequences with limits)**.**

Let $(x_n)_n$ be a sequence in a metric space (X,d) with limit $L := \lim\limits_{n\to\infty} x_n$. Then it is true for any subsequence $(x_{n_j})_j$, that $\lim\limits_{j\to\infty} x_{n_j} = L$.

PROOF: Let $\varepsilon > 0$. Because the sequence converges, there exists $N \in \mathbb{N}$ such that $d(x_j, L) < \varepsilon$ for all $j \geq N$. As $n_j \geq j$ for all j, we conclude that $n_j \geq N$ whenever $j \geq N$, hence $|x_{n_j} - L| < \varepsilon$ for all $j \geq N$. It follows that (x_{n_j}) has limit L. ■

The following is a simple corollary.

Proposition 12.11. [142]

Let x_n be a convergent sequence in a metric space (X,d) with limit $L \in X$. Let $K \in \mathbb{N}$. For $n \in \mathbb{N}$ let $y_n := x_{n+K}$. Then $\lim\limits_{n\to\infty} (y_n)_n = L$.

PROOF: The proof is left as exercise 12.10 on p.399. ■

Remark 12.8.

The following allows us to prove convergence of x_n to $L \in (X,d)$ by utilizing what we know about convergence in $(\mathbb{R}, d_| \cdot |)$.

$$\lim_{n\to\infty} x_n \ = \ L \quad \Leftrightarrow \quad \lim_{n\to\infty} d(x_n, L) \ = \ 0\,. \ \square$$

Remark 12.9 (Opposite of convergence)**.** 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.

[141] See cor.9.4 on p.270.

[142] See prop.9.14 on p.266.

The correct answer is as follows:

$\big[\ \lim_{k\to\infty} x_k\ =\ L$ is **NOT true** $\big]\ \Leftrightarrow$
$\big[$ 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 $d(x_j, L) \geq \varepsilon \big]$. □

It is easy to prove from the Remark 12.9 (Opposite of convergence) the following:

Proposition 12.12 (Opposite of convergence)**.**

*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 $n_1 < n_2 < n_3 < \dots \in \mathbb{N}$ 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: The proof is left as exercise 13.1. ■

12.5 Abstract Topological spaces

Theorem 12.2 on p.368 gives us a way of defining neighborhoods for sets which do not have a metric.

Definition 12.11 (Abstract topological spaces)**.** Let X be an arbitrary nonempty set and let $\mathfrak{U}$ be a set of subsets [143] of X whose members satisfy the properties a, b and c of (12.19) on p.368:

(12.22a) An arbitrary union $\bigcup\limits_{i\in I} U_i$ of sets $U_i \in \mathfrak{U}$ belongs to $\mathfrak{U}$,

(12.22b) $U_1, U_2, \dots, U_n \in \mathfrak{U}\ (n \in \mathbb{N}) \quad \Rightarrow \quad U_1 \cap U_2 \cap \ldots \cap U_n\ \in \mathfrak{U}$,

(12.22c) $X \in \mathfrak{U}$ and $\emptyset \in \mathfrak{U}$.

Then $(X, \mathfrak{U})$ is called a **topological space** The members of $\mathfrak{U}$ are called **open sets** of $(X, \mathfrak{U})$. The collection $\mathfrak{U}$ of open sets is called the **topology** of X. □

Remark 12.10. Let (X, d) be a metric space and let

(12.23) $$\mathfrak{U}_d\ :=\ \{U \subseteq X :\ U \text{ is an open subsets of } (X, d)\},$$

i.e., $U \in \mathfrak{U}_d\ \Leftrightarrow\ U$ 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 (12.7) on p.366). Then thm.12.2 on p.368 asserts the following.

Every metric space (X, d) is a topological space in the following sense: If $\mathfrak{U}_d$ denotes the open sets of (X, d) then $(X, \mathfrak{U}_d)$ is a topological space.

[143] We encountered subsets of 2^X with special properties previously when looking at rings of sets in Definition 8.4 (Rings, algebras, and σ–algebras of Sets) on p.235.

Remark 12.11. Let V be a vector space with a norm $\|\cdot\|$. We recall that this norm defines a metric $d_{\|\cdot\|}(\cdot,\cdot)$ via $d_{\|\cdot\|}(x,y) = \|x-y\|$ (see thm.12.1 on p.361). According to part **A** the norm $\|\cdot\|$ defines open sets

$$\mathfrak{U}_{\|\cdot\|} \;:=\; \mathfrak{U}_{d_{\|\cdot\|}} \tag{12.24}$$

in the metric space $(V, d_{\|\cdot\|})$.

Every normed vector space $(V, \|\cdot\|)$ is a topological space in the sense that If $\mathfrak{U}_d$ denotes the open subsets of a metric space (X,d) then $(V, \mathfrak{U}_d)$ is a topological space. □

We now discuss the terminology for topologies that are the open sets of metric spaces and, in particular, normed vector spaces.

Definition 12.12 (Metric Topology and Norm Topology). ⋆

(a) Let (X,d) be a metric space and let $\mathfrak{U}_d$ be as defined in (12.23). We say that $\mathfrak{U}_d$ is **induced by the metric** $d(\cdot,\cdot)$ or that it is **generated by the metric** $d(\cdot,\cdot)$. or that it is the **metric topology** of X. If it is clear which metric d on X we mean then we also simply refer to "the" metric topology.

(b) Let $(V, \|\cdot\|)$ be a normed vector space, and let $\mathfrak{U}_{\|\cdot\|}$ be as defined in (12.24), i.e., $\mathfrak{U}_{\|\cdot\|}$ is the topology defined by the metric $d_{\|\cdot\|}$. We say that this topology is **induced by the norm** $\|\cdot\|$ or that it is **generated by the norm** $\|\cdot\|$. If it is clear which norm on V we are studying then we call the topology associated with this norm the **norm topology** of V. □

We defined the discrete metric earlier in this chapter. (See Definition 12.3 on p.362.)

Definition 12.13 (Discrete topology). ⋆

Let X be a nonempty set with the discrete metric

$$d(x,y) \;=\; \begin{cases} 0 & \text{for } x = y, \\ 1 & \text{for } x \neq y. \end{cases}$$

We call the topology associated with the discrete metric the **discrete topology** of X. □

Proposition 12.13.

Let (X,d) be a metric space with the discrete metric. Then its associated topology is

$$\mathfrak{U}_d \;=\; 2^X \;=\; \{A : A \subseteq X\}.$$

PROOF: The proof is left as exercise 12.12. ■

Remark 12.12. It follows from prop.12.13 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 12.11 on p.372. □

We now discuss the example of a topology which is not generated by a metric.

Proposition 12.14.

Let X be an arbitrary nonempty set and let $\mathfrak{U} := \{\emptyset, X\}$. Then $(X, \mathfrak{U})$ is a topological space.

PROOF:

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). Moreover, any union of members of $\mathfrak{U}$ is either $\emptyset$ (if all members are $\emptyset$) or X (if at least one member is X). ■

Definition 12.14 (Indiscrete topology)**.**

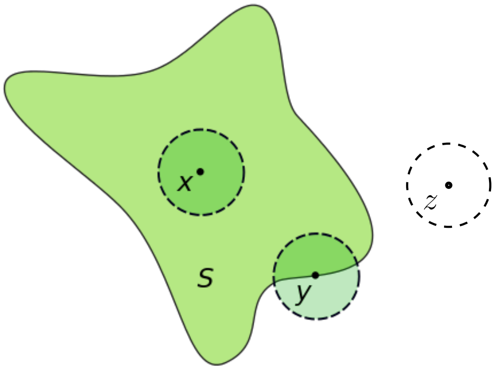
Let X be a nonempty set. The topology $\{\emptyset, X\}$ is called the **indiscrete topology** of X. □

Remark 12.13.

(a) It follows from prop.12.14 that the indiscrete topology is the smallest possible topology on X. i.e., the smallest possible collection of subsets of X whose members satisfy properties a, b, c of definition 12.11 on p.372.

(b) Prop.12.6 (Metric Spaces are Hausdorff Spaces) on p.368 guarantees that any two different points x and y in a metric space can be separated by appropriately chosen disjoint neighborhoods. This is not true for the indiscrete topology since the only superset of a nonempty open set is X, so the only neighborhood for x is X, and the same is true for y. □

Remark 12.14.



The picture to the right [144] demonstrates that there are exactly three mutually exclusive choices how a point in $(X, \mathfrak{U})$ is related to a subset S of X:

(a) either like the point x: There exists an open set U such that $x \in U \subseteq S$,

(b) or like the point z: There exists an open set U such that $z \in U \subseteq S^{\complement}$,

(c) or like the point y: There is no open set U such that $y \in U \subseteq S$ or $y \in U \subseteq S^{\complement}$, i.e., every open set that contains y intersects both S and $S^{\complement}$.

Thus we can classify any element $x \in X$ accordingly: Either x satisfies **(a)** or x satisfies **(b)** or x satisfies **(c)**. This leads to the definitions of interior points, exterior points, and boundary point of $S \subseteq X$.

[144] Source: Wikipedia, https://en.wikipedia.org/wiki/Interior_(topology). The author does not like to use the letter S for subsets of topological spaces, but it came with the picture.

Definition 12.15 (Neighborhoods and interior points in topological spaces)**.** Let $(X, \mathfrak{U})$ be a topological space, $x \in X$ and $S \subseteq X$. It is not assumed that S be open.

(a) S is called a **neighborhood** of x and x is called an **inner point** or **interior point** of S if there exists an open set U such that

$$x \in U \subseteq S.$$

We call the set $S^o := \{$ all interior points of S$\}$ the **interior** of S. An alternate but less commonly used notation for S^o is int(S).

(b) x is called an **exterior point** of S if x is an inner point of $S^\complement$, i.e., there exists an open set U' such that

$$x \in U' \subseteq S^\complement,$$

We call the set ext$(S) := \{$ all exterior points of S$\}$ the **open exterior** of S.

The expression ″open exterior″ has been adopted from Wikipedia. [145]

(c) x is called a **boundary point** of S if any neighborhood of x intersects both S and $S^\complement$. We call this set the **boundary** of S and denote it ∂S. □

Remark 12.15. Be sure you understand from the definitions of interior points and neighborhood above that the following is true:

If S is an arbitrary subset of X, U is an open subset of X, and $x \in X$, then

(a) x is an interior point of $S \Leftrightarrow S$ is a neighborhood of x.

(b) x is an interior point of $U \Leftrightarrow x \in U$.

(c) If $U \subseteq S$ then all elements of U are interior points of S, i.e., $U \subseteq S^o$.

To see that **(c)** is true observe that any $u \in U$ satisfies

$$u \in \text{open set } U \subseteq S\,.$$

In other words, u, U, and S satisfy the relationship (**??**) of the definition of interior points. □

Remark 12.16. For metric spaces (X, d) we first defined interior points, and afterward we defined an open subset as one which consists entirely of interior points. [146] Since openness is the defining property of topological spaces and thus at the very beginning it should not come as a surprise that for such spaces we had to proceed in reverse and define interior points and neighborhoods in terms of open sets. □

Proposition 12.15. *Let $(X, \mathfrak{U})$ be a topological space and let $A \subseteq X$. Then*

$$A^o = \bigcup \Big[U \in \mathfrak{U} : U \subseteq A \Big]\,. \tag{12.25}$$

In other words, the interior of A is the union of all open subsets of A.

[145] Source: https://en.wikipedia.org/wiki/Interior_(topology)

[146] See Definition 12.7 on p.366.

PROOF: For convenience we abbreviate $B := \bigcup \left[U \in \mathfrak{U} : U \subseteq A \right]$.

We first prove that A^o is an open set, i.e., for each $x \in A^o$ there is an open set U_x such that

$$\textbf{(a)} \qquad x \in U_x \subseteq A^o .$$

By definition of A^o, x is interior to A, thus there exists $U \in \mathfrak{U}$ such that $x \in U \subseteq A$. It follows from Remark 12.15**(c)** that $U \subseteq A_o$, thus $U_x := U$ satisfies **(a)**.

Next we show that $A^o \subseteq B$. Since A^o is open, $A^o \in \{U \in \mathfrak{U} : U \subseteq A\}$. But then,

$$A^o \subseteq \bigcup \left[U \in \mathfrak{U} : U \subseteq A \right], \quad \text{i.e.,} \quad A^o \subseteq B .$$

We finally prove that $B \subseteq A^o$. So let $x \in B$. We must show that $x \in A^o$.

By definition of B there exists $U \in \mathfrak{U}$ such that $U \subseteq A$ and $x \in U$. Again we conclude from Remark 12.15**(c)** that $U \subseteq A^o$, hence $x \in A^o$. ■

That last proposition shows that A^o is an open set which is, as a union of subsets of A, also a subset of A. Because A^o is the union of all such sets, we conclude that

The interior A^o of A is the largest of all open subsets of A.

Proposition 12.16.

Let $(X, \mathfrak{U})$ be a topological space.

If $A \subseteq B \subseteq X$ then $A^o \subseteq B^o$. □

PROOF: The proof is left as exercise 12.19. ■

The following proposition is worth while remembering: If we fix a subset A of a topological space X then each point $x \in X$ belongs either to the interior or the open exterior or the boundary of A.

Proposition 12.17.

Let $(X, \mathfrak{U})$ be a topological space and let $A \subseteq X$. Then,

$$X = A^o \biguplus ext(A) \biguplus \partial(A) . \tag{12.26}$$

Thus, X is partitioned into the interior, open exterior and boundary of anyone of its subsets.

PROOF: Obvious from the fact that any $x \in X$ falls into exactly one of the following categories:

- **(a)** either there exists an open set U such that $x \in U \subseteq S$, i.e., $x \in A^o$,
- **(b)** or there exists an open set U such that $z \in U \subseteq S^{\complement}$, i.e., $x \in$ ext(A),
- **(c)** or there is no open set U such that $x \in U \subseteq S$ or $x \in U \subseteq S^{\complement}$, i.e., every open set that contains x intersects both S and $S^{\complement}$, i.e., i.e., $x \in \partial(A)$.

See rem.12.14 on p.374. ■

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 12.17 (Hierarchy of topological spaces)**.** We have seen the following:

(a) $\mathbb{R}^n$, in particular $\mathbb{R} = \mathbb{R}^1$, is an inner product space (see prop.11.11 on p.342).
(b) All inner product spaces are normed spaces (see thm.11.3 on p.346).
(c) All normed spaces are metric spaces (see thm.12.1 on p.361).
(d) All metric spaces are topological spaces (see Definition 12.11 on p.372, Definition 12.12 on p.373).

12.6 Bases and Neighborhood Bases ⋆

This chapter has been marked as optional but we suggest that you skim its contents since some of the concepts taught here will be referred to in subsequent chapters.

Definition 12.16 (Base of the topology)**.**

Let $(X, \mathfrak{U})$ be a topological space. A subset $\mathfrak{B}$ of $\mathfrak{U}$ 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 $\mathfrak{B}$:

(12.27) $$U = \bigcup_{i \in I} B_i \quad (B_i \in \mathfrak{B} \text{ for all } i \in I)$$

where I is a suitable index set, which of course will in general depend on U. □

We note that, because X itself is open, (12.27) implies that $X = \bigcup [B : B \in \mathfrak{B}]$.

A base of the topology is a subset of that topology, i.e., a collection of open sets, which contains enough small open sets. We can localize that definition to a point x of X by looking at collections of open neighborhoods of x which contain enough small open neighborhoods of x and we arrive at the definition of neighborhood bases of x.

Definition 12.17 (Neighborhood base of a point)**.**

Let $(X, \mathfrak{U})$ be a topological space.

(a) The following set of subsets of X,

(12.28) $$\mathfrak{N}(x) := \{A \subseteq X : A \text{ is a neighborhood of } x\},$$

is called the **neighborhood system of** x

(b) 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 $A \in \mathfrak{N}(x)$, there exists a set $B \in \mathfrak{B}(x)$ 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 base has “enough” of those.

Definition 12.18 (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. □

Here are some propositions about bases, neighborhood bases, and first countability for metric spaces.

Proposition 12.18 (ε-neighborhoods are a base of the topology)**.**

Let (X, d) be a metric space. Then both

$$\mathscr{B}_1 \ := \ \{N_\varepsilon(x) : x \in X, \varepsilon > 0\} \qquad \textit{and} \qquad \mathscr{B}_2 \ := \ \{N_{1/n}(x) : x \in X, n \in \mathbb{N}\}$$

are bases for the topology of (X, d) (see 12.16 on p.377)

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 $1/n(x) \leq \varepsilon(x)$.

We observe that $U \ \subseteq \ \bigcup \left[N_{1/n(x)}(x) : x \in U \right] \subseteq U$.

The first inclusion follows from $\{x\} \subseteq N_{1/n(x)}(x)$ for all $x \in U$ and the second inclusion follows from $N_{1/n(x)}(x) \subseteq N_{\varepsilon(x)}(x) \subseteq U$ and the inclusion lemma (lemma 8.1 on p.232).

We obtain $U \ = \ \bigcup \left[N_{1/n(x)}(x) : x \in U \right]$ 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 12.4 (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

$$\mathfrak{B}(x) \ := \ \{ \ N_{1/n}(x) : n \in \mathbb{N} \ \}. \tag{12.29}$$

Then $\mathfrak{B}(x)$ is a neighborhood base of x because, by Definition 12.7 and Definition 12.9 (Interior points and neighborhoods in metric spaces), any neighborhood of x will contain one of the form $N_\varepsilon(x)$ and for any such $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$. ■

Proposition 12.19.

Let (X, d) be a metric space and let $\mathfrak{B} := \{N_{1/k}(x) : x \in X, k \in \mathbb{N}\}$. Then $\mathfrak{B}$ is a base of the topology for the associated topological space $(X, \mathfrak{U}_d)$.

PROOF: The proof is left as exercise 12.15 on p.400. ■

Definition 12.19 (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 $\mathfrak{U}$. □

The next theorem is related to the material in chapter 9.10 (Sequences that Enumerate Parts of $\mathbb{Q}$).

Theorem 12.5 (Euclidean space $\mathbb{R}^n$ is second countable)**.**

Let $\mathfrak{B}$ be the following collection of open subsets of $\mathbb{R}^n$:

$$\mathfrak{B} := \{\, N_{1/j}(\vec{q}) : \vec{q} \in \mathbb{Q}^n,\ j \in \mathbb{N}\,\}. \tag{12.30}$$

Here,

$$\mathbb{Q}^n = \{\vec{q} = (q_1, \dots, q_n) : q_j \in \mathbb{Q},\ 1 \le j \le n\}$$

is the set of all points in $\mathbb{R}^n$ with rational coordinates. Then $\mathfrak{B}$ is a countable base of $\mathbb{R}^n$.

PROOF (outline): Let $U \in \mathfrak{U}$ be an arbitrary open set in X. Any vector $\vec{x} \in U$ is interior point of U, hence we can find some $n_{\vec{x}} \in \mathbb{N}$ such that the entire $\frac{2}{n_{\vec{x}}}$–neighborhood $N_{2/(n_{\vec{x}})}(\vec{x})$ is contained within U.

As any vector can be approximated by vectors with rational coordinates, there exists $\vec{q} = \vec{q}_{\vec{x}} \in \mathbb{Q}^n$ such that $d(\vec{x}, \vec{q}_{\vec{x}}) < \frac{1}{n_{\vec{x}}}$, hence $\vec{x} \in N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}})$.

It follows from $N_{2/(n_{\vec{x}})}(\vec{x}) \subseteq U$ and prop.12.3 on p.367, applied to $\delta = \varepsilon = \frac{1}{n_{\vec{x}}}$, that

$$N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}}) \subseteq N_{2/n_{\vec{x}}}(\vec{x}) \subseteq U \text{ for all } \vec{x} \in U.$$
$$\text{Hence } U = \bigcup_{\vec{x} \in U} \{\vec{x}\} \subseteq \bigcup \left[\, N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}}) : \vec{x} \in U \,\right] \subseteq U.$$

We have managed to write the arbitrarily chosen open set U as a union of the sets $N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}})$ which belong to $\mathfrak{B}$. This proves that $\mathfrak{B}$ is a basis of the topology.

We recall from cor.7.6 on p. 229 that $\mathbb{Q}^n$is countable. For $j \in \mathbb{N}$ let $\mathfrak{B}_j := \{\, N_{1/j}(\vec{q}) : \vec{q} \in \mathbb{Q}^n\}$. Then each $\mathfrak{B}_j$ is countable because $\vec{q} \mapsto N_{1/j}(\vec{q})$ is a surjection from the countable set $\mathbb{Q}^n$ onto $\mathfrak{B}_j$. It follows that the base of the topology $\mathfrak{B} = \bigcup_j \mathfrak{B}_{j \in \mathbb{N}}$ is countable as the countable union of countable sets. ■

12.7 Metric and Topological Subspaces

It is often advantageous to focus our attention on a subset A of a metric space (X, d) or a topological space $(X, \mathfrak{U})$. It would be nice if one could find a way to define a metric d' (a topology $\mathfrak{U}'$) on A which coexists harmoniously with the metric d (the topology $\mathfrak{U}$) defined on X.

For example let X be the real numbers with the standard metric $d(x, y) = |b - a|$ and $A = [0, 1]$. This allows us, e.g., to talk of the assignment $x \mapsto \sqrt{x}$ which cannot be extended beyond A as a function $f : (A, d') \to (\mathbb{R}, d)$ for which both domain and codomain are metric spaces.

The solution to this problem is different for metric spaces and topological spaces, but both amount to the following:

A set U will be open in A if and only if $U = V \cap A$ for some suitable set V which is open in X.

Definition 12.20 (Metric subspaces)**.**

Given is a metric space (X, d) and a nonempty $A \subseteq (X, d)$. Let

$$d\big|_{A\times A} : A \times A \to \mathbb{R}_{\geq 0} \quad \text{be the restriction} \quad d\big|_{A\times A}(x, y) \ := \ d(x, y)(x, y \in A)$$

of the metric d to $A \times A$ (see Definition 5.15 on p.151).
It is trivial to verify that $(A, d\big|_{A\times A})$ is a metric space in the sense of Definition 12.1 on p.359.
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 12.18.

⚠ Metric subspaces come with their own collections of open and closed sets, neighborhoods, ε-neighborhoods, convergent sequences, ...
Watch out when looking at statements and their proofs whether those concepts refer to the entire space (X, d) or to the subspace $(A, d\big|_{AxA})$. □

Notation 12.1.

a) Because the only difference between d and $d_{A\times A}$ is the domain, it is customary to write d instead of $d\big|_{A\times A}$ to make formulas look simpler, if doing so does not give rise to confusion.

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", ... □

Example 12.5. Consider $A :=]0, 1] \cup]2, 3[\cup \{4\} \cup \{5\} \ \cup \ \{6\}$ as a subset of $(\mathbb{R}, d_{|.|})$, i.e., the real numbers with the Euclidean metric. Then $\{4, 5\}$ is OPEN in A and $\{1\}$ is interior to A. □

Definition 12.21 (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). 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 define

$$N_\varepsilon^A(a) \ = \ N_\varepsilon(a) \cap A. \tag{12.31}$$

i.e., $N_\varepsilon^A(a)$ is defined as the trace of $N_\varepsilon(a)$ in A. □

Proposition 12.20 (Open sets in metric subspaces are 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*

$$N_\varepsilon^A(a) \;=\; N_\varepsilon(a) \cap A \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \varepsilon\}\,, \tag{12.32}$$

i.e., $N_\varepsilon^A(a)$ is the "ordinary" ε–Neighborhood of a in the metric space $(A, d\big|_{A\times A})$ (as it was originally defined in Definition 12.6 on p.366). It thus follows from (12.31) that each ε-neighborhood in the subspace A is the trace of an ε-neighborhood in X.

(b) *Generalization: $U \subseteq A$ is open in $(A, d\big|_{A\times A}) \Leftrightarrow$ there is an open $V \subseteq (X, d)$ such that*

$$U \;=\; V \cap A\,. \tag{12.33}$$

In other words, U is the trace of a set V which is open in X.

PROOF of **(a)**: First we prove (12.32). As $d\big|_{A\times A}$ is the restriction of d to $A \times A$ it follows that

$$\begin{aligned} N_\varepsilon^A(a) \;&= N_\varepsilon(a) \cap A \;=\; \{x \in X : d(x,a) < \varepsilon\} \cap A \\ &= \{x \in A : d(x,a) < \varepsilon\} \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \varepsilon\}. \end{aligned}$$

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 $a \in U$. We must prove that a is an interior point of U with respect to $(A, d\big|_{A\times A})$.
Because $a \in V$ and V is open in X, there is $\varepsilon > 0$ such that $N_\varepsilon(a) \subseteq V$. It follows that $N_\varepsilon^A(a) = N_\varepsilon(a) \cap A \subseteq V \cap A = U$. As $N_\varepsilon^A(a)$ is open in A, a is an interior point of U with respect to the subspace $(A, d\big|_{A\times A})$.
Finally we prove that if $U \subseteq A$ is open in A then there is $V \subseteq X$ open in X such that $U = V \cap A$:
We can write $U \;=\; \bigcup \big[N_{\varepsilon(a)}^A(a) : a \in U\big]$ for suitable $\varepsilon(a) > 0$ (see the proof of prop.12.18 on p.378).
Let $V \;:=\; \bigcup \big[N_{\varepsilon(a)}(a) : a \in U\big]$. V is open in (X, d) as union of the open sets $N_{\varepsilon(a)}(a)$. Further,

$$\begin{aligned} V \cap A \;&= A \cap \bigcup \big[N_{\varepsilon(x)}(x) : x \in U\big] \;=\; \bigcup \big[N_{\varepsilon(x)}(x) \cap A : x \in U\big] \\ &= \bigcup \big[N_{\varepsilon(x)}^A(x) : x \in U\big] \;= U \end{aligned}$$

(the second equality follows from prop.8.1 on p.234). This finishes the proof. ■

Remark 12.19 (Convergence does not necessarily extend to metric subspaces)**.**

Let (X, d) be a metric space, $A \subseteq (X, d)$ and $a_n \in A$ for all $n \in \mathbb{N}$. Be aware that convergence of the sequence (a_n) in the space (X, d) (i.e., there exists $x \in X$ such that $x = \lim\limits_{n\to\infty} a_n$) does **NOT** imply convergence of the sequence in the subspace $(A, d\big|_{A\times A})$! Rather, we have the following dichotomy:

(a) $x \in A$: Then a_n converges to x in the subspace $(A, d\big|_{A\times A})$ (and also in (X, d)).

(b) $x \in A^\complement$: Then a_n converges to x in (X, d) but not in $(A, d\big|_{A\times A})$. □

Proposition 12.20 (Open sets in metric subspaces are traces of open sets in X) justifies to define subspaces of abstract topological spaces as follows.

Definition 12.22 (Topological subspaces). ⋆

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 $\mathfrak{U}_A$. In other words,

$$\mathfrak{U}_A = \{V \cap A : V \in \mathfrak{U}\}.$$

We call $(A, \mathfrak{U}_A)$ a **topological subspace** or also just a **subspace** of $(X, \mathfrak{U})$ and we call $\mathfrak{U}_A$ the **subspace topology induced by** $(X, \mathfrak{U})$ or the **subspace topology inherited from** $(X, \mathfrak{U})$. □

Proposition 12.21 (Topological subspaces are topological spaces).

Let $(X, \mathfrak{U})$ be a topological space, $A \subseteq X$, and let $\mathfrak{U}_A$ be the collection of all open sets in A. Then $(A, \mathfrak{U}_A)$ is a topological space, i.e., it satisfies Definition 12.11 on p.372 of an abstract topological space.

PROOF:

(a) Let $(U_i)_{i \in I}$ be a family of open sets in A. For each U_i there exists V_i open in X such that $U_i = V_i \cap A$. According to prop.8.1 (Distributivity of unions and intersections) on p.234 we obtain

$$A \cap \bigcup_{i \in I} V_i = \bigcup_{i \in I} (A \cap V_i) = \bigcup_{i \in I} U_i$$

and this proves that $\bigcup_{i \in I} U_i$ is the trace of the open set $\bigcup_{i \in I} V_i$ in A, hence open in A.

(b) Let $U_1, U_2, \dots U_n$ $(n \in \mathbb{N})$ be open in A. For each U_i there exists V_i open in X such that $U_i = V_i \cap A$. Because the intersection of sets is commutative we obtain

$$U_1 \cap \dots U_n = (V_1 \cap A) \cap \dots (V_n \cap A) = (A \cap \dots \cap A) \cap (V_1 \cap \dots \cap V_n) = A \cap (V_1 \cap \dots \cap V_n)$$

and this proves that $U_1 \cap \dots U_n$ is the trace of the open set $V_1 \cap \dots \cap V_n$ in A, hence open in A.

(c) It follows from $\emptyset = \emptyset \cap A$, $A = X \cap A$, and $X, \emptyset \in \mathfrak{U}$, that $\emptyset, A \in \mathfrak{U}_A$. ■

12.8 Contact Points and Closed Sets

If you look at any **closed interval** $[a, b] = \{y \in \mathbb{R} : a \leq y \leq b\}$ of real numbers, then all of its points are interior points, except for the end points a and b. Moreover a and b are contact points according to the following definition which makes sense for any abstract topological space.

Definition 12.23 (Contact points [147]). Given is a 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

(12.34) $A \cap N \neq \emptyset$ for any neighborhood N of x. □

[147] German: Berührungspunkt - see [16] Von Querenburg, p.21

Note 12.1. 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 U_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 neighborhood A_0 of 0 contains for some small enough $\delta > 0$ the entire interval $]-\delta, \delta[$. Let $x := \min(\delta/2, 1/2)$.
Clearly, $x \in]-\delta, \delta[\subseteq A_0$ and $x \in]0, 1[= A$.
It follows that $x \in A \cap A_0$. As A_0 was an arbitrary neighborhood of 0, we have proved that 0 is a contact point of A, even though $0 \notin A$.

(c) The above counterexample can be proven much faster if the criterion for contact points in metric spaces is employed: Let $x_n := 1/n$ $(n \geq 2)$ Then $x_n \in]0, 1[$ for all n and the sequence converges to 0. It follows that 0 is a contact point of $]0, 1[$. □

Definition 12.24 (Closed sets)**.** Let $(X, \mathfrak{U})$ be topological space and $A \subseteq X$. Let the set $\bar{A}$ be

(12.35) $$\bar{A} := \{x \in X : x \text{ is a contact point of } A\}.$$

We call $\bar{A}$ the **closure** of A. A set that contains all its contact points is called a **closed set**. □

Proposition 12.22. *If A is a subset of a topological space then*

(12.36) $$\bar{A} = A \cup \partial(A) = A^o \cup \partial(A).$$

PROOF of $\bar{A} = A^o \cup \partial(A)$:

We recall from prop.12.16 on p.376 that any $x \in X$ either belongs to the interior A^o or to the open exterior ext(A) or to the boundary $\partial(A)$. Since it is precisely the set ext(A) for whose elements one can find neighborhoods of x which do not intersect with A, we obtain

$$x \in \bar{A} \Leftrightarrow x \notin \text{ext}(A) \Leftrightarrow x \in A^o \cup \partial(A).$$

This proves the assertion.

PROOF of $\bar{A} = A \cup \partial(A)$:

It follows from the definitions of A^o and $\bar{A}$ that $A^o \subseteq A \subseteq \bar{A}$, thus

(12.37) $$A^o \cup \partial(A) \subseteq A \cup \partial(A) \subseteq \bar{A} \cup \partial(A).$$

Since $\bar{A} = A^o \cup \partial(A)$ formula (12.37) yields

(12.38) $$\bar{A} = A^o \cup \partial(A) \subseteq A \cup \partial(A) \subseteq \bar{A} \cup \partial(A) = \bar{A}.$$

We obtained the last equation from $\bar{A} = A^o \cup \partial(A)$ since this equation implies

$$\partial(A) \subseteq \bar{A}, \qquad \text{thus} \qquad \partial(A) \cup \bar{A} \subseteq \bar{A} \cup \bar{A} = \bar{A}.$$

Formula (12.38) shows that the set $A \cup \partial(A)$ is both subset and superset of $\bar{A}$. It follows that $A \cup \partial(A) = \bar{A}$. ■

Remark 12.20. It follows from note 12.1**(a)** that $A \subseteq \bar{A}$. ; □

The following theorem shows that we can characterize contact points of subsets of metric spaces by means of sequences.

Theorem 12.6 (Sequence criterion for contact points in metric spaces)**.**

> *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 "⇒": 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 $\frac{1}{N} < \varepsilon$. This is possible because $\mathbb{N}$ is not bounded (above) in $\mathbb{R}$.

For any $j \geq N$ we obtain $d(x_j, x) < 1/j \leq 1/N < \varepsilon$. This proves convergence $x_n \to x$.

PROOF of "⇐" Let $x \in X$ and assume there is $(x_n)_{n \in \mathbb{N}}$ such that $x_n \in A$ for all n 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 $n \geq N$, 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 $U_x \cap A \neq \emptyset$. ■

Proposition 12.23. *The complement of an open set is closed.*

PROOF of 12.23: Let A be an open set in a topological space $(X, \mathfrak{U})$. and assume $x \in X$ is a contact point of $A^\complement$. We want to prove that $A^\complement$ is a closed set, so we must show that $x \in A^\complement$.

We assume to the contrary that x is a contact point of $A^\complement$ such that $x \notin A^\complement$. Then $x \in A$.

A is open, so x is an interior point of A. Hence there is a neighborhood N_x that contains x and is entirely contained in A, hence $N_x \cap A^\complement = \emptyset$.

We also assumed that x is a contact point of $A^\complement$. This implies that $N_x \cap A^\complement \neq \emptyset$. We have reached a contradiction. ■

Proposition 12.24. *The complement of a closed set is open.*

PROOF: We will give two proofs of the above.

(a) First proof of prop.12.24, valid for all topological spaces:

Let A be closed set and $b \in A^\complement$.

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 Definition 12.23 there exists some neighborhood V of b such that $V \cap A = \emptyset$, i.e., $V \subseteq A^\complement$.

We have shown that an arbitrary $b \in A^\complement$ is an interior point of $A^\complement$, i.e., the complement of the closed set A is open. This proves the proposition.

(b) Alternate proof of prop.12.24, valid for metric spaces only, since it works with sequences. We give it to illustrate the use of Theorem 12.6, the sequence criterion for contact points.

Let $A \subset (X,d)$ be closed. If $A^\complement$ is not open then there must some be $b \in A^\complement$ which is not an interior point of $A^\complement$.

We show that this assumption leads to a contradiction. Because b is not an interior point of $A^\complement$, there is no δ-neighborhood, for whatever small δ, that entirely belongs to $A^\complement$. So, for each $j \in \mathbb{N}$, there is an $x_j \in N_{1/j}(b)$ which does not belong to $A^\complement$, i.e., $x_j \in A$.

We have constructed a sequence x_j which is entirely contained in A and which 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 $b \in A^\complement$ and we have a contradiction. ■

Theorem 12.7 (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.12.23 and prop.12.24 ■

Remark 12.21. Many books define 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 is then proven from those definitions that closed sets are exactly those that contain all their contact points. □

Here is an easy consequence of the fact that open sets are the complements of closed sets and vice versa.

Proposition 12.25.

Let $(X,\mathfrak{U})$ be a topological space. The closed sets of X satisfy the following property:

(12.39)
(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.*

The proofs of **(a)** and **(b)** follow easily from De Morgan's law (the duality principle for sets: see (8.1) on p.234). Observe that X plays the role of a universal set because all members U of $\mathfrak{U}$ and their complements $U^\complement$ are subsets of X.

PROOF of **(a)**: 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)**.

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 $C := \bigcup\limits_{1 \leq j \leq n} C_j$. 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

$$X^{\complement} = \emptyset \qquad \text{and} \qquad \emptyset^{\complement} = X. \ \blacksquare$$

We now derive some immediate properties of closures.

Proposition 12.26. *Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq B \subseteq X$. Then $\bar{A} \subseteq \bar{B}$.*

PROOF: The proof is left as exercise 12.18 on p.400. ■

Proposition 12.27.

Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq X$. Then,

$$\partial A = \bar{A} \cap \overline{A^{\complement}}. \tag{12.40}$$

In other words, $x \in X$ is a boundary point of A if and only if x is a contact point of both A and $A^{\complement}$.

PROOF: Left as exercise 12.17 on p.400. ■

Proposition 12.28 (Minimality of the closure of a set)**.**

Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq X$. Then

$$\bar{A} = \bigcap \Big[C \supseteq A : C \text{ is closed} \Big]. \tag{12.41}$$

In other words, the closure $\bar{A}$ of A is the smallest of all closed supersets of A.

PROOF: Let $\mathfrak{C} := \{C \supset A : C \text{ is closed}\}$ and let $F := \bigcap \mathfrak{C}$. We need to show that $\bar{A} = F$.

It follows from prop.12.25**(a)** that F is closed, hence $F = \bar{F}$. It follows from $C \supseteq A$ for all $C \in \mathfrak{C}$ that $F \supseteq A$, hence $F = \bar{F} \supseteq \bar{A}$.

It remains to be shown that $F \subseteq \bar{A}$. It is true that $\bar{A} \in \mathfrak{C}$ because $\bar{A}$ is a closed set which contains A, hence $\bar{A} \supseteq \bigcap \mathfrak{C} = F$. (See prop.12.26 on p.386). ■

Proposition 12.29 (Closure of a set as a hull operator [148])**.**

Let $(X, \mathfrak{U})$ be a topological space. Consider the closure of sets as a function

$$\bar{\ } : 2^X \longrightarrow 2^X; \qquad A \mapsto \bar{A}\,.$$

Then this function has the following properties for all $A, B \subseteq X$:

(a) $\bar{\emptyset} \;=\; \emptyset$, **(b)** $A \subseteq \bar{A}$, **(c)** $\bar{\bar{A}} = \bar{A}$, **(d)** $\overline{A \cup B} = \bar{A} \cup \bar{B}$.

PROOF: **(a)** follows from (12.39)**(c)** and **(b)** follows from remark 12.20.
The proof of **(c)** and **(d)** is left as exercise 12.20 on p.400.

Besides contact points there also is the concept of a limit point. We will not work with limit points in this document and only give its definition to make the reader aware that those two concepts are different and s/he must be mindful of this fact because many other writers work exclusively with limit points and often do not define contact points.
Here is the definition (see [13] Munkres, a standard book on topology):

Definition 12.25 (Contact points vs Limit points)**.** ★

Given is a topological space $(X, \mathfrak{U})$. Let $A \subseteq X$ and $x_0 \in X$. x_0 is called a **limit point** or **cluster point** or **point of accumulation** of A if every neighborhood U of x_0 intersects A in at least one point other than x_0, i.e.,

$$U \,\cap\, (A \setminus x_0) \;\neq\; \emptyset. \;\square$$

Remark 12.22. Not every element of a set $A \subseteq (X, \mathfrak{U})$ is necessarily a limit point of A. An example for this are the so called isolated points. [149] □

12.9 Bounded Sets and Bounded Functions in Metric Spaces

Definition 12.26 (bounded sets)**.** Given is a subset A of a metric space (X, d).

[148] This proposition states that the closure is a so–called **closure operator** which is defined to be a function

$$\mathcal{cl} : 2^X \to 2^X; \qquad A \mapsto \mathcal{cl}(A) := \bar{A}$$

on some abstract, nonempty set X (which need not be a topological space) such that the following are satisfied:

(a) $\mathcal{cl}(\emptyset) \;=\; \emptyset$, **(b)** $A \subseteq \mathcal{cl}(A)$, **(c)** $\mathcal{cl}(\mathcal{cl}(A)) = \mathcal{cl}(A)$, **(d)** $\mathcal{cl}(A \cup B) = \mathcal{cl}(A) \cup \mathcal{cl}(B)$.

It can be shown that if we define

$$\mathfrak{U} \;:=\; \{A^{\complement} : \mathcal{cl}(A) = A\}$$

then $(X, \mathfrak{U})$ satisfies the properties of a topological space.

[149] $a \;\in\; A$ is called an **isolated point** of A if there is a neighborhood U of a such that $U \cap A \;=\; \{a\}$, i.e., the "**punctured neighborhood**" $U \setminus \{a\}$ of a has empty intersection with A. Here is an example: Let $X := \mathbb{R}$ with the topology induced by the Euclidean metric $d(x, x') := |x - x'|$. Then the subset $A := [0,1] \cup \{3\} \cup \{4\}$ possesses 3 and 4 as isolated points. You should convince yourself that those two elements of A are NOT limit points of A.

The **diameter** of A is defined as

$$\text{diam}(\emptyset) := 0, \qquad \text{diam}(A) := \sup\{d(x,y) : x, y \in A\} \text{ if } A \neq \emptyset. \tag{12.42}$$

We call A a **bounded set** if $\text{diam}(A) < \infty$. □

Remark 12.23.

(a) Note that we needed a metric $d(x,y)$ to define the boundedness of a set. We cannot generalize this concept to topological spaces.

(b) A set can be bounded in one metric and unbounded in another. For example, let d be the Euclidean metric on $\mathbb{R}$ and let d' be the discrete metric on $\mathbb{R}$. Then each of the sets $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ is bounded in $(\mathbb{R}, d')$ (by the number 1), but it is unbounded in $(\mathbb{R}, d)$. □

Proposition 12.30.

Given is a metric space (X,d) and a nonempty subset A. The following are equivalent:

(12.43)

(a) *$\text{diam}(A) < \infty$, i.e., A is bounded.*

(b) *There exists $\gamma > 0$ and $x_0 \in X$ such that $A \subseteq N_\gamma(x_0)$.*

(c) *For all $x \in X$ there exists $\gamma > 0$ such that $A \subseteq N_\gamma(x)$.*

PROOF of "**(b)** ⇒ **(a)**": For any $x, y \in A$ we have

$$d(x,y) \leq d(x,x_0) + d(x_0,y) \leq 2\gamma$$

and it follows that $\text{diam}(A) \leq 2\gamma$.

PROOF of "**(a)** ⇒ **(b)**": Pick an arbitrary $x_0 \in A$ and let $\gamma := \text{diam}(A)$. Then for all $a \in A$

$$d(x_0,a) \leq \sup_{x \in A} d(x,a) \leq \sup_{x,z \in A} d(x,z) = \text{diam}(A) = \gamma.$$

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

$$d(y,z) \leq d(y,x_0) + d(x_0,z) \leq 2\gamma$$

and it follows that $\text{diam}(A) \leq 2\gamma < \infty$.

PROOF of "**(a)** ⇒ **(c)**": Given $x \in X$, pick an arbitrary $x_0 \in A$ and let $\gamma := d(x,x_0) + \text{diam}(A) + 1$. Then

$$\begin{aligned} y \in A \quad \Rightarrow \quad d(x,y) &\leq d(x,x_0) + d(x_0,y) \leq d(x,x_0) + \sup_{u \in A} d(u,y) \\ &\leq d(x,x_0) + \sup_{u,z \in A} d(u,z) = d(x,x_0) + \text{diam}(A) = \gamma. \end{aligned}$$

It follows that $A \subseteq N_\gamma(x)$. ■

Proposition 12.31.

Let (X, d) be a metric space. For $n \in \mathbb{N}$ let $A_n \subseteq X$ such that $\delta_n := diam(A_n) \to 0$ as $n \to \infty$. Let $A := \bigcap_n A_n$. Then,

$$\text{either} \quad A = \emptyset, \quad \text{or} \quad \text{there is some } a \in X \text{ such that} \quad A = \{a\}.$$

PROOF: Let $a, a' \in A$ and let $\delta := d(a, a')$. It follows from $A \subseteq A_n$ that

$$d(a, a') \leq \sup\{d(x, x') : x, x' \in A_n\} = \text{diam}(A_n), \quad \text{i.e.,} \quad d(a, a') \leq \delta_n$$

for all $n \in \mathbb{N}$. It follows from $\delta_n \to 0$ that $d(a, a') = 0$, i.e., $a = a'$. We have shown that A contains at most one element. ■

In metric spaces the points in the closure of a set A can be approaced by sequences that live in A. It should not come as a surprise that diam(A) does not increase when one replaces A by its closure.

Proposition 12.32.

Let (X, d) be a metric space and $A \subseteq X$. Then,

$$diam(A) = diam(\bar{A}) .$$

PROOF:

It follows from $A \subseteq \bar{A}$ that diam$(A) \leq$ diam$(\bar{A})$. It remains to prove that diam$(\bar{A}) \leq$ diam(A).

Nothing needs to be shown if A is unbounded, i.e., diam$(A) = \infty$, because $x \leq \infty$ is true for any $x \in \mathbb{R} \uplus \{\pm\infty\}$. We hence may assume that A is bounded.

Let $\varepsilon > 0$ and $x, y \in \bar{A}$. It follows from Thm.12.6 (Sequence criterion for contact points in metric spaces) on p.384 that there are sequences $(x_n)_n$ and $(y_n)_n$ in A such that $\lim\limits_{n\to\infty} x_n = x$ and $\lim\limits_{n\to\infty} y_n = y$. Thus there exist $N_x, N_y \in \mathbb{N}$ such that $d(x_j, x) < \frac{\varepsilon}{2}$ for all $j \geq N_x$ and $d(y_j, y) < \frac{\varepsilon}{2}$ for all $j \geq N_y$. Let $n := \max(N_x, N_y)$. Then

$$d(x, y) \leq d(x, x_n) + d(x_n, y_n) + d(y_n, y) < d(x_n, y_n) + \varepsilon \leq \text{diam}(A) + \varepsilon.$$

The inequality $d(x, y) \leq$ diam$(A) + \varepsilon$ is true for arbitrary $x, y \in \bar{A}$, hence diam$(A) + \varepsilon$ is an upper bound for the set $\{d(x, y) : x, y \in \bar{A}\}$. We conclude that

$$\text{diam}(\bar{A}) = \sup\left(\{d(x, y) : x, y \in \bar{A}\}\right) \leq \text{diam}(A) + \varepsilon$$

. The above holds for arbitrary $\varepsilon > 0$, and we conclude that diam$(\bar{A}) \leq$ diam(A). ■

Proposition 12.33.

Let (X, d) be a metric space. Let $A_1 \supseteq A_2 \supseteq \dots$ be subsets of X such that $diam(A_n) \to 0$ as $n \to \infty$ and let $A := \bigcap_j \bar{A}_j$. Let $x_n \in A_n$ for all n. Then,

- *$(x_n)_n$ converges if and only if A is not empty.*
- *If $A \neq \emptyset$, then A is the singleton set $A = \left\{\lim\limits_{n\to\infty} x_n\right\}$.*

xx

PROOF:

The proof is done in two stages.

(a) We first prove that if x_n has a limit $x \in X$ then $A = \{x\}$.

Let $n, k \in \mathbb{N}$ such that $k \geq n$. It follows from $x_k \in A_k \subseteq A_n$ that $x_k \in A_n$. Thm.12.6 (Sequence criterion for contact points in metric spaces) on p.384 yields $x = \lim\limits_{n\to\infty} x_n \in \bar{A}_n$. As n was arbitrary, we obtain $x \in \bigcap_j \bar{A}_j$, i.e., $x \in A$.

It follows from Proposition 12.32 and $\text{diam}(A_n) \to 0$, that $\text{diam}(\bar{A}_n) \to 0$. It now follows from Proposition 12.31 that the intersection A has size zero or one. Since $x \in A$, $A = \{x\}$. This concludes the proof of **b1**.

(b) It remains to prove that if $A \neq \emptyset$ then $(x_n)_n$ converges.

Let $x \in A$. Then $x \in \bar{A}_n$ for all n. We use again that $\text{diam}(A_n) = \text{diam}(\bar{A}_n)$ and obtain

$$d(x_n, x) \leq \text{diam}(\bar{A}_n) = \text{diam}(A_n) = \delta_n \to 0 \text{ as } n \to \infty. \tag{12.44}$$

Let $\varepsilon > 0$. It follows from $\lim\limits_{k\to\infty} \delta_k = 0$ that there exists $n_0 \in \mathbb{N}$ such that $\delta_k < \varepsilon$ for all $k \geq n_0$. We conclude from (12.44) that $d(x_k, x) < \varepsilon$ for all $k \geq n_0$, and hence that $\lim\limits_{k\to\infty} x_n = x$. ■

Remark 12.24. You may wonder whether it can ever happen that a sequence of nonempty sets $A_1 \supseteq A_2 \supseteq \cdots$ and $\text{diam}(A_n) \to 0$ as $n \to \infty$ can ever have a non–empty intersection A.

Yes, it can: Let (X, d) be the rational numbers with the standard metric $d(q, q') = |q - q'|$. Let

$$A_n := \{\, q \in]0, \infty[_{\mathbb{Q}} \colon 2 - \frac{1}{n} < q^2 < 2 + \frac{1}{n} \,\}$$

In other words, A_n is the open neighborhood of $\sqrt{2}$ with radius $1/n$. Clearly, $\bigcap_n A_n = \{\sqrt{2}\}$. Well, since $\sqrt{2} \in \mathbb{R}$, this would be correct in the metric space $(\mathbb{R}, d)$. But, since $\sqrt{2} \notin \mathbb{Q}$, $\bigcap_n A_n = \emptyset$! □

12.10 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 12.20 (page 369)

Definition 12.27 (Cauchy sequences). [150] Given is a metric space (X, d).

A sequence (x_n) in X is called a **Cauchy sequence** or, in short, it is Cauchy if for any $\varepsilon > 0$ (no matter how small), there exists some index $n_0 \in \mathbb{N}$ such that

$$d(x_i, x_j) < \varepsilon \qquad \text{for all } i, j \geq n_0 \tag{12.45}$$

This is called the **Cauchy criterion for convergence** of a sequence. □

Example 12.6 (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

$$|x_i - x_j| < \varepsilon \qquad \text{for all } i, j \geq n_0. \;\square \tag{12.46}$$

[150] so named after the great french mathematician Augustin–Louis Cauchy (1789–1857) who contributed massively to the most fundamental ideas of Calculus.

Proposition 12.34.

> *Let (X,d) be a metric space and $x_n \in X$ $(n \in \mathbb{N})$. Then the following are equivalent:*
> **(a)** *$(x_n)_n$ is Cauchy.*
> **(b)** *The diameters of the tail sets $T_n = \{x_j : j \geq n\}$ converge to zero.*
> **(c)** *There exists a nonincreasing sequence $A_1 \supseteq A_2 \supseteq \dots$ of subsets of X such that $x_n \in A_n$ and diam$(A_n) \to 0$ as $n \to \infty$.*

PROOF:

PROOF of **(a)** $\Rightarrow$ **(b)**:

Let $\varepsilon > 0$. It follows from the definition of Cauchy sequences that there exists $n_0 \in \mathbb{N}$ such that $d(x_i, x_j) < \varepsilon$ for all $i, j \geq n_0$. From this we obtain

$$\text{diam}(T_{n_0}) \ = \ \sup\{d(x_i,x_j) : i,j \geq n_0\} \leq \varepsilon.$$

It follows from prop.12.7 on p.370 that $\lim\limits_{n\to\infty} \text{diam}(T_n) = 0$.

PROOF of **(b)** $\Rightarrow$ **(c)**:

We choose $A_n := T_n$ as our nonincreasing sequence of sets.

PROOF of **(c)** $\Rightarrow$ **(a)**:

Let $\varepsilon > 0$. It follows from the definition of convergence diam$(A_n) \to 0$ that there exists $n_0 \in \mathbb{N}$ such that diam$(A_{n_0}) < \varepsilon$. Let $k \in \mathbb{N}, k \geq n_0$. Then $A_k \subseteq A_{n_0}$, hence diam$(A_k) \leq$ diam$(A_{n_0}) < \varepsilon$.

Let $i, j \in \mathbb{N}$ such that $i, j \geq n_0$. By assumption, $x_i \in A_i \subseteq A_{n_0}$ and $x_j \in A_j \subseteq A_{n_0}$, hence $x_i, x_j \in A_{n_0}$, hence $d(x_i, x_j) \leq$ diam$(A_{n_0}) < \varepsilon$. This proves that $(x_n)_n$ is Cauchy. ■

Proposition 12.35.

> *A Cauchy sequence in a metric space is bounded.*

PROOF: Let $(x_n)_n$ be a Cauchy sequence in a metric space (X,d). There is $N = N(1/2)$ such that $d(x_i,x_j) < 1/2$ for all $i,j \geq N$. In particular, $d(x_i,x_N) < 1/2$.

Let $M := \max\{d(x_j,x_N) : j < N\}$. We obtain for any two indices $i,j \in \mathbb{N}$ that

$$d(x_i,x_j) \ \leq \ d(x_i,x_N) + d(x_N,x_j).$$

$d(x_i,x_N)$ is bounded by M in case that $i < N$ and by $1/2$ if $i \geq N$; hence $d(x_i,x_N) < 1/2 + M$. We use the same reasoning to conclude that $d(x_N,x_j) < 1/2 + M$ and obtain $d(x_i,x_j) < 1 + 2M$. This proves the boundedness of $(x_n)_n$. ■

Theorem 12.8 (Convergent sequences are Cauchy)**.**

> *Let $(x_n)_n$ be a convergent sequence in a metric space (X,d). Then $(x_n)_n$ is Cauchy.*

PROOF: Let $L \in X$ and $x_n \to L$. Let $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that

$$k \geq N \ \ \Rightarrow \ \ d(x_k, L) \ < \ \varepsilon/2. \tag{12.47}$$

It follows from (12.47) that, for any $i,j \geq N$,

$$d(x_i,x_j) \ \leq d(x_i,L) + d(L,x_j) \ < \ \varepsilon/2 + \varepsilon/2 \ = \ \varepsilon.$$

It follows that the sequence satisfies (12.45) of the definition of a Cauchy sequence (def. 12.27 on p.390). ■

Proposition 12.36. *Let $(x_n)_n$ be a Cauchy sequence in a metric space (X, d).*

> *If some subsequence x_{n_j} converges to a limit x_0. Then*
> ***(a)*** *ANY subsequence of $(x_n)_n$ converges to L.*
> ***(b)*** *$(x_n)_n$ is a convergent sequence.*
> *Further, any subsequence y_{n_j} of a convergent sequence $(y_n)_n$ converges to the limit of $(y_n)_n$.*

PROOF of **(a)**: Let $n_1 < n_2 < n_3 \dots$ be such that x_{n_j} converges to L. For $k \in \mathbb{N}$ let $y_k := x_{n_k}$.
Let $\varepsilon > 0$. Convergence $y_j \to L$ implies that there is $N \in \mathbb{N}$ such that

$$d(y_j, L) < \varepsilon/2 \text{ for all } j \geq N. \tag{12.48}$$

Because (x_j) is Cauchy there also exists $N' \in \mathbb{N}$ such that

$$d(x_i, x_j) < \varepsilon/2 \text{ for all } i, j \geq N'. \tag{12.49}$$

Let $K := \max(n_N, N')$ and $j \geq K$. Then

$$d(x_j, L) \ \leq \ d(x_j, y_K) + d(y_K, L)$$

It follows from $n_K \geq K$ and $j \geq K$ and (12.49) that $d(x_j, y_K) = d(x_j, x_{n_K}) < \varepsilon/2$ and it follows from (12.48) that $d(y_K, L) < \varepsilon/2$. We conclude that $d(x_j, L) < \varepsilon$ for all $j \geq K$ and this proves convergence $x_j \to L$.
PROOF of **(b)**: This is trivial: The full sequence $x_1, x_2, \dots$ is a subsequence, and it converges by assumption to L
PROOF of the addendum: This is trivial, too: The convergent sequence $(x_n)_n$ is Cauchy, and the assumption now follows from part **(a)** ■

Here is the formal definition of a complete set in a metric space.

Definition 12.28 (Completeness in metric spaces)**.**

> A subset A of a metric space (X, d) is called **complete**, if any Cauchy sequence (a_n) with elements in A converges to some $a \in A$. □

Remark 12.25.
(a) It is **NOT sufficient** that $\lim\limits_{n\to\infty} x_n$ exists in X. It must not belong to the complement of A!
(b) In particular, X itself is complete iff any Cauchy sequence in X converges.
(c) A is complete as a subset of (X, d) iff the subspace $((A, d\big|_{A\times A})$ is complete "in itself". □

The following theorem of the completeness of the set of all real numbers [151] states that any Cauchy sequence converges to a real number. This is a big deal: To show that a sequence in $\mathbb{R}$ has a finite

[151] Remember the completeness axiom for $\mathbb{R}$ (axiom 9.1**(c)** on p.255) 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.9.14 (Characterization of limits via limsup and liminf) on p.290, a theorem which will be used in this chapter to prove the completeness of $\mathbb{R}$ as a metric space.

limit one need not provide the actual value of that limit. All one 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.

Theorem 12.9 (Completeness of the real numbers)**.**

> *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: It follows from prop.12.35 that x_n is bounded, hence $(x_n)_n$ possesses finite liminf and limsup. [152] 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 $|x_i - x_j| \leq \varepsilon$ for all $i, j \geq N$.

Let $T_n := \{x_j : j \geq n\}$ be the tail set of the sequence $(x_n)_n$. Let $\alpha_n := \inf T_n, \beta_n := \sup T_n$.

There is some $i \geq N$ such that $|x_i - \alpha_N| = x_i - \alpha_N \leq \varepsilon$ and there is some $j \geq N$ such that $|\beta_N - x_j| = \beta_N - x_j \leq \varepsilon$. It follows that

$$0 \ \leq \ \beta_N - \alpha_N \ = |\beta_N - \alpha_N| \ \leq \ |\beta_N - x_j| \ + \ |x_j - x_i| \ + \ |x_i - \alpha_N| \leq 3\varepsilon.$$

Further, if $k \geq N$ then $T_k \subseteq T_N$, hence $\alpha_k \geq \alpha_N$ and $\beta_k \leq \beta_N$. It follows that

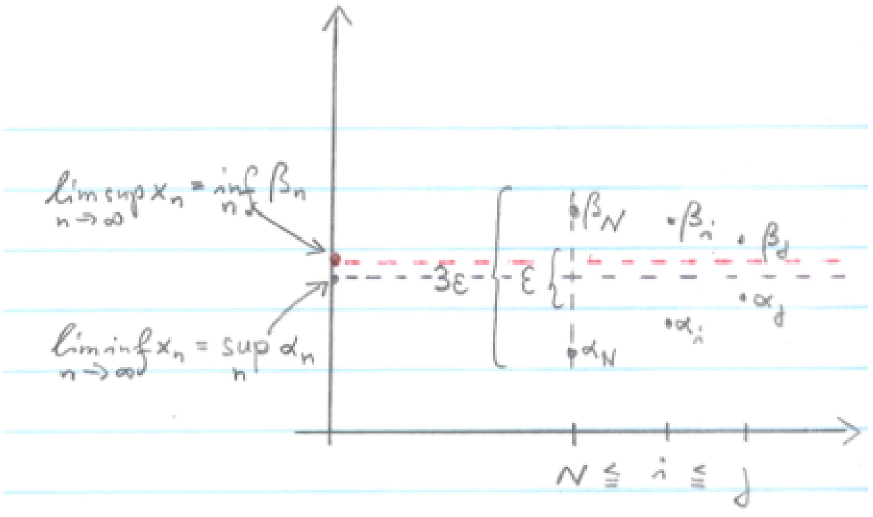
$$0 \leq \inf \beta_k - \sup \alpha_k \leq \beta_N - \alpha_N \leq 3\varepsilon.$$

But then

$$0 \ \leq \ \limsup_{k\to\infty} x_k - \liminf_{k\to\infty} x_k \ = \ \inf_k \beta_k - \sup_k \alpha_k \ \leq \ 3\varepsilon.$$

$\varepsilon > 0$ was arbitrary, hence $\limsup\limits_{k\to\infty} x_k \ = \ \liminf\limits_{k\to\infty} x_k$.

Figure 12.3: liminf = limsup for Cauchy sequences


$\limsup\limits_{n\to\infty} x_n = \inf\limits_n \beta_n$
$\liminf\limits_{n\to\infty} x_n = \sup\limits_n \alpha_n$
3ε
ε
β_N
β_i
β_j
α_i
α_j
α_N
$N \leq i \leq j$

.

[152] See ch.9.2 (Minima, Maxima, Infima and Suprema).

Part 3: It follows from theorem 9.14 on p.290 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 12.10 (Completeness of $\mathbb{R}^n$)**.**

Let $(\vec{x}_j)$ be a Cauchy sequence in $\mathbb{R}^n$. Then there exists $\vec{a} \in \mathbb{R}^n$ such that $\vec{a} = \lim\limits_{j\to\infty} \vec{x}_j$.

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 $i, j \geq K$ then $\|\vec{x}_i - \vec{x}_j\|_2 < \varepsilon$. Hence

$$|x_{i,k} - x_{j,k}| \ = \sqrt{|x_{i,k} - x_{j,k}|^2} \ \leq \ \sqrt{\sum_{k=1}^{n} |x_{i,k} - x_{j,k}|^2} \ = \ \|\vec{x}_i - \vec{x}_j\|_2 \ < \ \varepsilon.$$

It follows from the completeness of $\mathbb{R}$ as a metric space that there exist real numbers

$$a_1, a_2, a_3, \dots, a_n \qquad \text{such that } \ a_k \ = \lim_{n\to\infty} x_{n,k} \ (1 \leq k \leq n).$$

For a given $\varepsilon > 0$ we can find natural numbers $N_{0,1}, N_{0,2}, \dots, N_{0,n}$ such that

$$|x_{n,k} - a_k| \ < \ \frac{\varepsilon}{n} \quad \text{for all } n \geq N_{0,k} \text{ and for all } 1 \leq k \leq n.$$

Let $N^\star := \max(N_{0,1}, N_{0,2}, \dots, N_{0,n})$ and $\vec{a} := (a_1, a_2, \dots, a_n)$. It follows that

$$d(\vec{x}_j - \vec{a}) \ = \ \sqrt{\sum_{k=1}^{n} |x_{j,k} - a_k|^2} \ \leq \ \sqrt{n \cdot \left(\frac{\varepsilon}{n}\right)^2} \ = \ \frac{\varepsilon}{\sqrt{n}} \ \leq \ \varepsilon \qquad \text{for all } j \geq N^\star.$$

This proves convergence of $\vec{x}_j$ to $\vec{a}$.

You have learned in multivariable calculus that the limit of a sequence of vectors can be computed as the vector of the limits, taken separately for each coordinate. The proof is very similar to that of Theorem 12.10.

Proposition 12.37.

Let $\vec{x}_j = (x_{j,1}, x_{j,2}, \dots, x_{j,n})$ and $\vec{b} \in \mathbb{R}^n$. Then,

(12.50) $$\lim_{j\to\infty} \vec{x}_j \ = \ \vec{b} \ \Leftrightarrow \ \lim_{j\to\infty} x_{j,k} \ = \ b_k \textit{ for all } 1 \leq k \leq n\,.$$

PROOF of "⇒": If $\vec{x}_j$ converges to $\vec{b}$ then this sequence is Cauchy. We have seen in the proof of thm.12.10 that it has as limit a vector $\vec{a} := (a_1, a_2, \dots, a_n)$ whose k–th coordinate a_k was obtained as $a_k \ = \lim\limits_{j\to\infty} x_{j,k}$. In other words, $a_k = b_k$. This proves "⇒".

PROOF of "$\Leftarrow$": Assume that $\lim\limits_{j\to\infty} x_{j,k} \ = \ b_k$ for all $1 \leq k \leq n$. We copy word for word the second half of the proof of thm.12.10.

For a given $\varepsilon > 0$ we can find natural numbers $N_{0,1}, N_{0,2}, \dots, N_{0,n}$ such that

$$|x_{n,k} - b_k| \ < \ \frac{\varepsilon}{n} \quad \text{for all } n \geq N_{0,j} \text{ and for all } 1 \leq k \leq n.$$

Let $N^\star := \max(N_{0,1}, N_{0,2}, \dots, N_{0,n})$. It follows that

$$d(\vec{x}_j - \vec{b}) \ = \ \sqrt{\sum_{k=1}^{n} |x_{j,k} - b_k|^2} \ \leq \ \sqrt{n \cdot \left(\frac{\varepsilon}{n}\right)^2} \ = \ \frac{\varepsilon}{\sqrt{n}} \ \leq \ \varepsilon \qquad \text{for all } j \geq N^\star.$$

This proves convergence of $\vec{x}_j$ to $\vec{b}$ and hence "$\Leftarrow$". ■

Example 12.7 (Approximation of decimals)**.** The following illustrates Cauchy sequences and completeness in $\mathbb{R}$. We have seen in ch.9.6 (Decimal Expansions of Real and Rational Numbers) that any real number $x \geq 0$ can be written as a decimal

$$x \ = \ m \ + \ \sum_{j=1}^{\infty} d_j \cdot 10^{-j} \quad (m \in \mathbb{Z}_{\geq 0},\ d_j \in \{0, 1, 2, \dots, 9\}).$$

Further, any such infinite series is a real number since each partial sum $s_n = m \ + \ \sum\limits_{j=1}^{n} d_j \cdot 10^{-j}$ is bounded (above) by $m \ + \ 9 \sum\limits_{j=1}^{\infty} 10^{-j} \ = \ m + 1$, and thus, x is a real number as the supremum of the bounded and nondecreasing sequence $(s_n)_n$.

What just has been illustrated is that there a natural way to construct for a given $x \in \mathbb{R}$ Cauchy sequences of rational numbers that converge toward x. (Each s_n is rational as the sum of the finitely many rational numbers m and $\frac{d_j}{10^j}$. The completeness of $\mathbb{R}$ states that the reverse also is true: For any Cauchy sequence $s_n \in \mathbb{Q}$ there is an element $x \in \mathbb{R}$ toward which this sequence converges. □

The existence of irrational numbers tells us that the limit of a sequence of rational partial sums need not be rational. This can be used to construct metric spaces which are not complete.

Proposition 12.38. *The metric space $(\mathbb{Q}, d_{|\cdot|})$ (Euclidean metric) is not complete.*

PROOF: Let us work for the time being in the metric space $(\mathbb{R}, d_{|\cdot|})$ of all real numbers, not in the subspace $(\mathbb{Q}, d_{|\cdot|})$ which we are interested in.

Let $x \in \mathbb{R}$ be any positive irrational number, e.g., $x = \pi = 3.1415\dots$ or $x = \sqrt{2} = 1.414\dots$. x has a decimal representation $x = \ m + \sum\limits_{j=1}^{\infty} d_j 10^{-j}$ where each $d_j \in \mathbb{Z}_{\geq 0}$ is a digit, i.e., $0 \leq d_j \leq 9$. Let

$s_n \ = \ m + \sum\limits_{j=1}^{n} d_j 10^{-j}$ Then,

$$|x - s_n| \ = \ x - s_n \ = \ \sum_{j=n+1}^{\infty} d_j 10^{-j} \ \leq \ 9 \cdot \sum_{j=n+1}^{\infty} 10^{-j} \ = \ 10^{-n}\,.$$

It follows, not surprisingly, that $x = \lim\limits_{n\to\infty} x_n$.

We know from thm.12.8 (Convergent sequences are Cauchy) on p.391 that the sequence s_n is Cauchy in $(\mathbb{R}, d_{|\cdot|})$. But $s_n \in \mathbb{Q}$ for all n and the distance $d_{|\cdot|}(s_n, s_m) = |s_n - s_m|$ is the same in $(\mathbb{R}, d_{|\cdot|})$ and $(\mathbb{Q}, d_{|\cdot|})$.

It follows that s_n is Cauchy in $(\mathbb{Q}, d_{|\cdot|})$. We had constructed this sequence in such a way that it does not have a limit in $\mathbb{Q}$, and it follows that $(\mathbb{Q}, d_{|\cdot|})$ is not complete. ■

A byproduct of this next proposition is that the discrete metric is complete.

Proposition 12.39.

Let d be the discrete metric on a nonempty set X and let $(x_n)_n$ a sequence in X. Then,

$$(x_n)_n \text{ is Cauchy} \quad\Leftrightarrow\quad (x_n)_n \text{ converges} \quad\Leftrightarrow\quad (x_n)_n \text{ is constant eventually}.$$

PROOF:

We show the equivalence of **(a)** and **(c)**. It follows from the definition of Cauchy sequences that there exists $n_0 \in \mathbb{N}$ such that $d(x_i, x_j) < 1$ for all $i, j \geq n_0$ For the discrete metric $d(x_i, x_j) < 1$ means the same as $d(x_i, x_j) = 0$, thus $x_i = x_j$ for all $i, j \geq n_0$. This proves that x_n is eventually constant.

On the other hand, if x_n is eventually constant, then it follows from the definition of a property holding eventually that there exists $n_0 \in \mathbb{N}$ such that $x_i = x_j$ for all $i, j \geq n_0$ Thus $d(x_i, x_j) = 0$ for such i and j (in any metric!). Let $\delta > 0$. It follows that $d(x_i, x_j) < \delta$ for all $i, j \geq n_0$, i.e., $(x_n)_n$ is Cauchy. Matter of fact, since $d(x_{n_0}, x_j) = 0 < \delta$ for all $j \geq n_0$, it follows that the sequence is convergent with x_{n_0} as its limit. ■

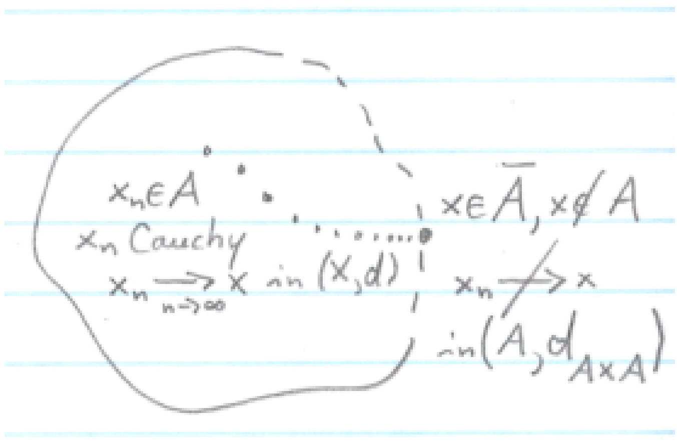
An easy corollary is

Corollary 12.2. *Discrete metric spaces are complete.*

PROOF: We must show that all Cauchy sequences converge in discrete metric spaces. This follows from the first equivalence of prop.12.39 above. ■

Theorem 12.11. *]* *Any complete subset of a metric space is closed.*

Figure 12.4: complete $\Rightarrow$ closed


$x_n \in A$
x_n Cauchy
$x_n \xrightarrow[n\to\infty]{} x$ in (X,d)
$x \in \bar{A}, x \notin A$
$x_n \not\to x$
in $(A, d_{A\times A})$

.

PROOF: Let (X,d) be a metric space and$A \subseteq X$. Let $a \in X$ be a contact point of A. The theorem is proved if we can show that $a \in A$.

a) We employ Definition 12.23 on p.382: 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 $N > \frac{2}{\varepsilon}$. if $j \in \mathbb{N}$ and $k \in \mathbb{N}$ both exceed N then

$$d(x_j, x_k) \;\leq\; d(x_j, a) + d(a, x_k) \;\leq\; \frac{1}{j} + \frac{1}{k} \;\leq\; \frac{1}{N} + \frac{1}{N} < \varepsilon.$$

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.12.3 on p.370).

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.12.11.

Theorem 12.12 (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, i.e., the metric subspace $(A, d\big|_{A\times A})$ is complete.*

PROOF: Let $(x_n)_n$ be Cauchy in A. We must show that there is $a \in A$ such that $x_n \to a$. Note that (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 $x_n \to x$. All x_n belong to A. According to thm.12.6 (Sequence criterion for contact points in metric spaces), x is a contact point of A.

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. ■

12.11 Exercises for Ch.12

12.11.0.1 Exercises for Ch.12.1 (Definition and Examples of Metric Spaces)

Exercise 12.1. Prove prop.12.1 on p.360: Let (X, d) be a metric space. Let $n \in \mathbb{N}$ and $x_1, x_2, \dots, x_n \in X$. Then

$$d(x_1, x_n) \leq \sum_{j=1}^{n-1} d(x_j, x_{j+1}) = d(x_1, x_2) + d(x_2, x_3) + d(x_{n-1}, x_n). \quad \square$$

Exercise 12.2. Prove thm.12.1 (Norms define metric spaces) on p.361: Let $(V, \|\cdot\|)$ be a normed vector space. Then the function

$$d_{\|\cdot\|}(\cdot,\cdot) : V \times V \to \mathbb{R}_{\geq 0}; \qquad (x, y) \mapsto d_{\|\cdot\|}(x, y) := \|y - x\|$$

defines a metric space $(V, d_{\|\cdot\|})$.

Hint: This proof is very easy. Even the triangle inequality for the metric $d(x, y) = \|x - y\|$ follows easily from the triangle inequality for the norm. $\square$

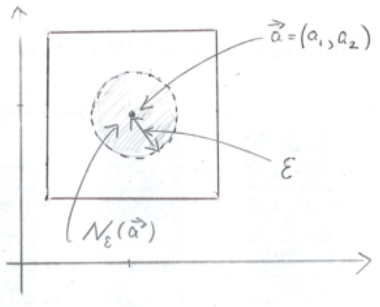
12.11.0.2 Exercises for Ch.12.2 (Measuring the Distance of Real–Valued Functions)

12.11.0.3 Exercises for Ch.12.3 (Neighborhoods and Open Sets)

Exercise 12.3. Prove prop.12.5 on p.367: Let $a, b \in \mathbb{R}$ such that $a < b$. Then the open interval $]a, b[$ is an open set in $(\mathbb{R}, d_{|\cdot|})$. $\square$

Exercise 12.4. Let $A := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$ be the first quadrant in the plane (the points on the coordinate axes are excluded). Prove that each element of A is an inner point, i.e., A is open in $\mathbb{R}^2$. See the picture for a hint.

Hint: Find for $\vec{a} = (a_1, a_2)$ small enough ε such that $N_\varepsilon(\vec{a}) \subseteq A$



$\square$

Exercise 12.5. Let $a, b \in \mathbb{R}$ such that $a < b$.
(a) The closed interval $[\,a, b\,]$ is not open in $(\mathbb{R}, d_{|\cdot|})$.
(b) The complement of the closed interval $[\,a, b\,]$ is open in $(\mathbb{R}, d_{|\cdot|})$. □

Exercise 12.6.
(a) Let $m \in \mathbb{Z}$, viewed as a subset of the metric space $(\mathbb{R}, d_{|\cdot|})$. Prove that m is a boundary point of $\mathbb{Z}$.
(b) Prove that the above also holds both for the set of rational numbers: $\mathbb{Q} \subseteq \partial(\mathbb{Q})$ and for the set of all irrational numbers: $\mathbb{R} \setminus \mathbb{Q} \subseteq \partial(\mathbb{R} \setminus \mathbb{Q})$. □

12.11.0.4 Exercises for Ch.12.4 (Convergence)

Exercise 12.7. 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 $T_{n_0} = \{x_j : j \geq n_0\}$ is contained in U (see Definition 9.20 (Tail sets of a sequence) on p.287.) □

Exercise 12.8. Prove remark 12.8 on p.371: Let (X, d) be a metric space and $x_n, L \in (X, d)$. Then

$$\lim_{n\to\infty} x_n \;=\; L \;\Leftrightarrow\; \lim_{n\to\infty} d(x_n, L) \;=\; 0\,. \;\square$$

Exercise 12.9. Prove prop.12.9 on p.371:
Let x_n, y_n be two sequences in a metric space (X, d). Assume there is $K \in \mathbb{N}$ such that $x_n = y_n$ for all $n \geq K$. Let $L \in X$ Then

$$\lim_{n\to\infty} x_n = L \;\Leftrightarrow\; \lim_{n\to\infty} y_n = L. \;\square$$

Exercise 12.10. Prove prop.12.11 on p.371:
Let x_n be a convergent sequence in a metric space (X, d) with limit $L \in E$. Let $K \in \mathbb{N}$. For $n \in \mathbb{N}$ let $y_n := x_{n+K}$. Then $\lim\limits_{n\to\infty} (y_n)_n = L$. □

Exercise 12.11. Let $f_n, f \in \mathscr{B}([0,1], \mathbb{R})$ $n \in \mathbb{N}$ be continuous such that $f = \lim\limits_{n\to\infty} f_n$ in $(\mathscr{B}([0,1], d_{\|\cdot\|_\infty})$(!) Prove $\lim\limits_{n\to\infty} \displaystyle\int_0^1 f_n(x)dx = \int_0^1 f(x)dx$. You must use the ε, N definition of convergence.
Hints: (a) No need to mention that continuous functions are both bounded and integrable and that they attain both max and min on closed and bounded intervals. **(b)** Use the mean value theorem: For cont. $h(\cdot)$ on $[0,1]$ let $\alpha := \min\limits_{x\in[0,1]} h(x), \beta := \max\limits_{x\in[0,1]} h(x)$. Then $\exists\ \lambda \in [\alpha, \beta]$ such that $\lim\limits_{n\to\infty} \displaystyle\int_0^1 h(x)dx \;=\; \lambda\,(=\;\lambda(1-0)$. **(c)** Use without proof that $\left|\displaystyle\int_0^1 h(x)dx\right| \;\leq\; \int_0^1 |h(x)|dx$ for any integrable $h(\cdot)$ on $[0,1]$ **(d)** Apply **(b)** and **(c)** to $h_n(x) = |f_n(x) - f(x)|$. (So you deal with $\alpha_n, \lambda_n, \beta_n$).
□

12.11.0.5 Exercises for Ch.12.5 (Abstract Topological spaces)

Exercise 12.12. It was stated in prop.12.13 on p.373 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 power set 2^X of X. Prove it. □

Exercise 12.13. Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq X$. Prove that the open exterior of A is

$$\text{ext}(A) \;=\; \left(\overline{A^{\complement}}\right)^o. \;\square$$

Exercise 12.14. Let X be a set that contains. at least two elements. [153] Prove that there is no metric d on X such that $\mathfrak{U}_d = \{\emptyset, X\}$, i.e., such that its only open sets are the empty set and X.. $\square$

12.11.0.6 Exercises for Ch.12.6 (Bases and Neighborhood Bases)

Exercise 12.15. Let (X, d) be a metric space and let $\mathfrak{B} := \{N_{1/k}(x) : x \in X, k \in \mathbb{N}\}$. Then $\mathfrak{B}$ is a base of the topology for the associated topological space $(X, \mathfrak{U}_d)$. $\square$

12.11.0.7 Exercises for Ch.12.7 (Metric and Topological Subspaces)

12.11.0.8 Exercises for Ch.12.9 (Bounded Sets and Bounded Functions)

12.11.0.9 Exercises for Ch.12.8 (Contact Points and Closed Sets)

Exercise 12.16. Let $A \subseteq \mathbb{R}$ be a closed, nonempty set which is bounded above. Prove that the maximum of A exists and that $\sup(A) = \max(A)$. $\square$

Exercise 12.17. Prove prop.12.27 on p.386: Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq X$. Then $\partial A = \bar{A} \cap \overline{A^{\complement}}$. $\square$

Exercise 12.18. Prove prop.12.26 on p.386: Let $(X, \mathfrak{U})$ be a topological space and $A \subseteq B \subseteq X$. Then $\bar{A} \subseteq \bar{B}$. $\square$

Exercise 12.19. Prove prop.12.16 on p.376: Let $(X, \mathfrak{U})$ be a topological space. If $A \subseteq B \subseteq X$ then $A^o \subseteq B^o$. $\square$

Exercise 12.20. Prove parts **(c)** and **(d)** of prop.12.29 (Closure of a set as a hull operator) on p.387: Let A and B be subsets of a topological space $(X, \mathfrak{U})$. Then **(c)** $\bar{\bar{A}} = \bar{A}$, **(d)** $\overline{A \cup B} = \bar{A} \cup \bar{B}$. $\square$

12.11.0.10 Exercises for Ch.12.10 (Completeness in Metric Spaces)

Exercise 12.21. Let (X, d) be a metric space and $A \subseteq X$, $A \neq \emptyset$. Let

$$\gamma := \gamma(A) := \inf\{d(x, y) : x, y \in A \text{ and } x \neq y\}.$$

(a) Prove that if $\gamma > 0$ then A is complete.

(b) The reverse is not true. Find a counterexample. $\square$

Exercise 12.22. Let (X, d) be a metric space and let $A \subseteq X$ be a finite subset. Prove that A is complete. $\square$

[153] See exercise 12.14 on p.400

Exercise 12.23. 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. $\square$

Exercise 12.24. 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. $\square$

13 Metric Spaces and Topological Spaces – Part II

13.1 Continuity

13.1.1 Definition and Characterizations of Continuous Functions

We have briefly discussed in ch.9.3 on p.263. the continuity of functions with arguments and values in $\mathbb{R}$. We now extend this definition to functions that map from metric spaces to metric spaces and, more generally, from topological spaces to topological spaces.

Definition 13.1 (Sequence continuity)**.**

Given are two metric spaces (X, d_1) and (Y, d_2). Let $A \subseteq X$, $x_0 \in A$ and let $f : A \to Y$ be a mapping from A to Y. We say that f is **sequence continuous at x_0** and we write

$$\lim_{x \to x_0} f(x) = f(x_0)\,, \tag{13.1}$$

if the following is true for any sequence (x_n) with values in A:

$$\text{if } x_n \to x_0 \text{ then } f(x_n) \to f(x_0). \tag{13.2}$$

In other words, the following must be true for any sequence (x_n) in A and $x_0 \in A$:

$$\lim_{n\to\infty} x_n = x_0 \quad \Rightarrow \quad \lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x_0). \tag{13.3}$$

We say that f is **sequence continuous** if f is sequence continuous at x_0 for all $x_0 \in A$. □

Remark 13.1. Important points to notice:

a) It is not enough for the above to be true for some sequences that converge to x_0. Rather, this 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 we need 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 13.2 (ε-δ continuity)**.**

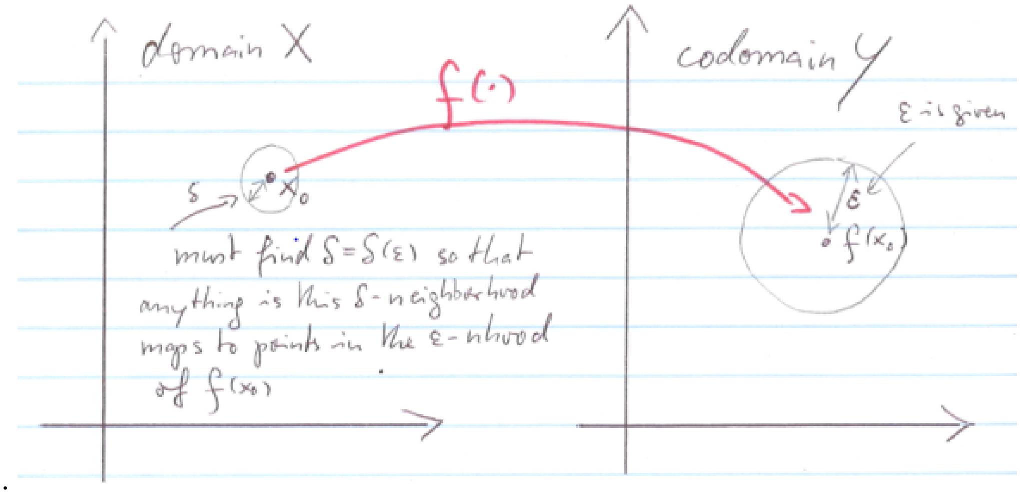
Given are two metric spaces (X, d_1) and (Y, d_2). Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. We say that $f(\cdot)$ is **ε–δ continuous at x_0** if the following is true: For any (whatever small) $\varepsilon > 0$ there exists $\delta > 0$ such that either one of the following equivalent statements is satisfied:

$$f\big(N_\delta(x_0) \cap A\big) \subseteq N_\varepsilon(f(x_0)), \tag{13.4}$$

$$d_1(x, x_0) < \delta \Rightarrow d_2(f(x), f(x_0)) < \varepsilon \quad \text{for all } x \in A. \tag{13.5}$$

We say that $f(\cdot)$ is ε-δ **continuous** if $f(\cdot)$ is ε-δ continuous at a for all $a \in A$. □

Figure 13.1: ε-δ continuity


domain X
$f(\cdot)$
codomain Y
ε is given
δ
x_0
ε
$f(x_0)$
must find $\delta = \delta(\varepsilon)$ so that anything in this δ-neighborhood maps to points in the ε-nhood of $f(x_0)$

Remark 13.2. We recall from thm.12.32 on p.381 that

$$N_\delta \cap A \;=\; N_\delta^A(a) \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \delta\}. \tag{13.6}$$

Hence (13.4) states that

f is ε-δ continuous at x_0 $\Leftrightarrow$ for all $\varepsilon > 0$ there exists $\delta > 0$ s.t. $f\big(N_\delta^A(x_0)\big) \subseteq N_\varepsilon(f(x_0))$. □

Theorem 13.1 (Continuity criterion)**.**

Let (X, d_1) and (Y, d_2) be two metric spaces. Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$. Then,

- *f is sequence continuous at x_0 $\Leftrightarrow$ f is ε-δ continuous at x_0.*
- *In particular f is sequence continuous (on A) if and only if f is ε-δ continuous.*

PROOF:

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 (13.4) nor the equivalent (**??**) is true for any $\delta > 0$.

a.1. In other words, No matter how small δ is choosen, there is at least one $x = x(\delta) \in A$ such that $d_1(x, x_0) < \delta$ but $d_2(f(x), f(x_0)) \geq \varepsilon$. In particular we obtain for $\delta := 1/m (m \in \mathbb{N})$ that

$$\text{there exists some } x_m \in N_{1/m}(x_0) \cap A \text{ such that } d_2(f(x_m), f(x_0)) \geq \varepsilon. \tag{13.7}$$

a.2. It follows from prop.12.7 on p.370 that the sequence $(x_m)_{m\in\mathbb{N}}$ converges to x_0.

a.3. Clearly $\big(f(x_m)\big)_{m\in\mathbb{N}}$ 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 (13.7) 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 $y_0 := f(x_0)$. We must prove that $y_n \to y_0$ as $n \to \infty$.

b.1. Let $\varepsilon > 0$. We can find $\delta > 0$ such that (13.4) and hence (**??**) is satisfied. Since we assumed that $x_n \to x_0$ there exists $N := N(\delta) \in \mathbb{N}$ such that $d_1(x_n, x_0) < \delta$ for all $n \geq N$.

b.2. It follows from (**??**) that $d_2(y_n, y) = d_2(f(x_n), f(x_0)) < \varepsilon$ for all $n \geq N$. In other words, $y_n \to y_0$ as $n \to \infty$ and the proof of "⇐" is finished.

It follows from the proofs of **(a)** and **(b)** that f is sequence continuous $\Leftrightarrow$ f is ε-δ continuous. ■

Definition 13.3 (Continuity in metric spaces)**.**

From now on we can use the terms "ε-δ continuous at x_0" and "sequence continuous at x_0" interchangeably for functions between metric spaces and we will simply speak about **continuity of f at x_0**. □

Remark 13.3 (continuity for real–valued functions of real numbers)**.** Let $(X, d_1) = (Y, d_2) = \mathbb{R}$. In this case equation (**??**) on p.**??** becomes

$$|x - x_0| < \delta \ \Rightarrow \ |f(x) - f(x_0)| < \varepsilon.$$

See thm.9.7 on p.273. □

We saw in the ε-δ continuity definition of a function with metric spaces for both domain and codomain and the subsequent remark 13.2 that continuity of $f : (A, d_1\big|_{A\times A}) \to (Y, d_2)$ in $x_0 \in A$ was equivalent to demanding that for any ε–neighborhood of $f(x_0)$ there is a δ–neighborhood of x_0 such that

$$f\big(N_\delta^A(x_0)\big) \ \subseteq \ N_\varepsilon(f(x_0)).$$

The fact that <u>any neighborhood</u> of a point z in a metric space contains a γ–neighborhood of z for suitably small γ, is at the basis of the following theorem.

Theorem 13.2 (Neighborhood 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 continuous at x_0 if and only if for any neighborhood $V_{f(x_0)}$ of $f(x_0)$, there exists a neighborhood U_{x_0} of x_0 in the metric space (X, d_1), such that

$$f(U_{x_0} \cap A) \ \subseteq \ V_{f(x_0)}\,. \tag{13.8}$$

Equivalently, (13.8) can be stated in terms of the subspace $(A, d_1\big|_{A\times A})$ as follows.

for any neighborhood $V_{f(x_0)}$ of $f(x_0)$ there exists a neighborhood $U_{x_0}^A$ of x_0 in the metric space $(A, d_1\big|_{A\times A})$ such that

$$f(U_{x_0}^A) \ \subseteq \ V_{f(x_0)}. \tag{13.9}$$

PROOF:

***a*) ⇒)**: Assume that f is continuous, i.e., ε-δ continuous at a. Let $V_{f(x_0)}$ be a neighborhood of $f(x_0)$.

Then $f(x_0)$ is interior point of $V_{f(x_0)}$ and we can find suitable $\varepsilon > 0$ such that $N_\varepsilon(f(x_0)) \subseteq V_{f(x_0)}$. ε-δ continuity at a implies the existence of $\delta > 0$ such that $f\big(N_\delta(x_0) \cap A\big) \subseteq N_\varepsilon(f(x_0))$, hence $f\big(N_\delta(x_0) \cap A\big) \subseteq V_{f(x_0)}$.

This proves both (13.8) (choose $U_{x_0} := N_\delta(x_0)$) and (13.9) (choose $U^A_{x_0} := N_\delta(x_0) \cap A$).

b) **⇐):** Assume that (13.8) is satisfied for any arbitrary neighborhood $V_{f(x_0)}$ of $f(x_0)$.

Let $\varepsilon > 0$. We need to show that there exists $\delta > 0$ such that

$$f(N_\delta(x_0) \cap A) \subseteq N_\varepsilon(f(x_0)). \tag{13.10}$$

$N_\varepsilon(f(x_0))$ is a neighborhood of $f(x_0)$. It follows from (13.8) that there exists a neighborhood U_{x_0} of x_0 such that

$$f(U_{x_0} \cap A) \subseteq N_\varepsilon(f(x_0)). \tag{13.11}$$

x_0 is interior point of any of its neighborhoods. In particular, it is interior to U_{x_0}.

Accordingly, there exists $\delta > 0$ such that $N_\delta(x_0) \subseteq U_{x_0}$, hence $N_\delta(x_0) \cap A \subseteq U_{x_0} \cap A$. It follows from the monotonicity of the direct image $\Gamma \mapsto f(\Gamma)$ that

$$f\big(N_\delta(x_0) \cap A\big) \subseteq f\big(U_{x_0} \cap A\big) \subseteq N_\varepsilon(f(x_0)). \tag{13.12}$$

The second inclusion relation follows from (13.11). We have proved the existence of $\delta > 0$ such that (13.10) is satisfied. This finishes the proof of "⇐)". ■

Before we generalize continuity to topological spaces we will now generalize thm.9.6 of ch.9.3 which was stated for real–valued function with domain $A \subseteq \mathbb{R}$. to real–valued function with domain $A \subseteq (X, d)$ where (X, d) is a metric space. The proof of this theorem demonstrates how to work with the definitions

Theorem 13.3 (Rules of arithmetic for continuous real–valued functions)**.**

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. Specifically, $\alpha f(\cdot) : x \mapsto \alpha \cdot f(x)$ where $\alpha \in \mathbb{R}$ is continuous at x_0. In particular ($\alpha = -1$) the function $-f(\cdot) : x \mapsto -f(x)$ is continuous at x_0.*

(c) *The sum $f + g(\cdot) : x \mapsto f(x) + g(x)$ is continuous at x_0.*

(d) *If $g(x_0) \neq 0$ then the quotient $f/g(\cdot) : x \mapsto f(x)/g(x)$ is continuous at x_0.*

(e) *Any linear combination* [154] $\sum\limits_{j=0}^{n} a_j f_j(\cdot) : x \mapsto \sum\limits_{j=0}^{n} a_j f_j(x)$ *is continuous in x_0.*

PROOF of **(a)**:

Let $f : A \to \mathbb{R};\ x \mapsto \alpha$ for some $\alpha \in \mathbb{R}$. Let $x_n \in A$ for all $n \in \mathbb{N}$ such that $x_n \to x_0$ as $n \to \infty$. Then $f(x_n) = f(x_0) = \alpha$ for all $n \in \mathbb{N}$, i.e., the sequence $f(x_n)_n$ is constant with value $f(x_0) = \alpha$, and it thus converges to $f(x_0)$ by prop.12.8 on p.370. This proves **(a)**.

PROOF of **(b)**: Since it follows from the already proven part **(a)** that the constant function $x \mapsto \alpha$ and thus in particular the function $x \mapsto -1$ are continuous everywhere on A it remains to prove that fg is continuous at x_0.

Let $x_n \in A$ for all $n \in \mathbb{N}$ such that $x_n \to x_0$ as $n \to \infty$. All we need to show is convergence $f(x_n)g(x_n) \to f(x_0)g(x_0)$. This follows from prop.9.17 (Rules of arithmetic for limits) on p.267, thus we have shown that fg is continuous at x_0. We have proven **(b)**.

PROOF of **(c)**: Let $x_n \in A$ for all $n \in \mathbb{N}$ such that $x_n \to x_0$ as $n \to \infty$. We must show convergence $f(x_n) + g(x_n) \to f(x_0) + g(x_0)$. This again follows from prop.9.17 and we have proved **(c)**.

proof of **(d)** (outline): The proof is done by (strong) induction.

Base case: For $n = 2$ the proof is obvious from parts **(a)**, **(b)** and **(c)**.

Induction step: Write

$$\sum_{j=0}^{n+1} a_j f_j(x) \;=\; \Big(\sum_{j=0}^{n} a_j f_j(x) \Big) \;+\; a_{n+1} f_{n+1}(x) \;=\; U \;+\; V.$$

The left term U is continuous by the induction assumption, thus the sum $U \,+\, V$ is continuous as the sum of two continuous functions (we showed this in **(c)**). This proves **(d)**. ■

The last theorem allows us to conclude that certain sets of continuous functions are vector spaces since sums $f + g$ and scalar products αf involving continuous functions are continuous.

Example 13.1 (Vector spaces of continuous real–valued functions)**.** Let (X, d) be a metric space. Then

$$\mathscr{C}(X, \mathbb{R}) \;:=\; \{f(\cdot) : f(\cdot) \text{ is a continuous real–valued function on } X\} \tag{13.13}$$

of all real continuous functions on X is a vector space. Note that we have seen this before in example 11.11 (Vector spaces of real–valued functions) on p.332 for the special case of $X \subseteq (\mathbb{R}, d_{|\cdot|})$.

The sup–norm

$$\|f(\cdot)\|_\infty \;=\; \sup\{|f(x)| : x \in X\}$$

(see (11.16) on p.344) is **not a real–valued function** on all of $\mathscr{C}(X, \mathbb{R})$ because $\|f(\cdot)\|_\infty = +\infty$ for any unbounded $f(\cdot) \in \mathscr{C}(X, \mathbb{R})$. To avoid complications from dealing with infinity, we often restrict the scope to the subspace

$$\mathscr{C}_{\mathscr{B}}(X, \mathbb{R}) \;:=\; \{h : h \text{ is a bounded continuous real–valued function on } X\}$$

(see prop.11.13 on p. 344) of the normed vector space $\mathscr{B}(X, \mathbb{R})$ of all bounded real–valued functions on X. On this subspace the sup–norm truly is a real–valued function since $\|f(\cdot)\|_\infty < \infty$. □

Remark 13.4. The equivalence of (13.8) and (13.9) in thm.13.2 (neighborhood characterization of continuity) has some profound consequences:

Assume that we have proven a statement about continuity at $x_0 \in X$ for all functions which have metric spaces as domain and codomain. Let's say we use the notation $f : (X, d) \to (Y, d')$. This statement then remains true for all functions $g : (A, d\big|_{A \times A}) \to (Y, d')$ as long as the proof does not make use of a property of X which its subset A does not satisfy.

A good example for this is thm.13.3 (rules of arithmetic for continuous real–valued functions). For example, if a function φ is continuous on all of X then its restriction $\varphi\big|_A$ to $A \subseteq X$ does not lose this property, and if it satisfies in addition $\varphi(x_0) \neq 0$ for some $x_0 \in A$ then it remains true that $\varphi\big|_A(x_0) \neq 0$.
Here is a somewhat contrived counterexample. If the assumptions state that X must be unbounded, i.e., diam$(X) = \infty$, then the validity of the statement does not necessarily extend to bounded subsets of X. [155] □

The last theorem allows us to generalize the notion of continuity to functions between abstract topological spaces.

Definition 13.4 (Continuity for topological spaces)**.** Given are two topological spaces $(X, \mathfrak{U}_1)$ and $(Y, \mathfrak{U}_2)$. Let $A \subseteq X$, $x_0 \in A$ and let $f : A \to Y$ be a mapping from A to Y.

We say that f is **continuous at x_0** if the following is true:
For any neighborhood $V_{f(x_0)}$ of $f(x_0)$, there exists a neighborhood U_{x_0} of x_0 in the topological space $(X, \mathfrak{U}_1)$, such that

$$f(U_{x_0} \cap A) \subseteq V_{f(x_0)} \,. \tag{13.14}$$

Equivalently, continuity at x_0 can be stated in terms of the subspace $(A, {\mathfrak{U}_1}_A)$ as follows. For any neighborhood $V_{f(x_0)}$ of $f(x_0)$ there is a neighborhood $U^A_{x_0}$ of x_0 in $(A, {\mathfrak{U}_1}_A)$ such that

$$f(U^A_{x_0}) \subseteq V_{f(x_0)}. \tag{13.15}$$

We say that f is **continuous** if f is continuous at a for all $a \in A$. □

Remark 13.5. Let (X, d) and (Y, d') be metric spaces with associated metric topologies $\mathfrak{U}_d$ and $\mathfrak{U}_{d'}$. [156] Let $A \subseteq X$ and $f : A \to Y$. Since the condition (13.8) for continuity at $x_0 \in A$ of f as a function between the metric spaces (X, d) and (Y, d') is identical to the condition (13.14) for continuity at $x_0 \in A$ of f as a function between the associated topological spaces $(X, \mathfrak{U}_d)$ and $(Y, \mathfrak{U}_{d'})$ it follows that any statement that we prove for continuity in topological spaces is automatically true for continuity in metric spaces. □

[2] B/G: Art of Proof defines in appendix A, p.136, continuity of a function f as follows: "$f^{-1}(\text{open}) = \text{open}$". The following proposition proves that their definition coincides with the one given here: the validity of (13.14) for all $x_0 \in X$.

Proposition 13.1 ("$f^{-1}(\text{open}) = \text{open}$" continuity)**.**

Let $(X, \mathfrak{U})$ and $(Y, \mathfrak{V})$ be two topological spaces and let $f : X \to Y$. Then,

- *f is continuous (on X) $\Leftrightarrow$ All preimages $f^{-1}(V)$ of open $V \subseteq Y$ are open in X.*

[155] Better counterexamples involve completeness and compactness, important subjects you will learn about later. It is possible for the entire space to be complete and/or compact and for certain subsets not to have that property.

[156] See Definition 12.12 on p.373.

PROOF of "⇒": 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 V because V is open. According to Definition 13.4 there exists a neighborhood U_a of a such that $f(U_a) \subseteq V$.

We conclude from the monotonicity of direct and inverse images and prop.8.1 on p.243 that

$$U_a \subseteq f^{-1}\big(f(U_a)\big) \subseteq f^{-1}(V) = U.$$

It follows that the arbitrarily chosen $a \in U$ is an interior point of U and this proves that U is open.

PROOF of "⇐": We now assume that all inverse images of open sets in Y are open in X.

Let $a \in X, b = f(a)$, and let V_b be a neighborhood of b. Any neighborhood of b contains an open neighborhood of b, hence we may assume that V_b is open. We are done if we can find an open neighborhood U_a of a such that

$$f(U_a) \subseteq V_b \tag{13.16}$$

Let $U_a := f^{-1}\big(V_b\big)$. Then U_a is open as the inverse image of the open set V_b It follows from the monotonicity of direct and inverse images and prop.8.8 on p.244 that

$$f(U) = f\big(f^{-1}(V_b)\big) = V_b \cap f(X) \subseteq V_b.$$

We have proved (13.16) ■

> Note that the previous proposition only addresses "global" continuity of f for **all** $x \in X$ and there is no local version which handles continuity at a specific x_0.

Note also that it is easily generalized to $f : A \to Y$ ($\emptyset \neq A \subseteq X$) by demanding that $f^{-1}(V)$ be open in $(A, \mathfrak{U}_A)$ for all $V \in \mathfrak{V}$.

Remark 13.6. Remark 13.4 on p.406 for metric spaces can be rephrased for topological spaces as follows:

In the interest of simplicity one may assume for statements involving continuity of a function f between topological spaces$(X, \mathfrak{U})$ and $(Y, \mathfrak{V})$ that f is defined on all of X rather than assuming more generally that f is defined (only) on some arbitrary subset A of X. The general case of $f : A \to Y$ is then covered by replacing $(X, \mathfrak{U})$ with $(A, \mathfrak{U}_A)$, i.e., we deal with $f : (A, \mathfrak{U}_A) \to (Y, \mathfrak{V})$ just as long as the proof does not make use of a property of X which its subset A does not satisfy.

It is easy to see that this condition is satisfied for prop.13.2, prop.13.3, and prop.13.4 below.

The next proposition was previously stated for real–valued functions of a real variable. See prop.9.23 on p.272.

Proposition 13.2 (The composition of continuous functions is continuous)**.**

> *Let $(X, \mathfrak{U})$, $(Y, \mathfrak{V})$ and $(Z, \mathfrak{W})$ be topological spaces. Let $f : X \to Y$ be continuous at $x_0 \in X$ and $g : Y \to Z$ continuous at $f(x_0)$.*
>
> - *Then, the composition $g \circ f : X \to Y$ is continuous at x_0.*

PROOF: The proof is left as exercise 13.4 (see p.432). ■

We now give some examples of continuous functions.

Proposition 13.3 (continuity of constant functions).

Let $(X,\mathfrak{U})$ and $(Y,\mathfrak{V})$ be topological spaces and $y_0 \in Y$.

- *Then, the constant function $f : x \mapsto y_0$ is continuous.*

PROOF: It suffices to show that inverse images of open sets are open. So let $V \in \mathfrak{V}$. Then either $x_0 \in V$ in which case $f^{-1}(V) = X$, or $x_0 \notin V$ in which case $f^{-1}(V) = \emptyset$. Since both X and $\emptyset$ are open in X it follows that f^{-1}(open) = open, hence f is continuous. ■

Proposition 13.4 (continuity of the identity mapping).

Let $(X,\mathfrak{U})$be a topological space and let

$$id_X : X \to X; \qquad x \mapsto x$$

be the identity function on X. Then id_X is continuous.

PROOF: It suffices to show that inverse images of open sets are open. So let $V \in \mathfrak{U}$. Then $id_X^{-1}(V) = V$, hence $id_X^{-1}(V)$ is open. This finishes the proof. ■

Remark 13.7. The proof just given also applies to metric spaces but it is instructive to give a direct proof of this proposition which works with the metric.

So let (X,d) be a metric space and let id_X be the identity function on X. Let $x \in X$ and $\varepsilon > 0$. let $\delta := \varepsilon$. If $x' \in X$ such that $d(x,x') < \delta$, then

$$d(id_X(x), id_X(x')) \;=\; d(x,x') \;<\; \delta \;=\; \varepsilon.$$

We have verified condition (**??**) of the ε–δ characterization of continuity and it follows that id_X is continuous at x. x was an arbitrary point in X, and it follows that the identity is continuous. [157] □

The next proposition gives a very simple example that the behavior of a function with respect to continuity strongly depends on the choice of metric on domain and/or codomain.

Proposition 13.5.

Let d be the standard Euclidean metric and let d′ be the discrete metric on the set $\mathbb{R}$ of all real numbers. Let

$$f : (\mathbb{R}, d') \to (\mathbb{R}, d); \quad x \mapsto x \qquad \textit{and} \qquad g : (\mathbb{R}, d) \to (\mathbb{R}, d'); \quad x \mapsto x$$

both be the identity function on $\mathbb{R}$. Then,

- *f is continuous at every point of $\mathbb{R}$*
- *g is not continuous anywhere on $\mathbb{R}$.*

The proof is left as exercise 13.6 (see p.433). ■

Because of their importance we state here once more rem.13.4, rem.13.5, and rem.13.6.

[157] Actually, we have proved a very strong form of continuity. Generally speaking, $\delta = \delta(\varepsilon, x_0)$ is tailored not only to the given ε, but also to the particular argument x_0 at which continuity needs to be verified. We were able to find δ which does not depend on the argument x_0 but only on ε. We will learn later that this makes id_X **uniformly continuous** on its domain X. See Definition 13.5 (Uniform continuity of functions) on p.410.

Remark 13.8.

(a) All statements about continuity proven for topological spaces are also true for the special case of metric spaces.

(b) One may assume for statements involving continuity of a function f between metric spaces (X,d) and (Y,d') or between topological spaces$(X,\mathfrak{U})$ and $(Y,\mathfrak{V})$ that f is defined on all of X rather than assuming more generally that f is defined (only) on some arbitrary subset A of X.
The general case of $f:A\to Y$ is then covered for metric spaces by replacing (X,d) with $(A,d\big|_{A\times A})$ (we deal with $f:(A,d\big|_{A\times A})\to(Y,d'))$, and it is covered for topological spaces by replacing $(X,\mathfrak{U})$ with $(A,\mathfrak{U}_A)$ (we deal with $f:(A,\mathfrak{U}_A)\to(Y,\mathfrak{V}))$, just as long as the proof does not make use of a property of X which its subset A does not satisfy. □

13.1.2 Uniform Continuity

It will be proved in theorem 14.13 (Uniform continuity on sequence compact spaces) on p.449 [158] that continuous real–valued functions on the compact set $[0,1]$ are uniformly continuous in the sense of the following definition. [159]

Definition 13.5 (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

$$d_2(f(x)-f(y))\ <\varepsilon \quad \text{for any } x,y\in A \text{ such that } d_1(x,y)\ <\ \delta. \qquad (13.17)$$ □

Remark 13.9. $f:(X,d_1)\to(Y,d_2)$ is uniformly continuous on $A\subseteq X$ if and only if the following is true: For all $\varepsilon>0$ there is $\delta>0$ such that

$$\left[\,B\subseteq A \text{ such that } \operatorname{diam}(B)\leq\delta\ \Rightarrow\ \operatorname{diam}\big(f(B)\big)\leq\varepsilon\,\right].$$ □

Remark 13.10 (Uniform continuity vs. continuity)**.** Note the following:

***a*.** Condition (13.17) for uniform continuity looks very close to the ε–δ characterization of ordinary continuity (**??**) on p.**??**. 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 must find a suitable δ. In other words, for plain continuity

$$\delta\ =\ \delta(\varepsilon,x_0).$$

[158] see chapter 14.4 (Continuous Functions and Compact Spaces) on p.447

[159] For the special case of $(X,d)=(\mathbb{R},d_{|\cdot|})$ where $d_{|\cdot|}(x,y)=|y-x|$, see [2] Beck/Geoghegan, Appendix A.3, "Uniform continuity".

In the case of uniform continuity all you get is ε. 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. It follows that uniform continuity implies continuity but the opposite need not be true.

c. Many concepts that are defined in metric spaces can be generalized to topological spaces Examples were neigborhoods, interior points and contact points, subspaces and continuity. Uniform continuity is not a concept that can be defined without a metric. [160] □

Example 13.2 (Uniform continuity of the identity mapping)**.** Let us have another look at rem.13.7 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. □

Example 13.3. Let $f(x) := \frac{1}{x}$ on $]0,1]$ with the Euclidean metric. Then f is NOT uniformly continuous on $]0,1]$. See exercise 13.2 □

Remark 13.11. Now that you have learned the definitions for both continuity and uniform continuity, have a closer look at example 4.28, p.112 in ch.4.5.3 (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. □

13.1.3 Continuity of Linear Functions

Lemma 13.1.

Let $f : (V, \|\cdot\|) \to (W, \,\big|\cdot\big|\,)$ be a linear function between two normed vector spaces. Let

$$\begin{aligned} a &:= \sup\{ \big| f(x) \big| : x \in V, \|x\| = 1\}, \\ b &:= \sup\{ \big| f(x) \big| : x \in V, \|x\| \leq 1\}, \\ c &:= \sup\{ \frac{\big| f(x) \big|}{\|x\|} : x \in V, x \neq 0\}. \end{aligned}$$

Then, $a \;=\; b \;=\; c$.

PROOF: We introduce the following three sets for this proof:

$$\begin{aligned} A &:= \{ \big| f(x) \big| : x \in V, \|x\| = 1\}, \\ B &:= \{ \big| f(x) \big| : x \in V, \|x\| \leq 1\}, \\ C &:= \Big\{ \frac{\big| f(x) \big|}{\|x\|} : x \in V, x \neq 0 \Big\}. \end{aligned}$$

[160] That is not entirely accurate: There is a notion of "uniform spaces" which generalize the concept of a metric but are less general than topological spaces and there is a notion of uniform continuity for those sets.

Proof that $a = b$:
It follows from $A \subseteq B$ that $a \leq b$. 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 $\|x\|^{-1} \geq 1$, hence

$$\left| f(y) \right| \;=\; \left| f\big(x/\|x\|\big) \right| \;=\; (1/\|x\|) \left| f(x) \right| \;\geq\; \left| f(x) \right| .$$

We conclude 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

$$\left| f(x) \right| /\|x\| \;=\; \left| f(x)/\|x\| \right| \;=\; \left| f\big(x/\|x\|\big) \right| \;=\; \left| f(y) \right| .$$

It follows that the sup over the bigger set C does not exceed the sup over A, hence $c = b$. ■

Definition 13.6 (norm of linear functions). $\boxed{\star}$ Let $f : (V, \|\cdot\|) \to (W, \left|\cdot\right|)$ be a linear function between two normed vector spaces. We denote the quantity $a = b = c$ from lemma 13.1 by $\|f\|$, i.e.,

(13.18)
$$\begin{aligned}\|f\| \;&=\; \sup\{ \left| f(x) \right| \,: x \in V, \|x\| = 1\}\\ &=\; \sup\{ \left| f(x) \right| \,: x \in V, \|x\| \leq 1\}\\ &=\; \sup\{ \frac{\left| f(x) \right|}{\|x\|} : x \in V, x \neq 0\}.\end{aligned}$$

$\|f\|$ is called the **norm of the linear function** f.

The justification for calling $f \mapsto \|f\|$ a norm [161] will be given in thm.13.5 on p.413.
We note that $\|f\|$ need not be finite. □

Theorem 13.4 (Continuity criterion for linear functions)**.**

Let $f : (V, \|\cdot\|) \to (W, \left|\cdot\right|)$ 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, such a continuous linear function satisfies the inequality

(13.19)
$$\left| f(x) \right| \;\leq\; \|f\| \cdot \|x\| , \qquad \textit{for all } x \in V.$$

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

(13.20)
$$\text{if } z \in V \text{ and } \|z\| < \delta \;\text{ then } \left| f(z) \right| \;=\; \left| f(z) - f(0) \right| < 1.$$

[161] 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!**

Let $x \in V$ such that $\|x\| \leq 1$. Then $\|\delta/2 \cdot x\| \leq \delta/2 < \delta$, hence, according to (13.20),

$$\delta/2 \cdot \big| f(x) \big| \;=\; \big| f(\delta/2 \cdot x) \big| \;<\; 1, \quad \text{hence} \quad \big| f(x) \big| < 2/\delta.$$

Because this last inequality is true for all $x \in V$ with norm bounded by 1, it follows that

$$\|f\| = \sup\{\, \big| f(x) \big| \; : x \in V, \|x\| \leq 1\} \;<\; 2/\delta < \infty.$$

We have proved that **A** ⇒ **D**.

We finally show **D** ⇒ **C** and we do this in two steps.

First we show **D** ⇒ (13.19). 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 (13.18).

Second step: Let $\varepsilon > 0$ and $\delta := \varepsilon / \|f\|$. Let $x, y \in V$ such that $\|x - y\| < \delta$. If we can prove that this implies $\big| f(x) - f(y) \big| < \varepsilon$, then f is indeed uniformly continuous and the proof is done. We show this as follows.

$$\big| f(x) - f(y) \big| \;=\; \big| f(x-y) \big| \;\overset{(13.19)}{\leq}\; \|f\| \cdot \|x - y\| \;<\; \|f\| \cdot \delta \;=\; \|f\| \cdot \varepsilon / \|f\| \;=\; \varepsilon. \quad \blacksquare$$

Theorem 13.5 ($\|f\|$ is a norm). $\boxed{\star}$ *Let*

(13.21)
$$\mathscr{C}_{lin}(V, W) \;:=\; \mathscr{C}_{lin}\big((V, \|\cdot\|), (W, \boldsymbol{|}\cdot\boldsymbol{|})\big) \;:=\; \{f : V \to W \;\; : f \textit{ is linear and continuous}\,\}.$$

Then, $\mathscr{C}_{lin}(V, W)$ *is a vector space and*

(13.22)
$$f \mapsto \|f\| \;=\; \sup\{\, \big| f(x) \big| \; : \|x\| = 1\}$$

defines a norm on $\mathscr{C}_{lin}(V, W)$.

PROOF:

In all of this proof let $A := x \in V : \|x\| = 1$.

(A) Proof that $\mathscr{C}_{lin}(V, W)$ is a vector space.

Let $f, g, \in \mathscr{C}_{lin}(V, W)$. We need to show that $f + g \in \mathscr{C}_{lin}(V, W)$, i.e., $f + g$ is both linear and continuous. Linearity is immediate. We now show continuity.

Let $x \in A$. Then

(13.23)
$$\big| f(x) + g(x) \big| \;\leq\; \big| f(x) \big| \;+\; \big| f(x) \big| \;\leq\; \|f\| + \|g\| \;<\; \infty.$$

The first inequality holds because the norm $\big| f(x) \big|$ satisfies the triangle inequality for norms. The second follows from (13.18) on p.412, and the finiteness of $\|f\| + \|g\|$ is, according to the continuity criterion for linear functions (thm.13.4 on p.412), equivalent to the continuity of both f and g.

We still must show that if $f \in \mathscr{C}_{lin}(V, W)$ and $\lambda \in \mathbb{R}$ then $\lambda f : x \mapsto \lambda f(x) \in \mathscr{C}_{lin}(V, W)$, i.e., we must show that this function is linear and continuous. Again, linearity is immediate. To show continuity we proceed as follows.

Let $x \in A$. $\boldsymbol{|}\cdot\boldsymbol{|}$ is absolutely homogeneous. Hence

(13.24)
$$\big| \lambda f(x) \big| \;=\; \big| |\lambda| f(x) \big| \;=\; |\lambda| \cdot \big| f(x) \big| \,.$$

It follows from prop.9.10 (positive homogeneity of inf and sup) on p.260 that

$$\|\lambda f\| = \sup\{\left|\lambda f(x)\right| : \|x\| = 1\} = \sup\{\left||\lambda| f(x)\right| : \|x\| = 1\} \tag{13.25}$$

$$= |\lambda| \cdot \sup\{\left|f(x)\right| : \|x\| = 1\} = |\lambda| \cdot \|f\| < \infty. \tag{13.26}$$

This proves that λf is continuous.

(B) Proof that $\|f\|$ is a norm on $\mathscr{C}_{fin}(V, W)$.
Because (13.23) is valid for all $x \in A$, we obtain

$$\|f + g\| = \sup\{\left|f(x) + g(x)\right| : x \in A\} \leq \|f\| + \|g\|. \tag{13.27}$$

This proves the triangle inequality.
Likewise, we obtain from the validity of (13.25) for all $x \in A$,

$$\|\lambda f\| = \sup\{\left||\lambda| f(x)\right| : x \in A\} = |\lambda| \sup\{\left|f(x)\right| : x \in A\} = |\lambda| \|f\|. \tag{13.28}$$

This proves absolute homogeneity.
Finally we show positive definiteness. Clearly $\|f\|$ is nonnegative as the sup of nonnegative numbers $\left|f(x)\right|$. Assume that $\|f\| > 0$. Then $\delta := \frac{1}{2}\|f\| > 0$ and there exists $x_0 \in A$ such that

$$\sup\{\left|f(x)\right| : x \in A\} - \left|f(x_0)\right| < \delta, \text{ i.e., } \|f\| - \left|f(x_0)\right| < \delta, \text{ hence } \left|f(x_0)\right| > \delta. \tag{13.29}$$

Positive definiteness of $\left|\cdot\right|$ implies that $f(x_0) \neq 0$ and hence $f \neq 0$. We have proved positive definiteness of $\|\cdot\|$. ■

13.2 Function Sequences and Infinite Series

13.2.1 Convergence of Function Sequences

Notation 13.1 (Functions with argument "$\cdot$")**.**

This chapter makes heavy use of the notation $f(\cdot)$ instead of f for a function $X \to \mathbb{R}$ to emphasize when sequences of functions $f_n(\cdot)$ are used and when function values (real numbers) $f_n(x)$ are used. □

Vectors are more complicated than numbers because an n–dimensional vector $v \in \mathbb{R}^n$ represents a list of only finitely many real numbers. Any such vector $(x_1, x_2, x_3, \cdots, x_n)$ can be interpreted as a real–valued function (remember: a real–valued function is one which maps it arguments into $\mathbb{R}$)

$$f(\cdot) : \{1, 2, 3, \cdots, n\} \to \mathbb{R} \qquad j \mapsto x_j \tag{13.30}$$

(see (11.4) on p.324).
Next come sequences $(x_j)_{j \in \mathbb{N}}$ which can be interpreted as real–valued functions

$$g(\cdot) : \mathbb{N} \to \mathbb{R} \qquad j \mapsto x_j. \tag{13.31}$$

Finally we deal with real–valued functions

$$h(\cdot) : X \to \mathbb{R} \qquad x \mapsto h(x) \tag{13.32}$$

which are defined on an arbitrary domain X as the most general case.

Now we add more complexity by not just dealing with one or two or three real–valued functions but with an entire sequence of functions

$$f_n(\cdot) : X \to \mathbb{R} \qquad x \mapsto f_n(x) \tag{13.33}$$

For any fixed argument x_0 we have a sequence $f_1(x_0), f_2(x_0), f_3(x_0), \cdots$ of real numbers which we can examine for convergence. This sequence may converge for some or all arguments $x_0 \in X$ to some limit $L = L(x_0) \in \mathbb{R}$. [162] Examination of the limit behavior of a function sequence is not only of interest if those functions are real–valued but also if their codomain is a metric space (Y, d).

It is time now for some definitions.

Definition 13.7 (Pointwise convergence of function sequences)**.**

Let X be a nonempty 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 nonempty subset of X.
We say that $f_n(\cdot)$ **converges pointwise** or, simply, **converges** to $f(\cdot)$ on A and we write $f_n(\cdot) \to f(\cdot)$ on A as $n \to \infty$, or simply $f_n(\cdot) \to f(\cdot)$ on A, if

$$f_n(x) \to f(x) \ \text{ as } n \to \infty \ \text{ for all } x \in A. \tag{13.34}$$

We omit the phrase "on A" if it is clear how A is defined, in particular if $A = X$. □

Definition 13.8 (Uniform convergence of function sequences)**.** Let X be a nonempty set, (Y, d) a metric space, let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$, and let $A \subseteq X$.

We say that $f_n(\cdot)$ **converges uniformly** to $f(\cdot)$ on A and we write

$$f_n(\cdot) \xrightarrow{uc} f(\cdot) \text{ on } A\,, \tag{13.35}$$

if, for each $\varepsilon > 0$ (no matter how small), there exists an index n_0 which can be chosen once and for all, <u>independently of the specific argument x</u>, such that

$$d(f_n(x), f(x)) < \varepsilon \quad \text{for all } x \in A \quad \text{and } n \geq n_0. \tag{13.36}$$

We omit the phrase "on A" if it is clear how A is defined, in particular if $A = X$. □

Note that the notation " $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ " is not very widely used.

Remark 13.12 (Uniform convergence implies pointwise convergence)**.** Take another look at definition Definition 12.10 (convergence of sequences in metric spaces) on p.369. Note that (13.36) implies, for any given $x \in A$, ordinary convergence $f(x) = \lim\limits_{n\to\infty} f_n(x)$. The reason is that the number $n_0 = n_0(\varepsilon)$ chosen in (13.36) will also satisfy (12.20) of that definition 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 $x \in A$. □

[162] We previously examined sequences of functions in ch.9.9 (Sequences of Sets and Indicator functions and their liminf and limsup) on p.296.

Example 13.4 (Constant sequence of functions)**.** Let X be a set and let $f : X \to \mathbb{R}$ be a real–valued 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!).

We obtain $d(f_n(x), f(x)) = 0 < \varepsilon$ for all $x \in X$ and $\varepsilon > 0$. Thus (13.36) in the definition of uniform convergence is trivially satisfied, hence $f_n(\cdot) \overset{uc}{\to} f(\cdot)$. □

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

$$|f_n(x) - f(x)| = 0 < \varepsilon \quad \text{for all } x \in A \text{ and } n > n_0 \;\; \blacksquare$$

The next proposition furnishes an example of a function sequence that converges pointwise, but not uniformly. The reader is suggested to draw a picture of the functions f_n.

Proposition 13.6. *Let* $X = [0, 1]$*,*
i.e., X *is the closed unit interval* $\{x \in \mathbb{R} : 0 \le x \le 1\}$ *. Let the functions* f_n *be defined as follows on* X*:*

$$f_n(x) = \begin{cases} n^2 x & \textit{for } 0 \le x \le \frac{1}{n} \\ \frac{1}{x} & \textit{for } \frac{1}{n} \le x \le 1 \end{cases}$$

Let the function $f(\cdot) : [0, 1] \to \mathbb{R}$ *be defined as follows.*

$$f(x) = \begin{cases} \frac{1}{x} & \textit{for } 0 < x \le 1 \\ 0 & \textit{for } x = 0 \end{cases}$$

Then the functions $f_n(\cdot)$ *converge pointwise but not uniformly to* $f(\cdot)$ *on the entire unit interval.* □

PROOF:

Before we start, note that both pieces of f_n fit together in the point $x = 1/n$ because the "$\frac{1}{x}$ definition" gives $f_n(a) = \frac{1}{1/n} = n$ and the "n^2x definition" gives the same value $n = n^2\frac{1}{n}$. We encourage you to draw a picture to convince yourself that $f_n(\cdot)$ is continuous at every point of $[0, 1]$. You are asked in exercise 13.3 on p.432 to give a proof of the continuity of f_n. Finally note that the limit function f is not continuous at all points of $[0, 1]$.

PROOF of pointwise convergence:
first we inspect the point $a = 0$. We have $f(0) = 0 = n^2 \cdot 0 = f^n(0)$ and the constant sequence of zeros certainly converges to zero. Now assume $a > 0$. If $n > 1/a$ then $f_n(a) = \frac{1}{a}$ for all such n. We have a constant sequence $(\frac{1}{a})$ except for the first finitely many n and this sequence converges to $\frac{1}{a} = f(a)$. See cor.9.4 on p.270. We have thus proved pointwise convergence.

PROOF that there is no uniform convergence:
To prove that (13.36) 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 $j > N$ such that $|f_j(x_N) - f(x_N)| \geq \varepsilon$. Let $N \in \mathbb{N}$ be any natural number and let $x_N := \frac{1}{N^2}$. Then

$$\begin{aligned} f_N(x_N) &= \frac{N^2}{N^2} = 1, \\ f_{2N}(x_N) &= \frac{(2N)^2}{N^2} = 4. \end{aligned}$$

Hence

$$\left|f_{2N}(x_N) - f_N(x_N)\right| = 3.$$

To recap: We found $\varepsilon > 0$ so that for each $N \in \mathbb{N}$ there is at least one $j \geq N$ and $x_N \in [0,1]$ such that $|f_j(X_N) - f_N(x_N)| > \varepsilon$: we chose

$$\varepsilon = 2, \quad j = 2N, \quad x_N = \frac{1}{N^2}$$

We have proved that convergence is pointwise but not uniform. ■

Let X be a nonempty set and $\mathscr{B}(X,\mathbb{R})$ the set of all bounded real–valued functions on X. We recall from Theorem 13.5 ($\|f\|$ is a norm) on p.413, that $\mathscr{B}(X,\mathbb{R})$ is a vector space with the norm

$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}$$

and it is a metric space with the corresponding metric

$$d_{\|\cdot\|_\infty}(f,g) = \sup\{|g(x) - f(x)| : x \in X\}$$

(see example 12.2 on p.361).

Proposition 13.7 (Uniform convergence is $\|\cdot\|_\infty$ convergence)**.**

The following is true for any a nonempty set X and $f_n, f \in \mathscr{B}(X,\mathbb{R})$:

$$\begin{aligned} f_n(\cdot) \overset{uc}{\to} f(\cdot) \quad &\Leftrightarrow \quad f_n(\cdot) \overset{\|\cdot\|_\infty}{\to} f(\cdot)\,, \quad \textit{i.e.,} \\ f_n(\cdot) \overset{uc}{\to} f(\cdot) \quad &\Leftrightarrow \quad f_n \textit{ converges to } f \textit{ in the metric space } \left(\mathscr{B}(X,\mathbb{R}), d_{\|\cdot\|_\infty}(\cdot,\cdot)\right). \end{aligned}$$

PROOF of "⇒": Assume that $f_n(\cdot) \overset{uc}{\to} f(\cdot)$. Let $\varepsilon > 0$. According to Definition 13.8 (Uniform convergence of function sequences) on p.415, there exists an index $n_0 = n_0(\varepsilon)$ (which does not depend on the function argument $x \in X$) such that

$$d(f_n(x), f(x)) = |f_n(x) - f(x)| < \varepsilon/2 \quad \text{for all } x \in X \quad \text{and } n \geq n_0.$$

Note that here the metric space Y in Definition 13.8 is $\mathbb{R}$, so $d(f_n(x), f(x))$ becomes $|f_n(x) - f(x)|$. We obtain

$$\|f_n - f\|_\infty = \sup\{|f_n(x) - f(x)| : x \in X\} \leq \varepsilon/2 \text{ for all } n \geq n_0,$$

i.e., $d_{\|\cdot\|_\infty}(f_n, f) < \varepsilon$ for all $n \geq n_0$. It follows that $f_n(\cdot) \overset{\|\cdot\|_\infty}{\to} f(\cdot)$.

PROOF of "⇐": Assume that $f_n \overset{\|\cdot\|_\infty}{\to} f$, i.e., $\lim\limits_{n\to\infty} f_n = f$ in the metric space $\left(\mathscr{B}(X,\mathbb{R}), d_{\|\cdot\|_\infty}\right)$.

Let $\varepsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that

$$d_{\|\cdot\|_\infty}(f_n, f) \;=\; \|f_n - f\|_\infty \;=\; \sup\{|f_n(x) - f(x)| : x \in X\} \;<\; \varepsilon \quad \text{for all } n \geq n_0$$

But then

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } x \in X \text{ and all } n \geq n_0.$$

This proves $f_n(\cdot) \overset{uc}{\to} f(\cdot)$. ■

The last proposition justifies the next definition.

Definition 13.9 (Norm and metric of uniform convergence). $\boxed{\star}$

We also call the sup–norm on $\mathscr{B}(X, \mathbb{R})$ the **norm of uniform convergence** on X and its associated metric $d_{\|\cdot\|_\infty}(\cdot, \cdot)$ the **metric of uniform convergence** on X. □

Theorem 13.6 (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 the following:

(a) *The functions $f_n(\cdot)$ are continuous at x_0 for all n.*

(b) *$f_n(\cdot) \overset{uc}{\to} f(\cdot)$ on V.*

Then, f is continuous at x_0

PROOF: Let $\varepsilon > 0$.

(A) Uniform convergence $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ on V guarantees the existence of some $N = N(\varepsilon)$ such that

$$d_2\big(f_n(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V \text{ and } n \geq \mathbb{N}.$$

In particular, for $n = N$,

$$d_2\big(f_N(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V\,. \tag{13.37}$$

(B) All functions f_n and in particular f_N are continuous in V. There is $\tilde{\delta} > 0$ such that

$$d_2(f_N(x), f_N(x_0)) < \frac{\varepsilon}{3} \text{ for all } x \in N_{\tilde{\delta}}(x_0). \tag{13.38}$$

(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 (13.37) and (13.38) both hold for any $x \in N_\delta(x_0)$. Because $x_0 \in N_\delta(x_0)$ we obtain

$$d(f(x), f(x_0)) \;\leq\; d(f(x), f_N(x)) \;+\; d(f_N(x), f_N(x_0)) \;+\; d(f_N(x_0), f(x_0)) \;<\; \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \;=\; \varepsilon.$$

The proof is finished. ■

For an example of uniform convergence we return to the n–th Bernstein Polynomials

$$B_n^f(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k},$$

which are defined for any $f : [0,1] \to \mathbb{R}$. It will be shown in ch.16.3 (The Weierstrass Approximation Theorem), that if f is **any** continuous function on the unitinterval, then $B_n^f(\cdot) \overset{uc}{\to} f(\cdot)$ on $[0,1]$ as $n \to \infty$.

We have done already most of the work to prove this for the three continuous functions $x \to 1$, $x \to x$, and $x \to x^2$.

Proposition 13.8. $\boxed{\star}$

Let $f : [0,1] \to \mathbb{R}$ be one of the functions

$$1 : x \mapsto 1; \qquad id : x \mapsto x; \qquad id^2 : x \mapsto x^2; \qquad (0 \leq x \leq 1).$$

Then,

$$B_n^f(\cdot) \overset{uc}{\to} f(\cdot) \text{ on } [0,1] \text{ as } n \to \infty.$$

PROOF: We derived in prop.6.23 on p.188 the formulas

$$B_n^1(x) \ = 1, \qquad B_n^{id}(x) \ = x, \qquad B_n^{id^2}(x) \ = \ \frac{1}{n}x \ + \ \frac{n-1}{n}x^2 \qquad\qquad (x \in \mathbb{R}).$$

Note that $\left(B_n^1\right)_n$ is the constant function sequence $p_n^1(\cdot) = 1$, and $\left(B_n^{id}\right)_n$ is the constant function sequence $B_n^{id}(\cdot) = id$. We have seen in example 13.4 (Constant sequence of functions) on p.416 that any constant function sequence has itself as uniform limit, thus the proposition is true for the functions 1 and id.

The function $id^2 : x \mapsto x^2$ needs a little more work. Let $\varepsilon > 0$. If $0 \leq x \leq 1$ then

$$\begin{aligned} d\big(B_n^{id^2}(x), id^2(x)\big) \ &= \left| \left(\frac{1}{n}x \ + \ \frac{n-1}{n}x^2 \right) \ - \ x^2 \right| \\ &= \left| \frac{1}{n}x \ - \ \frac{1}{n}x^2 \right| \ = \ \frac{1}{n} \cdot |x| \cdot |1-x| \ \leq \ \frac{1}{n}. \end{aligned}$$

We choose $n_0 \in \mathbb{N}$ such that $n_0 > \frac{1}{\varepsilon}$. This is always possible since the natural numbers are not bounded above in $\mathbb{R}$. Let $n \geq n_0$. Then $\frac{1}{n} \leq \frac{1}{n_0} < \varepsilon$, hence $d\big(B_n^{id^2}(x), id^2(x)\big) < \varepsilon$ for all $x \in [0,1]$. It follows that (13.36) in the definition of uniform convergence is satisfied, hence $B_n^{id^2} \overset{uc}{\to} id$. ■

Proposition 13.9.

Let X be a nonempty set, (Y,d) a metric space and let $f_n, f : X \to Y \ \ (n \in \mathbb{N})$. Then,

f is the uniform limit of the function sequence $(f_n)_n$

$\Leftrightarrow$ *there exists a sequence $\delta_n \geq 0$ such that* **1)** *$\delta_n \to 0$ as $n \to \infty$, and*

2) *$d\big(f_n(x), f(x)\big) \leq \delta_n$ for all $x \in X$ and $n \in \mathbb{N}$.*

PROOF:

(A) First we prove that uniform convergence $f_n \overset{uc}{\to} f$ implies that there are real numbers $\delta_n \geq 0$ that satisfy both **(1)** and **(2)**: It follows from Definition 13.8 on p.415 (Uniform convergence of function sequences) that the numbers $\delta_n := \sup\{d\big(f_n(x), f(x)\big) : x \in X\}$ converge to zero and thus define such a sequence.

(B) We now prove that the existence of a sequence $\delta_n \geq 0$ that satisfies both **(1)** and **(2)** implies $f_n \overset{uc}{\to} f$ on X. Let $\varepsilon > 0$. It follows from $\lim\limits_{k\to\infty} \delta_k = 0$ that there exists $n_0 \in \mathbb{N}$ such that $\delta_k < \varepsilon$ for all $k \geq n_0$. Thus $d\big(f_k(x), f(x)\big) \leq \delta_k < \varepsilon$ for all $x \in X$ and $k \geq n_0$. It follows from Definition 13.8 that $f_n \overset{uc}{\to} f$ on X. ■

13.2.2 Infinite Series

We start by repeating the definition of a sequence given in section 5.2 on p.131: (x_j) is nothing but a family of things x_j which are indexed by a consecutive set of integers, usually the natural numbers or the nonnegative integers. We make throughout this chapter on infinite series the following assumption:

Unless explicitly stated otherwise, sequences are always indexed $1, 2, 3, \dots$, i.e., the first index is 1 and, given any index, you obtain the next one by adding 1 to it. □

Proposition 13.10 (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:

(13.39a) $$\left|\sum_{k=n}^{\infty} a_k\right| < \varepsilon \qquad \textit{for all } n \geq n_0$$

(13.39b) $$\left|\sum_{k=n}^{m} a_k\right| < \varepsilon \qquad \textit{for all } m, n \geq n_0$$

PROOF: Write

(13.40) $$s \;=\; \sum_{k=1}^{\infty} a_k \;=\; \sum_{k=1}^{n} a_k + \sum_{k=n+1}^{\infty} a_k \;=\; s_n + \sum_{k=n+1}^{\infty} a_k$$

Remember the convergence criteria for real–valued 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. Expressed in terms of the distance to s this means there exists a suitable $n_0 \in \mathbb{N}$ such that

$$|s - s_n| < \varepsilon \qquad \text{for all } n \geq n_0$$

(see (12.10) on p.369). According to (13.40) we can write that as

$$\left|\sum_{k=n+1}^{\infty} a_k\right| < \varepsilon \qquad \text{for all } n \geq n_0,$$

which is the same as (13.39.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

$$|s_i - s_j| < \delta \text{ for all } i, j \geq n_0$$

(see (12.27) on p.390) which ensures convergence to some number s without specifying what it might actually be. Again we use (13.40) and obtain, assuming without loss of generality that $i < j$,

$$\Big|\sum_{k=i+1}^{j} a_k\Big| < \delta \text{ for all } j > i \geq n_0 \;\blacksquare$$

Corollary 13.1. *If a series $\sum a_j$ converges then $\lim\limits_{n\to\infty} a_n = 0$.*

PROOF: Let $\varepsilon > 0$. It follows from 13.39b that there is some $n_0 \in \mathbb{N}$ such that $|a_m - 0| = \Big|\sum\limits_{k=m}^{m} a_k\Big| < \varepsilon$ for all $m \geq n_0$. But this means that the sequence a_n converges to zero. $\blacksquare$

Here is a second corollary.

Corollary 13.2 (Dominance criterion for series [163]).

Let $N \in \mathbb{N}$ and let $\sum a_j$ and $\sum b_j$ be two series such that $|b_k| \leq a_k$ for all $k \geq N$.
It follows that if $\sum a_k$ converges, then $\sum b_k$ converges.
Moreover, if $|b_k| \leq a_k$ for all $k \in \mathbb{N}$, then, $\Big|\sum\limits_{k=1}^{\infty} b_j\Big| \leq \sum\limits_{k=1}^{\infty} a_j$

PROOF: Let $\varepsilon > 0$. It follows from 13.39b that there is some $n_0 \in \mathbb{N}$ such that $\Big|\sum\limits_{k=m}^{n} a_k\Big| < \varepsilon$ for all $m, n \geq n_0$. Let $M := \max(n_0, N)$. We obtain

$$\Big|\sum_{k=i+1}^{j} b_j\Big| \leq \sum_{k=i+1}^{j} |b_j| \leq \sum_{k=i+1}^{j} a_j < \varepsilon \quad \text{for all } j > i \geq M.$$

We conclude from (13.39b) that $\sum b_k$ converges.

Now assume that $|b_k| \leq a_k$ for all $k \in \mathbb{N}$. Let

$$s_n := \sum_{k=1}^{n} |a_k|, \quad s := \lim_{n\to\infty} s_n, \qquad t_n := \sum_{k=1}^{n} b_k, \quad t := \lim_n t_n \,.$$

It follows from the triangle inequality for real numbers that $|t_n| \leq s_n$ for all $n \in \mathbb{N}$. We apply prop.9.19 on p.269 to deduce that

$$|t| = \lim_n |t_n| \leq \lim_n s_n = s\,.$$

[163] This is a generalization of [2] B/G (Beck/Geoghegan) prop.12.3, p.115.

This completes the proof. ■

Remark 13.13. 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 $\sum\limits_{k=1}^{\infty} a_k \ = \infty$ or $\sum\limits_{k=1}^{\infty} a_k \ = -\infty$ or there is no limit at all. As an example for a series which has no limit, look at the oscillating sequence

$$a_0 = 1; \quad a_1 = -1; \quad a_2 = 1; \quad a_3 = -1; \cdots \quad s_n = \sum_{k=0}^{n} (-1)^n \tag{13.41}$$

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. [164]

Just so you understand the difference between limits and contact points (see (Definition 12.23) on p.382): Even though neither $(a_j)_j$ nor $(s_j)_j$ has a limit, the tail sets for both have two contact points each. The ones for $(a_j)_j$ have the contact points $\{1, -1\}$ and the ones for $(s_j)_j$ have the contact points $\{0, 1\}$. □

We now turn our attention to convergence properties of series. We copy from Chapter 7 (Cardinality I: Finite and Countable Sets) the notation $[N] = [1, N]_{\mathbb{N}}$ for $N \in \mathbb{N}$.

Definition 13.10 (Finite permutations). $\boxed{\star}$

Let $N \in \mathbb{N}$. A **permutation** of $[N]$ is a bijection

$$\pi(\cdot) : [N] \to [N]; \qquad j \mapsto \pi(j)\,.$$

As usual

$$\pi^{-1}(\cdot) : [N] \to [N]; \qquad \pi(j) \mapsto \pi^{-1}\pi(j) = j,$$

denotes the inverse function of $\pi(\cdot)$. We recall that it associates with each image $\pi(j)$ the unique argument j, which is mapped by $\pi(\cdot)$ to $\pi(j)$. It is customary to write

$$i_1 \text{ instead of } \pi(1), \quad i_2 \text{ instead of } \pi(2), \quad \ldots, \quad i_j \text{ instead of } \pi(j), \quad \ldots\,. \ \square$$

We extend the previous definition from $[N]$ to $\mathbb{N}$.

Definition 13.11 (Permutations of $\mathbb{N}$).

A **permutation** of $\mathbb{N}$ is a bijective function

$$\pi(\cdot) : \mathbb{N} \to \mathbb{N}; \qquad j \mapsto \pi(j). \ \square$$

[164] We could also have concluded as follows: $|s_j - s_{j-1}| = 1$ implies that the Cauchy formulation of the convergence criteria for series (see (13.39a) on p.420) is not satisfied, hence no convergence of the series.

Permutations are the means of describing 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 13.11.

Let (a_n) be a sequence of nonnegative real numbers. Exactly one of the following is true:

(a) *Either the series $\sum a_n$ converges (to a finite number). In that case,*

$$\sum_{n=1}^{\infty} a_n \;=\; \sum_{n=1}^{\infty} a_{\pi(n)} \quad \textit{for } \textbf{\textit{any}} \textit{ permutation } \pi(\cdot) \textit{ of } \mathbb{N}.$$

(b) *Or 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

(13.42) $$\alpha \;:=\; \max\{\pi(j) : j \leq N\} \qquad \text{and} \qquad \beta \;:=\; \max\{\pi^{-1}(k) : k \leq N\}.$$

Note that $\alpha \geq N$ and $\beta \geq N$. Because all terms a_j, b_k are nonnegative it follows that

$$\sum_{j=1}^{N} b_j \;=\; \sum_{j=1}^{N} a_{\pi(j)} \;\leq\; \sum_{k=1}^{\alpha} a_k \;\leq\; \sum_{k=1}^{\alpha} a_k \;+\; \sum_{k=\alpha+1}^{\infty} a_k \;=\; \sum_{k=1}^{\infty} a_k,$$
$$\sum_{k=1}^{N} a_k \;=\; \sum_{k=1}^{N} b_{\pi^{-1}(k)} \;\leq\; \sum_{j=1}^{\beta} b_j \;\leq\; \sum_{j=1}^{\beta} b_j \;+\; \sum_{j=\beta+1}^{\infty} b_j \;=\; \sum_{j=1}^{\infty} b_j.$$

We take limits as $N \to \infty$ and it follows from prop.9.19 on p.269 that

$$\sum_{j=1}^{\infty} b_j \;\leq\; \sum_{k=1}^{\infty} a_k \quad \text{and} \quad \sum_{k=1}^{\infty} a_k \;\leq\; \sum_{j=1}^{\infty} b_j, \quad \text{hence} \quad \sum_{k=1}^{\infty} a_k \;=\; \sum_{j=1}^{\infty} b_j.$$

This proves part **(a)** of the proposition.

PROOF of **(b)**: Assume that $\sum a_j$ diverges. Because all terms a_j are nonnegative, the sequence s_n of the partial sums is nondecreasing 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 $s \geq a_0 \geq 0$. 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 13.12 (absolutely convergent series)**.**

A series $\sum a_j$ is **absolutely convergent**, if the corresponding series $\sum |a_j|$ of its absolute values converges. □

Proposition 13.12.

Let $\sum a_k$ be an absolutely convergent series. Then $\sum a_k$ converges and

$$|\sum_{k=1}^{\infty} a_k| \leq \sum_{k=1}^{\infty} |a_k|. \tag{13.43}$$

PROOF: This follows from the dominance criterion (cor.13.2) ■

It follows from prop.13.11 on p.423 that if a series of nonnegative terms converges then its value is invariant under rearrangements of that series. The next theorem states that any absolutely convergent series also has that property and we will see later [165] that the reverse is also true: Any series whose value is invariant under rearrangements is absolutely convergent.

Theorem 13.7. [166]

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 $\sum a_k$. (Note that $\sum a_k$ converges according to Proposition 13.12.)

PROOF: As a first step we prove that $\sum b_k$ converges: Since $\sum |b_k|$ is a rearrangement of $\sum |a_k|$, $\sum |b_k|$ converges by Proposition 13.11 on p.423. By Proposition 13.12, $\sum b_k$ converges.

Let $s := \sum\limits_{k=1}^{\infty} a_k$, $t := \sum\limits_{k=1}^{\infty} b_k$. For $n \in \mathbb{N}$ let $s_n := \sum\limits_{k=1}^{n} a_k$ and $t_n := \sum\limits_{k=1}^{n} b_k$.

Let $\varepsilon > 0$. Since $\sum |a_k|$ converges, there exists $n_0 \in \mathbb{N}$ such that

$$\sum_{k=n_0+1}^{n_0+m} |a_k| \leq \sum_{k=n_0+1}^{\infty} |a_k| < \varepsilon \quad \text{for all } m \in \mathbb{N}. \tag{13.44}$$

Let $A := \{\pi(j) : 1 \leq j \leq n_0\}$ and $p_0 := \max(A)$. This maximum exists because the set A is finite.

Then $p_0 \geq n_0$. 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 $n, p \geq p_0$. 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.

$p_0 = \max(A)$ is 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 from (13.44) for big enough $m \in \mathbb{N}$ (the bigger of

[165] see cor.13.4 on p.432

[166] This was proved by the German mathematician Peter Gustav Lejeune Dirichlet (1805-1859).

$\max(\{\pi(j) : 1 \leq j \leq n\})$ and p) that

$$|s_n - t_p| \leq \sum_{k=n_0+1}^{n_0+m} |a_k| \ < \ \varepsilon.$$

This implies

$$|s - t_p| \leq |s - s_n| + |s_n - t_p| \ \leq \ |s - s_n| \ + \sum_{k=n_0+1}^{n_0+m} |a_k| \ < \ |s - s_n| + \varepsilon.$$

We had chosen $n \geq n_0$ and it follows from (13.44) that $|s - s_n| < \varepsilon$, hence $|s - t_p| < 2\varepsilon$.
But p could be any integer $\geq p_0$, and p_0 only depends (via n_0) on ε.
To summarize: for all $\varepsilon > 0$ there exists p_0 such that $p \geq p_0$ implies $|s - t_p| < 2\varepsilon$. But then $\lim\limits_{p\to\infty} t_p \ = \ s$.
On the other hand, $\lim\limits_{p\to\infty} t_p \ = \ t \ = \ \sum\limits_{p\to\infty} b_k$.

This concludes the proof that $\sum\limits_{p\to\infty} a_k \ = \ \sum\limits_{p\to\infty} b_k$. ■

Proposition 13.13.

Let $\sum a_n$ be an absolutely convergent series. Let $(a_{n_k})_k$ be a subsequence of $(a_n)_n$. Then, $\sum a_{n_k}$ converges absolutely.

PROOF: The proof is left as exercise 13.10. ■

The last proposition allows us to use the following simplified summation notation for absolutely convergent series.

Remark 13.14.

Assume that $\sum a_n$ is absolutely convergent. Let $n_1 < n_2 < \dots$ be a subsequence of all natural numbers and let $J := \{n_j : j \in \mathbb{N}\}$.

- Then we write $\sum\limits_{j \in J} a_{n_j} \ := \ \sum\limits_{j=1}^{\infty} a_{n_j}$.
- In particular, we write $\sum\limits_{j \in \mathbb{N}} a_j \ := \ \sum\limits_{j=1}^{\infty} a_j$, for the full sequence $n_j = j$ of indices. □

There are series which are convergent but not absolutely convergent. They are given a special name:

Definition 13.13 (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 13.14 (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 13.14 (Leibniz Test for Alternating Series).

Let $a_1 \geq a_2 \geq \ldots \downarrow 0$ be a nonincreasing sequence which decreases to zero. Then, the alternating series $\sum(-1)^k a_k$ converges.

PROOF: For each $n \in \mathbb{N}$ we have

$$\begin{aligned} s_{2n+1} &= s_{2n-1} + \big(s_{2n} - s_{2n+1}\big) \geq s_{2n-1}, \\ s_{2n+2} &= s_{2n} - \big(s_{2n+1} - s_{2n+2}\big) \leq s_{2n}, \\ s_{2n-1} &\leq s_{2n+1} = \big(s_{2n} - a2n+1\big) \leq s_{2n}. \end{aligned}$$

Hence, if $k, n \in \mathbb{N}$ such that $k \geq n$ then

$$s_{2n+1} \leq s_{2k+1} \leq s_{2k} \leq s_{2n}, \qquad \big|s_{2n} - s_{2n+1}\big| = s_{2n} - s_{2n+1} = a_{2n+1}. \tag{13.45}$$

(13.46)

Let $\varepsilon > 0$. It follows from $\lim\limits_{n\to\infty} a_n = 0$ that there exists $n_0 \in \mathbb{N}$ such that $a_j < \varepsilon$ for all $j \geq n_0$. Let $N := 2n_0 + 1$. Let $i, j \in \mathbb{N}$ such that $i \geq N$. Then either $i = 2k$ or $i = 2k+1$ for some suitable natural number $k \geq n_0$. Likewise, either $j = 2'$ or $j = 2k'+1$ for some suitable natural number $k' \geq n_0$. It follows from (13.45) that $s_{2n_0+1} \leq s_i, s_j \leq s_{2n_0}$. Because $|s_{2n_0} - s_{2n_0+1}| = a_{2n_0+1} < \varepsilon$, we have proven that the sequence s_n is Cauchy, hence converges because $\mathbb{R}$ is complete. ■

Example 13.5 (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 Rearrangement 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 will prove the following lemma.

Lemma 13.2. ⋆

Let $\sum a_k$ be a series. We split it into two series $\sum p_k$ and $\sum q_k$ as follows.

- *p_j denotes the jth strictly positive member of the sequence $(a_k)_k$.*
- *q_j denotes the jth strictly negative member of that sequence.*

Then, 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 conditionally convergent, then $\sum p_k$ has limit ∞ and $\sum q_k$ has limit $-\infty$.*

PROOF of **(a)**: Let $\alpha := \sum_{i=1}^{\infty} |a_i|$ and let $j \in \mathbb{N}$.
Let m be the index such that a_m is the jth (not mth!) strictly positive member of the sequence $(a_k)_k$. Then each p_i for $i \leq j$ is some $|a_k|$ for a suitable $k \leq m$. It follows from $m \geq j$ that

$$\sum_{i=1}^{j} p_i \;\leq\; \sum_{i=1}^{m} |a_i| \;\leq\; \sum_{i=1}^{\infty} |a_i| \;<\; \infty.$$

The above is true for all $j \in \mathbb{N}$ and it follows that $\sum\limits_{i=1}^{\infty} p_i \;<\; \infty$. The proof that $\sum q_k$ has a finite limit is similar.

PROOF of **(b)**: The proof will be done in three parts. In part 1 we will show that not both $\sum p_k$ and $\sum q_k$ can converge. In part 2 we will show that $\sum p_k = \infty$ and $\sum q_k \in \mathbb{R}$ leads to a contradiction. In part 3 we will show that $\sum q_k = -\infty$ and $\sum p_k \in \mathbb{R}$ leads to a contradiction.

Part 1: Let us assume that $\sum p_k < \infty$ and $\sum q_k > -\infty$.
For any $n \in \mathbb{N}$ we have

$$\sum_{k=1}^{n} |a_k| \;\leq\; \sum_{k=1}^{n} p_k \;+\; \sum_{k=1}^{n} (-q_k).$$

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 positive numbers $-q_1, \dots, -q_n$ or it is zero, in which case it contributes nothing to the series. Both series $\sum a_k$ and $\sum(-q_k)$ are nondecreasing, hence for each fixed n,

$$\sum_{k=1}^{n} |a_k| \;\leq\; \sum_{k=1}^{\infty} p_k \;-\; \sum_{k=1}^{\infty} q_k.$$

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.

Part 2: Let us assume that $\sum p_k = \infty$ and $\sum q_k \in \mathbb{R}$.
We fix $n \in \mathbb{N}$. Let M_n be the index of p_n, i.e., M_n is the smallest index j such that $a_j = p_n$. Note that

$$a_{M_n} \;=\; p_n \quad (\star) \qquad\qquad \text{and} \quad M_n \geq n. \quad (\star\star)$$

Let

$$I_n \;:=\; \{i \leq M_n : a_i > 0\}, \qquad J_n \;:=\; \{j \leq M_n : a_i < 0\}.$$

Then

$$\sum_{k=1}^{M_n} a_k \;=\; \sum_{i \in I_n} a_k \;+\; \sum_{j \in J_n} a_k \;\overset{(\star)}{=}\; \sum_{i=1}^{n} p_i \;+\; \sum_{j \in J_n} a_k \;\geq\; \sum_{i=1}^{n} p_i \;+\; \sum q_k. \tag{13.47}$$

It follows from $(\star\star)$ that if $n \to \infty$ then $M_n \to \infty$. Hence, from (13.47),

$$\sum a_k \;=\; \lim_{n\to\infty} \sum_{k=1}^{n} a_k \;=\; \lim_{n\to\infty} \sum_{k=1}^{M_n} a_k \;\geq\; \lim_{n\to\infty} \Big(\sum_{i=1}^{n} p_i \;+\; \sum q_k \Big) \;=\; \infty,$$

contrary to the assumption that $\sum a_k$ converges. We have reached a contradiction.

Part 3: Let us assume that $\sum q_k = \infty$ and $\sum p_k \in \mathbb{R}$.
We obtain a contradiction by applying part 2 to the series $\sum(-a_k)$. ■

Theorem 13.8 (Riemann's Rearrangement Theorem). [167]

Let $\alpha, \beta \in \mathbb{R}$ such that $\alpha \leq \beta$. and let the series $\sum a_k$ be conditionally convergent. Then a rearrangement $\sum b_k$ of $\sum a_k$ exists such that

$$\liminf_{n\to\infty} \sum_{k=1}^{n} b_k \;=\; \alpha \qquad \textit{and} \qquad \limsup_{n\to\infty} \sum_{k=1}^{n} b_k \;=\; \beta\,.$$

PROOF: $\boxed{\star}$

We may assume that $a_j \neq 0$ for all $j \in \mathbb{N}$ because terms of value zero do not contribute anything to the partial sums, hence omitting them leaves 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 13.2:

p_j is the jth strictly positive member of the sequence $(a_k)_k$;
q_j is the jth strictly negative member of $(a_k)_k$.

It was proved in lemma 13.2 that $\displaystyle\sum_{k=1}^{\infty} p_k \;=\; \infty$ and $\displaystyle\sum_{k=1}^{\infty} q_k \;=\; -\infty$.

case 1: $\beta \geq 0$.

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 $p_1, p_2, \dots, p_{u_1}$ the **first upcrossing** of the (unfinished) series $\sum b_k$.

We now construct the first piece of the desired rearrangement $\sum b_k$. Let

$$n_1 := u_1; \quad b_1 := p_1,\ b_2 := p_2,\ \dots,\ b_{n_1} := p_{u_1}; \quad \sigma_1 \;:=\; \sum_{j=1}^{n_1} b_j.$$

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 $q_1, q_2, \dots, q_{l_1}$ the **first downcrossing** of $\sum b_k$.

We add more terms to $b_1, b_2, \dots, b_{n_1}$.

$$n_2 := n_1 + l_1; \quad b_{n_1+1} := q_1, b_{n_1+2} := q_2, \dots, b_{n_2} := q_{l_1}; \quad \sigma_2 \;:=\; \sum_{j=1}^{n_2} b_j.$$

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 $U_2 := \left\{k \in \mathbb{N} : k > u_1 \text{ and } \sigma_2 + \displaystyle\sum_{j=u_1+1}^{u_1+k} p_j > \beta\right\}$. U_2 is not empty because $\displaystyle\sum_{j=u_1+1}^{\infty} p_j$ 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 $\sum b_k$.

[167] This was proved by the German mathematician Bernhard Riemann (1826-1866).

We add more terms to $b_1, b_2, \dots, b_{n_2}$.

$$n_3 := n_2 + u_2; \quad b_{n_2+1} := p_{u_1+1},\ b_{n_2+2} := p_{u_1+2}, \dots,\ b_{n_3} := p_{u_1+u_2}; \quad \sigma_3 \ := \sum_{j=1}^{n_3} b_j.$$

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 $L_2 := \Big\{ k \in \mathbb{N} : k > l_1$ and $\sigma_3 + \sum\limits_{j=l_1+1}^{l_1+k} q_j < \alpha \Big\}$. L_2 is not empty because $\sum\limits_{j=l_1+1}^{\infty} q_j$ 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 $b_1, b_2, \dots, b_{n_3}$.

$$n_4 := n_3 + l_2; \quad b_{n_3+1} := q_{l_1+1},\ b_{n_3+2} := q_{l_1+2}, \dots,\ b_{n_4} := q_{l_1+l_2}; \quad \sigma_4 \ := \sum_{j=1}^{n_4} b_j.$$

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

$$\begin{aligned} n_{(2N-1)} &:= n_{(2N-2)} + u_N; \\ b_{(n_{(2N-2)}+1)} &:= p_{(u_{(N-1)}+1)},\ b_{(n_{(2N-2)}+2)} := p_{(u_{(N-1)}+2)},\ \dots,\ b_{(n_{(2N-1)})} := p_{u_N}, \\ \sigma_{(2N-1)} \ &:= \sum_{j=1}^{n_{(2N-1)}} b_j. \end{aligned}$$

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

$$\begin{aligned} n_{(2N)} &:= n_{(2N-1)} + l_N; \\ b_{(n_{(2N-1)}+1)} &:= q_{(l_{(N-1)}+1)},\ b_{(n_{(2N-1)}+2)} := q_{(l_{(N-1)}+2)},\ \dots,\ b_{n_{(2N)}} := q_{l_N}, \\ \sigma_{2N} \ &:= \sum_{j=1}^{n_{(2N)}} b_j. \end{aligned}$$

We proceed to construct the $(N+1)$th upcrossing and the $(N+1)$th downcrossing as follows.

Let $U_{N+1} := \Big\{ k \in \mathbb{N} : k > u_N$ and $\sigma_{2N} + \sum\limits_{j=u_N+1}^{u_N+k} p_j > \beta \Big\}$. U_{N+1} is not empty because $\sum\limits_{j=u_N+1}^{\infty} p_j$ 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 $b_1, b_2, \dots, b_{n_{2N}}$.

$$\begin{aligned} n_{(2N+1)} &:= n_{(2N)} + u_{(N+1)}; \\ b_{(n_{(2N)}+1)} &:= p_{(u_N+1)},\ b_{(n_{(2N)}+2)} := p_{(u_N+2)},\ \dots,\ b_{(n_{(2N+1)})} := p_{u_{(N+1)}}, \\ \sigma_{(2N+1)} \ &:= \sum_{j=1}^{n_{(2N+1)}} b_j. \end{aligned}$$

Let $L_{N+1} := \Big\{k \in \mathbb{N} : k > l_N$ and $\sigma_{2N+1} + \sum\limits_{j=l_N+1}^{l_N+k} q_j < \alpha\Big\}$. L_{N+1} is not empty because $\sum\limits_{j=l_N+1}^{\infty} q_j$ 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 $b_1, b_2, \dots, b_{n_{(2N+1)}}$.

$$\begin{aligned} n_{(2(N+1))} &:= n_{(2N+1)} + l_{(N+1)};\\ b_{(n_{(2N+1)}+1)} &:= q_{(l_N+1)},\ b_{(n_{(2N+1)}+2)} := q_{(l_N+2)},\ \dots,\ b_{(n_{2(N+1)})} := q_{l_{(N+1)}},\\ \sigma_{2(N+1)} &:= \sum_{j=1}^{n_{2(N+1)}} b_j. \end{aligned}$$

We have defined by recursion $\sum\limits_{k=1}^{n_m} b_k$ for all $m \in \mathbb{N}$

We now show that the increasing sequence $(n_m)_{m\in\mathbb{N}}$ is not bounded above. We observe that $n_{(2m)}$ is the number of terms that belong to the first m upcrossings plus the first m downcrossings. 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 $n_{2m} \geq 2m$ and this proves that the sequence $(n_m)_{m\in\mathbb{N}}$ is indeed not bounded above.

It follows that $\sum b_k$ 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. This shows that each one of the terms of $\sum a_k$ has become part of $\sum b_k$ and it follows 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.9.13 (Characterization of limsup and liminf) on p.288 that

$$\liminf_{n\to\inf} s_n \leq \alpha \quad \text{and} \quad \limsup_{n\to\inf} s_n \geq \beta. \tag{13.48}$$

We now prove that for any $\varepsilon > 0$

$$\liminf_{n\to\inf} s_n \geq \alpha - \varepsilon \quad \text{and} \quad \limsup_{n\to\inf} s_n \leq \beta + \varepsilon. \tag{13.49}$$

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.13.1 on p.421). It follows that there exists $n_0 \in \mathbb{N}$ such that $|a_j| < \varepsilon$ for all $j \geq n_0$. We show next that

$$|p_j| \ = \ p_j < \varepsilon \text{ and } |q_j| \ = \ -q_j < \varepsilon \text{ for all } j \geq n_0. \tag{13.50}$$

$|p_j| = p_j < \varepsilon$ is true whenever $j \geq n_0$ because p_j is the jth positive member of $(a_n)_n$, hence $p_j = a_i$ for some $i \geq j \geq n_0$. Likewise, $|q_j| = -q_j < \varepsilon$ whenever $j \geq n_0$ because q_j is the jth negative member of $(a_n)_n$, hence $q_j = a_i$ for some $i \geq j \geq n_0$. We have proved (13.50).

We recall that $n_1, n_3, n_5, \dots$ are precisely the integers n for which $s_n > \beta$, so

$$s_{(n_1-1)} \leq \beta,\ s_{(n_3-1)} \leq \beta,\ \dots,\ s_{(n_{(2j-1)}-1)} \leq \beta,\ \dots.$$

But then $s_{(n_{(2j-1)})} \leq \beta + \varepsilon$ 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 jth upcrossing is less than ε

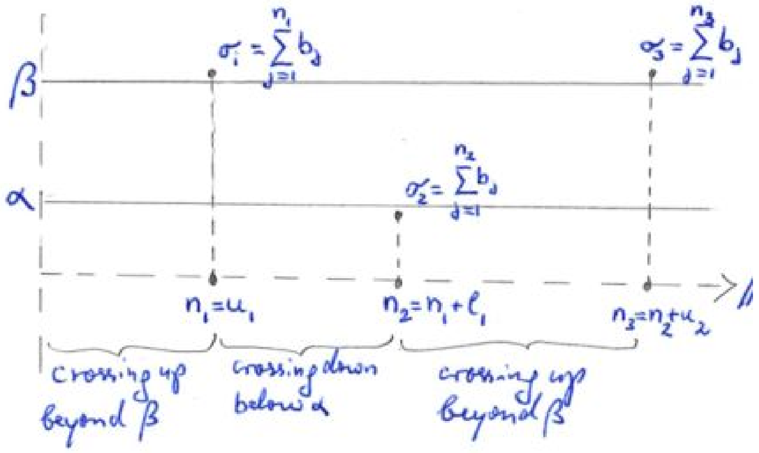
It follows from (13.50) that j is certainly big enough if $j \geq n_0$ because each upcrossing has size of at least 1. This shows that there are at most finitely many indices n such that $s_n > \beta + \varepsilon$ and we conclude that $\limsup_n s_n \leq \beta + \varepsilon$. A similar reasoning allows us to conclude that $\liminf_n s_n \geq \alpha - \varepsilon$.

We have proved (13.49) and this implies, together with (13.48), that

$$\liminf_{n\to\inf} s_n = \alpha \quad \text{and} \quad \limsup_{n\to\inf} s_n = \beta.$$

The picture to the right illustrates how the partial sums $\sigma_n = \sum\limits_{j=1}^{n} b_j$ alternatingly rise above β and fall below α.
$|p_j| = p_j$ and $|q_j| = -q_j \Rightarrow$ both $|p_j|$ and $|q_j|$ drop below ε eventually.
Thus $\beta \leq \limsup_n \sigma_n \leq \varepsilon + \beta$ and $\beta - \varepsilon \leq \liminf_n \sigma_n \leq \beta$ eventually, thus

$$\begin{cases} \limsup_n \sigma_n = & \beta, \\ \liminf_n \sigma_n = & \alpha. \end{cases}$$



We have proved the theorem for case 1: $\beta \geq 0$

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 $\sum_{j=1}^{n} a_j$ exceeds α when n is the last term of an upcrossing and it drops below β when n is the last term 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

$$\liminf_{n\to\inf} \sum_{k=1}^{n} c_k = \alpha \quad \text{and} \quad \limsup_{n\to\inf} \sum_{k=1}^{n} c_k = \beta. \ \blacksquare$$

Corollary 13.3.

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: We apply Riemann's Rearrangement Theorem to the special case $\beta = \alpha$: There is a rearrangement $\sum b_j$ of $\sum a_j$ such that

$$\liminf_{n\to\infty}\sum_{k=1}^{n} b_k \;=\; \alpha \quad \text{and} \quad \limsup_{n\to\infty}\sum_{k=1}^{n} b_k \;=\; \alpha.$$

It follows now from thm.9.14 on p.290 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 13.4.

> *Let $\sum a_k$ be a convergent series with limit $\alpha \in \mathbb{R}$ such that $\sum b_k = \alpha$, for each rearrangement. Then, $\sum a_k$ converge absolutely.*

PROOF: We assume to the contrary that the series $\sum a_k$ is not absolutely convergent, i.e., $\sum a_k$ is conditionally convergent. We apply Riemann's Rearrangement Theorem and find that there is a rearrangement of $\sum a_j$ which converges to a different real number, contrary to our assumption. ■

Corollary 13.5 (Dichotomy for convergent series)**.**

> *Let series $\sum a_k$ be a convergent series. Then*
> ***(a)*** *either all rearrangements of $\sum a_k$ converge to the same limit,*
> ***(b)*** *or, 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 Rearrangement Theorem or the series it is conditionally convergent and **(b)** is true according to cor.13.3. ■

13.3 Exercises for Ch.13

13.3.1 Exercises for Ch.13.1

Exercise 13.1. Prove prop.12.12 (Opposite of continuity) on p.372:
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. □

Exercise 13.2. Prove that $f(x) := \frac{1}{x}$ is **not** uniformly continuous on $]0,1]$. See example 13.3 on p.411. Hint: Examine the sequence $x_n := \frac{1}{n}$. □

Exercise 13.3. In prop13.6 on p.416 the functions $f_n(\cdot)$ were defined as follows on the closed unit interval $[\,0,1\,]$:

$$f_n(x) := \begin{cases} n^2 x & \text{for } 0 \leq x \leq \frac{1}{n} \\ \frac{1}{x} & \text{for } \frac{1}{n} \leq x \leq 1 \end{cases}$$

Prove that f_n is continuous for all $n \in \mathbb{N}$. □

Exercise 13.4. Prove prop.13.2 on p.408 of this document: Let $(X,\mathfrak{U})$, $(Y,\mathfrak{V})$ and $(Z,\mathfrak{W})$ be topological spaces. Let $f : X \to Y$ be continuous at $x_0 \in X$ and $g : Y \to Z$ continuous at $f(x_0)$. Then the composition $g \circ f : X \to Y$ is continuous at x_0. □

Exercise 13.5. Give alternate proofs of exercise 13.4 above in the special case of metric spaces by using the sequence continuity definition (Definition 13.1 on p.402): Let (X, d), (Y, d') and (Z, d'') be metric spaces. Let $f : X \to Y$ be continuous at $x_0 \in X$ and $g : Y \to Z$ continuous at $f(x_0)$. Then the composition $g \circ f : X \to Y$ is continuous at x_0. □

Exercise 13.6. Prove prop.13.5 on p.409 of this document: Let d be the standard Euclidean metric and let d' be the discrete metric on the set $\mathbb{R}$ of all real numbers. Let

$$f : (\mathbb{R},') \to (\mathbb{R}, d); \quad x \mapsto x \qquad \text{and} \qquad g : (\mathbb{R}, d) \to (\mathbb{R}, d'); \quad x \mapsto x$$

both be the identity function on $\mathbb{R}$. Then f is continuous at every point of $\mathbb{R}$, but g is not continuous anywhere on $\mathbb{R}$. □

Exercise 13.7. Let $X := [1, \infty[$ equipped with the standard Euclidean metric $d(x, x') = |x - x'|$.
Let $f_n : X \to \mathbb{R}; \quad x \mapsto \frac{nx+5}{(nx+3)^2}$. Prove that $f_n(\cdot) \overset{uc}{\to} 0$ on X. □

Exercise 13.8. Let $X := \mathbb{R}$, equipped with the Euclidean metric $d(x, x') = |x - x'|$. Let

$$f_n \,:\, \mathbb{R} \to \mathbb{R}; \qquad x \,\mapsto\, \frac{\sin(n^2 x)}{n}\,.$$

(a) Prove that $f_n(\cdot) \overset{uc}{\to} 0$ on $\mathbb{R}$.
(b) Prove that there is $x_0 \in \mathbb{R}$ such that the sequence $f'n(x_0)$ does not converge (pointwise). □

Exercise 13.9. Let $X := \mathbb{R}$ equipped with the standard Euclidean metric $d(x, x') = |x - x'|$.
Let $f_n : \mathbb{R} \to \mathbb{R}$ be the following sequence of functions:

$$f_n(x) \;:=\; \begin{cases} 0 & \text{if } |x| > \frac{1}{n}, \\ nx + 1 & \text{if } \frac{-1}{n} \leq x \leq 0, \\ -nx + 1 & \text{if } 0 \leq x \leq \frac{1}{n}, \end{cases}$$

i.e., the point $\big(x, f_n(x)\big)$ is on the straight line between $(-\frac{1}{n}, 0)$ and $(0, 1)$ for $\frac{-1}{n} \leq x \leq 0$, it is on the straight line between $(0, 1)$ and $(\frac{1}{n}, 0)$ for $0 \leq x \leq \frac{-1}{n}$, and it is on the x–axis for all other x. Draw a picture! Let $f(x) := 0$ for $x \neq 0$ and $f(0) := 1$.
(a) Prove that f_n converges pointwise to f on $\mathbb{R}$.
(b) Prove that f_n does not converge uniformly to f on $\mathbb{R}$. □
You may use without proof that each of the functions f_n is continuous on $\mathbb{R}$.

13.3.2 Exercises for Ch.13.2

Exercise 13.10. Prove prop.13.13 on p.425: Let $\sum a_n$ be an absolutely convergent series. Let $(a_{n_k})_k$ be a subsequence of $(a_n)_n$. Then $\sum a_{n_k}$ converges absolutely. □

14 Compactness

Let us say informally that a family $(U_i)_{i\in I}$ covers or is a cover of a set A if $A \subseteq \bigcup [U_i : i \in I]$.

This chapter will show that **(A)**, **(B)** and **(C)** below are equivalent statements for any subspace $(K, d\big|_{K\times K})$ of a metric space (X,d):

(A) Any sequence in K has a convergent subsequence with limit in K.
(B) K is complete and, given any $\varepsilon > 0$, no matter how small, K can be covered by finitely many ε–neighborhoods.
(C) Any open covering $(U_i)_i$ of K has a finite subcovering: one can find finitely many indices $i_1, \dots, i_n$ such that $K \subseteq U_{i_1} \cup \dots \cup U_{i_n}$

Such metric spaces K will be called "compact". Moreover, we will see that the following is true for the metric space $\mathbb{R}^n$ with the Euclidean metric: For $X = \mathbb{R}^n$, each of **(A)**, **(B)** and **(C)** is equivalent to

(D) K is bounded and closed.

Property **(C)** is the only one that makes sense for abstract topological spaces and will be used to define compactness for such spaces.

14.1 ε-Nets and Total Boundedness

Introduction 14.1. We start out with a few elementary observations for subsets of $\mathbb{R}^2$.

- **(a)** Let C be a square with side length $\varepsilon > 0$ and "edge points" $\vec{p}_1, \vec{p}_2, \vec{p}_3, \vec{p}_4$. Then each point in C belongs to one or more of the ε–neighborhoods $N_\varepsilon(\vec{p}_1), \dots, N_\varepsilon(\vec{p}_4)$, i.e., $C \subseteq \bigcup\limits_{j=1}^{4} N_\varepsilon(\vec{p}_j)$.
- **(b)** Let A be a bounded subset of $\mathbb{R}^2$, i.e., [168] there exists $\gamma > 0$ and $\vec{x_0} \in \mathbb{R}^2$ such that $A \subseteq N_\gamma(\vec{x_0})$. Then, if $\varepsilon > 0$, this circle of radius γ can be covered by a finite number of squares with side length ε.
- **(c)** Put **(a)** and **(b)** together: Any bounded set A can be covered by finitely many ε–neighborhoods.
- **(d)** Equivalently, for any bounded set $A \subseteq \mathbb{R}^2$ and $\varepsilon > 0$ there exists a finite set $G \subseteq \mathbb{R}^2$ such that $A \subseteq \bigcup\limits_{g\in G} N_\varepsilon(g)$

An exact proof will be given in Proposition 14.1 below that all of the above is true for all bounded subsets of $\mathbb{R}^n$, for any $n \in \mathbb{N}$. On the other hand, there are metric spaces (X,d) with bounded subsets which do not possess property **(d)**. □

Here is a simple counterexample to **(d)** of the introduction to this chapter.

Example 14.1. Let X be an infinite set, furnished with the discrete metric d. Then any subset of X, including X, is bounded, since $\operatorname{diam}(X) = 1$.

On the other hand, if $\varepsilon \leq 1$, then $N_\varepsilon(x) = \{x\}$ for all $x \in X$. thus X is not the union of a finite number of such ε–neighborhoods. □

[168] See Proposition 12.30 on p.388.

Considering this counterexample, it makes sense to give a special name for subsets of metric spaces which satisfy **(d)**. We will call such sets totally bounded and refer to G as an ε–net or ε–grid.

Definition 14.1 (ε–nets)**.**

Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \subseteq X$. let $G \subseteq X$ be a subset of X with the following property:

(14.1) For each $x \in A$ there exists $g \in G$ such that $x \in N_\varepsilon(g)$, i.e., $\bigcup\limits_{g \in G} N_\varepsilon(g) \ \supseteq A$.

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. □

The relation $\bigcup\limits_{g \in G} N_\varepsilon(g) \ \supseteq A$ asserts that the family $(N_\varepsilon(g))_{g \in G}$ is a collection of open sets which "covers" all of A We will later call a family of open sets $(U_i)_i$ which satisfies $\bigcup_i U_i \ \supseteq A$ an open cover of A.

Definition 14.2 (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 exists a finite(!) ε–grid for A. □

Remark 14.1.

(A) $A \subseteq (X, d)$ is totally bounded if and only if for each $\varepsilon > 0$ there is a finite collection $\mathscr{G}_\varepsilon = \{g_1, \dots g_n\}$ of points in X whose ε–balls $N_\varepsilon(g_j)$ cover A: For any $a \in A$ there is $j = j(a)$ such that $d(a, g_j) < \varepsilon$.

(B) Let $\varepsilon > 0$. Since all sets A satisfy $A = \bigcup\limits_{a \in A} \{a\} \subseteq \bigcup\limits_{a \in A} N_\varepsilon\{a\}$, all finite sets are totally bounded

(C) Note that the definition of total boundedness of a set A does not demand that the gridpoints are elements of A.

This allows us, for example, to accept the set $G := \{(i, j) : i, j \in \mathbb{Z}\}$ of all points in the plane with integer coordinates, as a 1–grid for the set $A := \{ (i + \pi, j + \pi) : i, j \in \mathbb{Z}\}$. □

Proposition 14.1 (ε–nets in $\mathbb{R}^n$)**.** ★ *Let (X, d) be $\mathbb{R}^n$ with the Euclidean metric.*

(A) *Let $\varepsilon > 0$. Then the set*

$$\varepsilon\mathbb{Z}^n \ = \ \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^n\} \ = \ \{(\varepsilon z_1, \dots, \varepsilon z_n) : z_j \in \mathbb{Z} \text{ for } j = 1, \dots, n\}$$

is an $(\varepsilon\sqrt{n})$–net for (any subset of) $\mathbb{R}^n$.

(B) *Let A be a bounded set in $\mathbb{R}^n$ and $\varepsilon > 0$. Then there is $k \in \mathbb{N}$ and $g_1, \dots g_k \in \varepsilon\mathbb{Z}^n$ such that*

$$A \ \subseteq \ N_\varepsilon(g_1) \ \cup \ \ N_\varepsilon(g_2) \ \cup \ \dots, \ \cup N_\varepsilon(g_k),$$

i.e., A is covered by finitely many ε–neighborhoods of points in the $(\varepsilon/\sqrt{n})$–grid $\varepsilon\mathbb{Z}^n$.

PROOF of **(A)**

Let $\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. For $j = 1, \dots, n$, let k_j be the integer such that

$$\varepsilon k_j \ \le \ x_j \ < \ \varepsilon(k_j + 1)\,. \tag{14.2}$$

It is obvious from (14.2) that $\vec{x} \in C$, the n–dimensional cube of side length ε, defined by

$$C \ := \ C(\vec{x}) \ := \ \{\vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n \ : \ \varepsilon k_j \ \le \ y_j \ < \ \varepsilon(k_j + 1)\}$$

Note that this is the following cube: [169] Its 2^n edgepoints are the vectors $(e_1, \dots, e_n)$ for which

$$\text{the } j\text{–th coordinate is either } \ e_j \ = \ \varepsilon k_j\,, \ \text{ or } \ e_j \ = \ \varepsilon(k_j + 1)\,. \tag{14.3}$$

Let $\vec{x}^\star = (x_1^\star, \dots, x_n^\star) \in \mathbb{R}^n$ be defined as follows. For $j = 1, \dots, n$, let

$$x_j^\star \ := \begin{cases} \varepsilon k_j & \text{if } x_j \le \varepsilon(k_j + 1/2)\,, \\ \varepsilon(k_j + 1) & \text{else}\,. \end{cases}$$

Thus $\vec{x}^\star$ is an edge point of the cube $C(\vec{x})$ since each $x_j^\star$ is of the form (14.3). Moreover, since each $x_j^\star$ satisfies $|x_j - x_j^\star| \le \varepsilon/2$,

$$d(\vec{x}, \vec{x}^\star) \ = \sqrt{\sum_{j=1}^{n} (x_j - x_j^\star)^2} \ \le \ \sqrt{n \cdot \left(\frac{\varepsilon}{2}\right)^2} \ = \ \frac{\varepsilon\sqrt{n}}{2} \ < \ \varepsilon\sqrt{n}\,. \tag{14.4}$$

We have found for arbitrary $\vec{x} \in \mathbb{R}^n$ a vector $\vec{x}^\star \in \varepsilon\mathbb{R}^n$ such that $\vec{x} \in N_{\varepsilon\sqrt{n}}(\vec{x}^\star)$. [170] This proves that $\varepsilon\mathbb{Z}^n$ is an $\varepsilon\sqrt{n}$–net in $\mathbb{R}^n$.

PROOF of **(B)**

Intuitively clear but very messy. Here is an outline.

For convenience, let $\varepsilon' := \varepsilon/\sqrt{n}$.

First we choose a radius R so big that $A \subseteq N_R(\vec{0})$. This is possible according to Proposition 12.30 on p.388. Next we choose $M \in \mathbb{N}$ which is so big that $M\varepsilon' > R$. Let $\vec{c}^{(1)}, \vec{c}^{(2)}, \dots, \vec{c}^{(2^n)}$ be the 2^n points in $\mathbb{R}^n$ for which each coordinate is either $M\varepsilon'$ or $-M\varepsilon'$. Those are the edge points of the n–dimensional cube

$$C \ := \ \{\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \ : \ |x_j| \le M\varepsilon \text{ for all } j = 1, \dots, n\}\,.$$

Since there are only $2M + 1$ integers m such that $|m| \le M$, there are only finitely many points $\vec{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ which belong to both C and $\varepsilon\mathbb{Z}^n$, i.e., such that each coordinate y_j is of the form $y_j = \varepsilon' m$ for some integer m which satisfies

$$-\varepsilon' M \ \le \ y_j \ \le \ \varepsilon' M\,, \quad \text{i.e.,} \quad -M \ \le \ m \ \le \ M\,.$$

[169] Draw pictures for dimensions $1, 2, 3$ with $\varepsilon = 1$!

[170] Here is an example. If $n = 5, \varepsilon = 1$, and $\vec{x} = (12.85, -12.35, \frac{1}{3}, 9, -\pi)$, then the associated grid point is $\vec{x}^\star = (13, -12, 0, 9, -3)$. The distance is:

$$d(\vec{x}, \vec{x}^\star) \ = \ \sqrt{.15^2 + .35^2 + (1/3)^2 + 0 + (\pi - 3)^2} \ \le \ \sqrt{1/2 + 1/2 + 1/2 + 0 + 1/2} \ \le \ \varepsilon \cdot \sqrt{n}.$$

We see that part ***A*** of the lemma is true for this specific example.

Those points $\vec{y}$ are the edge points of $(2M)^n$ n–dimensional cubes $C_1, C_2, \dots, C_{(2M)^n}$ of side length ε' whose union equals C.

We have seen in the proof of **(A)** that each C_i is covered by the $\sqrt{n} \cdot \varepsilon'$–neighborhoods of its 2^n edge points. Since $\sqrt{n} \cdot \varepsilon' = \varepsilon$, each C_i is covered by a finite number of ε–neighborhoods. Since

$$C = \bigcup \left[C_j \; : j = 1, \dots, (2M)^n \right],$$

C also is covered by a finite number of ε–neighborhoods. Since $A \subseteq C$, it follows that A is covered by a finite number of ε–neighborhoods. ■

Note for part **(A)** that$\varepsilon\mathbb{Z}^n$ is as intuitive a grid as you can think of, especially if you look at the 2–dimensional plane or 3–dimensional space and consider $\varepsilon = 1$.

Theorem 14.1. *Bounded subsets of $\mathbb{R}^n$ are totally bounded.*

PROOF: This is immediate from Proposition14.1**(B)**. ■

We have seen that all bounded subsets in $\mathbb{R}^n$ are totally bounded. In the remainder of this subchapter we will see that

- the reverse is true in any metric space: totally bounded subsets are always bounded,
- a subset A of a metric space is totally bounded if and only if all sequences in A possess subsequences which are Cauchy.

We start by proving the following:

(a) All sequences in a totally bounded set possess a subsequence which is Cauchy.
(b) All sequences in a bounded subset of $\mathbb{R}^n$ possess a subsequence which is Cauchy.

We saw earlier in this subchapter that all bounded subsets in $\mathbb{R}^n$ are totally bounded. Thus a proof of **(a)** also is a proof of **(b)**. We will prove **(b)** anyway and do so first, since this proof is easier to visualize than that of **(a)**

Proposition 14.2.

Let A be a bounded subset of $\mathbb{R}^n$ and let $(x_n)_n$ be a sequence such that $x_n \in A$ for all n. Then there exists a subsequence x_{n_j} which is Cauchy.

PROOF:

We will construct a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_j \downarrow 0$ and such that each A_j contains infinitely members of $(x_n)_n$. This allows us to find indices $n_0 < n_1 < \dots$ such that $x_{n_j} \in A_j$ for all j. It then follows from Proposition 12.34 in the chapter on completeness in metric spaces that $\left(x_{n_j}\right)_j$ is Cauchy.

Step 0: Since A_0 is bounded, there exists an n–dimensional cube C_0 such that $A_0 \subseteq C_0$. Let γ be the side length, δ_0 the diameter, and let $\vec{a}^{(0)} = (a_1^{(0)}, \dots, a_n^{(0)})$ be the center of that cube.

The following is a very generous estimate of δ_0. Let $\vec{x} = (x_1, \dots, x_n) \in C_0$. Since $\vec{a}^{(0)}$ is the center of

C_0, $|x_j - a_j^{(0)}| \leq \gamma/2$ for each j. Thus

$$d(\vec{x}, \vec{a}^{(0)}) = \sqrt{\sum_{j=1}^{n} (x_j - a_j^{(0)})^2} \leq \sqrt{n \cdot \left(\frac{\gamma}{2}\right)^2} = \frac{\gamma\sqrt{n}}{2} < \gamma\sqrt{n}\,. \tag{14.5}$$

It follows from $A_0 \subseteq C_0$ that

$$\delta_0 \leq \operatorname{diam}(C_0) \leq \gamma\sqrt{n}\,. \tag{14.6}$$

Step 1: We subdivide C_0 into 2^n cubes $C_{1,1}, \dots C_{1,2^n}$ of side length $\gamma/2 = 2^{-1}\gamma$.

It follows from $A_0 \subseteq C_0$ that $A_0 = (C_{1,1} \cap A_0) \cup \dots \cup (C_{1,2^n} \cap A_0)$. Since $(x_n)_n$ is an infinite sequence, there is at least one index $m_1 \in [1, 2^n]_{\mathbb{Z}}$ such that $C_{1,m_1} \cap A_0$ possesses infinitely many members

$$x_{1,1} := x_{n_1},\ x_{1,2} := x_{n_2},\ x_{1,3} := x_{n_3},\ \dots$$

of that sequence. Let $A_1 := C_{1,m_1} \cap A_0$ and $\delta_1 := \operatorname{diam}(A_1)$. Then (14.6) yields

$$\delta_1 \leq \operatorname{diam}(C_{1,m_1}) = \frac{1}{2} \cdot \operatorname{diam}(C_0) = \frac{\gamma\sqrt{n}}{2^1}\,.$$

Since $x_n \in A$ for all n, we also have $x_{1,n} \in A$ for all n. Thus the infinite subsequence $(x_{1,n})_n$ of the original sequence $(x_n)_n$ lives in the subset A_1 of A with diameter $\leq (\gamma\sqrt{n}) \cdot 2^{-1}$.

Step k**:** Assume that we have obtained sets $A_0 \supseteq A_1 \supseteq \dots \supseteq A_k$ such that, for each $j \in [1, k]_{\mathbb{Z}}$, the following holds true:

- **(1)** $A_j = C_{j,m_j} \cap A_{j-1}$ for a suitable cube C_{j,m_j} of side length $2^{-j}\gamma$
- **(2)** $\delta_j := \operatorname{diam}(A_j) \leq \frac{\gamma\sqrt{n}}{2^j}$, [171]
- **(3)** A_j contains an infinite subsequence $(x_{j,n})_n = x_{j,1}, x_{j,2}, \dots$ of the original sequence $(x_n)_n$.

Step $k+1$**:** We subdivide the cube C_{k,m_k} into 2^n cubes $C_{k+1,1}, \dots C_{k+1,2^n}$ of side length $\gamma/2 = 2^{-(k+1)}\gamma$.

It follows from $A_k \subseteq C_{k,m_k}$ that $A_k = (C_{k+1,1} \cap A_k) \cup \dots \cup (C_{k+1,2^n} \cap A_k)$.

Since $(x_{k,n})_n$ is an infinite sequence, there is at least one index $m_{k+1} \in [1, 2^n]_{\mathbb{Z}}$ such that $C_{k+1,m_{k+1}} \cap A_k$ possesses infinitely many members

$$x_{k+1,1} := x_{k,n_1},\ x_{k+1,2} := x_{k,n_2},\ x_{k+1,3} := x_{k,n_3},\ \dots$$

of that subsequence.

Let $A_{k+1} := C_{k+1,m_{k+1}} \cap A_k$ and $\delta_{k+1} := \operatorname{diam}(A_{k+1})$. Then (14.6) yields

$$\delta_{k+1} \leq \operatorname{diam}(C_{k+1,m_{k+1}}) = \frac{1}{2} \cdot \operatorname{diam}(C_{k,m_k}) = \frac{\gamma\sqrt{n}}{2^{k+1}}\,.$$

Since $x_n \in A$ for all n, we also have $x_{k+1,n} \in A$ for all n, thus We have an infinite subsequence $(x_{k+1,n})_n$ of the original sequence x_n which lives in the subset A_{k+1} of A with diameter $\leq (\gamma\sqrt{n}) \cdot 2^{-1}$.

[171] (this actually follows from **(1)** by means of a computation like the one done in (14.5)),

To summarize, we have achieved what we set out to do at the beginning of the proof: We have constructed a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_n \downarrow 0$ such that each A_j contains infinitely members of the sequence $(x_n)_n$.

- Since A_1 contains infinitely members of $(x_n)_n$, there exists $n_1 > n_0$ such that $x_{n_1} \in A_1$.
- Since A_2 contains infinitely members of $(x_n)_n$, there exists $n_2 > n_1$ such that $x_{n_2} \in A_2$.

  - -
- Since A_k contains infinitely members of $(x_n)_n$, there exists $n_k > n_{k-1}$ such that $x_{n_k} \in A_k$.

Let $z_k := x_{n_k}$. Then $(z_k)_k$ is a subsequence of $(x_n)_n$ such that $z_k \in A_k$. Since $\lim\limits_{k\to\infty} \delta_k = 0$, it follows from Proposition 12.34 on p.391 that $\left(x_{n_j}\right)_j$ is Cauchy. ■

The proof that allows us to extend Proposition 14.2 to totally bounded sets in arbitrary metric spaces is almost identical to the one given above. The main difference is that we no longer can subdivide the set A_k into 2^n subsets of smaller diameters from which to choose A_{k+1}. Rather, we must work directly with the definition of total boundedness to obtain A_k.

Theorem 14.2.

Let A be a totally bounded subset of a metric space (X, d). Let $(x_n)_n$ be a sequence such that $x_n \in A$ for all n. Then there exists a subsequence x_{n_j} which is Cauchy.

PROOF:

Let $A_0 := A$. As in the proof of Proposition 14.2, we will construct $A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_n :=$ diam $(A_n) \downarrow 0$ such that each A_j contains infinitely members of $(x_n)_n$.

Step 1: Let $\delta_1 := 2^{-1} = \frac{1}{2}$. Since A_0 is totally bounded, there exists a finite grid of length δ_1, i.e., there is a finite set $G_1 = \{g_{1,1},\, g_{1,2},\, \dots,\, g_{1,M_1}\}$ such that $A_0 \subset \bigcup [N_{\delta_1}(g_{1,j}) : j = 1, \dots, M_1]$.

Since $A_0 = \bigcup [N_{\delta_1}(g_{1,j}) \cap A_0 : j = 1, \dots, M_1]$ and A_0 contains the entire (infinite) sequence $(x_n)_n$, there exists $m_1 \in [1, M_1]_{\mathbb{Z}}$ such that $N_{\delta_1}(g_{1,m_1}) \cap A_0$ contains an infinite subsequence

$$x_{1,1} := x_{n_1},\ x_{1,2} := x_{n_2},\ x_{1,3} := x_{n_3},\ \dots$$

of that sequence.

Let $A_1 := A_0 \cap N_{\delta_1}(g_{1,m_1})$. Then diam$(A_1) \leq$ diam$(N_{\delta_1}(g_{1,k_1}))$, i.e., diam$(A_1) \leq \delta_1$.

Step 2: Let $\delta_2 := 2^{-2}$. Since A_0 is totally bounded, there exists a finite grid of length δ_2, i.e., there exists a finite set $G_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,M_2}\}$ such that $A_0 \subset \bigcup [N_{\delta_2}(g_{2,j}) : j = 1, \dots, M_2]$.

Since $A_1 \subseteq A_0$, it follows that $A_1 = \bigcup [N_{\delta_2}(g_{2,j}) \cap A_1 : j = 1, \dots, M_2]$. Since A_1 contains the infinite sequence $(x_{1,n})_n$, there exists $m_2 \in [1, M_2]_{\mathbb{Z}}$ such that $N_{\delta_2}(g_{2,m_2}) \cap A_1$ contains an infinite subsequence

$$x_{2,1} := x_{1,n_1},\ x_{2,2} := x_{1,n_2},\ x_{2,3} := x_{1,n_3},\ \dots$$

of that subsequence.

Let $A_2 := A_1 \cap N_{\delta_2}(g_{2,m_2})$. Then diam$(A_2) \leq$ diam$(N_{\delta_2}(g_{2,m_2}))$, i.e., diam$(A_2) \leq \delta_2$.

Step k: Assume that we have obtained sets $A_0 \supseteq A_1 \supseteq \cdots \supseteq A_k$ such that, for each $j \in [1, k]_{\mathbb{Z}}$, the following holds true:

(1) Let $\delta_j := 2^{-j}$. Then $A_j = N_{\delta_j}(g_{j,m_j}) \cap A_{j-1}$ for a suitable $g_{j,m_j} \in X$.

(2) A_j contains an infinite subsequence $(x_{j,n})_n \;=\; x_{j,1}, x_{j,2}, \dots$ of the original sequence $(x_n)_n$.

Step $k+1$**:** Let $\delta_{k+1} := 2^{-(k+1)}$. Since A_0 is totally bounded, there exists a finite grid of length δ_{k+1}, i.e., there exists a finite set $G_{k+1} = \{g_{k+1,1}, g_{k+1,2}, \dots, g_{k+1,M_{k+1}}\}$ such that
$A_0 \subset \bigcup \left[N_{\delta_{k+1}}(g_{k+1,j}) : j = 1, \dots, M_{k+1} \right]$.
Since $A_k \subseteq A_0$, it follows that $A_k = \bigcup \left[N_{\delta_{k+1}}(g_{k+1,j}) \cap A_k : j = 1, \dots, M_{k+1} \right]$. Since A_k contains the infinite sequence $(x_{k,n})_n$, there exists $m_{k+1} \in [1, M_2]_{\mathbb{Z}}$ such that $N_{\delta_{k+1}}(g_{k+1,m_{k+1}}) \cap A_k$ contains an infinite subsequence

$$x_{k+1,1} := x_{k,n_1},\ x_{k+1,2} := x_{k,n_2},\ x_{k+1,3} := x_{k,n_3},\ \dots$$

of that subsequence.
Let $A_{k+1} := A_k \cap N_{\delta_{k+1}}(g_{k+1,m_{k+1}})$. Then $\operatorname{diam}(A_{k+1}) \leq \operatorname{diam}\big(N_{\delta_{k+1}}(g_{k+1,m_{k+1}})\big)$. In other words, $\operatorname{diam}(A_{k+1}) \leq \delta_{k+1}$.

Since $x_n \in A$ for all n, we also have $x_{k+1,n} \in A$ for all n. This infinite subsequence $(x_{k+1,n})_n$ of the original sequence x_n lives in the subset A_{k+1} of A with diameter $\leq \delta_{k+1} = 2^{-k+1}$.

To summarize, we have achieved what we set out to do at the beginning of the proof: We have constructed a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_j \downarrow 0$ such that each A_j contains infinitely members of the sequence $(x_n)_n$.

- Since A_1 contains infinitely members of $(x_n)_n$, there exists $n_1 > n_0$ such that $x_{n_1} \in A_1$.
- Since A_2 contains infinitely members of $(x_n)_n$, there exists $n_2 > n_1$ such that $x_{n_2} \in A_2$.

- -

- Since A_k contains infinitely members of $(x_n)_n$, there exists $n_k > n_{k-1}$ such that $x_{n_k} \in A_k$.

Let $z_k := x_{n_k}$. Then $(z_k)_k$ is a subsequence of $(x_n)_n$ such that $z_k \in A_k$. Since $\lim\limits_{k\to\infty} \delta_k = 0$, it follows from Proposition 12.34 on p.391 that $\left(x_{n_j}\right)_j$ is Cauchy. ■

Proposition 14.3. *Totally bounded subsets of metric spaces are bounded.*

PROOF:

Let A be a subset of a metric space (X, d). We show the contrapositive: We assume that A is not bounded, i.e., $\operatorname{diam}(A) = \infty$, and we will show that A is not totally bounded.

We may assume that A is not empty because otherwise there is nothing to prove.

Step 1: We prove by induction that there exists a sequence $x_n \in A$ such that $d(x_i, x_j) \geq 1$ for any $i \neq j$.

Base case: Let $x_0 \in A$. Since A is not bounded, there exists $x_1 \in A$ such that $r_1 := d(x_0, x_1) \geq 1$.

Induction step: We assume that n elements $x_1, \dots x_n$ such that $d(x_i, x_j) \geq 1$ for any $1 \leq i < j \leq n$ have aready been chosen. Let

$$\beta \ := \ \max\{d(x_0, x_j) : j \leq n\}\,, \qquad r \ := \ \beta + 1\,.$$

Since A is not bounded, we can pick $x_{n+1} \in A \setminus N_r(x_0)$. We obtain for each $j \in [1, n]_{\mathbb{Z}}$,

$$\beta + 1 \ \leq d(x_{n+1}, x_0) \leq d(x_{n+1}, x_j) + d(x_j, x_0) \ \leq \ d(x_{n+1}, x_j) + \beta, \quad \text{i.e.,} \ \ 1 \ \leq \ d(x_{n+1}, x_j).$$

We have constructed a sequence (x_n) for which any two items have distance no less than 1.

Step 2: It follows that $(x_n)_n$ does not possess a Cauchy subsequence. According to Theorem 14.2, all sequences in totally bounded sets have Cauchy subsequences. It follows that A is not totally bounded. ■

Corollary 14.1. *If $A \subseteq \mathbb{R}^n$, then $\quad A$ is bounded $\Leftrightarrow$ A is totally bounded.*

PROOF: This is immediate from Proposition 14.3 above and Theorem 14.1 on p.437.

Next we prove the reverse of Theorem 14.2 on p.439.

Theorem 14.3.

Let A be a subset of a metric space (X,d) such that each sequence in A contains a Cauchy subsequence. Then A is totally bounded.

PROOF: [172] We prove the contrapositive: If a set is not totally bounded, then there exists a sequence without any Cauchy subsequences.

So assume that A is not totally bounded. Thus there is $\varepsilon > 0$ such that the following holds for any $n \in \mathbb{N}$: If $z_1, z_2, \dots z_n \in A$ then the union $\bigcup_{1 \le j \le n} N_\varepsilon(z_j)$ does not cover A: There exists $z \in A$ outside any one of those ε–neighborhoods, i.e., $z \in A \setminus \bigcup \left[N_\varepsilon(z_j) : 1 \le j \le n \right]$.

This allows us to create an infinite sequence $(x_j)_{j \in \mathbb{N}}$ such that $d(x_j, x_n) \ge \varepsilon$ for all $j, n \in \mathbb{N}$ such that $j \ne n$, say, $j < n$, as follows: We pick

$$x_1 \in A; \quad x_2 \in A \setminus N_\varepsilon(x_1); \quad x_3 \in A \setminus \big(N_\varepsilon(x_1) \cup N_\varepsilon(x_2)\big); \ \dots \quad x_n \in A \setminus \bigcup_{j<n} N_\varepsilon(x_j); \ \dots$$

Note that $x_n \in A \setminus \bigcup_{j<n} N_\varepsilon(x_j)$ implies $d(x_j, x_n) \ge \varepsilon$ for all indices $j < n$. Since this is true for arbitrary $n \in \mathbb{N}$, it is true that

$$d(x_i, x_j) \ \ge \ \varepsilon \quad \text{for all } i, j \in \mathbb{N} \text{ such that } i \ne j\,.$$

If $(x_{n_j})_j$ were a Cauchy subsequence of $(x_n)_n$ then there would be $n_0 \in \mathbb{N}$ such that

$$\textbf{(A)} \qquad d(x_i, x_j) \ < \ \varepsilon \quad \text{for all } i, j \in \mathbb{N} \text{ such that } i, j \ge n_0\,.$$

But the x_n were constructed such that $d(x_m, x_k) \ge \varepsilon$ for **all** $m \ne k$, in particular for $m := n_i$ and $k := n_j$ if $i \ne j$. Since $n_i \ne n_j$ whenever $i \ne j$, it is not possible to construct a subsequence $(x_{n_j})_j$ which satisfies **(A)**. We have shown that sets which are not totally bounded possess sequences without any Cauchy subsequences. ■

Corollary 14.2. *Let A be a subset of a metric space (X,d). Then,*

A is totally bounded $\quad\Leftrightarrow\quad$ every sequence in A possesses a Cauchy subsequence.

PROOF: This follows from Theorem 14.2 on p.439 and Theorem 14.3 above. ■

[172] The proof is similar to that of Proposition 14.3.

14.2 Sequence Compactness

We saw that totally bounded sets are those where every sequence possesses a Cauchy subsequence. Since all Cauchy sequences converge but the opposite usually is not true, total boundedness of a set A should be a weaker property than the following:

any sequence in A possesses a convergent subsequence.

In this subchapter we will examine sets with that property.

Definition 14.3 (Sequence compactness)**.** Let (X, d) be a metric space and let $A \subseteq X$.

We say that A is **sequence compact** or **sequentially compact** if it has the following property: Given any sequence (a_n) of elements of A, there exists $a \in A$ and a subset

$$n_1 < n_2 < \dots < n_j < \dots \quad \text{of indices such that} \quad a = \lim_{j\to\infty} a_{n_j},$$

In other words, there exists a subsequence [173] (a_{n_j}) which converges to a. □

Remark 14.2. It is important that you understand that is is not sufficient that $\lim\limits_{j\to\infty} a_{n_j} \in X$. Rather, we demand that this limit belongs to the smaller set A!

For example, the open unit interval $]0, 1[$ is totally bounded as a bounded subset of $\mathbb{R} = \mathbb{R}^1$. (Can you prove this directly?) But this set is not sequence compact, since the sequence $x_n := \frac{1}{n}$ does not possess a convergent subsequence: Any such convergent subsequence would have limit zero, and zero is not an element of $]0, 1[$. □

Proposition 14.4 (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: Let $(x_n)_n$ be a sequence in A. Since convergent subsequences are Cauchy, there exists a Cauchy subsequence. It follows from Corollary 14.2 on p.441 that A is totally bounded. ■

Proposition 14.5 (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. 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 from prop.12.36 on p.392 that the entire Cauchy sequence (x_n) converges to L. ■

The last two propositions have proved that any sequence compact set in a metric space is both totally bounded and complete. As the next theorem shows, the reverse is also true.

[173] See Definition 5.22 on p.158.

Theorem 14.4 (Sequence compact $\Leftrightarrow$ 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.14.4 on p.442 and prop.14.5 on p.442 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 $(x_n)_n$ be a sequence in A. Since A is totally bounded, we can extract a Cauchy subsequence $(x_{n_j})_j$. Since A is complete, this subsequence has a limit $L \in A$. Thus $(x_n)_n$ is a convergent subsequence of $(x_n)_n$. ■

Theorem 14.5 (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.*

PROOF: Sequence compact spaces are totally bounded and complete by Theorem 14.4 on p.443.

Since they are totally bounded, they also are bounded by Proposition 14.3 on p.440.

Since they are complete, they also are closed by Theorem 12.11 on p.396. ■

Remark 14.3. We obtain from the results of this and the previous subchapter the following:

> A subset of a metric space is sequentially compact
> $\Leftrightarrow$ it is totally bounded and complete
> $\Rightarrow$ it is bounded and closed. □

In subsets of $\mathbb{R}^n$ the last implication of Remark14.3 becomes an equivalence:

Theorem 14.6.

> *A subset of $\mathbb{R}^n$ is sequentially compact*
> $\Leftrightarrow$ *it is totally bounded and complete*
> $\Leftrightarrow$ *it is bounded and closed.*

PROOF:

Let $A \subseteq \mathbb{R}^n$. It suffices to prove that A is totally bounded and complete if A is bounded and closed since, as we mentioned in Remark14.3, everything else has already been established.

So assume that A is bounded and closed. Since A is closed and $\mathbb{R}^n$ is complete, A is complete by Theorem 12.12 on p.397. Since A is bounded, A also is totally bounded by Theorem 14.1. Thus A is both totally bounded and complete. ■

14.3 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 (5.20) on p.156).

Definition 14.4 (Coverings and open coverings)**.**

Let X be an arbitrary nonempty set and $A \subseteq X$. Let $U_i \in X$ $(i \in I)$ such that $A \subseteq \bigcup\limits_{i \in I} U_i$.

We call such a family a **covering** of A.

A **finite subcovering** of a covering $(U_i)_{i \in I}$ of the set A is a finite collection

$$U_{i_1}, \dots, U_{i_n} \quad (i_j \in I \quad \text{for } 1 \leq j \leq n) \quad \text{such that} \quad A \subseteq U_{i_1} \cup U_{i_2} \cup \ldots \cup U_{i_n}. \tag{14.7}$$

Assume in addition that X is a topological space, e.g., a normed vector space or a metric space. If all members U_i are open then we call $(U_i)_{i \in I}$ an **open covering** of A.

We also write **cover**, **finite subcover**, **open cover** instead of covering, finite subcovering, open covering □

Remark 14.4.

(a) Partitions [174] are coverings.

(b) Formula 14.1 (ε–nets definition, p.435) tells us that, if G is an ε–grid for $A \subseteq (X, d)$, then $\{N_\varepsilon(g)\}_{g \in G}$ is an open covering of A.

(c) If $(U_i)_{i \in I}$ is a covering of A then $(U_i \cap A)_{i \in I}$ is a covering of A which satisfies

$$\bigcup_{j \in I} (U_j \cap A) \ = \ A. \ \square \tag{14.8}$$

Definition 14.5 (Compact sets)**.**

Let $(X, \mathfrak{U})$ be a topological space and $K \subseteq X$.

- We call K **compact**, if K possesses the **"extract finite open subcovering" property**:

Given any **open** covering $(U_i)_{i \in I}$ of K, one can extract a finite subcovering. In other words, there is $n \in \mathbb{N}$ and indices

$$i_1, i_2, \dots, i_n \in I \qquad \text{such that} \quad A \subseteq \bigcup_{j=1}^{n} U_{i_j}. \ \square$$

Remark 14.5.

(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) Any subcovering of an open covering necessarily consists exclusively of open sets, i.e., it is again an open covering of A.

[174] see Definition 8.3 on p.233

(c) Let (X, d) be a metric space. Then

$$K \subseteq (X, d) \text{ is compact if and only if the metric subspace } (K, d\big|_{K \times K}) \text{ is compact,}$$

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 Definition 12.21 on p.380). □

Example 14.2. 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 (Definition 12.14 on p.374) is compact
- **(c)** A space with the discrete metric (Definition 12.3 on p.362) 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 a finite covering from the open covering $\left(]\frac{1}{n}, 1[\right)_{n \in \mathbb{N}}$. □

Example 14.3. for sequence compactness in metric spaces we have the following results which correspond to the previous example.

- **(a)** Any finite metric space is sequence compact.
- **(b)** Any metric space that only contains finitely many open sets is sequence compact. [175]
- **(c)** A space with the discrete metric is sequence 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. The next two theorems show that (subspaces of) metric spaces are compact if and only if they are sequentially compact.

Theorem 14.7 (Compact metric spaces are sequence compact)**.**

Let (X, d) be a compact metric space. Then X is sequence compact.

PROOF: We assume to the contrary that X is compact and that there is a sequence $(x_n)_n$ in X from which one cannot extract a convergent subsequence.

Let $F := \{x \in X : x = x_j \text{ for some } j \in \mathbb{N}\}$ [176] be the set of distinct(!) members of $(x_n)_n$. Let $z \in X$. There exists an open neighborhood U_z of z such that $U_z \cap F$ is finite, because otherwise one could construct a subsequence $(x_{j_m})_m$ of $(x_n)_n$ which converges to z. (See exercise 9.24 on p.304).

[175] We had to remove the example of the indiscrete topology because this topology does not come from a metric.

[176] 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. Note that F can be very small even if there are infinitely many indices j: If $x_j = (-1)^j$ then $F = \{-1, 1\}$ only contains two elements!

It follows from $\{z\} \subseteq U_z$ that $(U_z)_{z \in X}$ is an open covering of X. X is compact, thus we can extract a finite subcovering $U_{z_1}, U_{z_2}, \dots, U_{z_k}$.

$$F = F \cap X = F \cap \bigcup_{j=1}^{k} U_{z_j} = \bigcup_{j=1}^{k} (U_{z_j} \cap F)$$

is a finite union of the finite sets $U_{z_j} \cap F \subseteq U_{z_j}$ and thus finite. We conclude that the entire sequence $(x_n)_n$ consists of only finitely many distinct members.

But then at least one 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 constant subsequence converges (to $x_{k^\star}$). We have reached a contradiction. ■

The next proposition does most of the work in establishing the reverse direction: Sequence compactness implies compactness.

Proposition 14.6.

Let (X,d) be a sequence compact metric space. Let $(U_i)_{i \in I}$ be an open cover of X. Then, one can find for $(U_i)_{i \in I}$ a number $\rho > 0$ which possesses the following property:

- *For each $x \in X$ there exists $i \in I$ such that $N_\rho(x) \subseteq U_i$.*

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, so 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 there is at least one j such that $j > 2/\varepsilon$. It follows from $n_j \geq j$ 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 12.3 on p.367 that $N_{\varepsilon/2}(x_{n_j}) \subseteq N_{\varepsilon/2+\varepsilon/2}(x) = N_\varepsilon(x)$.

We apply first $(\star\star)$, then $(\star\star\star)$, then $(\star)$ and obtain

$$N_{1/n_j}(x_{n_j}) \subseteq N_{\varepsilon/2}(x_{n_j}) \subseteq N_\varepsilon(x) \subseteq U_{i_0}.$$

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. ■

Remark 14.6. Consider The number ρ of Proposition 14.6 above. Then $\lambda = 2\rho$ 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 cover, then any λ' which satisfies $0 < \lambda' < \lambda$ also is a Lebesgue number. of that same cover. □

We now can prove the converse of thm.14.7.

Theorem 14.8. *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.14.6 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.14.4 on p.443) 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 14.9 (Sequence compactness coincides with compactness in metric spaces)**.**

Let (X, d) be a metric space and let A be a subset of X. Then,

$$A \text{ is sequence compact} \Leftrightarrow A \text{ is compact}, \quad \text{i.e.,}$$
$$A \text{ is sequence compact} \Leftrightarrow \text{every open cover of } A \text{ possesses a finite subcover.}$$

PROOF: Theorems 14.7 and 14.8. ■

An easy consequence is the Heine–Borel theorem.

Theorem 14.10 (Heine–Borel Theorem)**.**

A subset of Euclidean space $\mathbb{R}^n$ is compact $\Leftrightarrow$ this set is closed and bounded.

PROOF: We have seen in thm.14.6 on p.443 that closed and bounded subsets of $\mathbb{R}^n$ are sequence compact. Since sequence compact metric spaces are compact (thm.14.8) it follows that closed and bounded subsets of $\mathbb{R}^n$ are compact.

On the other hand, let K be a compact subset of $\mathbb{R}^n$. Then K is sequence compact by Theorem 14.9 on p.447. Thus K is closed and bounded according to Theorem 14.5 on p.443. ■

14.4 Continuous Functions and Compact Spaces

Theorem 14.11 (Closed subsets of compact topological spaces are compact)**.**

Let A be a closed subset of a compact topological space $(X, \mathfrak{U})$. Then A is a compact subspace.

In other words, the open sets

$$\mathfrak{U}_A = \{V \cap A : V \in \mathfrak{U}\}$$

of the subspace $(A, \mathfrak{U}_A)$ possess the "extract finite open subcovering" property of Definition 14.5 on p.444.

PROOF: Let $(U_j)_{j\in J}$ be a family of sets open in A whose union is A. According to Definition 12.22 on p.382 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 $X \setminus A$ 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 $X \setminus A$ 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 (Definition 12.21 on p.380)

$$U_{i_1} = V_{i_1} \cap A,\ U_{i_2} = V_{i_2} \cap A,\ \dots,\ U_{i_n} = V_{i_n} \cap A$$

of those open sets in X are open in A and hence form an open covering of the subspace $(A, \mathfrak{U}_A)$. We have proved that the given open covering of A has contains a finite subcover of A. ■

Corollary 14.3 (Closed subsets of compact metric 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: Immediate from thm.14.11. ■

Let $(X, \mathfrak{U})$ and $(Y, \mathfrak{V})$ be topological spaces and $A \subseteq X$. We recall that continuity for functions $f : A \to (Y, \mathfrak{V})$ was defined in Definition 13.4 on p.407.

Theorem 14.12 (Continuous images of compact topological spaces are compact)**.**

> *Let $(X, \mathfrak{U})$ and $(Y, \mathfrak{V})$ be two topological spaces. and let $f : X \to Y$ be continuous on X.*
> - *If X is compact then the direct image $f(X)$ is compact.*

In other words, the topological subspace $(f(X), \mathfrak{V}_{f(X)})$ of Y is compact.

PROOF: Let $(V_j)_{j \in J}$ be a family of sets open in Y whose union contains $f(X)$. 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 $(f(X), \mathfrak{V}_{f(X)})$ 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 $U_j := f^{-1}(V_j)$. Then

$$\bigcup_{j \in J} U_j \;=\; \bigcup_{j \in J} f^{-1}(V_j) \;=\; f^{-1}\Big(\bigcup_{j \in J} V_j \Big) \;\supseteq\; f^{-1}\big(f(X)\big) \;\supseteq\; X. \tag{14.9}$$

The second equation above follows from prop. 8.4 (f^{-1} is compatible with all basic set ops) on p.240 and the last one follows from the fact that $f^{-1}\big(f((\Gamma)\big) \supseteq \Gamma$ for any subset Γ of the domain of f (see cor. 8.1 on p. 243). The "$\supseteq$" relation follows from the assumption that $\bigcup \Big[V_j : j \in J \Big] \;\supseteq\; f(X)$

According to prop.13.1 ("f^{-1}(open) $=$ open" continuity) on p.407, each U_j is open as the inverse image of the open set V_j under the continuous function f.

It follows from (14.9) 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 (8.20) on p.241) 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.8.8 on p. 244).

We have proved that the arbitrary open cover $(V_j)_{j \in J}$ of $f(X)$ contains a finite subcover $V_{j_1}, \dots, V_{j_n}$ and it follows that $f(X)$ is indeed a compact metric subspace of Y. ■

Corollary 14.4 (Continuous images of compact metric 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 its image $f(X)$ is compact, i.e., the metric subspace $(f(X), d_2)$ of Y is compact.*

PROOF: Immediate from thm.14.12 ■

Read the following remark for an easier way to prove the above theorem.

Remark 14.7. We could have proved the last two theorems 14.11 14.12 in the special case of metric spaces more easily using sequence compactness instead of covering compactness, but the following proofs do not generalize to abstract topological spaces.

Alternate proof of cor.14.3 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 $x_{n_j} \in A$ such that x_{n_j} converges to x. It follows from theorem 12.6 (Sequence criterion for contact points in metric spaces) on p.384 that x is a contact point of A and hence $x \in \bar{A} = A$. This proves that A is (sequence) compact. ■

Alternate proof of cor.14.4 which uses sequence compactness (outline).

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 $f(x) \in f(X)$. ■

Corollary 14.5.

Let $(X, \mathfrak{U})$ be a topological space, let (Y, d) be a metric space, and let $f : X \to Y$ be continuous.

- *If X is compact then f is bounded.*
- *In particular, any continuous function on a closed interval of real numbers is bounded.*

The proof is left as exercise 14.3 (see p.451). ■

Corollary 14.6 (Continuous real–valued functions attain max and min on a compact domain)**.**

Let $(X, \mathfrak{U})$ be a topological space, $A \subseteq X$ a compact subspace and $f : A \to \mathbb{R}$ continuous on A. Then there exist $x_\star, x^\star \in A$ such that

$$f(x_\star) = \min_{x \in A} f(x) \quad \text{and} \quad f(x^\star) = \max_{x \in A} f(x).$$

PROOF: It follows from thm.14.12 on p.448 and thm.14.5 on p.443 that $f(A)$ is closed and bounded in $\mathbb{R}$. It follows from exercise 12.16 on p.400 that $\min\big((f(A)\big)$ and $\max\big(f(A)\big)$ exist, i.e., according to the definition of preimages, there exist elements in the domain A of f which are mapped to those two values. ■

The following theorem relates compactness and uniform continuity. [177]

Theorem 14.13 (Uniform continuity on sequence compact spaces)**.**

Let (X, d_1), (Y, d_2) be metric spaces and let A be a compact subset of X. Then,

- *any continuous function $A \to Y$ is uniformly continuous on A.*

[177] See Definition 13.5 on p.410.

PROOF: Let us assume to the contrary that f is continuous but not uniformly continuous and find a contradiction. Because f is not uniformly continuous, there exists $\varepsilon > 0$ such that no $\delta > 0$, however small, will satisfy (13.17) on p.410 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 $x_j, x'_j \in A$ such that

$$d_1(x_j, x'_j) < \frac{1}{j} \quad \text{but} \quad d_2(f(x_j), f(x'_j)) \geq \varepsilon. \tag{14.10}$$

Because A is compact, it is sequence compact. There is a subsequence (x_{j_k}) of the x_j which converges to an element $x \in A$ We have

$$d_1(x'_{j_k}, x) \leq d_1(x'_{j_k}, x_{j_k}) + d_1(x_{j_k}, x) \leq \frac{1}{j_k} + d_1(x_{j_k}, x). \tag{14.11}$$

Both right–hand terms converge to zero as $k \to \infty$. This is obvious for $1/j_k$ because $j_k \geq k$ for all k and it is true for $d_1(x_{j_k}, x)$ because x_{j_k} converges to x.
It follows from (14.11) that (x'_{j_k}) also converges to x. The (ordinary) continuity of f gives us

$$f(x) = \lim_{k\to\infty} f(x'_{j_k}) = \lim_{k\to\infty} f(x_{j_k}).$$

Since $\lim\limits_{k\to\infty} f(x_{j_k}) = f(x)$ and $\lim\limits_{k\to\infty} f(x'_{j_k}) = f(x)$ there exist $N, N' \in \mathbb{N}$ such that

$$d_2(f(x), f(x_{j_k})) < \frac{\varepsilon}{2} \text{ for } k \geq N; \quad d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} \text{ for } k \geq N'.$$

Both inequalities are true whenever $k \geq \max(N, N')$. It follows for all such k that

$$d_2(f(x_{j_k}), f(x'_{j_k})) < d_2(f(x_{j_k}), f(x)) + d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and we have a contradiction to (14.10). ■

Corollary 14.7 (Uniform continuity on closed intervals). *Let a, b be two real numbers such that $a \leq b$.*

Any continuous real–valued function on the closed interval $[a, b]$ is uniformly continuous:
For any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon \quad \text{for all } x, y \in [a, b] \text{ such that } |x - y)| < \delta \tag{14.12}$$

PROOF: This follows from the previous theorem (14.13) because closed intervals $[a, b]$ are closed and bounded sets and, in $\mathbb{R}$, any closed and bounded set is sequence compact . ■

14.5 Exercises for Ch.14

Exercise 14.1. Let $N \in \mathbb{N}$. Let $X := \{x_1, x_2, \dots, x_N\}$ be a finite set with a metric $d(\cdot, \cdot)$ (so (X, d) is a metric space). Prove that X is compact three different ways:

(a) Show sequence compactness to prove that X is compact.
(b) Show that X has the "extract finite open subcovering" property to prove that it is compact.
(c) Show that X is complete and totally bounded to prove that it is compact. □

Hints:

(a) ANY sequence in X possesses a constant subsequence (WHY?)

(b) If $(U_i)_i$ covers X then for each x there exists (at least one) i such that $x \in U_i$ (WHY?) How many of those U_i do you need to cover X if X has only N elements?

(c) Prop.12.36 on p.392 should prove useful.

Exercise 14.2. Prove the following which was used in the proof of thm.14.7 (Compact metric spaces are sequence compact) on p.445: Let (X, d) be a metric space and let $(x_n)_n$ be a sequence in X. Let $F := \{x \in X : x = x_j \text{ for some } j \in \mathbb{N}\}$. Let $z \in X$ be such that any neighborhood U of z contains infinitely many points of F. Then one can extract a subsequence $(x_{n_j})_j$ of $(x_n)_n$ such that $d(x_{n_j}, z) < \frac{1}{j}$. □

Exercise 14.3. Prove Corollary 14.5 on p.449 of this document:

Let $(X, \mathfrak{U})$ be a topological space, let (Y, d) be a metric space, and let $f : X \to Y$ be continuous. If X is compact then f is bounded.

In particular, if $h : [a, b] \to \mathbb{R}$ is a continuous function on the closed interval $[a, b]$ of real numbers (and both $[a, b]$ and $\mathbb{R}$ carry the Euclidean metric), then h is bounded. □

15 Applications of Zorn's Lemma

15.1 More on Partially Ordered Sets

Some of the following was copied almost literally from [9] Dudley.

Definition 15.1.

Let $(X, \preceq)$ be a POset (partially ordered set), $A \subseteq X$, and $m \in A$.

- m is called **maximal** for A iff there is no $a \in A$ such that $a \neq m$ and $m \preceq a$. m is called a **maximum** of A if $a \in A$ and $a \preceq m$ for all $a \in A$.
- m is called **minimal** for A iff there is no $a \in A$ such that $a \neq m$ and $m \succeq a$. m is called a **minimum** of A if $a \in A$ and $a \succeq m$ for all $a \in A$.

Proposition 15.1 below shows that such a maximum or minimum is unique. Thus, we may write $\max(A)$ for the maximum of A and $\min(A)$ for the minimum of A. □

It will be proved in prop.15.1 below that if A possesses a maximum and/or a minimum then it is unique. Thus we may write $\max(A)$ for the maximum of A and $\min(A)$ for the minimum of A.

Proposition 15.1.

Let $(X, \preceq)$ be a nonempty POset and $A \subseteq X$. If A has a maximum then it is unique.

PROOF: The proof is left as exercise 15.1. ■

Note 15.1 (Notes on maximal elements and maxima)**.**

(a) If $(X, \preceq)$ is not linearly ordered, then its subsets 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) Let $A \subseteq X$. If $m \in A$ is a maximum of A then this implies that m must be related to all other elements of A. □

For the following example we recall from Definition 5.5 (Linear orderings) on p.131 that a chain C in a POset $(X, \preceq)$ is a subset $C \subseteq X$ which is totally ordered, i.e., for any $x, x' \in C$ at least one of $x \preceq x'$ or $x' \preceq x$ is true.

Example 15.1 (Maximal elements, maxima and chains)**. (A)** Let X be the collection of all intervals $[a, b]$ of length $b - a \leq 2$ such that $a, b \in \mathbb{R}$ and $a \leq b$. These intervals are partially ordered by inclusion. Any interval of length equal to 2 is a maximal element in $(X, \subseteq)$. There is no maximum.

(B) Let

$$\begin{aligned} A &:= \{\, [3 + 1/n,\, 5 - 1/n] \;:\; n \in \mathbb{N} \}\,, \\ B &:= \{\, [4 + 1/n,\, 5 + 1/n] \;:\; n \in \mathbb{N} \}\,, \\ C &:= \{\, [8 - 1/n,\, 8 + 1/n] \;:\; n \in \mathbb{N} \}\,. \end{aligned}$$

Then A and C are chains in $(X, \subseteq)$, but B is not a chain in $(X, \subseteq)$. □

A hundred years ago the following was seen as extremely controversial by mathematicians who specialize in the foundations of mathematics.

Axiom 15.1 (Zorn's Lemma)**.**

Zorn's Lemma: Let $(X, \preceq)$ be a partially ordered set with the **ZL property**:

$$\text{Every chain } C \subseteq X\text{, possesses an upper bound } u \in X\text{, i.e., } c \preceq u \text{ for all } c \in C. \qquad \textbf{(ZL)}$$

Then X has a maximal element. □

Remark 15.1. 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 Zorn's Lemma 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 proven that if one accepts (rejects) Zorn's Lemma as a mathematical tool then this is equivalent to accepting (rejecting) the **Axiom of Choice** which states the following.

Let $\mathscr{A}$ be a collection of nonempty sets and let Ω be a set such that $\bigcup[A : A \in \mathscr{A}] \subseteq \Omega$. Then there exists a choice function on $\mathscr{A}$, i.e., a function $c : \mathscr{A} \to \Omega$ that satisfies $c(A) \in A$ for all $A \in \mathscr{A}$, $c(\cdot)$ picks or chooses an element $c(A)$ for each argument $A \in \mathscr{A}$: See Definition 5.23 on p.161).

(c) Moreover the Axiom of Choice, hence Zorn's Lemma, is equivalent to prop.5.8**(b)** on p.148: For **arbitrary**, not empty sets A, B the following is true. If $\varphi : A \to B$ is surjective then φ has a right inverse, i.e., a function $\psi : B \to A$ such that $\varphi \circ \psi = id_B$. For a proof see the optional Chapter 5.3 (Right Inverses and the Axiom of Choice). □

We will see now how Zorn's Lemma allows a surprisingly simple proof to the effect that **any** vector space has a basis.

15.2 Existence of Bases in Vector Spaces

The following is thematically a continuation of the material in chapter 11 (Vectors and vector spaces).

We will prove that every vector space, **even if it does not possess a finite subset which spans the entire space**, possesses a basis (see Definition 11.11 (Basis of a vector space) on p.338).

For the remainder of this chapter we assume that V is a vector space and define

$$\mathfrak{B} := \{A \subseteq V : A \text{ is linearly independent }\}. \tag{15.1}$$

Obviously $\mathfrak{B}$ is a partially ordered set with respect to set inclusion. The next lemma allows us to apply Zorn's Lemma. For the definition of a chain, see Definition 5.5 on p.452.

Lemma 15.1. *Every chain $\mathfrak{C}$ in $(\mathfrak{B}, \subseteq)$ possesses an upper bound.*

PROOF: Let $U := \bigcup \big[C : C \in \mathfrak{C}\big]$. We will show that U is linearly independent, i.e., $U \in \mathfrak{B}$. 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

$$\sum_{j=1}^{k} \alpha_j x_j \;=\; 0. \tag{15.2}$$

We must show that each α_j is zero. For each $0 \leq j \leq k$ 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 $0 \leq i, j \leq k$. But then there exists an index j_0 such that $C_{j_0} \supseteq C_j$ for all j, hence $x_1, x_2, \dots x_k \in C_{j_0}$. The set C_{j_0} is linearly independent because $C_{j_0} \in \mathfrak{C} \subseteq \mathfrak{B}$. It follows that $\alpha_1 = \dots = \alpha_k = 0$. ■

Theorem 15.1. *Every vector space V has a basis.*

PROOF: It follows from lemma 15.1 and Zorns Lemma (axiom 15.1 on p. 453) 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 $\mathit{span}(B) = V$.

Let us assume to the contrary that there exists $y \in \mathit{span}(B)^{\complement}$. It follows from lemma 11.2 on p.338 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 $(\mathfrak{B}, \subseteq)$. ■

15.3 The Cardinal Numbers are a totally ordered set

As another application of Zorn's Lemma we now prove thm.10.4, p.312, of ch.10.2 (Cardinality as a Partial Ordering).

Theorem 15.2. *Let $X, Y \subseteq \Omega$. Then $\mathbf{card}(X) \leq \mathbf{card}(Y)$ or $\mathbf{card}(Y) \leq \mathbf{card}(X)$*

PROOF: [178] The result is immediate if $X = \emptyset$ or $Y = \emptyset$. Assume X and Y are not empty. Let

$$\mathscr{F} \;:=\; \{D_f \xrightarrow{f} C_f : D_f \subseteq X,\ C_f \subseteq Y, \text{ and } f \text{ is bijective}\}$$

be the set of all bijective functions with domain contained in X and codomain contained in Y. We define a partial order $\preceq$ on $\mathscr{F}$ as follows: Let $D_f \xrightarrow{f} C_f$ and $D_g \xrightarrow{g} C_g$. Then

$$f \preceq g \quad \text{if and only if} \quad D_f \subseteq D_g,\ C_f \subseteq C_g, \quad \text{and } g\Big|_{D_f} = f.$$

We will prove that $\mathscr{F}$ has the ZL property: Let $\mathfrak{C}$ be a chain in $\mathscr{F}$. Let $D_u \xrightarrow{u} C_u$ be defined as follows:

$$D_u := \bigcup [D_f : f \in \mathfrak{C}], \quad C_u := \bigcup [C_f : f \in \mathfrak{C}], \quad u(x) := g(x) \text{ for } x \in D_g.$$

Note that the assignment $u(x) = g(x)$ for $x \in D_g$ is unambiguous: If there also is $g' \in \mathfrak{C}$ such that $x \in D_{g'}$ then we obtain from the total ordering of $\mathfrak{C}$ that $g' \preceq g$, i.e., g is an extension of g', or $g \preceq g'$, i.e., g' is an extension of g. In either case it follows that $g'(x) = g(x)$.

[178] See [11] Haaser/Sullivan: Real Analysis.

Moreover u is injective: If $x_1, x_2 \in D_u$ and $x_1 \neq x_2$ then there exist $f, g \in \mathfrak{C}$ such that $x_1 \in D_f$ and $x_2 \in D_g$. Since f extends g or vice versa, say, f extends g, we may assume that both $x_1, x_2 \in D_f$. It follows from the injectivity of f that $f(x_1) \neq f(x_2)$, hence $u(x_1) \neq u(x_2)$.

Also note that u is surjective: If $y \in C_u$ then $y \in C_f$ for some $f \in \mathfrak{C}$ and surjectivity gives us $x \in D_f$ such that $f(x) = y$. But then $x \in D_u$ and $u(x) = f(x) = y$.

A bijective function u has been constructed in such a fashion that it extends any $g \in \mathfrak{C}$ and hence is an upper bound of $\mathfrak{C}$. It follows that the partially ordered set $(\mathscr{F}, \preceq)$ possesses the ZL property (see axiom 15.1 (Zorn's Lemma) on p.453), hence there exists a maximal element $D_h \xrightarrow{h} C_h$ in $\mathscr{F}$.

We claim that $D_h = X$ or $C_h = Y$ (or both). Otherwise there would be $x_0 \in D_h^{\complement}$ and $y_0 \in C_h^{\complement}$ and the function

$$\psi : D_h \cup \{x_0\} \longrightarrow C_h \cup \{y_0\}, \qquad \begin{cases} x \mapsto h(x) & \text{if } x \in D_h, \\ x \mapsto y_0 & \text{if } x = x_0 \end{cases}$$

is a bijective extension of h, thus, $\psi \in \mathscr{F}$, and $h \preceq \psi$. Since $\psi \neq h$, this contradicts the maximality of h.

Case 1: $D_h = X$. Then changing C_h to Y makes h an injective function which maps X into Y, i.e., $\mathbf{card}(X) \leq \mathbf{card}(Y)$.

Case 2: $C_h = Y$. Then $h : D_h \xrightarrow{\sim} Y$ is a bijective function whose inverse $h^{-1} : Y \xrightarrow{\sim} D_h$ is an injection from Y into the subset D_h of X, i.e., $\mathbf{card}(Y) \leq \mathbf{card}(X)$. ■

15.4 Extensions of Linear Functions in Arbitrary Vector Spaces

We now turn our attention to extending a linear real–valued function f from a subspace $F \subsetneq V$ to the entire vector space V. Note that setting $f(x) = 0$ for all $x \in F^{\complement}$ does not yield a linear extension of f. See exercise 15.3 on p.466.

Lemma 15.2.

Let V be a vector space and let F be a (linear) subspace of V. Let $f : F \to \mathbb{R}$ be linear. Let

$$\mathscr{G} := \{(W, f_W) : W \text{ is a subspace of } V,\ W \supseteq F,\ f_W : W \to \mathbb{R} \text{ is a linear extension of } f \text{ to } W\}.$$

Then the following defines a partial ordering on $\mathscr{G}$:

$$(U, f_U) \preceq (W, f_W) \ \Leftrightarrow\ V \subseteq W \text{ and } f_W\big|_U = f_U .$$

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 $x \in W$ 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 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 $\mathscr{C}$. ■

Theorem 15.3 (Extension theorem for linear real–valued 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 $(U, f_U) \preceq (W, f_W) \Leftrightarrow U \subseteq W$ and $f_W\big|_U = f_U$.

We have seen in lemma 15.2 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 15.1 on p.453) and conclude that $\mathscr{G}$ possesses a maximal element (F', f') .

We show that $F' = V$.

If this was not true then we could find $a \in V \setminus F'$ and, according to prop.11.9 on p.340, applied with $V' = \mathbb{R}$ and $y_0 = \alpha$, extend f' to a linear function $\tilde{f}$ on $span(F' \uplus \{a\})$. It follows that $\big(span(F)' \uplus \{a\}, \tilde{f}\big) \in \mathscr{G}$ and $(F', f') \precneqq \big(span(F)' \uplus \{a\}, \tilde{f}\big)$.

This contradicts the maximality of (F', f'). and we have reached a contradiction. ■

Next, we will make a digression and talk about dual vector spaces and dual linear functions. All the material is $\boxed{\star}$ optional reading.

Definition 15.2 (Dual vector space)**.** $\boxed{\star}$ Let V and W be vector spaces, and let $L : V \to W$ be linear.

(a) We call $V^* := \{f : f$ is a linear function $V \to \mathbb{R}\}$ the **dual** or **algebraic dual** of V.

(b) We call $L^* : W^* \to V^*$, defined by $L^*(f) := f \circ L$,
i.e., $L^*(f)(x) = f(Lx) \;\forall\, x \in V$, the **dual function** or **dual mapping** of L. □

This is not a course about linear algebra, so we give here some elementary properties of algebraic duals of vector spaces [179] and dual functions without proof.

Proposition 15.2. ⋆

For the following see Definition 11.2 (Transposed matrix) on p.324.

(a) *V^* is a vector space, i.e., $f, g, \in V^*$ and $\alpha, \beta \in \mathbb{R} \Rightarrow \alpha f + \beta g \in V^*$.*
(b) *Since V^* is a vector space, its dual $V^{**} := \left(V^*\right)^*$ exists.*
(c) *Assume that $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. For every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ there exists a matrix $A = \big((a_{ij})\big)$ such that for every column vector $\vec{x}$, $L(\vec{x}) = A\vec{x}$, i.e., the function value $\vec{y} = L(\vec{x})$ has coordinates $y_i = \sum_{j=1^n} a_{ij}x_j$.*

(d) *If V is a finite dimensional vector space, then there is a bijection $V \to V^*$ which is linear in both directions. [180] This allows us to "identify" $\left(\mathbb{R}^n\right)^*$ with $\mathbb{R}^n$, thus, the dual function of L from Definition 15.2 is a linear function $L^* : \mathbb{R}^m \to \mathbb{R}^n$ (Careful: Switched dimensions!).*
*According to part **(c)** of this remark, there exists a matrix $A^* = \big((a_{k\ell})\big)$ such that the following is true. If $\vec{y}^* \in \mathbb{R}^m$ and $\vec{x}^* \in \mathbb{R}^n$ are column vectors such that $\vec{x}^* = L^*\big(\vec{y}^*\big)$, then $\vec{x}^* = A^*\vec{y}$, the product of the matrix A^* and the column vector $\vec{y}$.*
This matrix A^* is the transpose $A^\top$ of A: $a^*_{k\ell} = a_{\ell k}$ for $k = 1, \dots n$ and $\ell = 1, \dots, m$. □

PROOF: To be found in text books on linear algebra. ■

The following can be found as an exercise in [10] Friedberg, Insel, Spence: Linear Algebra.

Theorem 15.4. ⋆

Let $L : V \to W$ be a linear function between two vector spaces V and W. Let $L^* : W^* \to V^*$ be the associated dual function of L. Then,

$$L \text{ is injective } \Leftrightarrow L^* \text{ is surjective}\,; \qquad L^* \text{ is injective } \Leftrightarrow L \text{ is surjective}\,.$$

PROOF:" We will show that L^* is injective $\Leftrightarrow$ L is surjective.

Be aware that 0 can be the zero vector in any of V, W, V^*, W^*. In the latter two cases 0 denotes the (linear) zero function which maps any vector to zero.

PROOF of "$\Rightarrow$):"

We prove the contrapositive. We assume that L is not surjective and will show that L^* is not injective. Let $R := L(V)$ be the range of L. Since L is not surjective, $R \subsetneq W$. Let $w_0 \in W \setminus R$. Let $\psi \in W^*$ be defined as follows.

$$\psi(w) = 0 \text{ for } w \in R, \quad \psi(\alpha w_0) = \alpha\,.$$

Then ψ is defined, so far, on $\mathit{span}(W \uplus \{w_0\})$. According to Theorem 15.3 (Extension theorem for linear real–valued functions), ψ can be extended to a real–valued, linear function on all of W, which we call again ψ.

[179] If you work with vector spaces that carry a topology, e.g., normed vector spaces $(V, \|\cdot\|)$, then there also is the concept of the topological dual $V' := \{\, f \in V^* : f \text{ is continuous} \,\}$. If V is a finite dimensional normed space then one can show that $V' = V^*$, but this will not necessarily be true for normed spacea of infinite dimension,

Obviously, $\psi \neq 0$, since $\psi(\alpha) = 1$. On the other hand. for any $v \in V$,

$$L^*(\psi)(x) \;=\; (\psi \circ L)(x) \;=\; \psi\big(L(x)\big) \;=\; 0\,.$$

since $L(x) \in L(V) = R$, and we assumed that $\psi = 0$ on R. Thus $L^*(\psi) = 0$. Since also $L^*(0) = 0$ and $\psi \neq 0$, we conclude that L^* is not injective.

PROOF of "$\Leftarrow$):"

We show directly that L is surjective $\Rightarrow$ L^* is injective. Let $\varphi, \psi \in W^*$ such that $L^*(\varphi) = l^*(\psi)$. We must show that $\varphi = \psi$. It follows from $L^*(\varphi) = l^*(\psi)$ and the definition of the dual function that, for all $v \in V$,

$$\textbf{(A)} \qquad \varphi\big(L(v)\big) \;= L^*(\varphi)(v) \;=\; L^*(\psi)(v) \;=\; \psi\big(L(v)\big)\,.$$

Let $w \in W$. Since L is surjective, there exists $v \in V$ such that $w = L(v)$. It follows from **(A)** that

$$\varphi(w) \;=\; \varphi\big(L(v)\big) \;=\; \psi\big(L(v)\big) \;=\; \psi(w)\,.$$

Since $w \in W$ was arbitrary, this shows that $\varphi = \psi$. This concludes the proof. ■

Corollary 15.1.

*Let $A = \big((a_{ij})\big)$ be a matrix with m rows and n columns. Then **(a)** $\Leftrightarrow$ **(b)**, where*

(a) *The set of m linear equations in n unknowns $\vec{x} = (x_1, \dots, x_n)^\top$,*

$$A\,\vec{x} \;=\; \vec{y}\,,$$

has a solution $\vec{x}$ for any choice of right hand side $\vec{y} = (y_1, \dots, y_m)^\top$.

(b) *the set of n linear equations in m unknowns $\vec{\xi} = (\xi_1, \dots, \xi_m)^\top$,*

$$A^\top\,\vec{\xi} \;=\; \vec{\eta}\,,$$

has at most one solution $\vec{\xi}$ for any $\vec{\eta} = (\eta_1, \dots, \eta_n)^\top$.

PROOF: This follows from Theorem 15.4 and the fact that the linear function $L : \vec{x} \mapsto A\,\vec{x}$ has as its dual the function $L^* : \vec{\xi} \mapsto \vec{\xi}\,A^\top$. ■

15.5 The Hahn-Banach Extension Theorem ⋆

Note that this chapter is starred, hence optional. The proof given here is a more detailed version of the one found in [7] Choquet.

Let V be a vector space and let F be a (linear) subspace of V. Let $f : F \to \mathbb{R}$ be linear function. The Hahn-Banach Extension Theorem shows how to extend f from its domain F to a linear function on entire space V, subject to some majorization condition.

If V is a normed space (hence a metric space) and if f is continuous on F then this majorization condition can be chosen in such a fashion that the linear extension will be continuous on all of V.

In preparation for this subject matter we must study sublinear functions, which generalize both linear functions and norms.

15.5.1 Sublinear Functionals

Definition 15.3 (Sublinear functionals)**.** Let V be a vector space and $p : V \to \mathbb{R}$ such that

(a) if $\lambda \in \mathbb{R}_{\geq 0}$ and $x \in V$ then $p(\lambda x) \; = \; \lambda p(x)$ (positive homogeneity)
(b) if $x, y \in V$ then $p(x+y) \; \leq \; p(x) + p(y)$ (subadditivity)
Then we call p a **sublinear functional** on V. □

The term "functional" has several meanings in mathematics. We use it here to denote a real–valued function whose domain is a vector space.

Proposition 15.3.

Let V be a vector space and $p : V \to \mathbb{R}$ sublinear. Let $x \in V$. Then
(a) $p(0) = 0$,
(b) $-p(x) \leq p(-x)$,

PROOF of **(a)**: $p(0) = p(0 \cdot 0) = 0 \cdot p(0) = 0$.

PROOF of **(b)**: This follows from $0 = p(0) = p(x + (-x)) \; \leq p(x) + p(-x)$. ■

Example 15.2 (Norms are sublinear)**.** Let $(V, \|, \cdot\|)$ be a normed vector space. Then the function $p(x) := \|x\|$ is sublinear.

Indeed, norms are absolutely homogenous: We have $\|\lambda x\| \; = \; |\lambda| \cdot \|x\|$ not only for $\lambda \geq 0$ but for all $\lambda \in \mathbb{R}$. Further, subadditivity is just the validity of the triangle inequality. □

Example 15.3 (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 $\lambda \geq 0$ but for all $\lambda \in \mathbb{R}$.

Further linear functions satisfy additivity: $f(x+y) = f(x) + f(y)$,
hence also subadditivity $f(x+y) \leq f(x) + f(y)$. □

More about sublinearity can be found in chapter 15.6 on p.463

15.5.2 The Hahn-Banach extension theorem and its Proof

This chapter follows closely [7] Choquet.

As mentioned previously, the subject of this chapter is the extension of a linear, real–valued function from a subspace to the entire vector space in such a fashion that some majorization condition will be preserved. The Hahn-Banach extension theorem (theorem 15.5 below) states that if p is a sublinear functional defined on all of a vector space V and if a linear, real–valued function f is defined on a subspace F of V such that $f \leq p$ on F, then f can be linearly extended to V in such a way that p dominates this extension everywhere on V. Once we have that, it is not very difficult to prove what we truly want, thm.15.6 (Continuous extensions of continuous linear functions).

The following remark is about first extending f to "one more dimension".

Remark 15.2. 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 $a \in V \setminus F$. We saw in prop.11.9 on p.340 that any linear extension $\tilde{f}$ of f to $\mathit{span}(F \uplus \{a\})$ is uniquely determined by its value $k := \tilde{f}(a)$.

Indeed, any $x \in \mathit{span}(F \uplus \{a\})$ 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

$$\tilde{f}(x + \lambda a) \ = \tilde{f}(x) + \lambda \tilde{f}(a) \ = \ f(x) + \lambda k. \ \square \tag{15.3}$$

Theorem 15.5 (Hahn–Banach extension theorem)**.**

Let V be a vector space and $p : V \to \mathbb{R}$ a sublinear function.
Suppose F is a (linear) subspace of V and $f : F \to \mathbb{R}$ is a linear mapping such that $f \leq p$ on F.
Then there is an extension of f to a linear map $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f} \leq p$ on V.

Before proving this theorem, first we prove two lemmata.

Lemma 15.3. *Suppose F is a subspace of V, $f : F \to \mathbb{R}$ is a linear mapping, $a \in V \setminus F$, and $k \in \mathbb{R}$. Let $\tilde{f}$ be the linear extension of f to $\mathit{span}(F \biguplus \{a\})$ given in prop.11.9 on p.340, choosing $V' = \mathbb{R}$ and $y_0 = k$:*

$$\tilde{f}(x + \lambda a) \ := \ f(x) + \lambda k, \quad \textit{i.e.}, \quad \tilde{f}(a) \ = \ k. \tag{15.4}$$

Then,

$$k \leq \inf_{u \in F} \{p(u + a) - f(u)\} \ \Leftrightarrow \ \tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a) \ \textit{ for all } \lambda > 0 \textit{ and } x \in F, \tag{15.5}$$

$$k \geq \sup_{v \in F} \{f(v) - p(v - a)\} \ \Leftrightarrow \ \tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a) \ \textit{ for all } \lambda < 0 \textit{ and } x \in F. \tag{15.6}$$

Further,

$$\begin{aligned} &\sup_{v \in F} \{f(v) - p(v - a)\} \ \leq k \leq \ \inf_{u \in F} \{p(u + a) - f(u)\} \\ &\qquad\qquad \Leftrightarrow \ \tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a) \ \textit{ for all } \lambda \in \mathbb{R} \textit{ and } x \in F, \end{aligned} \tag{15.7}$$

Proof of (15.5), ⇒): Let us assume that $\lambda > 0$ and $x \in F$. Then $u := \frac{x}{\lambda} \in F$ because F is a subspace. On account of the left–hand side of (15.5),

$$\begin{aligned} \tilde{f}(x + \lambda a) \ &= f(x) + \lambda k \ = \ \lambda\big(f(x/\lambda) + k\big) \ = \ \lambda\big(f(u) + k\big) \ \leq \ \lambda\Big(f(u) + \big(\ p(u + a) - f(u)\big)\Big) \\ &= \lambda p(u + a) \ = \ \lambda p(x/\lambda + a) \ = \ p(x + \lambda a) \end{aligned}$$

The inequality follows from the left–hand side of (15.5), and we used the positive homogeneity of p for the last equation.

Proof of (15.5), ⇐): We assume $\tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a)$ for all $\lambda > 0$ and $x \in F$. We will show that $k = \tilde{f}(a) \ \leq \ p(u + a) - f(u)$ for all $u \in F$.

$$p(u + a) - f(u) \ \geq \tilde{f}(u + a) - f(u) \ = \ \tilde{f}(u) + \tilde{f}(a) - f(u) \ = \ f(u) + \tilde{f}(a) - f(u) \ = \ \tilde{f}(a) \ = \ k.$$

Proof of (15.6), ⇒): Let us assume that $\lambda < 0$ and $x \in F$. Then $v := \frac{x}{\lambda} \in F$ because F is a subspace.

Because of the left–hand side of (15.6) and $\lambda < 0$ and positive homogeneity of p,

$$\begin{aligned} k \ \geq \ f(v) - p(v-a) \ &\Rightarrow \lambda k \ \leq \ f(\lambda v) - \lambda p(v-a) \\ &\Rightarrow \ -f(\lambda v) + \lambda k \ \leq (-\lambda)p(v-a) \ = \ p\big((-\lambda)(v-a)\big) \ = \ p\big((-\lambda)v + \lambda a\big). \end{aligned}$$

Since $x = \lambda v$, and since linearity of f implies $-f(x) = f(-x)$, that last inequality yields

$$f(-x) + \lambda k \ = \ -f(x) + \lambda k \ \leq p(-x + \lambda a), \ \text{ hence } \ \tilde{f}(-x + \lambda a) \ = \ f(-x) + \lambda k \ \leq \ p(-x + \lambda a)$$

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 (15.6), ⇐): We assume $\tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a)$ for all $\lambda < 0$ and $x \in F$. We now show that $k = \tilde{f}(a) \ \geq \ f(v) - p(v - a)$ for all $v \in F$.

We apply $\tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a)$ with $x := v$ and $\lambda := -1$ and obtain

$$-p(v-a) + f(v) \ \leq \ -\tilde{f}(v-a) + f(v) \ = \ \tilde{f}(a-v) + f(v) \ = \ \tilde{f}(a) - \tilde{f}(v) + f(v) \ = \ \tilde{f}(a) \ = \ k.$$

Proof of (15.7), ⇒): Let us assume that $\lambda \in \mathbb{R}$ and $x \in F$. For $\lambda \neq 0$, validity of the right–hand side of (15.7) follows from (15.5) and (15.6). If $\lambda = 0$ then we must show that $\tilde{f}(x) \leq p(x)$. This is true because $x \in F$ implies $\tilde{f}(x) = f(x)$ and we assumed that $f \leq p$ on F.

Proof of (15.7), ⇐): This is immediate from (15.5) and (15.6). ■

Lemma 15.4. *Let V be a vector space and $p : V \to \mathbb{R}$ a sublinear function. Let $F \subsetneq V$ be a genuine subspace of V and $a \in V \setminus F$. Let $f : F \to \mathbb{R}$ be a linear mapping with $f \leq p$ on 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 such that $\tilde{f} \leq p$ on G,*
- ***(b)*** *This extension is unique if and only if $\sup_{v\in F}\{f(v) - p(v-a)\} \ = \ \inf_{u \in F}\{p(u+a) - f(u)\}$.*

Proof of a. For $u, v \in F$ we have

$$f(u) + f(v) \ = f(u+v) \leq p(u+v) \ = \ p\big((u+a) + (v-a)\big) \ \leq \ p(u+a) + p(v-a),$$

and hence $f(v) - p(v-a) \ \leq \ p(u+a) - f(u)$. Therefore

$$\sup_{v\in F}\{f(v) - p(v-a)\} \ \leq \ \inf_{u\in F}\{p(u+a) - f(u)\}. \tag{15.8}$$

For a fixed $k \in \mathbb{R}$, we define $\tilde{f}(x + \lambda a) \ = \ f(x) + \lambda k$. We claim that $\tilde{f} \leq p$ if and only if we have

$$\sup_{v\in F}\{f(v) - p(v-a)\} \ \leq k \ \leq \ \inf_{u\in F}\{p(u+a) - f(u)\} \tag{15.9}$$

which will conclude the proof of **(a)** since such a k exists by 15.8.

Our claim holds because $f(x) + \lambda k \ = \ \tilde{f}(x + \lambda a) \ \leq \ p(x + \lambda a)$ for all λ if and only if

$$\begin{aligned} k \ &\leq p(u+a) - f(u) \quad \text{for all } u \in F \\ \text{and } \ k \ &\geq f(v) - p(v-a) \quad \text{for all } v \in F \end{aligned}$$

(the cases $\lambda > 0$ and $\lambda < 0$ respectively). This was proved in lemma 15.3.

Proof of b. From (15.9) we deduce that k is unique if and only if

$$\sup_{v\in F}\{f(v)-p(v-a)\} \;=\; \inf_{u\in F}\{p(u+a)-f(u)\}.$$

Because the extension $\tilde{f}(x+\lambda a) \;=\; f(x)+\lambda k$ of f to G is uniquely determined by k and vice versa, we have proven b. ■

PROOF of thm.15.5 (Hahn–Banach Extension Theorem):

Let $\mathscr{G} \;=\; \{(V',g):\; V'$ is a subspace of $V, V'\supseteq F,\; g: V'\to\mathbb{R}$ is linear, $g\big|_F = f,$ and $g\;\leq p$ on $V'\}$.
We define a partial order "$\preceq$" on $\mathscr{G}$ as follows:

(15.10) $$(V_1,g_1)\preceq(V_2,g_2) \;\Leftrightarrow\; V_1\subseteq V_2 \text{ and } g_2 \text{ is an extension of } g_1.$$

Note that $\mathscr{G}$ is not empty because $(F,f)\in\mathscr{G}$.

(a) We first prove that any chain $\mathscr{C}\subseteq(\mathscr{G},\preceq)$ has an upper bound: Let $W := \bigcup\big[V' : (V',g)\in\mathscr{C}\big]$. Then W is a subspace of V 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, and we conclude that $x+\lambda y\in V_2$, hence $x+\lambda y\in W$. It follows that W is a subspace.

Next we construct a linear $h: W\to\mathbb{R}$ such that $h\leq p$ 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 $h(x)=g_1(x)\leq p(x)$, i.e., $h\leq p$ 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 to conclude that both x,y belong to some subspace V' such that $(V',g)\in\mathscr{C}$. 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, $V'\subseteq W = \bigcup\big[U : U\in\mathscr{C}\big]$. We have seen that h is linear, dominated by p, and 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) It follows from **(a)** that we can apply Zorn's Lemma (axiom 15.1 on p.453), and hence conclude that $(\mathscr{G},\preceq)$ possesses a maximal element (M,m).

We show that $M = V$. Assume to the contrary that there exists $a\in V\setminus M$. According to lemma 15.4 on p.461, we can extend m to a linear function $\tilde{m}$ on $M\uplus\{a\}$ in such a fashion that $\tilde{m}\leq p$ on $M\uplus\{a\}$. It follows that $(M\uplus\{a\},\tilde{m})\in\mathscr{G}$. Further, $(M,m)\precneqq(M\uplus\{a\},\tilde{m})$ because $(M\subsetneqq M\uplus\{a\}$. This contradicts the maximality of (M,m). We conclude that $M=V$.

The proof is finished: m is the linear extension of f to V we were looking for. ■

Theorem 15.6 (Continuous extensions of continuous linear functions).

Let $(V,\|\cdot\|)$ be a normed vector space. Let F be a (linear) subspace of V. Then,

- *any continuous, linear $f: F\to\mathbb{R}$ possesses a continuous, linear extension $\tilde{f}: V\to\mathbb{R}$.*

PROOF: Let $p(x) := \|f\| \cdot \|x\|$ [181] Because p is a positive multiple of the norm $\|\cdot\|$ on all of V, it also is a norm on V (see Remark 11.7) on p.345), hence sublinear by example 15.2 on p.459. According to the Hahn-Banach extension theorem there exists a linear extension $\tilde{f}$ of f to all of V such that

$$\tilde{f}(x) \leq p(x) \text{ for all } x \in V. \tag{15.11}$$

We replace x with $-x$ and obtain from the linearity of $\tilde{f}$ that $-\tilde{f}(x) = \tilde{f}(-x) \leq p(-x)$.
We note that $p(x) = p(-x)$ because p is a norm. Hence

$$-\tilde{f}(x) \leq p(-x) = p(x), \text{ i.e., } \tilde{f}(x) \geq -p(x).$$

Together with (15.11) this shows that

$$-p(x) \leq \tilde{f}(x) \leq p(x) \text{ for all } x \in V,$$

and thus

$$|\tilde{f}(x)| \ \leq \ p(x) \ = \ \|f\| \cdot \|x\| \text{ for all } x \in V \tag{15.12}$$

It follows from (15.12) that f has been extended in such a way that $\|\tilde{f}\| \leq \|f\|$. We aply the continuity criterion for linear functions (thm.13.4 on p.412) 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 $\tilde{f}$. ■

15.6 Convexity ★

Note that this chapter is starred, hence optional.

Definition 15.4 (Concave-up and convex functions)**.**

Let $-\infty \leq \alpha < \beta \leq \infty$ and let $I :=]\alpha, \beta[$ be the open interval of real numbers with endpoints α and β. Let $f : I \to \mathbb{R}$.

- **(a)** The **epigraph** of f is the set $epi(f) := \{(x_1, x_2) \in I \times \mathbb{R} : x_2 \geq f(x_1)\}$ 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 $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leq \lambda \leq 1$ is contained in $epi(f)$. See Figure 15.1.
- **(c)** Let f be differentiable at all points $x \in I$. Then f is **concave-up**, if the function $f' : x \mapsto f'(x) = \frac{df}{dx}(x)$ is increasing. □

Proposition 15.4 (Convexity criterion)**.**

f is convex if and only if the following is true: For any

$$\alpha \ < \ a \ \leq \ x_0 \ \leq \ b \ < \ \beta$$

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

$$f(x_0) \ \leq \ S(x_0). \tag{15.13}$$

[181] See Definition 13.6 on p.412 for the definition of $\|f\|$

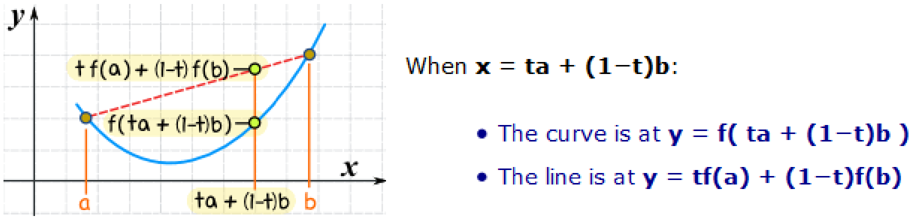

y
x
t f(a) + (1-t) f(b)
f(ta + (1-t)b)
a
ta + (1-t)b
b
When x = ta + (1−t)b:
• The curve is at y = f(ta + (1−t)b)
• The line is at y = tf(a) + (1−t)f(b)

Figure 15.1: Convex function
182

Note that any x_0 between a and b can be written as $x_0 = \lambda a + (1-\lambda)b$ for some $0 \le \lambda \le 1$ 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 $a < b$ such that $a, b \in I$ and $0 \le \lambda \le 1$ it is true that

$$f(\lambda a + (1-\lambda)b) \ \le \lambda f(a) + (1-\lambda)f(b). \tag{15.14}$$

PROOF of "⇒": Any line segment S that connects the points $(a, f(a))$ and $(b, f(b))$ in such a way that S is entirely contained in the epigraph of f will satisfy $(x_0, S(x_0)) \in epi(f)$ and hence $f(x_0) \le S(x_0)$ for all $a \le x_0 \le b$. It follows that convexity of f implies (15.13).

PROOF of "⇐": Let (15.13) be valid for all $a, b \in I$. Let $\vec{a} = (a_1, a_2), \vec{b} = (b_1, b_2) \in epi(f)$. Then

$$a_2 \ge f(a_1) \ \text{ and } \ b_2 \ge f(b_1). \tag{15.15}$$

We must show that the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \le \lambda \le 1$ is contained in $epi(f)$.

Let $\vec{a}' := (a_1, f(a_1))$. Let $S' := \{\lambda\vec{a}' + (1-\lambda)\vec{b} : 0 \le \lambda \le 1\}$ 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 $S'' := \{\lambda\vec{a}' + (1-\lambda)\vec{b}' : 0 \le \lambda \le 1\}$ 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 $a \le x_0 \le b$ we have $f(x_0) \le S''(x_0)$. We conclude from $S(\cdot) \ge S'(\cdot) \ge S''(\cdot)$ that $S(x_0) \ge f(x_0)$, i.e. $(x_0, S(x_0)) \in epi(f)$. As this is true for any $a \le x_0 \le b$ it follows that the line segment S is entirely contained in the epigraph of f. ■

Proposition 15.5 (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 there are $a, b, x_0 \in I$ such that $a < x_0 < b$ and $f(x_0) > S(x_0)$. Here $S(x_0)$ denotes 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 $(b, f(b))$.
Let m be the slope of the linear function $S(\cdot) : x \mapsto S(x)$, i.e.,

$$m \ = \ \frac{S(b) - S(a)}{b - a}.$$

It follows that

$$(15.16) \qquad m \ = \frac{S(b) - S(x_0)}{b - x_0} > \frac{S(b) - f(x_0)}{b - x_0} = \frac{f(b) - f(x_0)}{b - x_0} = f'(\xi)$$

for some $x_0 < \xi < b$ (according to the mean value theorem for derivatives). Further

$$(15.17) \qquad m \ = \frac{S(x_0) - S(a)}{x_0 - a} < \frac{f(x_0) - S(a)}{x_0 - a} = \frac{f(x_0) - f(a)}{x_0 - a} = f'(\eta)$$

for some $a < \eta < x_0$ (according to the mean value theorem for derivatives).
Because f is concave up we have

$$f'(a) \ \leq \ f'(\eta) \ \leq \ f'(x_0) \ \leq \ f'(\xi) \ \leq \ f'(b).$$

From (15.16) and (15.17) we obtain

$$m \ < \ f'(\eta) \ \leq \ f'(x_0) \ \leq \ f'(\xi) \ < \ m,$$

and we have reached a contradiction. ■

Proposition 15.6 (Sublinear functions are convex). *Let $f : \mathbb{R} \to \mathbb{R}$ be sublinear. Then f is convex.*

PROOF: Let $0 \leq \lambda \leq 1$ and $x, y \in \mathbb{R}$. Then

$$(15.18) \qquad p(\lambda x + (1 - \lambda)y) \ \leq p(\lambda x) \ + \ p((1 - \lambda)y) \ = \ \lambda p(x) \ + \ (1 - \lambda)p(y).$$

It follows from prop.15.4 that f is concave-up. ■

15.7 Exercises for Ch.15

Exercise 15.1. Prove prop.15.1 on p.452: Let $(X, \preceq)$ be a nonempty POset and $A \subseteq X$. If A has a maximum then it is unique. □

Exercise 15.2. Let $A \subseteq \mathbb{R}^2$ and $\vec{x}_1 = (x_1, y_1), \vec{x}_2 = (x_2, y_2) \in A$. Let

$$\vec{x}_1 \ \preceq \ \vec{x}_2 \ \Leftrightarrow \ x_1 \leq x_2 \text{ and } y_1 \leq y_2.$$

(a) Prove that "$\preceq$" defines a partial order on A.
(b) Prove that no maximal elements exist if $A = \mathbb{R}^2$.
(c) What are the maximal elements of $A = \ \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\|_2 = 1\}$? □

Exercise 15.3. The following was stated at the beginning of ch.15.4 (Extensions of Linear Functions in Arbitrary Vector Spaces) on p.455.

Let $F \subseteq V$ be a subspace of a vector space V and let $f : F \to \mathbb{R}$ be linear. Let

$$g : F \to \mathbb{R}; \qquad x \mapsto \begin{cases} f(x) & \text{if } x \in F, \\ 0 & \text{if } x \in F^{\complement}. \end{cases}$$

Then g is linear only if $f(x) = 0$ for all $x \in F$. Prove it. □

16 Approximation theorems $\boxed{\star}$

Note that this chapter is starred, hence optional,

Introduction 16.1. Everyone who knows about limits is familiar with the concept of approximations. For example the sequence $\left(\frac{1}{n}\right)_n$ becomes arbitrarily close to the number zero, i.e., it approximates zero since $\lim\limits_{n\to\infty}\frac{1}{n}=0$. Another example is the sequence $\big(s_n(x)\big)_n$ which we define for $x\in\mathbb{R}$ as $s_n(x):=\sum\limits_{j=0}^{n}\frac{x^j}{j!}$. It converges for each x to the number e^x and thus approximates that number. We can rephrase this last example as the approximation of the function $f: x\mapsto e^x$ by the sequence of functions $s_n: x\mapsto\sum\limits_{j=0}^{n}\frac{x^j}{j!}$.

Here is another example about the approximation of functions with sequences of functions. For Let U be an open subset of $\mathbb{R}$ and let $f: U\to\mathbb{R}$ be a function which is differentiable for each $x\in U$. For each $n\in\mathbb{N}$ let

$$\triangle_n^f: U\to\mathbb{R};\qquad x\mapsto\triangle_n^f(x):=\begin{cases}\frac{1}{n}\big(f(x+\frac{1}{n})-f(x)\big) & \text{if } x+\frac{1}{n}\in U,\\ 0 & \text{otherwise.}\end{cases}$$

Since U is open x is an interior point of U, thus $x+\frac{1}{n}\in U$ eventually, thus $\lim\limits_{n\to\infty}\triangle_n^f(x)=f'(x)$, the derivative of f at x

The convergence $\lim\limits_{n\to\infty}\sum\limits_{j=0}^{n}\frac{x^j}{j!}\ =e^x$ is much better behaved than the convergence $\lim\limits_{n\to\infty}\triangle_n^f(x)=f'(x)$ since one can prove that it is uniform on each interval $[a,b]$ where $a,b\in\mathbb{R}$. Matter of fact the following is true for any "power series" $s(x):=\sum\limits_{j=0}^{\infty}c_j(x-x_0)^j$ where $x_0\in\mathbb{R}$ and the "coefficients" $c_j\in\mathbb{R}$ for all j. If $r,a,b\in\mathbb{R}$ such that $x_0-r<a<b<x_0+r$ (thus $r>0$!) and such that the partial sum functions $s_n(x):=\sum\limits_{j=0}^{n}c_j(x-x_0)^j$ converge pointwise to $s(x)$ for $a\le x\le b$ then this convergence is uniform in x.

On the other hand the convergence of the difference quotient functions $\triangle_n^f(\cdot)$ to the derivative f' will generally not be uniform on such closed and bounded intervals, and how could it? The function f is continuous on U since it is differentiable at each $x\in U$, thus $\triangle_n^f(\cdot)$ is continuous as the scalar multiple of the difference of the two continuous functions $f(\cdot+\frac{1}{n})$ and f. But it follows from thm.13.6 on p.418 that uniform limits of continuous functions are continuous, and there are plenty of differentiable functions which are not continuous at all points.

Here is an example.

$$\text{Let }\ f(x)\ :=\begin{cases}-x^2 & \text{if } x<0,\\ x^2 & \text{if } x\ge 0.\end{cases}\qquad \text{Then }\ f'(x)\ =\ \begin{cases}-2x & \text{if } x<0,\\ \ \ 2x & \text{if } x\ge 0\end{cases}\ =\ 2|x|$$

is not differentiable at $x=0$. $\square$

So the question arises what kind of functions are the uniform limit of functions of a more specialized nature. Weierstrass found in the late 19th century a very general answer in his approximation theorem:

For any continuous function f on a closed and bounded interval $[a,b]$ one can find a sequence of polynomials $p_n(x)$ such that f is the uniform limit of those polynomials.

We will prove Weierstrass's approximation theorem in this chapter.

16.1 The Positive, Linear Operators $f \mapsto B_n^f$

Introduction 16.2. We examined in ch.6.7 (Bernstein Polynomials) the n–th Bernstein polynomial

$$B_n^f : \mathbb{R} \to \mathbb{R}; \qquad x \mapsto B_n^f(x) \ = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}$$

which one can associate with any function $f : [0,1] \to \mathbb{R}$, no matter how badly it may behave.

If we think of f as the argument of an assignment $f \mapsto B_n^f$ then this mapping defines a function $B_n : \mathscr{F}([0,1],\mathbb{R}) \to \mathscr{F}([0,1],\mathbb{R})$ from the set $\mathscr{F}([0,1],\mathbb{R})$ of all real–valued functions on the unit interval to itself.

We will see in prop.16.2 below that each of those functions B_n is a linear function from the vector space $\mathscr{F}([0,1],\mathbb{R})$ to itself. This proposition further shows that each B_n is positive in the sense that if the argument f is nonnegative, i.e., $f(x) \geq 0$ for all $0 \leq x \leq 1$ then its image B_n^f also is nonnegative.

The functions B_n^f are continuous since all polynomials are continuous, and this would allow us to shrink the codomain of B_n to the set $\mathscr{C}([0,1],\mathbb{R})$ of all continuous, real–valued functions on the unit interval, or even to the set of all polynomials $p(x)$ where $0 \leq x \leq 1$. We prefer not to do so for the following reason:

It is customary to call a linear function $F : \mathscr{F} \to \mathscr{F}$ which possesses a vector space $\mathscr{F}$ of real–valued functions both as domain and codomain a linear operator on $\mathscr{F}$, and to call a linear operator on $\mathscr{F}$ which assigns nonnegative functions to nonnegative functions a positive linear operator on $\mathscr{F}$. In short, we will see in this chapter that the assignments $B_n : \mathscr{F}([0,1],\mathbb{R}) \to \mathscr{C}([0,1],\mathbb{R})$ are positive linear operators for each $n \in \mathbb{N}$. □

Definition 16.1 (Positive linear operators)**.**

Let (X,d) be a metric space, and let $\mathscr{F}$ be a subspace of the vector space $\mathscr{F}(X,\mathbb{R})$, i.e., with any two functions $f(\cdot), g(\cdot) \in \mathscr{F}$ their sum $f+g$ also belongs to $\mathscr{F}$ and that the function λf $(\lambda \in \mathbb{R})$ also belongs to $\mathscr{F}$.

A ***linear operator*** T on $\mathscr{F}$ is a linear function [183] $T : \mathscr{F} \to \mathscr{F}$

A ***positive linear operator*** T on $\mathscr{F}$ is a linear operator on $\mathscr{F}$ with the following property:

$$(16.1) \quad f \geq 0 \ \Rightarrow \ Tf \geq 0, \quad \text{i.e.,} \quad f(x) \geq 0 \text{ for all } x \in X \ \Rightarrow \ Tf(x) \geq 0 \text{ for all } x \in X.$$

Proposition 16.1 (Properties of positive linear operators)**.**

Let T be a positive linear operator on a subspace $\mathscr{F}$ of $\mathscr{F}(X, \mathbb{R})$. Then,

- **(a)** *T is **monotone increasing**, i.e., for any two functions $f, g \in \mathscr{F}$ such that $f \leq g$ it is true that $T(f) \leq T(g)$. In other words,*

$$f(x) \leq g(x) \text{ for all } x \in X \quad \Rightarrow \quad T(f)(x) \leq T(g)(x) \text{ for all } x \in X. \tag{16.2}$$

- **(b)** *Assume that $T(|f|)$ is defined for a function $f \in \mathscr{F}$. Then $|T(f)| \leq T(|f|)$. In other words,*

$$|T(f)(x)| \ \leq \ T(|f|)(x) \qquad \text{for all } x \in X. \tag{16.3}$$

Proof of **(a)**:

If $f \leq g$ then $g - f \geq 0$, hence $T(g - f) \geq 0$ since T is positive. Linearity of T then yields $T(g) - T(f) = T(g - f) \geq 0$. But $T(g) - T(f) \geq 0$ means the same as $T(f) \leq T(g)$, and we have proven part **(a)**.

Proof of **(b)**:

We have $f \leq |f|$ since $f(x) \leq |f(x)|$ for all $x \in X$. It follows from **(a)** that $T(f) \leq T(|f|)$. We also have $-f \leq |f|$ since $-f(x) \leq |f(x)|$ for all $x \in X$. It now follows from **(a)** that $T(-f) \leq T(|f|)$. But T is linear, so $T(-f) = -T(f)$, thus we have -$T(f) \leq T(|f|)$. In summary we have shown that both

$$T(f)(x) \leq T(|f|)(x) \ \text{ and } \ -T(f)(x) \leq T(|f|)(x) \ \text{ for all } x \in X.$$

It now follows from prop.3.54 on p.76 that $|T(f)(x)| \leq T(|f|)(x)$ for all $x \in X$, i.e., $|T(f)| \leq T(|f|)$. and we have proven part **(a)**. We have proven part **(b)**. ■

Proposition 16.2 (Linearity and positivity of Bernstein polynomial assignments).

- **(a)** *Let $f(\cdot), g(\cdot)$ be two real–valued functions on $[0,1]$ and $\alpha, \beta \in \mathbb{R}$. Let $h : [0,1] \to \mathbb{R}$ be defined as*

$$\begin{aligned} h \ &:= \ \alpha f + \beta g, \ \textit{i.e.}, \ h(x) \ = \ \alpha f(x) + \beta g(x) \ (0 \leq x \leq 1). \\ \textit{Then} \ \ B_n^h \ &= \ \alpha B_n^f + \beta B_n^g, \ \textit{i.e.}, \ B_n^h(x) \ = \ \alpha B_n^f(x) + \beta B_n^g(x) \ (x \in \mathbb{R}). \end{aligned}$$

  To express this more succinctly:

$$B_n^{\alpha f + \beta g} \ = \ \alpha B_n^f + \beta B_n^g. \tag{16.4}$$

- **(b)** *Let f be a real–valued function on $[0,1]$ which is nonnegative, i.e., $f(x) \geq 0$ for $0 \leq x \leq 1$. Then $B_n^f(\cdot) \geq 0$ on $[0,1]$ (but not necessarily for $x \notin [0,1]$).*

PROOF of **(a)**: For any $x \in [0,1]$ we have

$$\begin{aligned}
B_n^h(x) &= \sum_{k=0}^{n} \binom{n}{k} h(\tfrac{k}{n})\, x^k\, (1-x)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} \Big(\alpha f(\tfrac{k}{n}) + \beta g(\tfrac{k}{n})\Big)\, x^k\, (1-x)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} \alpha f(\tfrac{k}{n})\, x^k\, (1-x)^{n-k} \;+\; \sum_{k=0}^{n} \binom{n}{k} \beta g(\tfrac{k}{n})\, x^k\, (1-x)^{n-k} \\
&= \alpha \sum_{k=0}^{n} \binom{n}{k} f(\tfrac{k}{n})\, x^k\, (1-x)^{n-k} \;+\; \beta \sum_{k=0}^{n} \binom{n}{k} g(\tfrac{k}{n})\, x^k\, (1-x)^{n-k} \\
&= \alpha B_n^f(x) \;+\; \beta B_n^g(x)
\end{aligned}$$

PROOF of **(b)**: For any $x \in [0,1]$ we have both $0 \leq x \leq 1$ and $0 \leq (1-x) \leq 1$. It follows that both x^k and $(1-x)^{n-k}$ are nonnegative as products of nonnegative numbers.

Note that $0 \leq k \leq n$ implies that $\frac{k}{n} \in [0,1]$. We assumed that $f \geq 0$ on $[0,1]$, thus all numbers $f(k/n)$ are nonnegative. Finally, any binomial coefficient is nonnegative because it is defined as $\frac{n!}{k!(n-k)!}$, and the numbers $n!, k!$ and $(n-k)!$ all are nonnegative.

Thus each summand $\binom{n}{k} f\left(\frac{k}{n}\right) x^k\, (1-x)^{n-k}$ is nonnegative as a product of nonnegative numbers.

It follows that $B_n^f(x)$ is nonnegative as the sum of nonnegative items.

To summarize: If restricted to the continuous functions $\mathscr{C}([0,1],\mathbb{R})$, each $B_n(\cdot)$ is a positive linear operator on $\mathscr{C}([0,1],\mathbb{R})$. ■

Corollary 16.1. *Let $n \in \mathbb{N}$. Then $B_n(\cdot)$ is a positive linear operator on $\mathscr{C}([0,1],\mathbb{R})$.*

PROOF:

Since $B_n^f\big(\mathscr{C}([0,1],\mathbb{R})\big) \subseteq B_n^f\big(\mathscr{F}([0,1],\mathbb{R})\big) \subseteq B_n^f\big(\mathscr{C}([0,1],\mathbb{R})\big)$ it follows from prop.16.2 above that the restriction of B_n^f to $\mathscr{C}([0,1],\mathbb{R})$ is a linear positive operator on $\mathscr{C}([0,1],\mathbb{R})$.

■

16.2 Korovkin's First Theorem

Introduction 16.3. The main task of this chapter will be the proof of Korovkin's first theorem. This theorem gives a simple condition which guarantees that a sequence T_n of positive linear operators defined on $\mathscr{C}([a,b],\mathbb{R})$ has the property that T_n^f converges uniformly to f for <u>any</u> continuous function f defined on the closed and bounded interval $[a,b]$ of two real numbers a and b. The proof given here follows Bauer [1]. □

Unless stated differently we assume the following for all of this subchapter:
a and b are two real numbers such that $a < b$, and

$$T_n(\cdot) : \mathscr{C}([a,b],\mathbb{R}) \to \mathscr{C}([a,b],\mathbb{R}); \qquad f(\cdot) \mapsto T_n^f(\cdot) = T_n(f)(\cdot)$$

is a sequence of positive linear operators on $\mathscr{C}([a,b],\mathbb{R})$. This means in particular that for each continuous real–valued function $f(\cdot)$ on $[a,b]$ the image

$$T_n^f : x \mapsto T_n^f(x)$$

is itself a continuous, real–valued function on $[a,b]$.

Before we state and prove Korovkin's First Theorem, there are two technical issues which we will prove separately in form of two lemmata so that the proof of the theorem itself does not become to lengthy.

Lemma 16.1. *Let there be real numbers $a < b$, and let f be a continuous, real–valued function on $[a,b]$. Let $\varepsilon > 0$. Then there exists (a potentially very large) number $\alpha \in \mathbb{R}$ such that*

$$|f(x) - f(y)| \ < \ \varepsilon \ + \ \alpha(x-y)^2 \qquad \textit{for all } x,y \in [a,b]. \tag{16.5}$$

This number α can be chosen once and for all, independently of x and y.

PROOF: The interval $[a,b]$ is a closed and bounded subset of $\mathbb{R}$ and hence compact. This follows from the Heine–Borel Theorem (14.10 on p.447).

Since f is continuous it follows from cor.14.6 on p.449 that the set $f([a,b])$ possesses a min and a max, in particular that that this set is bounded: If $\gamma = \max(|a|,|b|)$ then

$$|f(x)| \ < \ \gamma \quad \text{for all } x \in [a,b]. \tag{16.6}$$

It further follows from cor.14.7 on p.450 that the continuous function f is uniformly continuous on $[a,b]$. Thus we can find for any $\varepsilon > 0$, no matter how small, some $\delta^\star > 0$ such that

$$|f(x) - f(y)| \ < \ \varepsilon \quad \text{for all } x,y \in [a,b] \text{ such that } |x-y| < \delta^\star.$$

Let $\delta := \delta^{\star 2}$. The uniform continuity characterization above then reads

$$|f(x) - f(y)| \ < \ \varepsilon \quad \text{for all } x,y \in [a,b] \text{ such that } (x-y)^2 < \delta. \tag{16.7}$$

Let γ be the constant from (16.6), i.e., $|f(z)| < \gamma$ for all $z \in [a,b]$. We will show that

$$\alpha \ := \ \frac{2\gamma}{\delta}$$

satisfies (16.5) by looking separately at the two cases $(x-y)^2 \leq \delta$ and $(x-y)^2 > \delta$.
Case 1: Assume that $(x-y)^2 \leq \delta$.
Observe that it is always true that

$$\varepsilon \ < \ \varepsilon \ + \ \alpha(x-y)^2$$

since $\alpha(x-y)^2 \geq 0$. The rest is easy. According to (16.7), we have

$$(x-y)^2 \leq \delta \ \Rightarrow\ |f(x)-f(y)| \ < \ \varepsilon \ \leq \ \varepsilon \ + \ \alpha(x-y)^2,$$

and **case 1** is proven.
Case 2: Assume that $(x-y)^2 > \delta$, i.e., $\delta < (x-y)^2$.
γ was chosen such that $|f(x)| < \gamma$ on $[a,b]$ (see (16.6)). Thus we have

$$\begin{aligned} |f(x)-f(y)| \ &\leq |f(x)| + |f(y)| \\ &< \gamma + \gamma \ = \ \alpha\delta \ < \ \alpha(x-y)^2 < \varepsilon \ + \ \alpha(x-y)^2, \end{aligned} \tag{16.8}$$

i.e., **case 2** also holds, and hence α satisfies (16.5). This concludes the proof of the lemma. ■

Lemma 16.2. *If the sequence of positive linear operators T_n on $\mathscr{C}([a,b],\mathbb{R})$ satisfies*

$$\textit{\textbf{i)}}\ T_n^1 \xrightarrow{uc} 1, \qquad \textit{\textbf{ii)}}\ T_n^{id} \xrightarrow{uc} id, \qquad \textit{\textbf{iii)}}\ T_n^{id^2} \xrightarrow{uc} id^2.$$

Then $\lim\limits_{n\to\infty} \|T_n^{id^2} - 2idT_n^{id} + id^2T_n^1\|_\infty \ = \ 0.$

PROOF: Let $A_n := T_n^{id^2} - 2idT_n^{id} + id^2T_n^1$. Then

$$\begin{aligned} A_n \ &= \ T_n^{id^2} - id^2T_n^1 + 2id^2T_n^1 - 2idT_n^{id} \\ &= \ (T_n^{id^2} - id^2T_n^1) \ + \ 2(id^2T_n^1 - idT_n^{id}) \ = \ B_n + 2C_n \end{aligned}$$

where we define $B_n := T_n^{id^2} - id^2T_n^1$ and $C_n := id^2T_n^1 - idT_n^{id}$. Since

$$0 \ \leq \ \|B_n + 2C_n\|_\infty \ \leq \ \|B_n\|_\infty \ + \ 2\|C_n\|_\infty$$

it suffices to prove that $\lim\limits_{n\to\infty} \|B_n\|_\infty \ = \ \lim\limits_{n\to\infty} \|C_n\|_\infty = 0$.
Proof that $\lim\limits_{n\to\infty} \|B_n\|_\infty \ = \ 0$:
We rewrite $B_n \ = \ (T_n^{id^2} - id^2) \ + \ id^2(1 - T_n^1)$. The continuous function $id^2 : x \mapsto x^2$ assumes a maximum on $[a,b]$ which we call γ. Thus $\|B_n\|_\infty \ \leq \ \|T_n^{id^2} - id^2)\|_\infty \ + \ \gamma\|1 - T_n^1\|_\infty$. Uniform convergence $T_n^1 \xrightarrow{uc} 1$ and $T_n^{id^2} \xrightarrow{uc} id^2$ yields

$$\lim_{n\to\infty} \|T_n^{id^2} - id^2)\|_\infty = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \gamma\|1 - T_n^1\|_\infty = \gamma \lim_{n\to\infty} \|1 - T_n^1\|_\infty = 0,$$

hence $\lim\limits_{n\to\infty} \|B_n\|_\infty = 0$
Proof that $\lim\limits_{n\to\infty} \|C_n\|_\infty \ = \ 0$:

$$\begin{aligned} C_n \ &= id^2T_n^1 - idT_n^{id} = \ (id^2T_n^1 - id^2 + id^2) - (idT_n^{id} + id^2 - id^2) \\ &= \ id^2(T_n^1 - 1) + id^2 \ - \ id(T_n^{id} - id) - id^2 \\ &= \ id^2(T_n^1 - 1) \ - \ id(T_n^{id} - id) \end{aligned}$$

hence

$$\|C_n\|_\infty \ \leq \ \|id^2(T_n^1 - 1)\|_\infty \ + \ \|id(T_n^{id} - id)\|_\infty.$$

But $|id(x)| = |x| \leq \gamma' := \max(|a|,|b|)$, and we saw that $|id^2(x)| = x^2 \leq \gamma$ for all $a \leq x \leq b$. Thus

$$\|C_n\|_\infty \;\leq\; \gamma\|(T_n^1 - 1)\|_\infty \;+\; \gamma'\|(T_n^{id} - id)\|_\infty.$$

It now follows from $T_n^1 \xrightarrow{uc} 1$ and $T_n^{id} \xrightarrow{uc} id$ that

$$\lim_{n\to\infty} \gamma\|T_n^1 - 1)\|_\infty = 0 = \gamma \lim_{n\to\infty} \|T_n^1 - 1)\|_\infty \quad = 0 \quad \text{and} \quad \lim_{n\to\infty} \gamma'\|T_n^{id} - id\|_\infty = \gamma' \lim_{n\to\infty} T_n^{id} - id\|_\infty = 0,$$

hence $\lim\limits_{n\to\infty} \|C_n\|_\infty = 0$.

We noted earlier that the convergence of $\|B_n\|_\infty$ and $\|C_n\|_\infty$ was sufficient to prove the lemma. ■

We have now assembled everything to formulate and prove Korovkin's First Theorem.

Theorem 16.1 (Korovkin's First Theorem)**.** *Assume we have uniform convergence* [184] $T_n^f(\cdot) \xrightarrow{uc} f(\cdot)$ *for the following three elements f of $\mathscr{C}([a,b],\mathbb{R})$:*

$$\begin{aligned} 1(\cdot) :& x \mapsto 1 && \textit{the constant function } 1,\\ id(\cdot) :& x \mapsto x && \textit{the identity on } [a,b],\\ id^2(\cdot) :& x \mapsto x^2. \end{aligned}$$

Then $T_n^f \xrightarrow{uc} f$ for all $f \in \mathscr{C}([a,b],\mathbb{R})$.

PROOF: Let $\varepsilon > 0$. According to lemma 16.1 on p.471 there exists $\alpha > 0$ such that

$$|f(x) - f(y)| \;<\; \varepsilon \;+\; \alpha(x-y)^2 \qquad \text{for all } x,y \in [a,b]. \tag{16.9}$$

We interpret the above as an inequality where x acts as a variable (a function argument) and y is "fixed but arbitrary". We thus obtain for each y the inequality of functions,

$$|f(\cdot) - f(y)| \;<\; \varepsilon \;+\; \alpha(id(\cdot) - y)^2 \qquad \text{for all } y \in [a,b]. \tag{16.10}$$

According to (16.3) on p.469, $|T_n^h(\cdot)|$ does not exceed $T_n|h|$ for any $h \in \mathscr{C}([a,b],\mathbb{R})$. We thus obtain for $h : x \mapsto f(x) - f(y)$

$$\begin{aligned} |T_n f(\cdot) - f(y) T_n 1(\cdot)| \; &= |T_n\big(f(\cdot) - f(y)1(\cdot)\big)|\\ &\leq T_n(|f - f(y)1|)\\ &= T_n(|f - f(y)|) \end{aligned} \tag{16.11}$$

which is, according to (16.9), smaller than

$$\begin{aligned} T_n(\varepsilon \;+\; \alpha(id - y)^2 \; &= T_n\varepsilon \;+\; \alpha T_n(id - y\cdot 1)^2\\ &= \varepsilon T_n(1) \;+\; \alpha\Big(T_n(id^2) - 2yT_n(id) + y^2 T_n(1)\Big). \end{aligned} \tag{16.12}$$

We combine (16.11) and (16.12) and obtain

$$|T_n f - f(y) T_n 1| \;\leq\; \varepsilon T_n 1 \;+\; \alpha\Big(T_n id^2 - 2yT_n id + y^2 T_n 1\Big).$$

[184] see (13.35) on p.415

For the next step we recall that each of the expressions $T_nf, -T_n1, T_nid^2, T_nid$ is a function on the interval $[a,b]$. We evaluate them at the argument y and obtain

$$(16.13)\qquad |T_nf(y) - f(y)T_n1(y)| \ \le\ \varepsilon T_n1(y) \ +\ \alpha\Big(T_nid^2(y) - 2yT_nid(y) + y^2T_n1(y)\Big).$$

This last inequality we can in turn interpret as one for the functions

$$\begin{aligned} y \mapsto& |T_nf(y) - f(y)T_n1(y)|\\ y \mapsto& \varepsilon T_n1(y) \ +\ \alpha\Big(T_nid^2(y) - 2id(y)T_nid(y) + id(y)^2T_n1(y)\Big). \end{aligned}$$

It then follows from (16.13) that

$$(16.14)\qquad |T_nf - fT_n1| \ \le\ \varepsilon T_n1 \ +\ \alpha\Big(T_nid^2 - 2idT_nid + id^2T_n1\Big).$$

The function f is continuous and hence bounded on the closed interval $[a,b]$. Thus there exists $\gamma > 0$ such that $|f(x)| < \gamma$ for all $x \in [a,b]$. (See Corollary 14.5 on p.449 of this document>)
We use the triangle inequality twice in the following:

$$(16.15)\qquad \begin{aligned} |T_nf - f| &= |T_nf - fT_n1 + fT_n1 - f|\\ &= |T_nf - fT_n1 + f(T_n1 - 1)|\\ &\le |T_nf - fT_n1| \ +\ |f(T_n1 - 1)|\\ &\le |T_nf - fT_n1| \ +\ \gamma|(T_n1 - 1)|. \end{aligned}$$

We combine (16.15) and (16.14) and obtain

$$\begin{aligned} |T_nf - f| &\le |T_nf - fT_n1| \ +\ \gamma|(T_n1 - 1)|\\ &\le \varepsilon T_n1 \ +\ \alpha\Big(T_nid^2 - 2idT_nid + id^2T_n1\Big) \ +\ \gamma|(T_n1 - 1)|\\ &\le \varepsilon\|T_n1\|_\infty \ +\ \alpha\|T_nid^2 - 2idT_nid + id^2T_n1\|_\infty \ +\ \gamma\|(T_n1 - 1)\|_\infty. \end{aligned}$$

This inequality is true for each argument $a \le x \le b$, thus

$$\sup_{a\le x\le b} |T_nf(x) - f(x)| \le \varepsilon\|T_n1\|_\infty \ +\ \alpha\|T_nid^2 - 2idT_nid + id^2T_n1\|_\infty \ +\ \gamma\|(T_n1 - 1)\|_\infty,$$

i.e.,

$$(16.16)\qquad \|T_nf - f\|_\infty \le \varepsilon\|T_n1\|_\infty \ +\ \alpha\|T_nid^2 - 2idT_nid + id^2T_n1\|_\infty \ +\ \gamma\|(T_n1 - 1)\|_\infty.$$

It follows from $T_n1 \xrightarrow{uc} 1$ that there exists $N_1 \in \mathbb{N}$ such that $\|T_n1 - 1\|_\infty \le 1$ for all $n \ge N_1$, thus

$$(16.17)\qquad \|\varepsilon T_n1\|_\infty = \|\varepsilon \ +\ \varepsilon(T_n1 - 1)\|_\infty \le \|\varepsilon\|_\infty \ +\ \varepsilon\|(T_n1 - 1)\|_\infty \le 2\varepsilon \ \text{ for all } n \ge N_1.$$

It also follows from $T_n1 \xrightarrow{uc} 1$ that $\lim\limits_{n\to\infty} \|(T_n1 - 1)\|_\infty = 0$, hence

$$\lim_{n\to\infty} \gamma|(T_n1 - 1)| \le \gamma \lim_{n\to\infty} \|(T_n1 - 1)\|_\infty = 0.$$

Thus there exists $N_2 \in \mathbb{N}$ such that

$$\gamma|(T_n 1 - 1)| \ \leq \ \varepsilon \ \text{ for all } n \geq N_2. \tag{16.18}$$

It moreover follows from lemma 16.2 above that

$$\lim_{n\to\infty} \alpha\|T_n^{id^2} - 2idT_n^{id} + id^2 T_n^1\|_\infty = \ \alpha \lim_{n\to\infty} \|T_n^{id^2} - 2idT_n^{id} + id^2 T_n^1\|_\infty \ = \ 0.$$

Thus there exists $N_3 \in \mathbb{N}$ such that

$$\alpha\|T_n^{id^2} - 2idT_n^{id} + id^2 T_n^1\|_\infty \ \leq \ \varepsilon \ \text{ for all } n \geq N_3. \tag{16.19}$$

Let $N := \max(N_1, N_2, N_3)$. It follows from (16.16), (16.17), (16.18) and (16.19) that

$$\|T_n f - f\|_\infty \ \leq \ 4\varepsilon \ \text{ for all } n \geq N_3.$$

Of course the choice of N will depend on the choice of ε, but the fact of importance is that such $N = N(\varepsilon)$ can be found for any $\varepsilon > 0$, no matter how small. It follows that

$$\lim_{n\to\infty} \|T_n f - f\|_\infty = 0, \qquad \text{i.e.,} \quad T_n^f \xrightarrow{uc} f.$$

Since $f \in \mathscr{C}([a,b],\mathbb{R})$ was arbitrarily chosen we are finished with the proof of Korovkin's First Theorem. ■

16.3 The Weierstrass Approximation Theorem

Introduction 16.4. We have seen earlier that the mappings

$$B_n : \mathscr{C}([0,1],\mathbb{R}) \to \mathscr{C}([0,1],\mathbb{R}); \qquad f(\cdot) \mapsto B_n^f(\cdot)$$

which assign to a function $f : [0,1] \to \mathbb{R}$ it's n^{th} Bernstein polynomial

$$x \ \mapsto \ B_n^f(x) \ = \ \sum_{k=0}^{n} \binom{n}{k} f(\frac{k}{n}) \ id(\cdot)^k \ (1 - id(\cdot))^{n-k}$$

have the following properties:

(a) They are linear positive operators on the space $\mathscr{C}([0,1],\mathbb{R})$ of all continuous real–valued functions on the unit interval $[0,1]$. (see prop.16.2 on p.469).

(b) We have uniform convergence $B_n^f(\cdot) \overset{uc}{\to} f(\cdot)$ for the three functions

$$1 : x \mapsto 1; \qquad id(\cdot) : x \mapsto x; \qquad id^2(\cdot) : x \mapsto x^2; \qquad (0 \leq x \leq 1).$$

(see proposition (13.8) on p.419). We will obtain as an easy consequence of Korovkin's First Theorem that any continuous real–valued function defined on the unit interval is the uniform limit of a sequence of polynomials, and we then generalize this result to any closed and bounded interval $[a,b]$. This result was first proven (by entirely different means by Weierstrass in the 1880s. □

Proposition 16.3 (Weierstrass Approximation Theorem on $[0,1]$). *Any continuous real–valued function on the unit interval $[0,1]$ can be uniformly approximated by a sequence of polynomials.* ■

PROOF: We have seen in cor.16.1 on p.470 that the Bernstein polynomial assignments $B_n(\cdot)$ form a sequence of positive linear operators on $\mathscr{C}([0,1],\mathbb{R})$. We also have seen that if f is one of the three continuous functions $1 : x \mapsto 1$, $id : x \mapsto x$, $id^2 : x \mapsto x^2$ then $B_n^f(\cdot) \overset{uc}{\to} f(\cdot)$ It follows from Korovkin's First Theorem that this uniform convergence extends to all continuous functions on the unit interval. This proves the theorem since all functions $x \mapsto B_n^f(x)$ are Bernstein polynomials and hence polynomials. ■

The trick that we will use to generalize this last result from $[0,1]$ to $[a,b]$ consists of using the fact that those intervals can be bijected by functions of the form $\psi(x) = mx + b$, and that uniform convergence $f_n \xrightarrow{uc} f$ implies that of $f_n \circ \varphi \xrightarrow{uc} f \circ \varphi$ and vice versa.

Lemma 16.3 (B/G prop.2.34)**.** *Let $n \in [0,\infty[_{\mathbb{Z}}$, $\alpha_j, m, b \in \mathbb{R}$, $\alpha_n \neq 0$.*

Let $p : \mathbb{R} \to \mathbb{R}$ be a polynomial $p(x) = \sum\limits_{j=0}^{n} \alpha_j x^j$, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be defined as $\varphi(x) = mx + b$.

Then $\quad p \circ \varphi : \mathbb{R} \to \mathbb{R}; \quad x \;\mapsto\; \sum\limits_{j=0}^{n} \alpha_j (mx+b)^j$ is a polynomial.

The proof is left as exercise 16.1 (see p.477). ■

Proposition 16.4. *Let $A \subseteq \mathbb{R}$, $\varphi : \mathbb{R} \to \mathbb{R}$ defined as $\varphi(x) = mx + b$ $(m, b \in \mathbb{R})$, and let $f_n, f \in \mathscr{C}(\varphi(A),\mathbb{R})$ $(n \in \mathbb{N})$, i.e. f_n and f are continuous functions defined on $\varphi(A) = \{\varphi(x) : x \in A\}$. Assume further that $f_n \xrightarrow{uc} f$ on $\varphi(A)$. Then $f_n \circ \varphi \xrightarrow{uc} f \circ \varphi$ on A.*

PROOF:
For any $a \in A$ we have $\varphi(a) \in \varphi(A)$, thus

$$|f_n \circ \varphi(a) - f \circ \varphi(a)| \;\leq\; \sup_{y \in \varphi(A)} |f_n(y) - f(y)| \;=\; \|f_n - f\|_\infty.$$

But $a \in A$ was arbitrary, thus $\|f_n \circ \varphi - f \circ \varphi\|_\infty \leq \|f_n - f\|_\infty$ It follows from $f_n \xrightarrow{uc} f$ on $\varphi(A)$ that $\lim\limits_{n\to\infty} \|f_n - f\|_\infty = 0$, hence $\lim\limits_{n\to\infty} \|f_n \circ \varphi - f \circ \varphi\|_\infty = 0$, hence $f_n \circ \varphi \xrightarrow{uc} f \circ \varphi$ on A. ■

Corollary 16.2. *Let $a, b \in \mathbb{R}$ such that $a < b$ and $\varphi : [a,b] \to [0,1]$ defined as $\varphi(x) := \frac{x-a}{b-a}$.*
***(a)**. φ is a bijection $[a,b] \xrightarrow{\sim} [0,1]$.*
***(b)**. Let $h_n, h \in \mathscr{C}([0,1],\mathbb{R})$ $(n \in \mathbb{N})$ such that $h_n \xrightarrow{uc} h$ on $[0,1]$. Then $h_n \circ \varphi \xrightarrow{uc} h \circ \varphi$ on $[a,b]$.*

PROOF of **(a)**: It is easily verified that $\psi : [0,1] \to [a,b]$ defined as $\psi(t) := a + t(b-a)$ defines the inverse of φ.
PROOF of **(b)**: This follows from prop.16.4 on p.476 setting $A := [a,b]$ since then $\varphi(x) = \frac{1}{b-a} \cdot x - \frac{a}{b-a}$ is of the form $mx + b$, and since it follows from **(a)** that $\varphi(A) = \varphi([a,b]) = [0,1]$. ■
The Weierstrass approximation theorem for continuous functions on the unit interval (prop.16.3 on p.475) is now easily generalized to continuous functions on an arbitrary closed and bounded interval.

Theorem 16.2 (Weierstrass Approximation Theorem)**.** *Let $a, b \in \mathbb{R}$ such that $a < b$. Then any continuous real–valued function on $[a,b]$ can be uniformly approximated by a sequence of polynomials.* ■

PROOF: Let $f \in \mathscr{C}([a,b], \mathbb{R}$, and let $\varphi : [a,b] \xrightarrow{\sim} [0,1]$ be defined as $\varphi(x) := \frac{x-a}{b-a}$. We recall from cor.16.2 on p.476 that φ is bijective. Let $h := f \circ \varphi^{-1}$. Then h is continuous on the unit interval and thus, according to prop.16.3, f is the uniform limit of a sequence of polynomials h_n (we might choose, e.g., the Bernstein polynomials $h_n := B_n^{f \circ \varphi^{-1}}$). It follows from prop.16.4 on p.476 that $h_n \circ \varphi \xrightarrow{uc} h \circ \varphi$ on $\varphi^{-1}[0,1]$. But $h \circ \varphi = f \circ \varphi^{-1} \circ \varphi = f$ and $\varphi^{-1}[0,1] = [a,b]$, thus $h_n \circ \varphi \xrightarrow{uc} f$ on $[a,b]$.

We finally note that it follows from lemma 16.3 on p.476 that the functions $p_n := h_n \circ \varphi$ are polynomials since the h_n are polynomials. Since $p_n \xrightarrow{uc} f$ on $[a,b]$ we have proven the theorem. ■

16.4 Exercises for Ch.16

Exercise 16.1. Prove lemma 16.3 on p.476 of this document: Let $n \in [0, \infty[_{\mathbb{Z}}$, $\alpha_j, m, b \in \mathbb{R}$, $\alpha_n \neq 0$.

Let $p : \mathbb{R} \to \mathbb{R}$ be a polynomial $p(x) = \sum_{j=0}^{n} \alpha_j x^j$, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be defined as $\varphi(x) = mx + b$.

Then $p \circ \varphi : \mathbb{R} \to \mathbb{R}; \quad x \mapsto \sum_{j=0}^{n} \alpha_j (mx+b)^j$ is a polynomial.

Hint: Apply the binomial theorem to $(mx+b)^j$. □

16.5 Blank Page after Ch.16

This page is intentionally left blank.

17 Algebraic Structures ⋆

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 [2] B/G forms a group.

Note that this chapter is starred and hence optional.

17.1 More on Groups (⋆)

This chapter adds to the material of ch.3.1 (Semigroups and Groups) on p.50.

Example 17.1. 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 3.4 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 $(\mathbb{Q},\cdot)$. However the set $\mathbb{C}^\star$ of all non-zero complex numbers is an abelian group. □

We now turn our attention to functions which map from a group to another group in such a way that they are, in a sense, compatible with the binary operations on their domain and codomain.

Example 17.2. Let $(G,\diamond)$ and $(H,\bullet)$ be defined as follows:

$$G := \{g\in\mathbb{R} : g = e^x \text{ for some } x\in\mathbb{R}\},\quad e^x\diamond e^y := e^x\cdot e^y = e^x + y, \tag{17.1}$$

$$H := \{h\in\mathbb{R} : h = \ln(x) \text{ for some } u \in]0,\infty\},\quad \ln u\bullet\ln v := \ln u+\ln v = \ln(xy). \tag{17.2}$$

(a) Both $(G,\diamond)$ 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:

$$\varphi : (G,\diamond)\to(H,\bullet),\quad \varphi(g) := \ln g, \tag{17.3}$$

$$\psi : (H,\bullet)\to(G,\diamond),\quad \psi(h) := e^h. \tag{17.4}$$

Then φ and ψ satisfy the following:

$$\varphi(g_1\diamond g_2) = \varphi(g_1)\bullet\varphi(g_2),\quad \varphi(1) = 0,\quad \varphi(g^{-1}) = \varphi(g)^{-1}, \tag{17.5}$$

$$\psi(h_1\bullet h_2) = \psi(h_1)\diamond\psi(h_2),\quad \psi(0) = 1,\quad \psi(h^{-1}) = \psi(h)^{-1}. \tag{17.6}$$

Further, the functions φ and ψ are inverse to each other, i.e.,

$$\psi\big(\varphi(g)\big) = g \quad \text{and} \quad \varphi\big(\psi(h)\big) = h \tag{17.7}$$

for all $g \in G$ and $h \in H$. □

If you talk about φ and ψ as "the functions" and $\diamond$ and $\bullet$ as "the operations" you might state the results (17.5) and (17.6) as follows:

The functions are structure compatible with the group operations on their domains and codomains: See also Example 3.7 on p.58.

17.2 More on Commutative Rings and Integral Domains

This chapter adds to the material of ch.3.2 (Commutative Rings and Integral Domains) on p.60.

The following is an example of a commutative ring with unit which contains zero divisor. This is not a document on algebra and we only give an outline. See, e.g., [2] Beck/Geoghegan ch.6 for details.

Theorem 17.1 (Division Algorithm for Integers)**.** *See [2] Beck/Geoghegan, thm.6.13.*
*Let $n \in \mathbb{N}$ and $m \in \mathbb{Z}$. There exist two integers q (the **quotient**) and r (the **remainder**) such that*

$$m = qn + r \tag{17.8}$$

$$0 \leq r < n \tag{17.9}$$

q and r are uniquely determined by m and n.

PROOF: Will not be given. ■

Example 17.3. The following examples illustrate the division algorithm.

- **(a)** $m = 63, n = 10$: Then $q = 6$ and $r = 3$ because $63 = 6 \cdot 10 + 3$ and $0 \leq 3 < 10$.
- **(b)** $m = 63, n = 13$: Then $q = 4$ and $r = 11$ because $63 = 4 \cdot 13 + 11$ and $0 \leq 11 < 13$.
- **(c)** $m = -63, n = 13$: Then $q = -5$ and $r = 2$ because $-63 = (-5) \cdot 13 + 2$ and $0 \leq 2 < 13$.
- **(d)** $m = 63, n = 7$: Then $q = 9$ and $r = 0$ because $63 = 9 \cdot 7 + 0$ and $0 \leq 0 < 7$. □

18 Construction of the Number Systems $\boxed{\star}$

18.1 The Peano Axioms

Note that this chapter is starred, hence optional.

Definition 18.1 (Set of nonnegative integers)**.** We define the set $\mathbb{N}_0$ (the nonnegative 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 σ),
  Ax.2.3 Induction axiom: Let $U \subseteq \sigma(\mathbb{N}_0)$ such that **(a)** $0 \in U$, **(b)** If $n \in U$ then $\sigma(n) \in U$.
  It then follows that $U = \mathbb{N}_0$.

We define $\mathbb{N} := \mathbb{N}_0 \setminus \{0\}$. □

Definition 18.2 (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 18.1. *f^n is defined for all $n \in \mathbb{N}_0$.*

PROOF: Let $U := \{k \in \mathbb{N}_0 : f^k \text{ is defined }\}$. 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 18.1 ($\sigma(\cdot)$ as successor function)**.** Of course the meaning of $\sigma(n)$ will be that of $n+1$:

$$0 \overset{\sigma}{\mapsto} 1 \overset{\sigma}{\mapsto} 2 \overset{\sigma}{\mapsto} 3 \overset{\sigma}{\mapsto} \dots \quad \square$$

Definition 18.3 (Addition and multiplication on $\mathbb{N}_0$)**.** Let $m, n \in \mathbb{N}_0$. Let

$$m + n := \sigma^n(m), \tag{18.1}$$

$$m \cdot n := (\sigma^m)^n(0). \tag{18.2}$$

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.18.1 the meaning of f^n, i.e., $(\sigma^m)^n$. □

Proposition 18.2. *Addition and multiplication satisfy all rules of arithmetic we learned in high school such as*

$$m + n = n + m \qquad \textit{commutativity of addition} \tag{18.3}$$

$$k + (m + n) = (k + m) + n \qquad \textit{associativity of addition} \tag{18.4}$$

$$m \cdot n = n \cdot m \qquad \textit{commutativity of multiplication} \tag{18.5}$$

$$k \cdot (m \cdot n) = (k \cdot m) \cdot n \qquad \textit{associativity of multiplication} \tag{18.6}$$

$$k \cdot (m + n) = k \cdot m + k \cdot n \qquad \textit{distributivity of addition} \tag{18.7}$$

$$n \cdot 1 = 1 \cdot n = n \qquad \textit{neutral element for multiplication} \tag{18.8}$$

(18.9)

Here 1 is defined as $1 = \sigma(0)$.

PROOF: Drudge work. ■

Definition 18.4 (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 $m \leq n$ 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 $m \geq n$ if $n \leq m$. □

Proposition 18.3. *"$<$" and "$\leq$" 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. ■

18.2 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 18.5 (Integers as equivalence classes)**.** We define the following equivalence relation $(m_1, n_1) \sim (m_2, n_2)$ on the cartesian product $\mathbb{N}_0 \times \mathbb{N}_0$:

$$(m_1, n_1) \sim (m_2, n_2) \;\Leftrightarrow\; m_1 + n_2 = n_1 + m_2 \tag{18.10}$$

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 (18.10).

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 zeros 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 18.4 (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 18.2. (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 (18.10) 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 18.6 (Addition, multiplication and subtraction on $\mathbb{Z}$)**.** Let $[(m_1, n_1)]$ and $[(m_2, n_2)] \in \mathbb{Z}$. We define

$$-[(m_1, n_1)] := [n_1, m_1], \tag{18.11}$$

$$[(m_1, n_1)] + [(m_2, n_2)] := [(m_1 + m_2, n_1 + n_2)] \tag{18.12}$$

$$[(m_1, n_1)] \cdot [(m_2, n_2)] := [(m_1 m_2 + n_1 n_2, m_1 n_2 + n_1 m_2)] \tag{18.13}$$

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 $[(m_1, n_1)] \leq [(m_2, n_2)]$ 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 $[(m_1, n_1)] \geq [(m_2, n_2)]$ if $[(m_2, n_2)] \leq [(m_1, n_1)]$ and we then say that $[(m_1, n_1)]$ is greater than or equal to $[(m_2, n_2)]$.

We write $\mathbb{Z}_{\geq 0}$ for the set of all integers z such that $z \geq 0$ and $\mathbb{Z}_{\neq 0}$ for the set of all integers z such that $z \neq 0$. You should convince yourself that $\mathbb{Z}_{\geq 0} = \mathbb{N}_0$. □

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 18.5. *Let $m, n \in \mathbb{N}_0$. Then*

$$[(m, n)] \; + \; [(0, 0)] \; = \; [(0, 0)] \; + \; [(m, n)] \; = \; [(m, n)], \tag{18.14}$$

$$\big(- [(m, n)]\big) \; + \; [(m, n)] \; = \; [(m, n)] \; + \; \big(- [(m, n)]\big) \; = \; [0, 0] \tag{18.15}$$

$$[(m, n)] \; \cdot \; [(1, 0)] \; = \; [(1, 0)] \; \cdot \; [(m, n)] \; = \; [(m, n)], \tag{18.16}$$

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 18.3. 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 Definition 18.6:

Addition is defined such that $(m_1 - n_1) + (m_2 - n_2) = (m_1 + m_2) - (n_1 + n_2)$
and multiplication: $(m_1 - n_1) \cdot (m_2 - n_2) = \big(m_1 m_2 + (-n_1)(-n_2)\big) \; - \; \big(m_1 n_2 + n_1 m_2\big)$. □

18.3 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 18.7 (Fractions as equivalence classes)**.** We define the following equivalence relation $(p,q) \sim (r,s)$ on the cartesian product $\mathbb{Z} \times \mathbb{Z}_{\neq 0}$:

$$(p,q) \sim (r,s) \;\Leftrightarrow\; p \cdot s \;=\; q \cdot r \tag{18.17}$$

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 (18.17).

We "embed" $\mathbb{Z}$ into $\mathbb{Q}$ with the injective function $e : \mathbb{Z} \to \mathbb{Q}$ defined as $e(z) := [(z,1)]$. □

Remark 18.4. (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 (18.17) 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 18.8 (Addition, multiplication, subtraction and division in $\mathbb{Q}$)**.** Let $[(p_1,q_1)]$ and $[(p_2,q_2)] \in \mathbb{Q}$. We define

$$-[(p_1,q_1)] \;:= [(-p_1,q_1)], \tag{18.18}$$

$$[(p_1,q_1)] + [(p_2,q_2)] \;:= [(p_1q_2 + q_1p_2, n_1n_2)] \tag{18.19}$$

$$[(p_1,q_1)] - [(p_2,q_2)] \;:= [(p_1,q_1)] + \big(-[(p_2,q_2)]\big) \tag{18.20}$$

$$[(p_1,q_1)] \cdot [(p_2,q_2)] \;:= [(p_1p_2, q_1q_2)] \tag{18.21}$$

$$[(p_1,q_1)]^{-1} \;:= [(1,1)]/[(p_1,q_1)] \;:=\; [(q_1,p_1)] \text{ (if } p_1 \neq 0), \tag{18.22}$$

$$[(p_1,q_1)]/[(p_2,q_2)] \;:= [(p_1q_2, q_1p_2)] \;=\; [(p_1,q_1)] \cdot [(p_2,q_2)]^{-1} \text{ (if } p_2 \neq 0) \quad \square \tag{18.23}$$

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. [185]

Further we have

Proposition 18.6 (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. ■

[185] This was shown for multiplication $[(p_1,q_1)] \cdot [(p_2,q_2)] \;=\; [(p_1p_2, q_1q_2)]$ in exercise 5.16 on p.165.

18.4 Constructing the Real Numbers via Dedekind Cuts

The material presented here, including the notation, follows [14] 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

$$\Phi : x \mapsto \Phi(x) := \ \{q \in \mathbb{Q} : q < x\} \tag{18.24}$$

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 18.9 (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. □

Theorem 18.1. *(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 $q \leq p$. 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 Definition 18.9.b that $q \in \alpha$, contrary to our assumption. ■

Theorem 18.2. *(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 Definition 18.9.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 Definition 18.9.b: Let $q < p$ and $p \in r^{\star}$. Then also $q \in r^{\star} = \{p' \in \mathbb{Q} : p' < r\}$.

PROOF of Definition 18.9.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 18.10 (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 18.5. If we define intervals in $\mathbb{Q}$ in the usual way for $p, q \in \mathbb{Q}$:

$$]p, q[:= \{r \in \mathbb{Q} : p < r < q\}, \quad [p, q] := \{r \in \mathbb{Q} : p \leq r \leq q\}, \text{ etc.}$$

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 18.11 (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 $\alpha \leq \beta$ if $\alpha < \beta$ or $\alpha = \beta$, i.e., $\alpha \subseteq \beta$. □

Proposition 18.7 (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 $a \leq q$ then $a \in \alpha$ (we applied Definition 18.9.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 18.3 (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

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

PROOF: Not given here. ■

Theorem 18.4 (Multiplication of two cuts)**.** *Let $\alpha \geq 0^\star, \beta \geq 0^\star$ be two nonnegative cuts. Let*

$$\alpha \cdot \beta := \begin{cases} \{q \in \mathbb{Q} : q < 0\} \cup \{ab : a \in \alpha, b \in \beta\} & \text{if } \alpha \geq 0^\star, \beta \geq 0^\star, \\ -|\alpha| \cdot |\beta| & \text{if } \alpha < 0^\star, \beta \geq 0^\star \text{ or } \alpha \geq 0^\star, \beta < 0^\star, \\ |\alpha| \cdot |\beta| & \text{if } \alpha < 0^\star, \beta < 0^\star. \end{cases}$$

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 18.5 (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.18.7 on p.486 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 18.6. *(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 18.7. *(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 $q \geq p$ and $q \in \alpha$ we obtain $p \in \alpha$ from Definition 18.9.b. ■

Theorem 18.8 (Dedekind's Theorem)**.** *(Rudin thm.1.32) Let $\mathbb{R} = A \biguplus B$ a partitioning of $\mathbb{R}$ such that*

(a) *$A \neq \emptyset$ and $B \neq \emptyset$*
(b) *$\alpha \in A, \beta \in B \;\Rightarrow\; \alpha < \beta$ (i.e., $\alpha \subsetneq \beta$).*

Then there exists a unique cut $\gamma \in \mathbb{R}$ such that if $\alpha \in A$ then $\alpha \leq \gamma$ and if $\beta \in B$ then $\gamma \leq \beta$.

PROOF: We first prove uniqueness and afterwards the existence of γ.

PROOF of uniqueness: Assume there is $\gamma'' \in \mathbb{R}$ which satisfies $\alpha \leq \gamma''$ for all $\alpha \in A$ and $\gamma'' \leq \beta$ for all $\beta \in B$.

We may assume that $\gamma < \gamma''$. It follows from thm.18.6 on p.487 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 $\gamma \;:=\; \bigcup \big[\alpha : \alpha \in A\big]$.

Step 1: We now show that γ is a cut.

We first show that Definition 18.9.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 $\alpha \subseteq \gamma = \bigcup \big[\alpha : \alpha \in A\big]$ 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 Definition 18.9.a is satisfied.

Next we show the validity of Definition 18.9.b. Let $p \in \gamma$, i.e., $p \in \alpha_0$ for some $\alpha_0 \in A$. Let $q < p$. Then $q \in \alpha_0 \subseteq \bigcup \big[\alpha : \alpha \in A\big]$, i.e., $q \in \gamma$. We conclude that Definition 18.9.b is satisfied.

Now we show the validity of Definition 18.9.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 Definition 18.9.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 $\alpha \leq \gamma \leq \beta$ for all $\alpha \in A$ and $\beta \in B$. It is trivial that $\alpha \leq \gamma$ for all $\alpha \in A$ because $\gamma := \bigcup \big[\alpha : \alpha \in A\big]$.

To show that $\gamma \leq \beta$ for all $\beta \in B$ we prove that the opposite statement that

$$\gamma > \beta, \quad \text{i.e.,} \quad \gamma \setminus \beta \neq \emptyset \quad \text{for some cut } \beta \in B \tag{18.25}$$

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 18.1. *Let $\mathbb{R} = A \uplus B$ be a partitioning of $\mathbb{R}$ such that*

(a) *$A \neq \emptyset$ and $B \neq \emptyset$*
(b) *$\alpha \in A, \beta \in B \Rightarrow \alpha < \beta$ (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.18.8 there exists $\gamma \in \mathbb{R}$ such that if $\alpha \in A$ then $\alpha \leq \gamma$ and if $\beta \in B$ then $\gamma \leq \beta$. 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 18.9 (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 $b \geq x$ for all $x \in E$. Then B is not empty by assumption. Let $A := B^{\complement} = \{\alpha \in \mathbb{R} : \alpha < x$ i.e., $\alpha \subsetneq x$ 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 $x \leq \alpha$ 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 $x \leq \beta$ 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.18.8). Hence there exists $\gamma \in \mathbb{R}$ such that $\alpha \leq \gamma \leq \beta$ for all $\alpha \in A$ and $\beta \in B$.

We now show 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.18.6 on p. 487 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 $\gamma \geq a$ for all $a \in A$.

It follows that $\gamma \notin A$, i.e., $\gamma \in B$ and we conclude from cor.18.1 that $\gamma = \min(B)$, i.e., $\gamma = \sup(E)$. ■

18.5 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 $i, j, k, m, n \in \mathbb{N}$, $\varepsilon, p, q, r, s, p_n, p_{i,j}, \cdots \in \mathbb{Q}$, $x, y, z, x_n, x_{i,j}, \cdots \in \mathbb{R}$.

- **(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 $n \geq N$ 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 $i, j \geq N$ 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., $\lim\limits_n p_n \neq 0$), 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 $q_n - p_n \geq \varepsilon\ \forall\ n \geq N$.
- **g2.** Define $[(p_n)_n] \leq [(q_n)_n]$ iff $\forall\ \varepsilon > 0$ *exists* $N \in \mathbb{N}$ such that $q_n - p_n \geq -\varepsilon\ \forall\ n \geq N$.
- **g3.** show that $[(p_n)_n] < [(q_n)_n]$ iff $[(p_n)_n] \leq [(q_n)_n]$ 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 $x_n \leq \bar{q}_n \leq x_n + 1/n$. Now show that **(1)** $(q_n)_n$ is Cauchy and then **(2)** $\lim\limits_n x_n = [(q_n)_n]$.
- **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 $j \leq n\}$. 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) $\leq 1/n$, $[(u_n)_n]$ 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 [11] Haaser/Sullivan: Real Analysis.

19 Measure Theory $\boxed{\star}$

Note that this entire section is starred, hence optional.

Introduction:

The following are the best known examples of measures ($a_j, b_j \in \mathbb{R}$):

$$\begin{aligned} &\text{Length}: \quad \lambda^1([a_1,b_1]) \;:=\; b_1 - a_1, \\ &\text{Area}: \quad \lambda^2([a_1,b_1]\times[a_2,b_2]) \;:=\; (b_1-a_1)(b_2-a_2), \\ &\text{Volume}: \quad \lambda^3([a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]) \;:=\; (b_1-a_1)(b_2-a_2)(b_3-a_3). \end{aligned}$$

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.

19.1 Basic Definitions

Definition 19.1 (Extended real–valued functions)**.**

$$\overline{\mathbb{R}}_+ \;:=\; \mathbb{R}_+ \,\cup\, \{+\infty\} \;=\; \{x \in \mathbb{R} : x \geq 0\} \,\cup\, \{+\infty\}$$

is the set of all nonnegative real numbers augmented by the elements ∞ and $-\infty$.
A function $F : X \to Y$ whose codomain Y is a subset of

$$\overline{\mathbb{R}} \;:=\; \mathbb{R} \,\cup\, \{\infty\} \,\cup\, \{-\infty\}$$

is called an **extended real–valued 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:

$$0 \cdot \pm\infty \;=\; \pm\infty \cdot 0 \;=\; 0. \;\square \tag{19.1}$$

This convention is very convenient, but it comes at a price: it is no longer true that all sequences $(a_n)_n$ and $(b_n)_n$ of real numbers that have limits $a = \lim\limits_{n\to\infty} a_n$, $b = \lim\limits_{n\to\infty} b_n$, satisfy $\lim\limits_{n\to\infty} a_n b_n = ab$.
Counterexample: $a_n = n, b_n = \frac{1}{n}$.

The definition of a σ–algebra was given previously in Definition 8.4 (Rings, algebras, and σ–Algebras of Sets) on p.235. We repeat it here in equivalent terms for your convenience.

Definition 19.2 (σ–algebras)**.** Let Ω be a nonempty 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:

$$\emptyset \in \mathfrak{F}, \tag{19.2a}$$

$$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F} \tag{19.2b}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F} \tag{19.2c}$$

- The pair $(\Omega, \mathfrak{F})$ is called a **measurable space**.
- The elements of $\mathfrak{F}$ (they are set!) are called **measurable sets**. □

Remark 19.1. If$\mathfrak{F}$ is a σ–algebra then

$$\emptyset \in \mathfrak{F} \quad \text{and} \quad \Omega \in \mathfrak{F} \tag{19.3a}$$

$$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F} \tag{19.3b}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \text{and} \quad \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F} \tag{19.3c}$$

The last assertion is a consequence of De Morgan's laws (Theorem 8.1 on p.234).

If $\mathfrak{F}$ is a σ–algebra then If countably many (i.e., a finite or infinite sequence of) operations are performed that involve
• unions, • intersections, • complements, • set differences, • symmetric differences
of elements of $\mathfrak{F}$ then the resulting set also belongs to $\mathfrak{F}$. □

Proposition 19.1 (Minimal sigma–algebras)**.** *Let Ω be a nonempty set.*

***A**: The intersection of arbitrarily many σ–algebras is a σ–algebra.*

***B**: Let $\mathfrak{E}$ be a set which contains subsets of Ω. It is not assumed that $\mathfrak{E}$ is a σ–algebra. Then there exists a σ–algebra which contains $\mathfrak{E}$ and is minimal in the sense that it is contained in any other σ–algebra that also contains $\mathfrak{E}$. We name this σ–algebra $\sigma(\mathfrak{E})$ because it clearly is uniquely determined by $\mathfrak{E}$. It is constructed as follows:*

$$\sigma(\mathfrak{E}) = \bigcap \{\mathfrak{F} : \mathfrak{F} \supseteq \mathfrak{E} \textit{ and } \mathfrak{F} \textit{ is a } \sigma\textit{–algebra for } \Omega\}. \tag{19.4}$$

PROOF of ***A***:
We must prove (19.2a), (19.2b) and (19.2c). Let $(\mathfrak{F}_\alpha)_\alpha$ be an arbitrary family of σ–algebras for Ω. Let

$$\mathfrak{F} := \bigcap_\alpha \mathfrak{F}_\alpha.$$

$\emptyset$ and Ω belong to each σ–algebra according to (19.2a). It follows that they both belong to the intersection $\bigcap_\alpha \mathfrak{F}_\alpha$, i.e., $\mathfrak{F}$ satisfies (19.2a). Let $A \in \mathfrak{F}$. Then $A \in \mathfrak{F}_\alpha$ for each α. $\complement A$ belongs to each σ–algebra according to (19.2b). It follows that $\complement A \in \bigcap_\alpha \mathfrak{F}_\alpha$, i.e., $\mathfrak{F}$ satisfies (19.2b). Finally, let $A_n \in \mathfrak{F}$ for all $n \in \mathbb{N}$. Then $A_n \in \mathfrak{F}_\alpha$ 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 (19.2c). It follows that they both belong to the intersection $\bigcap_\alpha \mathfrak{F}_\alpha$, i.e., $\mathfrak{F}$ satisfies (19.2c). It follows that $\mathfrak{F}$ is a σ–algebra.

PROOF of $\boldsymbol{B}$:

First of all, we know that $\sigma(\mathfrak{E})$ is an intersection of σ–algebras and, according to part $\boldsymbol{A}$ of this proposition, really is a σ–algebra. We now prove that $\sigma(\mathfrak{E})$ contains $\mathfrak{E}$ and is the minimal σ–algebra with that property. First let us prove that $\sigma(\mathfrak{E}) \supseteq \mathfrak{E}$. But that is obvious because it is the intersection of sets all of which contain $\mathfrak{E}$. On the other hand, $\sigma(\mathfrak{E})$ is the intersection of **all** σ–algebras that contain $\mathfrak{E}$, so it is impossible for any other σ–algebra to both be a strict subset of $\sigma(\mathfrak{E})$ and also contain $\mathfrak{E}$. ■

Definition 19.3 (Abstract measures)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space.

A **measure** on $\mathfrak{F}$ is an extended real–valued function

$$\mu(\cdot) : \mathfrak{F} \to \overline{\mathbb{R}}_+; \quad A \mapsto \mu(A) \qquad \text{such that}$$

(19.5) $\quad \mu(\emptyset) \;=\; 0 \qquad$ (positivity)

(19.6) $\quad A, B \in \mathfrak{F} \text{ and } A \subseteq B \;\Rightarrow\; \mu(A) \leq \mu(B) \qquad$ (monotony)

(19.7) $\quad (A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \text{ disjoint} \;\Rightarrow\; \mu\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n) \qquad$ (σ–additivity)

- The triplet $(\Omega, \mathfrak{F}, \mu)$ is called a **measure space**
- We call μ a **finite measure** on $\mathfrak{F}$ if $\mu(\Omega) < \infty$.
- If $\mu(\Omega) = 1$ then $\mu(\cdot)$ is called a **probability measure**. □

Disjointness in (19.7) means that $A_i \cap A_j = 0$ for any $i, j \in \mathbb{N}$ such that $i \neq j$ (see Definition 2.6 on p.17).

A measure space can support many different measures.

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 19.1 (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

$$\begin{aligned} \lambda^1([a,b]) \;&:=\; b - a, \\ \lambda^2([a_1,b_1] \times [a_2,b_2]) \;&:=\; (b_1 - a_1)(b_2 - a_2), \\ \lambda^3([a_1,b_1] \times [a_2,b_2] \times [a_3,b_3]) \;&:=\; (b_1 - a_1)(b_2 - a_2)(b_3 - a_3), \\ \lambda^n([a_1,b_1] \times \cdots \times [a_n,b_n]) \;&:=\; (b_1 - a_1)(b_2 - a_2)\dots(b_n - a_n) \end{aligned} \tag{19.8}$$

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 [186] λ^n is called n**-dimensional Lebesgue measure**

[186] This is not entirely correct: we must demand that the measure is σ-finite, i.e., there are measurable sets with finite

Note that Lebesgue measure is not finite. □

Example 19.2. You can easily verify that the following set function defines a measure on an arbitrary nonempty 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 19.2 (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 property (19.2c) 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 $A_{N+1} = A_{N+2} = \ldots = 0$:

(19.9)
$$\begin{aligned} &A_1, A_2, \ldots, A_N \in \mathfrak{F} \text{ mutually disjoint} \\ &\quad \Rightarrow \mu(A_1 \uplus A_2 \uplus \ldots \uplus A_N) = \mu(A_1) + \mu(A_2) + \ldots + \mu(A_N) \quad \text{(additivity).} \end{aligned}$$

In the case of only two disjoint measurable sets A and B the above simply becomes

$$\mu(A \uplus B) = \mu(A) + \mu(B). \ \Box$$

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 19.4 (Contents as additive measures)**.** Let Ω be a nonempty set and let $\mathfrak{F}$ be a σ–algebra for Ω.

A **content** on $\mathfrak{F}$ is a real–valued function $m(\cdot) : \mathfrak{F} \to \mathbb{R}$, $A \mapsto m(A)$ which satisfies

(19.10a) $m(\emptyset) = 0$ (positivity)

(19.10b) $A, B \in \mathfrak{F}$ and $A \subseteq B \Rightarrow m(A) \leq m(B)$ (monotony)

(19.10c) $A_1, A_2, \ldots, A_N \in \mathfrak{F}$ 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 19.2 (Simple properties of measures)**.** *Let $A, B, \in \mathfrak{F}$ and let μ be a measure on $\mathfrak{F}$. Then*

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 $\lambda^n(A_K) = (2k)^n < \infty$.

$$\mu(A) \geq 0 \quad \textit{for all } A \in \mathfrak{F}, \tag{19.11a}$$
$$A \subseteq B \;\Rightarrow\; \mu(B) \;=\; \mu(A) + \mu(B \setminus A), \tag{19.11b}$$
$$\mu(A \cup B) + \mu(A \cap B) \;=\; \mu(A) + \mu(B). \tag{19.11c}$$

If μ is finite then also

$$A \subseteq B \;\Rightarrow\; \mu(B \setminus A) \;=\; \mu(B) - \mu(A), \tag{19.12a}$$
$$\mu(A \cup B) \;=\; \mu(A) + \mu(B) - \mu(A \cap B). \tag{19.12b}$$

PROOF: The first property follows from the fact that $\mu(\emptyset) = 0$, $\emptyset \subseteq A$ for all $A \in \mathfrak{F}$ and (19.6.
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 (19.11c) is obviously true. Hence we may assume that $\mu(A \cup B) < \infty$. We have

$$A \cup B \;= (A \cap B) \uplus (B \setminus A) \uplus (A \setminus B) \tag{19.13a}$$
$$A \cup B \;= A \uplus (B \setminus A) \;=\; B \uplus (A \setminus B) \tag{19.13b}$$

It follows from (19.13a) that

$$\mu(A \cup B) \;=\; \mu(A \cap B) + \mu(B \setminus A) + \mu(A \setminus B) \tag{19.14}$$

It follows from (19.13b) that

$$2 \cdot \mu(A \cup B) \;=\; \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) \tag{19.15}$$

We subtract the left and right sides of (19.14) from those of (19.15) 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. ■

19.2 Sequences of Sets – limsup and liminf

Assumption 19.1 (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 19.5 (Monotone set sequences)**.** A sequence A_k of arbitrary subsets of X is called

(19.16a) **nondecreasing** if $A_1 \subseteq A_2 \subseteq \dots$

(19.16b) **nonincreasing** if $A_1 \supseteq A_2 \supseteq \dots$

(19.16c) **strictly increasing** if $A_1 \subsetneq A_2 \subsetneq \dots$

(19.16d) **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 19.6 (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

(19.17a) $$A = \lim_{n\to\infty} A_n \quad \text{or} \quad A_n \uparrow A \text{ for } n \to \infty,$$

(19.17b) $$B = \lim_{n\to\infty} B_n \quad \text{or} \quad B_n \downarrow B \text{ for } n \to \infty. \ \square$$

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 19.7 (lim inf and lim sup of set sequences)**.** Given are sets $A_n, B_n \subseteq X$ $(n \in \mathbb{N})$. Let

(19.18) $$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k \quad \textbf{(limit inferior)}$$

(19.19) $$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k \quad \textbf{(limit superior)}$$

In general those two will not coincide. But if they do then we define

(19.20) $$\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$$

We call $\lim\limits_{n\to\infty} A_n$ the **limit** of the sequence (A_n) and we write

$$A_n \to A \quad \text{for } n \to \infty. \ \square$$

The following comments should make matters easier to understand if you abbreviate

Lemma 19.1 (lim inf and lim sup as monotone limits)**.** *Given are sets* $A_n, B_n \subseteq X$ $(n \in \mathbb{N})$*. Let*

(19.21) $$A_{\star n} := \bigcap_{k\geq n} A_k \qquad \textit{Then} \quad A_{\star n} \uparrow \liminf_{n\to\infty} A_n$$

(19.22) $$A^{\star}{}_n := \bigcup_{k\geq n} A_k \qquad \textit{Then} \quad A^{\star}{}_n \downarrow \limsup_{n\to\infty} A_n$$

PROOF: Let $m, n \in \mathbb{N}$ such that $m < n$. Then

$$\begin{aligned} A_{\star m} &= \bigcap_{k=m}^{n-1} A_k \cap \bigcap_{k \geq n} A_k &= \bigcap_{k=m}^{n-1} A_k \cap A_{\star n} &\subseteq A_{\star n} \\ A^{\star}{}_m &= \bigcup_{k=m}^{n-1} A_k \cup \bigcup_{k \geq n} A_k &= \bigcup_{k=m}^{n-1} A_k \cup A^{\star}{}_n &\supseteq A^{\star}{}_n \end{aligned}$$

This proves that $A_{\star n}$ is nondecreasing and $A * \star_n$ is nonincreasing. By the very definition of the limit of a monotone sequence of sets it is true that

$$\begin{aligned} \lim_{n \to \infty} A_{\star n} &= \bigcup_{n \in \mathbb{N}} A_{\star n} = \liminf_{n \to \infty} A_n \\ \lim_{n \to \infty} A^{\star}{}_n &= \bigcap_{n \in \mathbb{N}} A^{\star}{}_n = \limsup_{n \to \infty} A_n \end{aligned}$$

■

19.3 Conditional Expectations as Generalized Averages

EMPTY!!
EMPTY!!
EMPTY!!

20 Appendix: Addenda to Beck/Geoghegan's "The Art of Proof"

This chapter contains extensions of material found in [2] Beck/Geoghegan, the book which is meant to be read in conjunction with these lecture notes. Some of this material is referenced in earlier chapters.

Notation 20.1.

Even though the subject matter of this chapter is primarily the book [2] Beck/Geoghegan: The Art of Proof, a reference such as prop.9.7 will refer to proposition 9.7 of this document. Proposition 9.7 of the Beck Geoghegan book will be referenced as "B/G prop.9.7" or "[2] B/G prop.9.7" or something similar. □

20.1 AoP Ch.1: Integers

Note that B/G ch.1, axioms 1.1 – 1.5 for the set $\mathbb{Z}$ of the integers match the definition of an integral domain which was given in Definition 3.10 on p.63 of this document. Does that mean that "integral domain" is just a fancy name for the set $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$? The answer is No! We have seen in prop.3.12 of ch.3 (The Axiomatic Method) that not only the integers, but also the rational numbers and the real numbers with the binary operations of addition and multiplication are integral domains.

So what is going on here? The answer: B/G chose to specify the set $\mathbb{Z}$ in stages. The first set of axioms, the one just mentioned, specifies the algebraic properties of addition and multiplication. B/G axioms 2.1(i) – 2.1(iv) are added in their second chapter to tell us that the integral domain $\mathbb{Z}$ is an ordered integral domain with positive cone $\mathbb{N}$. See Definition 3.11 on p.68. Finally, B/G ax.2.15, the induction axiom, is added. Only at this point is $\mathbb{Z}$ completely specified. We chose in this document to define the integers "in one shot" rather than piecemeal. See axiom 6.1 on p.6.1. We duplicated the material of B/G ch.1 in ch.refssec:arithm-integral-domains (Arithmetic in Integral Domains) on p.64 and the material of B/G ch.2.1 and 2.2 in ch.3.4 (Order Relations in Integral Domains) on p.68.

20.1.1 Ch.1.1 – Axioms

There are no addenda at this point in time.

20.2 AoP Ch.2: Natural Numbers and Induction

The following remark is also part of rem.3.4 on p.64.

Remark 20.1. B/G ch.1, axioms 1.1 – 1.5 plus 2.1(i) – 2.1(iv) for the set $\mathbb{Z}$ of the integers state that $\mathbb{Z}$ is an ordered integral domain with positive cone $\mathbb{N}$. See Definition 3.11 on p.68. □

Remark 20.2.

It follows from the last remark that all of the material in B/G ch.2.1 and 2.2, i.e., all of the propositions and corollaries and definitions inbetween B/G prop.2.1 and prop.2.13, extend to any ordered integral domain $(R, \oplus, \odot, P)$. An example would be prop.3.33 on p.70 which corresponds to B/G ch.1 prop.2.2. □

20.2.1 AoP Ch.2.2 (Ordering the Integers)

The material given here complements ch.2.2 (Ordering the Integers) of [2] Beck/Geoghegan.

Remark 20.3. B/G Axioms 1.1 – 1.5 together with axiom 2.1 do not suffice to characterize the integers; a symbol different from $\mathbb{Z}$ might have been more appropriate. These axioms are also satisfied by the set $\mathbb{Q}$ of all rational numbers (fractions), provided one interprets the set $\mathbb{N}$ described by axiom 2.1 as the set of all (strictly) positive fractions. They are also satisfied by the set $\mathbb{R}$ of all real numbers (decimals), provided one interprets $\mathbb{N}$ as the set of all (strictly) positive decimals. Only addition of the induction axiom (B/G axiom 2.15) excludes $\mathbb{Q}$ and $\mathbb{R}$. □

Accordingly, all propositions and theorems of the B/G text before the induction axiom apply not only to integers but to rational and real numbers as well.

20.2.2 AoP Ch.2.3 (Induction)

There are no addenda at this point in time.

20.2.3 Bounded Sets in $\mathbb{Z}$

There are no addenda at this point in time.

20.2.4 Exercises for Ch.20.2

There are no exercises at this point in time.

20.3 AoP Ch.3: Some Points of Logic

Here are some references for items discussed in ch.3 of the B/G text to where they appear in ch.4 (Logic) on p.84 of this document.

(a) Universal quantifier $\forall$, existental quantifier $\exists$, unique existental quantifier $\exists!$: ch.4.5.1 on p.108.

(b) "$(\forall\ x)(\forall\ y)$ has the same meaning as $(\forall\ x$ **and** $y)$": part a of Definition 4.17 (Doubly quantified expressions) in ch.4.5.2 on p.110.

We recommend that you look at prop.4.1 and the note which precedes it. The latter is reproduced here:

(1) The order in which the qualifiers are applied is important. $\forall x \exists y$ generally does not mean the same as $\exists y \forall x$.

(2) Interchanging variable names in the qualifiers is not OK. $\forall x \exists y$ generally does not mean the same as $\forall y \exists x$.

(c) **Skip part (c) if you have no knowledge of logic beyond what is in ch.3 of B/G.**
If you have had some training in logic you may have learned to express "if P (is true) then Q (is true)" as "$P \to Q$" rather than "$P \Rightarrow Q$". There is a difference: Proving a statement of the form "if P then Q" means to show that the statements P and Q are related in a fashion that makes it "logically impossible", [187] for P to be true and Q to be false. "$P \Rightarrow Q$" has the combination $\big($ P is true, Q is false $\big)$ marked as irrelevant (logically impossible); the remaining three combinations give the same outcome as for $P \to Q$, i.e., they all evaluate to **true**.

Example: Let us assume that x is an integer. Let P be the statement P : "$x = 100$", and let Q be the statement Q : "$x > 10$". Then it is correct to write $P \Rightarrow Q$ because, no matter the value of x, it is not possible that P is true and Q is false.

(d) The equivalent forms of "if P then Q" such as "Q whenever P" are listed in ch.4.2.5 (Arrow and Implication Operators) on p.95. There you also find the definition of the **converse** and the **contrapositive** of an implication.

Here are some notes about B/G ch.3.3: Negations.

(a) The negation of "A **and** B" is "(**not** A) **or** (**not** B)", and the negation of "A **or** B" is "(**not** A) **and** (**not** B)". This is known as "De Morgan's law" for statements (see thm.4.3 on p.101). Do you see the connection to De Morgan's law for sets? (see thm.8.1 on p.234).

(b) **Skip part (b) if you have no knowledge of logic beyond what is in ch.3 of B/G.**
B/G states that the negation of "$P \Rightarrow Q$" is "P **and not** Q". This should be stated more appropriately as follows: The negation of "$P \Rightarrow Q$" is "P **and** (**not** Q)". The reason for this equivalence is that "$P \Rightarrow Q$" is defined as "**not** (P **or** Q)": we can apply De Morgan's laws to obtain the negation.

(c) We chose to write parentheses in the above expression to avoid ambiguity. But note that there is a "binding power" or preference for the logical operators: The negation "**not** P" of a statement P has higher preference than "P **and** Q" and "P **or** Q":
The meaning of "**not** P **or** Q" is "(**not** P) **or** Q", not "**not** (P **or** Q)."

20.4 AoP Ch.4: Recursion

There are no addenda at this point in time.

20.5 AoP Ch.5: Underlying Notions in Set Theory

There are no addenda at this point in time.

[187] See ch.4.2.2, Definition 4.7 on p.90.

20.6 AoP Ch.6: Equivalence Relations and Modular Arithmetic

20.6.1 Equivalence Relations

Remark 20.4. B/G defines in this chapter the absolute value for integers as usual:

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

See Definition 2.20 on p.28. Note that this document has generalized this concept to integral domains. See Definition 3.13 on p.75. □

20.6.2 The Division Algorithm

There are no addenda at this point in time.

20.6.3 The Integers Modulo n

There are no addenda at this point in time.

20.6.4 Prime Numbers

There are no addenda at this point in time.

20.6.5 Exercises for Ch.20.6

There are no exercises at this point in time.

20.7 AoP Ch.7: Arithmetic in Base Ten

20.7.1 Base–Ten Representation of Integers

You will find the material of this B/G chapter represented somewhat differently in ch.6.13 (The Base–β Representation of the Integers) of this document.

20.8 AoP Ch.8: Real Numbers

B/G ch.8 axiomatically defines the set $\mathbb{R}$ of all real numbers. It was stated in ch.2.3 (Numbers) on p.24 that real numbers are the same as decimal numbers. A proof of this will be given in B/G ch.12 (Decimal Expansions).

20.8.1 Axioms

Remark 20.5.

> Note that B/G axioms 8.1 – 8.5 and axiom 8.26 do not suffice to determine what we think of as the real numbers because the set $\mathbb{Q}$ also satisfies each one of them. Only the addition of axiom 8.52 (completeness axiom) will accomplish this.

We encountered a similar situation in the first two chapters of B/G with the set $\mathbb{Z}$ of all integers where axioms 1.1 – 1.5 and 2.1 also are valid for $\mathbb{Q}$ and axiom 2.15 (induction axiom) was needed to completely determine the set $\mathbb{Z}$. □

Remark 20.6. For the following see also rem20.2 on p. 497.

Note that B/G axioms 8.1 – 8.5 in ch.8.1 for the set $\mathbb{R}$ of the real numbers together with B/G prop.8.7 (the cancellation rule holds in $\mathbb{R}$) imply that $\mathbb{R}$ is an integral domain, but one with the additional property that $(\mathbb{R}_{\neq 0}, \cdot)$ is an abelian group [188] (see Definition 3.2 on p.52). B/G prop.8.8 – prop.8.24 only depend on the integral domain properties of $\mathbb{R}$, just as the corresponding propositions of B/G ch.1 only depend on the integral domain properties of $\mathbb{Z}$,

This explains the remark preceding B/G prop.8.8 in which it is reasoned that the proofs of B/G prop.8.8 through B/G prop.8.24 in B/G ch.8.1 are literally the same as those for the corresponding propositions in B/G ch.1. S □

20.9 AoP Ch.9: Embedding $\mathbb{Z}$ in $\mathbb{R}$

Remark 20.7. Note that B/G prop.9.10 and prop.9.12 are covered by thm.5.2 on p.149. □

20.10 AoP Ch.10: Limits and Other Consequences of Completeness

There are no addenda at this point in time.

20.11 AoP Ch.11: Rational and Irrational Numbers

There are no addenda at this point in time.

20.12 AoP Ch.12: Decimal Expansions

There are no addenda at this point in time.

20.13 AoP Ch.13: Cardinality

There are no addenda at this point in time.

20.14 Exercises for Ch.20

All exercises appear at the end of the individual subchapters.

[188] such an integral domain is called a **field**.

21 Exam Preparation

Last update: February 23, 2018.

Most of this chapter features problems which are typical for what you might find on one of my Math 330 exams. You will also find here a list of definitions which you **need NOT** to learn by heart. Be aware though that those definitions may be referenced later in the text.

I plan to add to those lists in the future and also include Sample Problems for more topics.

Note that solutions have been written here for many but not all problems!

21.1 Sample Problems for Induction

Problem 21.1. (**Induction**). Let $x_1 = 1, x_2 = 1 + \frac{1}{2}, \dots, x_k = \sum_{j=1}^{k} \frac{1}{j}$ $(k \in \mathbb{N})$.
Prove by induction that $\sum_{k=1}^{n} x_k \; = \; (n+1)x_n - n$ $(n \in \mathbb{N})$.

Solution to #21.1: See Grimaldi Discrete Math 4ED, Exercise 4.1, # 2c, p.176 ■

Problem 21.2. (**Induction**). Prove by induction that $\sum_{j=1}^{n} j(j!) \; = \; (n+1)! - 1$ $(n \in \mathbb{N})$.

Solution to #21.2:
Base case $n = 1$: LS = $1 \cdot (1!) = \cdot 1 = 1 = 2 - 1 = (2!) - 1$ = RS.
Induction assumption $(\star)$: $\sum_{j=1}^{n} j(j!) \; = \; (n+1)! - 1$ for all $j \in \mathbb{Z}$ such that $0 \leq j < n$.
Need to show $(\star\star)$: $\sum_{j=1}^{n+1} j(j!) \; = \; (n+2)! - 1$.

$$\begin{aligned} \text{LS of } (\star\star) \; &= \sum_{j=1}^{n} j(j!) + (n+1)(n+1)! \overset{(\star)}{=} (n+1)! - 1 + (n+1)(n+1)! \\ &= (1)(n+1)! + (n+1)(n+1)! - 1 \; = \; (n+2)(n+1)! - 1 \; = \; \text{RS of } (\star\star). \end{aligned}$$

Thus $(\star\star)$ is valid. This finishes the proof by induction. ■

Problem 21.3. (**Strong Induction**). Let $x_0 = 1, \; x_1 = 2, \; x_2 = 3, \; \dots, \; x_n = x_{n-1} + x_{n-2} + x_{n-3}$ $(n \in \mathbb{N}, n \geq 3)$. Prove by strong induction that $x_n \leq 3^n$ for all $n \in \mathbb{Z}_{\geq 0}$.

Solution to #21.3:
Base cases are $n = 0, n = 1, n = 2$. (We need three base cases because the recursion formula) $x_n = x_{n-1} + x_{n-2} + x_{n-3}$ requires the knowledge of three predecessors.)
$n = 0$ is valid since $x_0 = 1 = 3^0$.
$n = 1$ is valid since $x_1 = 2 < 3 = 3^1$.
$n = 2$ is valid since $x_2 = 3 < 9 = 3^2$.
Induction assumption $(\star)$: $x_j \leq 3^j$ for all $j \in \mathbb{Z}$ such that $0 \leq j < n$.

Need to show $(\star\star)$: $x_n \leq 3^n$ for all $n \geq 3$.
It follows from $(\star)$ that $x_{n-3} \leq 3^{n-3}$, $x_{n-2} \leq 3^{n-2}$, $x_{n-1} \leq 3^{n-1}$.
Thus **(a)** $x_{n-1} + x_{n-2} + x_{n-3} \leq 3^{n-3} + 3^{n-2} + 3^{n-1} \leq 3 \cdot 3^{n-1}$.

Since $n \geq 3$, LS of $(\star\star)$ $= x^n = x_{n-1} + x_{n-2} + x_{n-3} \overset{\textbf{(a)}}{\leq} 3 \cdot 3^{n-1} = 3^n =$ RS of $(\star\star)$.
Thus $(\star\star)$ is valid. This finishes the proof by induction. ■

Problem 21.4. (**Strong Induction**).
Let $x_0 = 2,\ x_1 = 4,\ x_{n+1} = 3x_n - 2x_{n-1}$ for $n \in \mathbb{N}$. Prove by strong induction that $x_n = 2^{n+1}$ for every integer $n \geq 0$. Hint: Is one number enough for the base case?

Solution to #21.4:
Base cases: $n = 0, 1$: $x_0 = 2 = 2^{0+1}$. Further, $x_1 = 4 = 2^{1+1}$. This proves the base cases.
Induction assumption $(\star)$: Let $n \in \mathbb{N}$. Assume that $x_j = 2^{j+1}$ for **all** $0 \leq j \leq n$.
Need to show $(\star\star)$: $x_{n+1} = 2^{n+2}$.

$$\begin{aligned}\text{LS of } (\star\star) &= x_{n+1} = 3x_n - 2x_{n-1} \quad \text{(the recursive definition)}\\ &= 3(2^{n+1}) - 2(2^n) \quad ((\star) \text{ was applied both to } j = n \text{ and } j = n-1\,)\\ &= 6 \cdot 2^n - 2 \cdot 2^n = 4 \cdot 2^n = 2^{n+2} = \text{RS of } (\star\star).\end{aligned}$$

Thus $(\star\star)$ is valid. This finishes the proof by induction. ■

Problem 21.5. (**Strong Induction**).
Let $x_0 = 1,\ x_1 = 3,\ x_{n+1} = 2x_n + 3x_{n-1}$ for $n \in \mathbb{N}$. Prove by strong induction that $x_n = 3^n$ for every integer $n \geq 0$. Hint: Is one index enough for the base case?

Solution to #21.5:
Base cases: $n = 0, 1$: $x_0 = 1 = 3^0$. Further, $x_1 = 3 = 3^1$. This proves the base cases.
Induction assumption $(\star)$: Let $n \in \mathbb{N}$. Assume that $x_j = 3^j$ for **all** $0 \leq j \leq n$.
Need to show $(\star\star)$: $x_{n+1} = 3^{n+1}$..

$$\begin{aligned}\text{LS of } (\star\star) &= x_{n+1} = 2x_n + 3x_{n-1} \quad \text{(the recursive definition)}\\ &= 2(3^n) + 3(3^{n-1}) \quad ((\star) \text{ was applied both to } j = n \text{ and } j = n-1\,)\\ &= 2 \cdot 3^n + 1 \cdot 3^n = 3 \cdot 3^n = 3^{n+1} = \text{RS of } (\star\star).\end{aligned}$$

Thus $(\star\star)$ is valid. This finishes the proof by induction. ■

Problem 21.6. (**Recursion**).
Let $x_1 = 3,\ x_{n+1} = x_n + 2n + 3$ ($n \in \mathbb{N}$. Prove by induction that $x_n = n(n+2)$ ($n \in \mathbb{N}$).

Solution to #21.6:
Base case $n = 1$: LS = 3 $= 1(1+2) = 1 =$ RS.
Induction assumption $(\star)$: $x_n = n(n+2)$.

Need to show $(\star\star)$: $x_{n+1} = (n+1)(n+3)$

$$\begin{aligned} \text{LS of } (\star\star) &= x_{n+1} \overset{\text{def.}}{=} x_n + 2n + 3 \\ &\overset{(\star)}{=} n(n+2) + 2n + 3 \ = \ n^2 + 4n + 3 \ = \ (n+1)(n+3) \ = \ \text{RS of } (\star\star). \end{aligned}$$

Thus $(\star\star)$ is valid. This finishes the proof by induction. ■

21.2 Sample Problems for Functions and Relations

Problem 21.7. (**Partial order relations**). Remark 5.5.**D** on p.130 states that if $(X, \preceq)$ is a POset and if $A \subseteq X$ then the relation $\preceq_A$ on A defined as $x \preceq_A y$ if and only if $x \preceq y$ $(x, y \in A)$ is a partial ordering on A. Prove it.

Solution to #21.7:

We must prove reflexivity, antisymmetry, and transitivity.

(a) Proof of reflexivity: Let $a \in A$. We must prove that $a \preceq_A a$. Since $x \in X$ and "$\preceq$" is reflexive as a partial ordering on X we obtain $a \preceq a$, i.e., $a \preceq_A a$.

(b) Proof of antisymmetry: Let $a, b \in A$ such that $a \preceq_A b$ and $b \preceq_A a$. We must prove that $a = b$. Since $a, b \in X$, "$\preceq$" is antisymmetric as a partial ordering on X, $a \preceq_A b \Rightarrow a \preceq b$, and $b \preceq_A a \Rightarrow b \preceq a$, we obtain $a = b$.

(c) Proof of transitivity: Let $a, b, c \in A$ such that $a \preceq_A b$ and $b \preceq_A c$. We must prove that $a \preceq_A c$. Since $a, b, c \in X$, "$\preceq$" is transitive as a partial ordering on X. Moreover $a \preceq_A b \Rightarrow a \preceq b$, and $b \preceq_A c \Rightarrow b \preceq c$, thus $a \preceq c$, i.e., $a \preceq_A c$. ■

Problem 21.8. (**Functions**). Given is a function $f : A \longrightarrow B$ $(A, B \neq \emptyset)$. Give the definitions of each of the following:

(a) f is injective.
(b) f is surjective.
(c) f is bijective.
(d) f has a left-inverse g.
(e) f has a right-inverse h.

For **(d)** and **(e)**, give the "arrow diagram" which show domain and codomain for each function involved. In both cases it will like the one to the left. Each symbol **S** denotes a (possibly different) set and each symbol φ denotes a (possibly different) function.

$$\begin{tikzcd} \mathbf{S} \arrow[r, "\varphi"] \arrow[dr, "\varphi"'] & \mathbf{S} \arrow[d, "\varphi"] \\ & \mathbf{S} \end{tikzcd}$$

Solution to #21.8:

Solution to problems a,b,c:

Injective means one-one: If $a_1, a_2 \in A$ and $f(a_1) = f(a_2)$ then $a_1 = a_2$.

Surjective means onto: If $b \in B$ then there is $a \in A$ such that $f(a) = b$.

Bijective means both injective and surjective.

Solution to problem d: If this diagram commutes:

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[dr, "1_A"'] & B \arrow[d, "g"] \\
& A
\end{tikzcd}$$

i.e., $g \circ f = 1_A$, then we call g a **left inverse** of f ("to the left of the reference object f"), i.e., f has a left inverse (namely the function g).

Solution to problem e: If this diagram commutes:

$$\begin{tikzcd}
B \arrow[r, "h"] \arrow[dr, "1_B"'] & A \arrow[d, "f"] \\
& B
\end{tikzcd}$$

i.e., $f \circ h = 1_B$, then we call h a **right inverse** of f ("to the right of the reference object f)", i.e., f has a right inverse (namely the function h). ■

Problem 21.9. (**Equivalence relations and partial order relations**).

(a) Let $a, b \in \mathbb{Z}$. State as precisely as possible the definition of $a \mid b$.

(b) Is the relation $a \sim b \;\Leftrightarrow\; a \mid b$ **reflexive**? **symmetric**? **antisymmetric**? **transitive**? If true, prove it. If false, give a counterexample.

Solution to #21.9:

(a) There is $z \in \mathbb{Z}$ such that $az = b$.

(b) Is the relation $a \sim b \;\Leftrightarrow\; a \mid b$

- **reflexive**: TRUE: If $a \in \mathbb{Z}$ then $a = 1 \cdot a$, hence $a \mid a$, hence $a \sim a$.
- **symmetric**: FALSE: Counterexample: $5 \mid 10$ but $10 \nmid 5$.
- **antisymmetric**: FALSE: Counterexample: $-1 \mid 1$ and $1 \mid -1$, but $1 \neq -1$.
- **transitive**: TRUE: Let $a, b, c \in \mathbb{Z}$ such that $a \sim b$ and $b \sim c$.
  Then $a \mid b$, i.e., $b = ma$, and $b \mid c$, i.e., $c = nb$ for suitable $m, n \in \mathbb{Z}$.
  Thus $c = (mn)a$. Since $(mn) \in \mathbb{Z}$, $a \mid c$, hence $a \sim c$. ■

Problem 21.10. (**Functions and equivalence relations**).
Let $f : X \to Y (X, Y \neq \emptyset)$. Prove that $a \sim b \;\Leftrightarrow\; f(a) = f(b)$ is an equivalence relation on X.

The solution to this problem is not given here. ■

21.3 Sample Problems for Convergence and Uniform Convergence

Problem 21.11. (**Convergence in Metric Spaces**).
Prove closed book prop.13.9 on p.419.

Problem 21.12. (**Convergence in Metric Spaces**).
Prove closed book thm.12.3 on p.370.

Problem 21.13. (**Convergence in Metric Spaces**).
Prove closed book prop.12.7 on p.370.

Problem 21.14. (**Convergence in Metric Spaces**).
Prove closed book prop.12.8 on p.370.

Problem 21.15. (**Uniform convergence**).

Let $(X, d) := ([0,1], d_{|\cdot|})$ be the unit interval, equipped with the standard Euclidean metric $d(x, x') = |x - x'|$, and let $(Y, d') := ([0,1], d_{\|\cdot\|_{L^2}})$ be the same set, but equipped with the metric derived from the L^2–norm $\|f\|_{L^2} = \sqrt{\int_a^b f(x)^2 dx}$. (See Definition 11.21 on p.349).

For $n \in \mathbb{N}$ let $f_n(x) := (-1)^n$ if $0 \le x \le \frac{1}{n}$ and 0 if $\frac{1}{n} < x \le 1$.

(a) Prove that $f_n(\cdot) \to 0$ on (Y, d').

(b) Prove that $f'n(x)$ does not converge pointwise on (X, d). Hint: What exactly must you show?

Solution to #21.15:

PROOF of **(a)**: Let $\delta > 0$. Then $d'(f_n, 0) = \sqrt{\int_0^{1/n} 1^2 dt + \int_{1/n}^1 0^2 dt} = \sqrt{\frac{1}{n}}$, and this expression is less than δ whenever $n > \frac{1}{\delta^2}$. Let $N := \min\{j \in \mathbb{N} : j > \frac{1}{\delta^2}\}$. It follows that $d'(f_n, 0) < \delta$ for all $n \ge N$. This proves that $f_n(\cdot) \to 0$ on (Y, d').

PROOF of **(b)**: We have $f_n(0) = (-1)^n$ for all $n \in \mathbb{N}$. This sequence is not Cauchy in (X, d) since $d(f_j(0), f_{j+1}(0) = 2$ for all $j \in \mathbb{N}$ and hence not convergent (see thm.12.8 (Convergent sequences are Cauchy) on p.391). ■

21.4 Other Topics

Problem 21.16. (**Logic**). Given a function $f : X \to Y$, negate the following statements:

(a) There exists $x \in X$ and $y \in Y$ such that $f(x) = y$,

(b) For all $x \in X$ there exists $y \in Y$ such that $f(x) = y$,

(c) $\exists x \in X$ such that $\forall y \in Y$ such that $f(x) \ne y$.

(d) $\forall x_1, x_2 \in X$: if $x_1 \ne x_2$ then $f(x_1) \ne f(x_2)$.

Solution to #21.16:

(a) $\forall x \in X, \forall y \in Y : f(x) \ne y$,

(b) $\exists x \in X$ such that $\forall y \in Y : f(x) \ne y$,

(c) $\forall x \in X \; \exists y \in Y$ such that $f(x) = y$,

(d) $\exists x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$. ■

Problem 21.17. (**Sets**). Prove AND REMEMBER the set identities **2.a** through **2.g** of prop.8.3 on p.236:

(b) $A \triangle \emptyset = \emptyset \triangle A = A$ (neutral element $\emptyset$ for $\triangle$)

(c) $A \triangle A = \emptyset$ (inverse element $\emptyset$ for $\triangle$) [189]

(d) $A \triangle B = B \triangle A$ (commutativity of $\triangle$)

(e) $(A \cap B) \cap C = A \cap (B \cap C)$ (associativity of $\cap$)

(f) $A \cap \Omega = \Omega \cap A = A$ (neutral element Ω for $\cap$)

(g) $A \cap B = B \cap A$ (commutativity of $\cap$)

Solution to #21.17:

Trivial (we omitted the tough one, **2.h**). ■

Problem 21.18. (**Set functions**). Given is an arbitrary collection of sets $(A_j)_{j \in J}$. Determine for each

[189] The inverse element for A in the sense of Definition 3.3 on p.53. is A itself!

assertion below whether it is true or false. If it is true, prove it. If it is false, give a counterexample.

(a) $f(\bigcup\limits_{j\in J} A_j) \subseteq \bigcup\limits_{j\in J} f(A_j)$; **(b)** $\bigcup\limits_{j\in J} f(A_j) \subseteq f(\bigcup\limits_{j\in J} A_j)$;

(c) $f(\bigcap\limits_{j\in J} A_j) \subseteq \bigcap\limits_{j\in J} f(A_j)$; **(d)** $\bigcap\limits_{j\in J} f(A_j) \subseteq f(\bigcap\limits_{j\in J} A_j)$;

You may use the fact that the direct image is increasing with its argument: $A \subseteq B \Rightarrow f(A) \subseteq f(B)$.

Solution to #21.18:

Solution to a:

$$\begin{aligned} y \in f(\bigcup_{j\in J} A_j) &\Rightarrow \exists x \in \bigcup_{j\in J} A_j \text{ such that } f(x) = y \quad \text{(def. direct image)}\\ &\Rightarrow \exists j_0 \in J \text{ such that } x \in A_{j_0} \quad \text{(def. union)}\\ &\Rightarrow y = f(x) \in f(A_{j_0}) \quad \text{(def. direct image)}\\ &\Rightarrow y = f(x) \in \bigcup_{j\in J} f(A_{j_0}) \quad \text{(def. union)} \end{aligned}$$

Solution to b: As the direct image is increasing with its argument and $A_i \subseteq \bigcup_{j\in J} A_j$ for all $i \in J$ it follows that $f(A_i) \subseteq f\big(\bigcup_{j\in J} A_j\big)$ for all $i \in J$. Hence

$$\bigcup_{i\in J} f(A_i) \subseteq \bigcup_{i\in J} \Big(f(\bigcup_{j\in J} A_j)\Big) = f(\bigcup_{j\in J} A_j)$$

The equality on the right hand side results from the fact that the set $f\big(\bigcup_{j\in J} A_j\big)$ does not depend on the index variable $i \in J$ with respect to which the "outer" union takes place. ■

Problem 21.19. (**Cardinality**).
Give an alternate proof of thm.9.12 (The real numbers are uncountable) on p.285 which is based on the fact that the cardinality of a set is less than that of its power set (thm.10.1 on p.308). Hint: Find bijections $f : \{x \in \mathbb{R} : x = \sum\limits_{j=1}^{\infty} d_j 10^{-j} \text{ and } d_j = 3 \text{ or } 4\ \forall j\} \xrightarrow{\sim} \{3,4\}^{\mathbb{N}}$ and $g : \{3,4\}^{\mathbb{N}} \xrightarrow{\sim} 2^{\mathbb{N}}$.

Solution to #21.19:

Let $\Gamma := \{x \in \mathbb{R} : x = \sum\limits_{j=1}^{\infty} d_j 10^{-j} \text{ and } d_j = 3 \text{ or } 4\ \forall j\}$.

Both $f : \Gamma \xrightarrow{\sim} \{3,4\}^{\mathbb{N}}$; $\sum\limits_{j=1}^{\infty} d_j 10^{-j} \mapsto (d_j)_{j\in\mathbb{N}}$ and $g : \{3,4\}^{\mathbb{N}} \xrightarrow{\sim} 2^{\mathbb{N}}$; $(d_j)_{j\in\mathbb{N}} \mapsto \{j \in \mathbb{N} : d_j = 4\}$

are bijections, thus $g \circ f$ is a bijection from $\Gamma \subseteq \mathbb{R}$ to $2^{\mathbb{N}}$. According to thm.10.1 the cardinality of the latter exceeds that of $\mathbb{N}$. Since $\mathbb{N}$ is countably infinite we obtain that Γ and hence its superset $\mathbb{R}$ is uncountable. ■

Problem 21.20. (**Continuity**). Let $a, b, c, d \in \mathbb{R}$ such that $a < b$ and $c < d$. Let $f :]a, b[\to]c, d[$ be bijective and strictly monotone, i.e., strictly increasing or decreasing. Prove that both f and f^{-1} are continuous.

Hint: Use ε–δ continuity.

The solution to this problem is not given here. ■

Problem 21.21. (**Continuity**).

Let $(X, \mathfrak{U})$ and $(Y, \mathfrak{V})$ be topological spaces and $f : X \to Y$. Prove that f is continuous if and only if the preimage $f^{-1}(F)$ of any closed $F \subseteq Y$ is closed in X.

Solution to #21.21:

PROOF: We utilize for both directions prop.13.1 on p.407: f is continuous if and only if the preimage $f^{-1}(V)$ of any open $V \subseteq Y$ is open in X, and also the fact that

$$\text{if } B \in Y \text{ then } \left(f^{-1}(B^{\complement})\right)^{\complement} = f^{-1}(B) \qquad .(\star)$$

This equation follows from 8.12 on p.240: the complement of the inverse image is the inverse image of the complement.

(a) PROOF of f is continuous $\Rightarrow$ f^{-1} (closed) = closed:

Let $F \in Y$ be closed. We must prove that $f^{-1}(F)$ is closed. Complements of closed sets are open and vice versa, hence $F^{\complement}$ is open, hence continuity of f implies that $f^{-1}(F^{\complement})$ is open. Thus $\left(f^{-1}(F^{\complement})\right)^{\complement}$ is closed According to $(\star)$ that closed set equals $f^{-1}(F)$. This proves **(a)**.

(b) PROOF of f^{-1} (closed) = closed $\Rightarrow$ f is continuous:

It suffices to show that $f^{-1}(V)$ is open for any open $V \subseteq Y$ because this implies the continuity of f. So let $V \subseteq Y$ be open. Then $V^{\complement}$ is closed, hence $f^{-1}(V^{\complement})$ is closed by our assumption, hence its complement $\left(f^{-1}(V^{\complement})\right)^{\complement}$ is open According to $(\star)$ that open set equals $f^{-1}(V)$. This proves **(b)**. ■

Problem 21.22. (**Continuity**). Let $X := \mathbb{R}$, equipped with the standard Euclidean metric $d(x, x') = |x - x'|$. Let $f_n : \mathbb{R} \to \mathbb{R}; \quad x \mapsto \frac{\sin(n^2 x)}{n}$.

- **(a)** Prove that $f_n(\cdot) \overset{uc}{\to} 0$ on $\mathbb{R}$.
- **(b)** Prove that $f'n(x)$ does not converge pointwise on $\mathbb{R}$. Hint: What exactly must you show?

Solution to #21.22:

PROOF of **(a)**: Let $\delta > 0$. It follows from $\lim\limits_{n\to\infty} \frac{1}{n} = 0$ that there exists $N \in \mathbb{N}$ such that $\frac{1}{n} = |\frac{1}{n} - 0| < \delta$ for all $n \geq N$. We conclude from $|\sin t| \leq 1$ for all $t \in \mathbb{R}$ that $|f_n(x) - 0| = \frac{\sin(n^2 x)}{n} \leq \frac{1}{n} < \delta$ for all $n \geq N$ and also for all $x \in X$. This proves that $f_n(\cdot) \overset{uc}{\to} 0$ on X.

PROOF of **(b)**: Let $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Then $f'(x) = \frac{n^2 \cos(n^2 x)}{n} = n \cos(n^2 x)$. Thus $f'_n(0) = n \cos(0) = n$. It follows that $\lim\limits_{n\to\infty} f'_n(0) = \infty$, thus $f'_n(x)$ does converge at $x = 0$. ■

Problem 21.23. (**Compactness**).

Let $(X, \mathfrak{U})$ be a topological space. Prove that X is compact if and only if every family $(F_i)_{i\in I}$ of closed sets has the **finite intersection property** (short: **fip**), i.e., $(F_i)_{i\in I}$ satisfies the following: If every finite selection $F_{i_1}, F_{i_2}, \dots, F_{i_k}$ of members of $(F_i)_i$ has nonempty intersection then $\bigcap\limits_{i\in I} F_i \neq \emptyset$.

Note that an equivalent formulation is the contrapositive: If $\bigcap\limits_{i\in I} F_i = \emptyset$ then there must be finitely many members $F_{i_1}, F_{i_2}, \dots, F_{i_n}$ such that $F_{i_1} \cap \dots \cap F_{i_n} = \emptyset$.

Solution to #21.23:

PROOF: We recall the definition of compactness: Every family $(U_i)_{i\in I}$ of open sets which covers X possesses a finite selection $U_{i_1}, U_{i_2}, \dots, U_{i_k}$ which covers X.

(a) PROOF of X is compact $\Rightarrow$ every family $(F_i)_{i\in I}$ of closed sets has the fip:

We work with the contrapositive: Assume that $\bigcap\limits_{i\in I} F_i = \emptyset$ For each $i \in I$ let $U_i := F_i^{\complement}$. Then each U_i is open as the complement of a closed set. It follows from de Morgan that $\bigcup\limits_{i\in I} U_i = \emptyset^{\complement} = X$, i.e., the family $(U_i)_i$ is an open covering of X. Since X is compact we can extract a finite subcover $U_{i_1}, \dots, U_{i_k}$: $U_{i_1} \cup \dots \cup U_{i_k} = X$. We obtain from de Morgan that $F_{i_1} \cup \dots \cup F_{i_k} = X^{\complement} = \emptyset$. It follows that $(F_i)_{i\in I}$ has the fip, and this proves **(a)**. Since $\bigcup\limits_{i\in I} U_i = X$ de Morgan yields $\bigcup\limits_{i\in I} U_i = X$

(b) PROOF that if every family $(F_i)_{i\in I}$ of closed sets has the fip then X is compact:

Let $(U_i)_{i\in I}$ be a family of open sets which covers X. We must extract a finite covering. For each $i \in I$ let $F_i := U_i^{\complement}$. Then F_i is closed as the complement of an open set. Since $\bigcup\limits_{i\in I} U_i = X$ de Morgan yields $\bigcap\limits_{i\in I} F_i = X^{\complement} = \emptyset$. But the family of closed sets $(F_i)_{i\in I}$ possesses the fip, thus there exist finitely many members $F_{i_1}, \dots, F_{i_k}$ such that $F_{i_1} \cap \dots \cap F_{i_k} = \emptyset$. From de Morgan we obtain $U_{i_1} \cup \dots \cup U_{i_k} = \emptyset^{\complement} = X$. We have extracted a finite covering from the family $(U_i)_{i\in I}$, thus X is compact. This proves **(b)**. ■

21.5 Non–essential Definitions

A non–essential Definition is one which the student need not remember in the sense that it will not occur in a quiz or exam. It is possible though that such a definition will be referenced in later parts of the document.

For example, the non–essential term "abelian group", defined in Definition 3.2 on p.52, is referenced in example 3.4 which can be found on p.53.

This chapter contains the beginnings of a list of non–essential definitions. It is broken down into several lists on a chapter by chapter basis.

> Generally speaking, any definition that is given in an optional (starred) chapter or in a construct other than a proper definition, e.g., in a footnote or a remark, is considered non–essential, and it very likely will not be included in those lists

Ch.3: The Axiomatic Method:
binary operation (not essential until its formal definition in ch.5) • semigroup, monoid • abelian group (but remember commutative group) • linear function, additivity, homogeneity • commutative ring with unit

Ch.5: Functions and Relations:
linear/total ordering • linearly/totally ordered set • inverse relation • maps to operator (but remember assignment operator)

22 Other Appendices

22.1 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		

22.2 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.

Notation 22.1. a) If two subsets A and B of a space Ω are disjoint, i.e., $A \cap B = \emptyset$, then we often write $A \biguplus B$ 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 $]0, +\infty[$, $\mathbb{R}_{\geq 0}$ 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 $\mathbb{Z}_{\geq 0}$ for $\mathbb{N} \biguplus \{0\}$. $\mathbb{Z}_{\geq 0}$ is the B/G notation. It is very unusual but also very intuitive. □

Definition 22.1. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. We call that sequence **nondecreasing** or **increasing** if $x_n \leq x_{n+1}$ 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 **nonincreasing** or **decreasing** if $x_n \geq x_{n+1}$ for all n.

We call it **strictly decreasing** if $x_n > x_{n+1}$ for all $n \in \mathbb{N}$. □

References

[1] Heinz Bauer. Approximation and abstract boundaries. American Mathematics Monthly Vol.85 No.8 p.632-647, 1978.

[2] Matthias Beck and Ross Geoghegan. The Art of Proof. Springer, 1st edition, 2010.

[3] Regina Brewster and Ross Geoghegan. Business Calculus, a text for Math 220, Spring 2014. Binghamton University, 11th edition, 2014.

[4] Matthew Brin and Gerald Marchesi. Linear Algebra, a text for Math 304, Spring 2016. Binghamton University, 13th edition, 2016.

[5] John Bryant and Penelope Kirby. Course Notes for MAD 2104 Discrete Mathematics I. Florida State University.

[6] G. Chartrand, A. Polimeni, and Ping Zhang. Mathematical Proofs: A Transition to Advanced Mathematics. Pearson, 4th edition, 2018.

[7] Gustave Choquet. Lectures on Analysis, Vol. 2: Representation Theory. Benjamin, New York, 1st edition, 1969.

[8] John P. D'Angelo and Douglas B. West. Mathematical Thinking: Problem-Solving and Proofs - Pearson Mode Pearson, 2nd edition, 2018.

[9] Richard M. Dudley. Real Analysis and Probability. Cambridge University Press, Cambridge, New York, 2nd edition, 2002.

[10] S.H. Friedberg, A.J. Insel, and L.E. Spence. Linear Algebra. Pearson, 5th edition, 2018.

[11] Norman B. Haaser and Joseph A. Sullivan. Real Analysis. publisher = Van Nostrand, year = 1971, 1st edition.

[12] A.N. Kolmogorov and S.V. Fomin. Introductory Real Analysis. Dover, Mineola, 1st edition, 1975.

[13] James R. Munkres. Topology. Prentice-Hal, 1st edition, 2000.

[14] Walter Rudin. Principles of Mathematical Analysis. McGraw-Hill, New York, San Francisco, Toronto, London, 2nd edition, 1964.

[15] James Stewart. Single Variable Calculus. Thomson Brooks Cole, 7th edition, 2012.

[16] Boto Von Querenburg. Mengentheoretische Topologie. Springer, Berlin, Heidelberg, New York, 1st edition, 1973.

[17] 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