

# Math 330 - Additional Material

Student edition with proofs

Michael Fochler  
Department of Mathematics  
Binghamton University

Last update: April 14, 2018

Version: **2018-04-14**

©Michael Fochler 2005-2018

# Contents

<b>1</b>	<b>Before You Start</b>	<b>7</b>
1.1	About This Document . . . . .	7
1.2	How to Properly Write a Proof ( <b>Study this!</b> ) . . . . .	9
<b>2</b>	<b>Preliminaries about Sets, Numbers and Functions</b>	<b>11</b>
2.1	Sets and Basic Set Operations . . . . .	11
2.2	Numbers . . . . .	15
2.3	A First Look at Functions . . . . .	19
2.4	Proof by Induction . . . . .	21
2.5	Exercises for Ch.2 . . . . .	23
	2.5.1 Exercises for Sets . . . . .	23
	2.5.2 Exercises for Proofs by Induction . . . . .	24
2.6	Addenda to Ch.2 . . . . .	25
<b>3</b>	<b>The Axiomatic Method</b>	<b>27</b>
3.1	Semigroups and Groups . . . . .	27
3.2	Commutative Rings and Integral Domains . . . . .	32
3.3	Addenda to Ch.3 . . . . .	36
<b>4</b>	<b>Logic (Skim this!)</b>	<b>39</b>
4.1	Statements and Statement Functions . . . . .	39
4.2	Logic Operations and their Truth Tables . . . . .	42
	4.2.1 Overview of Logical Operators . . . . .	42
	4.2.2 Negation and Conjunction, Truth Tables and Tautologies ( <b>Understand this!</b> ) . . . . .	43
	4.2.3 Biconditional and Logical Equivalence Operators – Part 1 . . . . .	48
	4.2.4 Inclusive and Exclusive Or . . . . .	49
	4.2.5 Arrow and Implication Operators . . . . .	50
	4.2.6 Biconditional and Logical Equivalence Operators – Part 2 ( <b>Understand this!</b> ) . . . . .	56
	4.2.7 More Examples of Tautologies and Contradictions ( <b>Understand this!</b> ) . . . . .	57
4.3	Statement Equivalences ( <b>Understand this!</b> ) . . . . .	58
4.4	The Connection Between Formulas for Statements and for Sets ( <b>Understand this!</b> ) . . . . .	61
4.5	Quantifiers for Statement Functions . . . . .	63
	4.5.1 Quantifiers for One–Variable Statement Functions . . . . .	64
	4.5.2 Quantifiers for Two–Variable Statement Functions . . . . .	65
	4.5.3 Quantifiers for Statement Functions of more than Two Variables . . . . .	67
	4.5.4 Quantifiers and Negation ( <b>Understand this!</b> ) . . . . .	68
4.6	Proofs ( <b>Understand this!</b> ) . . . . .	69
	4.6.1 Building Blocks of Mathematical Theories . . . . .	69
	4.6.2 Rules of Inference . . . . .	72
	4.6.3 An Example of a Direct Proof . . . . .	74
	4.6.4 Invalid Proofs Due to Faulty Arguments . . . . .	76
4.7	Categorization of Proofs ( <b>Understand this!</b> ) . . . . .	77
	4.7.1 Trivial Proofs . . . . .	78
	4.7.2 Vacuous Proofs . . . . .	78
	4.7.3 Direct Proofs . . . . .	78

4.7.4	Proof by Contrapositive . . . . .	78
4.7.5	Proof by Contradiction (Indirect Proof) . . . . .	79
4.7.6	Proof by Cases . . . . .	79
<b>5</b>	<b>Functions and Relations</b>	<b>81</b>
5.1	Cartesian Products and Relations . . . . .	81
5.2	Functions (Mappings) and Families . . . . .	85
5.2.1	Some Preliminary Observations about Functions . . . . .	85
5.2.2	Definition of a Function and Some Basic Properties . . . . .	87
5.2.3	Examples of Functions . . . . .	90
5.2.4	Injective, Surjective and Bijective functions . . . . .	93
5.2.5	Binary Operations and Restrictions and Extensions of Functions . . . . .	98
5.2.6	Operations on Real Valued Functions . . . . .	99
5.2.7	Sequences, Families and Functions as Families . . . . .	101
5.3	Exercises for Ch.5 . . . . .	105
5.3.1	Exercises for Functions and Relations . . . . .	105
5.4	Addenda to Ch.5 . . . . .	107
<b>6</b>	<b>More on Sets (Understand this!)</b>	<b>109</b>
6.1	More on Set Operations (Study this!) . . . . .	109
6.2	Cartesian Products of More Than Two Sets . . . . .	113
6.3	Rings and Algebras of Sets ( $\star$ ) . . . . .	116
6.4	Exercises for Ch.6 . . . . .	117
<b>7</b>	<b>Sets and Functions, Direct and Indirect Images (Study this!)</b>	<b>118</b>
7.1	Direct Images and Indirect Images (Preimages) of a Function . . . . .	118
7.2	Indicator Functions . . . . .	125
7.3	Exercises for Ch.7 . . . . .	127
<b>8</b>	<b>Some Miscellaneous Topics</b>	<b>130</b>
8.1	Cardinality - Alternate Approach to Beck/Geoghegan (Study this!) . . . . .	130
8.1.1	Cardinality as a Partial Ordering . . . . .	141
8.2	Exercises for Ch.8 . . . . .	143
8.3	Addenda to Ch.8 . . . . .	144
<b>9</b>	<b>The Completeness of the Real Numbers System (Study this!)</b>	<b>145</b>
9.1	Minima, Maxima, Infima and Suprema . . . . .	145
9.2	Convergence and Continuity in $\mathbb{R}$ . . . . .	150
9.3	Geometric Series . . . . .	158
9.4	Limit Inferior and Limit Superior . . . . .	159
9.5	Sequences of Sets and Indicator functions and their $\liminf$ and $\limsup$ . . . . .	167
9.6	Sequences that Enumerate Parts of $\mathbb{Q}$ . . . . .	171
9.7	Exercises for Ch.9 . . . . .	172
<b>10</b>	<b>Vectors and Vector spaces (Understand this!)</b>	<b>174</b>
10.1	$\mathbb{R}^n$ : Euclidean Space . . . . .	174
10.1.1	$n$ -Dimensional Vectors . . . . .	174
10.1.2	Addition and Scalar Multiplication for $n$ -Dimensional Vectors . . . . .	175

10.1.3	Length of $n$ -Dimensional Vectors and the Euclidean Norm	176
10.2	General Vector Spaces	178
10.2.1	Vector spaces: Definition and Examples	178
10.2.2	Normed Vector Spaces ( <b>Study this!</b> )	188
10.2.3	The Inequalities of Young, Hoelder, and Minkowski ( $\star$ )	196
10.3	Exercises for Ch.10	202
<b>11</b>	<b>Metric Spaces</b>	<b>204</b>
11.1	The Topology of Metric Spaces	204
11.1.1	Definition and Examples of Metric Spaces	204
11.1.2	Measuring the Distance of Real Valued Functions	207
11.1.3	Neighborhoods and Open Sets	210
11.1.4	Convergence	213
11.1.5	Abstract Topological spaces	214
11.1.6	Bases and Neighborhood Bases	217
11.1.7	Metric and Topological Subspaces	219
11.1.8	Bounded Sets and Bounded Functions	221
11.1.9	Contact Points and Closed Sets	222
11.1.10	Completeness in Metric Spaces	226
11.2	Continuity ( <b>Study this!</b> )	233
11.2.1	Definition and Characterizations of Continuous Functions	233
11.2.2	Continuity of Constants and Sums and Products	238
11.2.3	Continuity of Polynomials ( <b>Understand this!</b> )	239
11.2.4	Uniform Continuity	241
11.2.5	Continuity of Linear Functions ( <b>Understand this!</b> )	242
11.3	Function Sequences and Infinite Series	244
11.3.1	Convergence of Function Sequences ( <b>Study this!</b> )	244
11.3.2	Infinite Series	248
11.4	Exercises for Ch.11	259
11.4.1	Exercises for Ch.11.1	259
11.4.2	Exercises for Ch.11.2	262
<b>12</b>	<b>Compactness</b>	<b>263</b>
12.1	Introduction: Closed and Bounded sets in Euclidean space	263
12.2	Four Definitions of Compactness	265
12.3	$\varepsilon$ -Nets and Total Boundedness	265
12.4	Sequence Compactness	268
12.5	Open Coverings and the Heine–Borel Theorem	271
12.6	Continuous Functions and Compact Spaces	274
12.7	Exercises for Ch.12	277
<b>13</b>	<b>Applications of Zorn’s Lemma</b>	<b>278</b>
13.1	More on Partially Ordered Sets ( <b>Study this!</b> )	278
13.2	Existence of Bases in Vector Spaces ( <b>Study this!</b> )	279
13.3	The Cardinal Numbers are a totally ordered set	280
13.4	Extensions of Linear Functions in Arbitrary Vector Spaces	281
13.5	The Hahn-Banach Extension Theorem ( $\star$ )	282

13.5.1	Sublinear Functionals	282
13.5.2	The Hahn-Banach extension theorem and its Proof	282
13.6	Convexity (★)	286
13.7	Exercises for Ch.13	288
<b>14</b>	<b>Algebraic Structures</b>	<b>289</b>
14.1	More on Groups (★)	289
14.2	More on Commutative Rings and Integral Domains	290
<b>15</b>	<b>Construction of the Number Systems</b>	<b>292</b>
15.1	The Peano Axioms (★)	292
15.2	Constructing the Integers from $\mathbb{N}_0$	293
15.3	Constructing the Rational Numbers from $\mathbb{Z}$	294
15.4	Constructing the Real Numbers via Dedekind Cuts	295
15.5	Constructing the Real Numbers via Cauchy Sequences	299
<b>16</b>	<b>Measure Theory (★)</b>	<b>301</b>
16.1	Basic Definitions	301
16.2	Sequences of Sets – limsup and liminf	304
16.3	Conditional Expectations as Generalized Averages	306
<b>17</b>	<b>Appendix: Addenda to Beck/Geoghegan’s “The Art of Proof”</b>	<b>307</b>
17.1	AoP Ch.1: Integers	307
17.1.1	Ch.1.1 – Axioms	307
17.2	AoP Ch.2: Natural Numbers and Induction	308
17.2.1	AoP Ch.2.2 (Ordering the Integers)	309
17.2.2	AoP Ch.2.3 (Induction)	310
17.2.3	Bounded Sets in $\mathbb{Z}$	310
17.2.4	Exercises for Ch.17.2	312
17.3	AoP Ch.3: Some Points of Logic	312
17.4	AoP Ch.4: Recursion	313
17.4.1	Generalization of Some Recursive Definitions	313
17.5	AoP Ch.5: Underlying Notions in Set Theory	314
17.6	AoP Ch.6: Equivalence Relations and Modular Arithmetic	315
17.6.1	Equivalence Relations	315
17.6.2	The Division Algorithm	315
17.6.3	The Integers Modulo $n$	317
17.6.4	Prime Numbers	317
17.6.5	Exercises for Ch.17.6	318
17.7	AoP Ch.7: Arithmetic in Base Ten	319
17.7.1	Base-Ten Representation of Integers	319
17.7.2	The Addition Algorithm for Two Nonnegative Numbers (Base 10)	321
17.8	AoP Ch.8: Real Numbers	321
17.8.1	Axioms	322
17.9	AoP Ch.9: Embedding $\mathbb{Z}$ in $\mathbb{R}$	322
17.10	AoP Ch.10: Limits and Other Consequences of Completeness	322
17.11	AoP Ch.11: Rational and Irrational Numbers	324

17.12AoP Ch.12: Decimal Expansions . . . . .	324
17.13AoP Ch.13: Cardinality . . . . .	328
17.14Exercises for Ch.17 . . . . .	328
<b>18 Other Appendices</b>	<b>329</b>
18.1 Greek Letters . . . . .	329
18.2 Notation . . . . .	329
<b>References</b>	<b>330</b>
<b>List of Symbols</b>	<b>331</b>
<b>Index</b>	<b>334</b>

# 1 Before You Start

## Errors detected by Math 330 students, Spring 2017:

Date	Topic
2017-01-26	Error in def.5.2. <b>Incorrect version:</b> A relation is symmetric if $x_1 R x_2$ implies $x_1 R x_2$ for all $x_1, x_2 \in X$ . <b>Correct version:</b> 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 <b>Brad Whistance</b> .

## History of Updates:

Date	Topic
2017-12-21	Moved all additions from separate "Addenda to ch.nn" sections to their proper place in preparation for the Spring 2018 semester.
2017-01-30	Mainly reformatting: increased use of tables. Significant additions to ch.17 (Appendix: Addenda to Beck/Geoghegan's "The Art of Proof").
2017-01-10	Many updates during Fall 2016. Significant streamlining and reorg of ch.8 (liminf, limsup, ...) through ch.13 (Zorn's Lemma).

## 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:

- Here you find background material on topics that cannot be found in sufficient detail or not at all in the textbook [1] 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.
- This document covers material which is beyond the scope of [1] B/G such as
  - material on lim inf and lim sup
  - convergence, continuity and compactness in metric spaces
  - applications of Zorn's lemma

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

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

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

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

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

For example, the reference to ch.9 (The Completeness of the Real Numbers System) has been tagged with (**Understand this!**), the first subchapter ch.9.1 (Minima, Maxima, Infima and Suprema) has been tagged with (**Study this!**), the subchapter ch.9.5 (Sequences of Sets and Indicator functions and their  $\liminf$  and  $\limsup$ ) has no tag.

**Notation Alert:**

All directives apply to the entire subtree and a lower level directive overrides the “parent directives”. Example: the “Understand this!” directive of subsection 11.2.3: Continuity of Polynomials overrides the “Study this!” directive of subsection 11.2 on Continuity. Accordingly, when you do not see any comment, back up in the table of contents: first to the parent entry, then to its parent entry ... until you find one.

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

b. “**Understand this**” directive: When you read “Understand this”, you should know the definitions, propositions and theorems without worrying about proofs. It is quite likely that this kind of material will be referenced in more important sections of this document.

c. “**Skim this**” directive: When you read “Skim this”, you should understand how the material is structured. You may find it easier to do some of your homework. A good example is chapter 4 on logic. It has been marked as “Skim this” but some of the later subchapters override this directive as “Understand this”. Understanding the material of those subchapters will be problematic unless you can find your way around in the preceding “Skim this” parts.

d. “**Skip this**” directive: When you read “Skip this”, you need not worry about the content.

e. “(\*)” directive: Chapters marked “(\*)” are optional.

**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 so that I can move along at a fast pace except for some topics that I will focus on in detail. For this document it means that you should do the “Study this” and “Understand this” material as was indicated above.

□

**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 the one given in the [1] 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. 7.2 on p.120, make use of " $\Leftrightarrow$ " to show that two sets are equal. You should study this technique but, as you will hear me say many times in class, I recommend that you abstain from using " $\Leftrightarrow$ " between statements in your proofs. 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.  $\square$

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

## 1.2 How to Properly Write a Proof (Study this!)

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

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

**Method a.**

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

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:

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

Right side:

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

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.**



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

**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  $-2 = 2$ :

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

We are done. ■

Second proof that  $-2 = 2$ :

$$\begin{aligned} -2 &= 2 \quad (\text{want to prove}) \\ (-2)^2 &= 2^2 \quad (\text{multiply both sides from the right w. } 0) \\ 4 &= 4 \quad ((B/G \text{ ax.1.2 (additive neutral element)}) \end{aligned}$$

We are done. ■

Now you know why you must never use "method **c**" for a proof. <sup>1</sup>

<sup>1</sup> You will learn later in this chapter about injective functions which guarantee that if you do an operation (apply a

## 2 Preliminaries about Sets, Numbers and Functions

### 2.1 Sets and Basic Set Operations

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

An entire book can be filled with a mathematically precise theory of sets.<sup>2</sup> 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 **setbuilder notation**, i.e., instructions that specify what belongs to the set, is used instead.

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

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

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

**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?  $\square$

**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  $X$ :

$$\begin{aligned} &\{x : x \text{ is an integer and } 18 \leq x \leq 24\}, \\ &\{\alpha : \alpha \text{ is an integer and } 18 \leq \alpha \leq 24\}, \\ &\{\mathfrak{J} : \mathfrak{J} \text{ is an integer and } 18 \leq \mathfrak{J} \leq 24\} \end{aligned}$$

---

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  $x^2 = y^2$ .

<sup>2</sup> 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.

$$(2.1) \quad A := \{B : B \text{ is a set and } B \notin B\}$$

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. □

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

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

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

As you will learn later, **c1+c2+c3** means that " $\sim$ " is an equivalence relation (see def.5.3 on p.82) and **d:** means that the empty set is a subset (see the next definition) of any other set. □

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

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

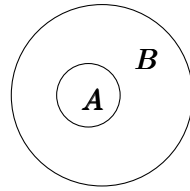


Figure 2.1: Set inclusion:  $A \subseteq B$ ,  $B \supseteq A$

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

“iff” is a short for “if and only if”:  $P$  iff  $Q$  for two statements  $P$  and  $Q$  means that if  $P$  is valid then  $Q$  is valid and vice versa.<sup>3</sup>

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

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

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

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

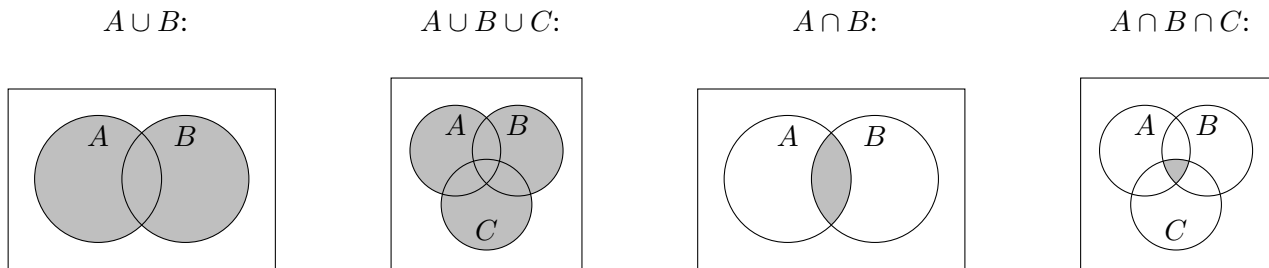


Figure 2.2: Union and intersection of sets

<sup>3</sup> A formal definition of “if and only if” will be given in def.4.10 on p.48 where we will also introduce the symbolic notation  $P \Leftrightarrow Q$ . Informally speaking, a statement is something that is either true or false.

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

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  $B$ :

$$(2.2) \quad A \setminus B := \{x \in A : x \notin B\}$$

The **symmetric difference**  $A \Delta B$  (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  $B$ :

$$(2.3) \quad A \Delta B := (A \cup B) \setminus (A \cap B) \quad \square$$

**Definition 2.6** (Universal set). Usually there always is a big set  $\Omega$  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.  $\square$

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

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

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

$$(2.4) \quad A^c := \complement A := \Omega \setminus A = \{\omega \in \Omega : \omega \notin A\} \quad \square$$

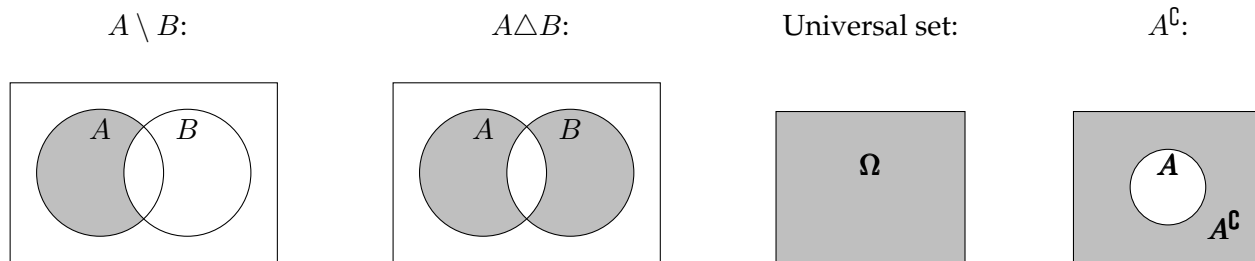


Figure 2.3: Difference, symmetric difference, universal set, complement

**Remark 2.4.** Note that for any kind of universal set  $\Omega$  it is true that

$$(2.5) \quad \Omega^c = \emptyset, \quad \emptyset^c = \Omega. \quad \square$$

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

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

<sup>4</sup>  $\mathbb{R}$  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.2.2 ("Classification of numbers") on p.15).

the  $\delta$ -neighborhood <sup>5</sup> 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

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

Draw some Venn diagrams to visualize the following formulas.

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

- (2.7a)  $A \cup \emptyset = A; \quad A \cap \emptyset = \emptyset$   
(2.7b)  $A \cup \Omega = \Omega; \quad A \cap \Omega = A$   
(2.7c)  $A \cup A^c = \Omega; \quad A \cap A^c = \emptyset$   
(2.7d)  $A \Delta B = (A \setminus B) \uplus (B \setminus A)$   
(2.7e)  $A \setminus A = \emptyset$   
(2.7f)  $A \Delta \emptyset = A; \quad A \Delta A = \emptyset$   
(2.7g)  $X \Delta A = X \setminus A$   
(2.7h)  $A \cup B = (A \Delta B) \uplus (A \cap B)$   
(2.7i)  $A \cap B = (A \cup B) \setminus (A \Delta B)$   
(2.7j)  $A \Delta B = \emptyset$  if and only if  $B = A$

The proof is left as exercise 2.1.

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

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

of a set  $\Omega$  is the set of all its subsets.  $\square$

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

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

## 2.2 Numbers

We start with an informal classification of numbers. It is not meant to be mathematically exact. For this consult, e.g., [1] B/G (Beck/Geoghegan).

**Definition 2.9** (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 written as a finite string of digits, possibly preceded by a minus sign, **integers**. In particular we demand the absence of a decimal point. Examples of integers are

$$(2.8) \quad 3, -29, 0, 3 \cdot 10^6, -1, 12345678901234567890, -2018.$$

<sup>5</sup> Neighborhoods of a point will be discussed in the chapter on the topology of  $\mathbb{R}^n$  (see (11.6) on p.210) In short, the  $\delta$ -neighborhood of  $a$  is the set of all points with distance less than  $\delta$  from  $a$ .

(Note that  $3 \cdot 10^6 = 3000000$  is a finite string of digits). We write  $\mathbb{Z}$  for the set of all integers.

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

A **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

$$(2.9) \quad 3.0, -29.0, 0.0, -0.75, \bar{3}, 2.74\bar{9}, \pi = 3.141592\dots, -34.56.$$

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.,  $\bar{3} = 0.3333333333\dots$  and  $1.234\bar{567} = 1.234567567567\dots$

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  $1.234\bar{567}$  from above has period length 3 and the number  $0.14\bar{5}$  has period length 2.

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

$$d_1 d_2 d_3 \dots d_k 00 \dots,$$

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 (2.8) are equal in value to the first three decimals of (2.9). The mathematician says that we **identify** the integer  $\pm d_1 d_2 d_3 \dots d_k$  and the decimal numeral  $\pm d_1 d_2 d_3 \dots d_k .0$ , i.e., we do not distinguish those expressions and we consider them as equal, just as we would "six" and "half a dozen".  $\square$

*We are ready to give an informal definition of the most important kind of numbers.*

**Definition 2.10** (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 def.2.9 that integers, in particular natural numbers, are real numbers. Thus we have the following set relations:

$$(2.10) \quad \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{R}. \quad \square$$

**Definition 2.11** (Rational numbers). A number that is an integer or can be written as a fraction is called a **rational number** and we write  $\mathbb{Q}$  for the set of all rational numbers. Examples of rational numbers are

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

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:

$$(2.11) \quad 0.\bar{3} = \bar{3} = 0.3333333333\dots = \frac{1}{3} = \frac{-1}{-3} = \frac{2}{6},$$



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

$$(2.12) \quad -4 = -4.000 = -3.\bar{9} = -\frac{12}{3} = \frac{4}{-1} = \frac{-4}{1} = \frac{12}{-3} = -\frac{400}{100}.$$

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  $d \in \mathbb{N}$ .  $\square$

**Definition 2.12** (Irrational numbers). There are numbers which cannot be expressed as integers or fractions or numbers with a finite amount of decimals to the right of the decimal point. Examples are  $\sqrt{2}$  and  $\pi$ . Those **irrational numbers** (that is their official name) fill the gaps between the rational numbers. In fact, there is a simple way (but not easy to prove) of characterizing irrational numbers: Rational numbers are those that can be expressed with at most finitely many digits to the right of the decimal point, including repeating decimals. You can find the underlying theory and exact proofs in B/G ch.12. Irrational numbers must then be those with infinitely many decimal digits without any continually repeating patterns.  $\square$

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

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

It follows that  $x = 144/990$ , and that is certainly a fraction.  $\square$

**Remark 2.6.** We will see in ch.8.1 (Cardinality - Alternate Approach to Beck/Geoghegan) on p.130 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.  $\square$

**Definition 2.13** (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**.  
 $\mathbb{R} := \{\text{all integers or decimal numbers with finitely or infinitely many decimal digits}\}$  denotes the set of all **real numbers**.  
 $\mathbb{R} \setminus \mathbb{Q} = \{\text{all real numbers which cannot be written as fractions of integers}\}$  denotes the set of all **irrational numbers**. There is no special symbol for irrational numbers. Example:  $\sqrt{2}$  and  $\pi$  are irrational.

The following are 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 non-negative integers,  
 $\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$  denotes the set of all non-negative real numbers,  
 $\mathbb{R}^+ := \mathbb{R}_{> 0} := \{x \in \mathbb{R} : x > 0\}$  denotes the set of all positive real numbers,  
 $\mathbb{R}^* := \mathbb{R}_{\neq 0} := \{x \in \mathbb{R} : x \neq 0\}$ .  $\square$

**Definition 2.14** (Intervals of real numbers). We use the following notation for intervals of real numbers  $a, b$ :

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

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

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

We further define the following intervals of “infinite length”:

$$(2.13) \quad \begin{aligned} ]-\infty, a] &:= \{x \in \mathbb{R} : x \leq a\}, & ]-\infty, a[ &:= \{x \in \mathbb{R} : x < a\}, \\ ]a, \infty[ &:= \{x \in \mathbb{R} : x > a\}, & [a, \infty[ &:= \{x \in \mathbb{R} : x \geq a\}, & [-\infty, \infty[ &:= \mathbb{R} \end{aligned}$$

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

**Definition 2.15** (Absolute value). For a real number  $x$  we define its **absolute value** as

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

$\square$

**Example 2.4.**  $|3| = 3; |-3| = 3; |-5.38| = 5.38$ .  $\square$

**Remark 2.7.** For any real number  $x$  we have

$$(2.14) \quad \sqrt{x^2} = |x|. \quad \square$$

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

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

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  $-3$ .  $\square$

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

$$(2.15) \quad \text{Triangle Inequality : } |\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$$

*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 non-negative, both are negative or where one of each is positive and negative. ■*

## 2.3 A First Look at Functions

The material on functions presented in this section will be discussed again and in greater detail in chapter 5 (Functions and Relations) on p.81.

**Note 2.1** (Motivation for a good function definition). 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  $(x, \sqrt{x})$  and sometimes it just refers to the function value  $\sqrt{x}$  for a “fixed but arbitrary” number  $x$ . In case of the function  $f_2(x, y)$ : Sometimes  $f_2(x, y)$  means the entire graph, i.e., the entire collection of pairs  $((x, y), \ln(x - y))$  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.85 which precedes the mathematically precise definition of a function (def.5.6 on p.88). 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 get to a usable definition of a function there are several things to consider. In the following  $f_1(x)$  and  $f_2(x, y)$  again denote the functions  $f_1(x) = \sqrt{x}$  and  $f_2(x, y) = \ln(x - y)$ .

- a. The source of all allowable arguments ( $x$ -values in case of  $f_1(x)$  and  $(x, y)$ -values in case of  $f_2(x, y)$ ) will be called the **domain** of the function. The domain is explicitly specified as part of a function definition and it may be chosen for whatever reason to be only a subset of all arguments for which the function value is a valid expression. In case of the function  $f_1(x)$  this means that the domain must be restricted to a subset of the interval  $[0, \infty[$  because the square root of a negative number cannot be taken. In case of the function  $f_2(x, y)$  this means that the domain must be restricted to a subset of  $\{ (x, y) : x, y \in \mathbb{R} \text{ and } x - y > 0 \}$  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 superset of the set of all function values for which the argument belongs to the domain of the function. In case of the function  $f_1(x)$  this means that we are OK if the codomain is a superset of the interval  $[0, \infty[$ . Such a set is big enough because square roots are never negative. It is OK to specify the interval  $] - 3.5, \infty[$  or even the set  $\mathbb{R}$  of all real numbers as the codomain. In case of the function  $f_2(x, y)$  this means that we are OK if the codomain contains  $\mathbb{R}$ . Not that it would make a lot of sense, but the set  $\mathbb{R} \cup \{ \text{all inhabitants of Chicago} \}$  also is an acceptable choice for the codomain.
- c. A function  $y = f(x)$  is not necessarily something that maps (assigns) numbers or pairs of numbers to numbers but domain and codomain can be a very different kind of animal. In chapter ?? on logic you will learn about statement functions  $A(x)$  which assign arguments  $x$  from some set  $\mathcal{U}$ , called the universe of discourse, to statements  $A(x)$ , i.e., sentences that are either true or false.

- d. Considering all that was said so far you can think of the graph of a function  $f(x)$  with domain  $D$  and codomain  $C$  (see earlier in this note) as the set

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

Alternatively you 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 you do not write about functions in general but about specific functions such as  $f_1(x) = \sqrt{x}$  and  $f_2(x, y) = \ln(x - y)$ . We further write

$$f : C \longrightarrow D$$

as a short way of saying that the function  $f(x)$  has domain  $C$  and codomain  $D$ .

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

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

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

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

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

**Definition 2.16** (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

$$(2.16) \quad f : X \rightarrow Y, \quad x \mapsto f(x).$$

We read " $a \mapsto b$ " as " $a$  is assigned to  $b$ " or " $a$  maps to  $b$ " and refer to  $\mapsto$  as the **maps to operator** or **assignment operator**. The **graph** of such a function is the collection of pairs

$$(2.17) \quad \Gamma_f := \{(x, f(x)) : x \in X\}. \quad \square$$

**Remark 2.8.** The name given to the argument variable is irrelevant. Let  $f_1, f_2, X, Y, U, V$  be as defined in d of note 2.1. The function

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

is identical to the function  $f_1$ . The function

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

is identical to the function  $f_2$  and so is the function

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

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

We next define cartesian products of sets.<sup>6</sup> 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 2.17** (Preliminary definition: cartesian product). Let  $X$  and  $Y$  be two sets The set

$$(2.18) \quad X \times Y := \{(x, y) : x \in X, y \in Y\}$$

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

$$(2.19) \quad X_1 \times X_2 \cdots \times X_n := \{(x_1, x_2, \dots, x_n) : x_j \in X_j \text{ for each } j = 1, 2, \dots, n\}$$

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

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

**Example 2.5.** The graph  $\Gamma_f$  of a function with domain  $X$  and codomain  $Y$  (see def.2.17) is a subset of the cartesian product  $X \times Y$ .  $\square$

**Example 2.6.** The domains given in **a** and **d** of note 2.1 are subsets of the cartesian product

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

.

## 2.4 Proof by Induction

Let  $k_0 \in \mathbb{Z}$  and let

$$A_{k_0} := \{k \in \mathbb{Z} : k \geq k_0\} = k_0, k_0 + 1, k_0 + 2, \dots$$

be the set of all integers at least as big as  $k_0$ . We give the following property of  $A_{k_0}$  without proof.<sup>7</sup>

<sup>6</sup> See ch.5.1 (Cartesian products and relations) on p.81 for the real thing and examples.

<sup>7</sup> This is the more or less the induction axiom in ch.2 of [1] B/G: The Art of Proof. The first two chapters of this book use the “axiomatic” method to develop the mathematical structure of integers and natural numbers and give an exact proof of the induction principle. See also Ax.2.3 in def.15.1 in ch.15.1 (The Peano Axioms) on p.292.

**Theorem 2.1** (Induction Axiom Property). Let  $A \subseteq \mathbb{Z}$  such that

- a.  $k_0 \in A$ ,
- b.  $k \in A$  implies  $k + 1 \in A$ .

Then  $A \supseteq A_{k_0}$ .  $\square$

**Theorem 2.2** (Principle of proof by 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.** 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:**

A proof of this theorem will not be given here. Its validity follows from the Induction Axiom Property theorem above.  $\blacksquare$

Either you have been explained this principle before and say “Oh, that – what’s the big deal?”, or you will be mighty confused. You will see how this principle works by walking through a proof by induction of the triangle inequality for  $n$  real numbers.

**Proposition 2.3** (The Triangle Inequality for  $n$  real numbers). Let  $n \in \mathbb{N}$  such that  $n \geq 2$ . Let  $a_1, a_2, \dots, a_n \in \mathbb{R}$ . Then

$$(2.20) \quad |a_1 + a_2 + \dots + a_n| \leq |a_1| + |a_2| + \dots + |a_n|$$

**PROOF:**

Note that this proposition generalizes prop. 2.15 on p.316 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 2.20 was already shown (see (2.15) on p.18).

**B. Induction step:** Let us assume that 2.20 is true for some  $k \geq 2$ . This is our induction assumption. We now must prove the inequality for  $k + 1$  terms  $a_1, a_2, \dots, a_k, a_{k+1} \in \mathbb{R}$ . We abbreviate

$$A := a_1 + a_2 + \dots + a_k; \quad B := |a_1| + |a_2| + \dots + |a_k|$$

then our induction assumption for  $k$  numbers is that  $|A| \leq B$ . We know from (2.15) 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

$$(2.21) \quad |A + a_{k+1}| \leq |A| + |a_{k+1}| \leq B + |a_{k+1}|$$

In other words,

$$(2.22) \quad |(a_1 + a_2 + \dots + a_k) + a_{k+1}| \leq B + |a_{k+1}| = (|a_1| + |a_2| + \dots + |a_k|) + |a_{k+1}|,$$

and this is (2.20) 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  $k \geq k_0 = 2$ .  $\blacksquare$

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 2.9** (Why induction works). But how can we from all of the above conclude that the triangle inequality works for all  $n \in \mathbb{N}$  such that  $n \geq k_0 = 2$ ? This is how:

- Step 1: We know that the statement is true for  $k_0 = 2$  because that was proved in A.
- Step 2: But according to B, 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 B, 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 B, 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 B, 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  $n \geq k_0$ .  $\square$

**Remark 2.10.** The above theorem 2.2 is often stated for the special case  $k_0 = 1$ <sup>8</sup>

It is easy to prove the general case from that special case with the help of the following simple proposition  $\square$

**Proposition 2.4.** Let  $i, j, n \in \mathbb{N}$ . Then  $n + i \in A_i \Leftrightarrow n + j \in A_j$ .

*PROOF:* We have

$$n + i \in A_i \Rightarrow n + i \geq i \Rightarrow n \geq 0 \Rightarrow n + j \geq j \Rightarrow n + j \in A_j.$$

We reverse the roles of  $i$  and  $j$  and obtain  $n + j \in A_j \Rightarrow n + i \in A_i$  This finishes the proof.  $\blacksquare$

**Corollary 2.1.** Let  $k_0, n \in \mathbb{Z}$ . Then  $n \in A_{k_0}$  if and only if  $n - k_0 + 1 \in \mathbb{N}$

*PROOF:* Let  $m := n - 1$ . We obtain from prop.2.4

$$n = m + 1 \in \mathbb{N} \Leftrightarrow n + k_0 - 1 = m + k_0 \in A_{k_0}.$$

This proves the corollary.  $\blacksquare$

## 2.5 Exercises for Ch.2

### 2.5.1 Exercises for Sets

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

**Exercise 2.2.** 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.  $\{y\} \subseteq X$   $\square$

<sup>8</sup> [1] 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.

For the subsequent exercises refer to example 5.4 for the preliminary definition of the size  $|A|$  of a set  $A$  and to def.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.81

**Exercise 2.3.** Find the size of each of the following sets:

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

**Exercise 2.4.** Let  $X = \{x, y, \{x\}, \{x, y\}\}$  and  $Y = \{x, \{y\}\}$ . True or false?

$$\begin{array}{llll} \text{a. } x \in X \cap Y & \text{c. } x \in X \cup Y & \text{e. } x \in X \setminus Y & \text{g. } x \in X \Delta Y \\ \text{b. } \{y\} \in X \cap Y & \text{d. } \{y\} \in X \cup Y & \text{f. } \{y\} \in X \setminus Y & \text{h. } \{y\} \in X \Delta Y \quad \square \end{array}$$

**Exercise 2.5.** Let  $X = \{1, 2, 3, 4\}$  and let  $Y = \{x, y\}$ .

$$\begin{array}{llll} \text{a. What is } X \times Y? & \text{c. What is } |X \times Y|? & \text{e. Is } (x, 3) \in X \times Y? & \text{g. Is } 3 \cdot x \in X \times Y? \\ \text{b. What is } Y \times X? & \text{d. What is } |X \times Y|? & \text{f. Is } (x, 3) \in Y \times X? & \text{h. Is } 2 \cdot y \in Y \times X? \quad \square \end{array}$$

**Exercise 2.6.** Let  $X = \{8\}$ . What is  $2^{(2^X)}$ ?

**Exercise 2.7.** Let  $A = \{1, \{1, 2\}, 2, 3, 4\}$  and  $B = \{\{2, 3\}, 3, \{4\}, 5\}$ . Compute the following.

$$\text{a. } A \cap B \quad \text{b. } A \cup B \quad \text{c. } A \setminus B \quad \text{d. } B \setminus A \quad \text{e. } A \Delta B \quad \square$$

## 2.5.2 Exercises for Proofs by Induction

**Exercise 2.8.** Let  $n \in \mathbb{N}$ . Then  $n^2 + n$  is even, i.e., this expression is an integer multiple of 2.

**Solution:**

The base case ( $n_0 = 1$ ) holds because  $1^2 + 1 = 2$ , and this is an even number.

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

$$(2.23) \quad \text{Induction assumption: } k^2 + k \text{ is even, i.e., } k^2 + k = 2j \text{ for some suitable } j \in \mathbb{N}.$$

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

$$(k+1)^2 + k+1 = k^2 + 2k + 1 + k + 1 = (k^2 + k) + 2(k+1) \stackrel{(2.23)}{=} 2j + 2(k+1).$$

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

**Exercise 2.9.** Use the result from exercise 2.8 above to prove by induction that  $n^3 + 5n$  is an integer multiple of 6 for all  $n \in \mathbb{N}$ .

**Solution:**

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

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

$$(2.24)$$

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

We must show that there exists  $j' \in \mathbb{Z}$  such that  $(k+1)^3 + 5(k+1) = 6j'$ . We know from exercise 2.8 that  $3(k^2 + k) = 3 \cdot 2 \cdot i$  for a suitable  $i \in \mathbb{Z}$ , hence

$$\begin{aligned} (k+1)^3 + 5(k+1) &= k^3 + 3k^2 + 3k + 1 + 5k + 5 = (k^3 + 5k) + 3(k^2 + k) + 6 \\ &= (k^3 + 5k) + 6i + 6 \stackrel{(2.24)}{=} 6(j + i + 1). \end{aligned}$$

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



## 2.6 Addenda to Ch.2

**Definition 2.18** (Partition). Let  $\Omega$  be a set and  $\mathfrak{A} \subseteq 2^\Omega$ . We call  $\mathfrak{A}$  a **partition** or a **partitioning** of  $\Omega$  if

- a.  $A \cap B = \emptyset$  for any two  $A, B \in \mathfrak{A}$ , i.e.  $\mathfrak{A}$  consists of mutually disjoint subsets of  $\Omega$  (see def.2.4),
- b.  $\Omega = \bigcup [A : A \in \mathfrak{A}]$ .  $\square$

**Example 2.7.**

- 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.2.18 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  $[1, \infty[$ .  $\square$

**Remark 2.11.** 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 [11] 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  $f : ]a, b[ \rightarrow \mathbb{R}$  be a function which is continuous on  $]a, b[$ . Then

- a.  $f$  is integrable, i.e., for any  $\alpha, \beta \in \mathbb{R}$  such that  $a < \alpha < \beta < b$  the **definite integral**  $\int_\alpha^\beta f(u)du$  exists. For a definition of integrability see section 4.1 of [11] Stewart, J: Single Variable Calculus.
- b. Integration is “linear”, i.e., it is additive:  $\int_\alpha^\beta (f(u) + g(u))du = \int_\alpha^\beta f(u)du + \int_\alpha^\beta g(u)du$ ,  
and you also can “pull out” constant  $\lambda \in \mathbb{R}$ :  $\int_\alpha^\beta \lambda f(u)du = \lambda \int_\alpha^\beta f(u)du$ .
- c. Integration is “monotonic”:  
If  $f(x) \leq g(x)$  for all  $\alpha \leq x \leq \beta$  then  $\int_\alpha^\beta (f(u))du \leq \int_\alpha^\beta g(u)du$ .
- d.  $f$  has an **antiderivative** : There exists a function  $F : ]a, b[ \rightarrow \mathbb{R}$  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  $F(\beta) - F(\alpha) = \int_\alpha^\beta f(u)du$  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  $\int f(x)dx = F(x) + C$  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.,  $G = F$ .  $\square$

### 3 The Axiomatic Method

The purpose of this chapter is to familiarize you 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 you have the following to work with:

- a. *Axioms: mathematical statements that you may use unquestioningly.*
- 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.*
- c. *Propositions, theorems and lemmata: mathematical statements that you may use because they were previously proven.*

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 allow the axiomatic definition of the integers, as can be found in the first chapter of [1] Beck/Geoghegan: *The Art of Proof*.

#### 3.1 Semigroups and Groups

**Definition 3.1** (Semigroups and monoids). 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  $u := s \diamond t \in S$ .<sup>9</sup> The pair  $(G, \diamond)$  is called a **semigroup** if the operation  $\diamond$  satisfies

$$(3.1) \quad \text{associativity: } (s \diamond t) \diamond u = s \diamond (t \diamond u) \text{ for all } s, t, u \in S.$$

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

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

is called a **monoid**.

We can write  $S$  instead of  $S, \diamond$  if it is clear which binary operation on  $S$  is represented by  $\diamond$ .  $\square$

**Example 3.1.**  $(\mathbb{Z}, +)$  (the integers with addition) and  $(\mathbb{Z}, \cdot)$  (the integers with multiplication) are monoids: Both  $+$  and  $\cdot$  are associative and addition has zero, multiplication has 1 as neutral element.

This is also true for the other number systems:  $(\mathbb{N}, +)$  and  $(\mathbb{N}, \cdot)$  (natural numbers),  $(\mathbb{Q}, +)$  and  $(\mathbb{Q}, \cdot)$  (rational numbers),  $(\mathbb{R}, +)$  and  $(\mathbb{R}, \cdot)$  (real numbers), all are monoids.  $\square$

**Example 3.2.** If you do not know from linear algebra or ch.10.2 on p.178 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?)  $\square$

---

<sup>9</sup> In other words, we have a function  $\diamond : S \times S \rightarrow S$ ,  $(s, t) \mapsto \diamond(s, t) := s \diamond t$  in the sense of def.5.6 on p.88. or def.2.16 on p.20.

The next example is so important that we state it as a proposition. You should refresh your memory about function composition (chain rule in calculus!) to understand it.

**Proposition 3.1.** Let  $A$  be a nonempty set and let  $S := \{f : f \text{ is a function } A \rightarrow A\}$ .<sup>10</sup>

We define a binary operation  $\circ$  on  $S$  as follows.  $(f, g) \mapsto g \circ f$  assigns to two functions  $f, g : A \rightarrow A$  the function<sup>11</sup>

$$g \circ f : A \rightarrow A; \quad x \mapsto g \circ f(x) := g(f(x)).$$

$(S, \circ)$  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

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

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

We now prove the existence of a neutral element. Let  $id_A : A \rightarrow A; x \mapsto x$  be the function which does nothing with its arguments.<sup>12</sup> We have

$$(id_A \circ f)(x) = id_A(f(x)) = f(x) = f(id_A(x)) = (f \circ id_A)(x)$$

for all  $x \in A$ . It follows that the three assignments  $x \mapsto (id_A \circ f)(x)$ ,  $x \mapsto (f \circ id_A)(x)$ , 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. ■

**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

$$(3.3) \quad s \diamond e = e \diamond s = s$$

$$(3.4) \quad s \diamond e' = e' \diamond s = s$$

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

PROOF: We have

$$e \stackrel{(14.10)}{=} e' \diamond e \stackrel{(14.9)}{=} e'.$$

Here we applied (14.10) with  $s = e$  and then (14.9) with  $s = e'$ . ■

<sup>10</sup> 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  $g(x) = 7x + 5e^x$ .

<sup>11</sup> See def.5.7 (Function composition) on p.90. Example: If  $f(x) = 3x^2$  and  $g(x) = 7x + 5e^x$  then  $g \circ f(x) = g(f(x)) = g(3x^2) = 21x^2 + 5e^{3x^2}$ .

<sup>12</sup>  $id_A$  is called the **identity function** or just the **identity** on  $A$ .

**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. Note that 0 is a neutral element for  $(\mathbb{Z}_{\geq 0}, \diamond)$ , because  $n \diamond 0 = |n| = n$  for all  $n \in \mathbb{Z}_{\geq 0}$ .  $\square$

**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

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

Then  $(G, \diamond)$  is called a **group**.

Assume moreover that the operation  $\diamond$  satisfies

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

Then  $G$  is called a **commutative group** or **abelian group**.<sup>13</sup> We write  $G$  instead of  $(G, \diamond)$  if it is clear which binary operation on  $G$  is represented by  $\diamond$ .  $\square$

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

- |  |                         |
|--|-------------------------|
| a. If $g, h \in G$ then $g \diamond h \in G$   | <i>binary operation</i> |
| b. If $g, h, k \in G$ then $(g \diamond h) \diamond k = g \diamond (h \diamond k)$             | <i>associativity</i>    |
| c. There exists $e \in G$ such that $g \diamond e = e \diamond g = g$ for all $g \in G$        | <i>neutral element</i>  |
| d. For each $g \in G$ there exists $g' \in G$ such that<br>$g \diamond g' = g' \diamond g = e$ | <i>inverse element</i>  |
| $G$ is a <b>commutative group (abelian group)</b> if moreover                                  |                         |
| e. If $g, h \in G$ then $g \diamond h = h \diamond g$  | <i>commutativity</i>    |

**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

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

and this proves uniqueness.  $\blacksquare$

<sup>13</sup> named so after the Norwegian mathematician Niels Henrik Abel who lived in the first half of the 19th century and died at age 26.

**Definition 3.3** (inverse element  $g^{-1}$ ). Usually we write  $g^{-1}$  for the unique element of  $G$  that is associated with the given  $g \in G$  by means of the formula (3.5). We call  $g^{-1}$  the **inverse** element of  $g$ .  $\square$

**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  $k$ . 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  $\mathbb{Z}_{\neq 0}$ .  $\square$

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

$$(3.7) \quad (g^{-1})^{-1} = g,$$

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

**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  $(g^{-1})^{-1}$  of  $g^{-1}$ . We have shown (3.7).

**PROOF** of (3.8): We have

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

This chain of equations proves that  $g^{-1} \diamond h^{-1}$  is an inverse of  $h \diamond g$ . It follows from prop.3.2 that  $g^{-1} \diamond h^{-1}$  is the unique inverse  $(h \diamond g)^{-1}$  of  $h \diamond g$ . We have shown (3.8).  $\blacksquare$

**Example 3.5.** Let  $S := \{f : f \text{ is a function } \mathbb{R} \rightarrow \mathbb{R}\}$  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 def.3.2,  $S$  is a group if and only if each function  $f : \mathbb{R} \rightarrow \mathbb{R}$  possesses an inverse element  $f^{-1} : \mathbb{R} \rightarrow \mathbb{R}$  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

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

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 def.2.16 (preliminary definition of a function) on p.20) in at most one point.<sup>14</sup> 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  $n \in \mathbb{Z}$ .  $\square$

**Proposition 3.3.** Let  $G := \{f : \mathbb{R} \rightarrow \mathbb{R} : f(x) = ax + b \text{ for some } a, b \in \mathbb{R} \text{ where } a \neq 0\}$  be the set of all polynomials of degree 1. These are precisely the functions whose graph is a straight line in the  $x, y$ -plane. 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 of all we must 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 def.3.1 (semigroups and monoids) on p.27). In other words, we must prove 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 prop.3.1 that  $(G, \circ)$  is a monoid. We only have to notice 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 that 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  $f(x) = ax + b$  ( $a \neq 0$ ). We have

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

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

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

We have

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

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.  $\blacksquare$

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

---

<sup>14</sup> 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. 5.10 (horizontal and vertical line tests) on p.95).

**Definition 3.4** (Linear functions on  $\mathbb{R}$ ). A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a **linear function on  $\mathbb{R}$**  if the following is true for all  $x, y, \lambda^{15} \in \mathbb{R}$ :

$$(3.9) \quad f(x + y) = f(x) + f(y) \quad \textbf{(additivity)},$$

$$(3.10) \quad f(\lambda x) = \lambda f(x) \quad \textbf{(homogeneity)}. \quad \square$$

You will learn later about the general definition of a linear function. See def.10.8 (linear mappings) on p.184.

**Theorem 3.3.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ . 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  $x \in \mathbb{R}$ :

$$(3.11) \quad f(1) = f\left(\frac{y-x}{y-x}\right) \stackrel{(3.10)}{=} \frac{f(y-x)}{y-x} \stackrel{(3.9)}{=} \frac{f(y) - f(x)}{y-x}$$

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

Next we observe that  $f(0) = f(2 \cdot 0) \stackrel{(3.10)}{=} 2f(0)$ , hence  $f(0) = 0$ .

We substitute  $y = 0$  in (3.11) 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 3.9 and 3.10 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  $f$ . ■

## 3.2 Commutative Rings and Integral Domains

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 specialize from the outset to commutative rings with unit and integral domains.

**Definition 3.5** (Commutative rings with unit). Let  $R$  be a nonempty set with two binary operations

$$\oplus : (a, b) \mapsto a \oplus b, \text{ called } \textbf{addition}, \quad \text{and} \quad \odot : (a, b) \mapsto a \odot b, \text{ called } \textbf{multiplication},$$

---

<sup>15</sup>  $\lambda$  (pronounced lambda) is the greek version of the letter l. Chapter 18.1 (Greek Letters) on p.329 contains a list of the most commonly used Greek letters.



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 with respect to addition by  $0$  and the inverse element of  $a \in R$  with respect to addition by  $\ominus a$ .
- b.  $(R, \cdot)$  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:

$$(3.12) \quad a \odot (b \oplus c) = (a \odot b) \oplus (a \odot c) \text{ for all } a, b, c \in R.$$

- d.  $1 \neq 0$ .

The triplet  $(R, \oplus, \odot)$  is called a **commutative ring with unit**.

We simply write  $R$  instead of  $(R, \oplus, \odot)$  if it is clear which binary operations on  $R$  are represented by  $\oplus$  and by  $\odot$ .  $\square$

**Remark 3.1.** Recall from thm.14.1 and thm.3.2 that the neutral elements  $0$  and  $1$  and the additive inverse  $\ominus b$  are uniquely determined ( $b \in R$ ).

$\square$

**Notations 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.14.1 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:

$$(3.13) \quad a \ominus b := a \oplus (\ominus b). \quad \square$$

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

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 3.4.** Let  $(R, \oplus, \odot)$  be a nonempty set with two binary operations  $\oplus$  and  $\odot$  which satisfies **a, b, c** of def.3.5, 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  $a \in R$ ,

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

PROOF of **b**:

$$(3.14) \quad a \odot 0 \stackrel{3.2}{=} a(0 \oplus 0) \stackrel{3.12}{=} a \odot 0 \oplus a \odot 0, \text{ hence}$$

$$(3.15) \quad \begin{aligned} 0 &\stackrel{3.5}{=} a \odot 0 \oplus (\ominus(a \odot 0)) \stackrel{3.14}{=} (a \odot 0 \oplus a \odot 0) \oplus (\ominus(a \odot 0)) \\ &\stackrel{3.1}{=} a \odot 0 \oplus (a \odot 0) \oplus (\ominus(a \odot 0)) \stackrel{3.5}{=} a \odot 0 \oplus 0 \stackrel{3.2}{=} a \odot 0. \end{aligned}$$

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

**Proposition 3.5.**

- a. The set  $R := \{0\}$  satisfies conditions **a**, **b**, **c** of def.3.5,
- b. Let  $(R, \oplus, \odot)$  be a nonempty set with two binary operations  $\oplus$  and  $\odot$  which satisfies **a**, **b**, **c** of def.3.5. 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:

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

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

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

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

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

hence 0 is neutral element for  $\oplus$  and the semigroup  $(R, \oplus)$  is a monoid.

**Proof of b, “ $\Rightarrow$ ”** Let  $a \in R$ . Then  $a \times 1 = a$  because 1 is the neutral element for multiplication, and  $a \oplus 0 = a$  because 0 is the neutral element for addition. We obtain

$$\begin{aligned} a &= a \odot 1 && (1 \text{ is the neutral element for } \odot) \\ &= a \odot 0 && (\text{we assumed } 1 = 0) \\ &= 0 && (\text{see prop.3.4}). \end{aligned}$$

We selected an arbitrary item  $a \in R$  and showed that  $a = 0$ . This proves that  $R = \{0\}$ .

**Proof of b, “ $\Leftarrow$ ”** This follows trivially from **a** and the fact that  $R$  does not contain any two different elements. ■

**Definition 3.6** (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  $a \neq 0$ :

$$(3.16) \quad \text{If } a \odot b = a \odot c \text{ then } b = c. \quad \square$$

For an example of a commutative ring with unit which contains zero divisors see ch.14.2 (More on Commutative Rings and Integral Domains) on p.290

**Definition 3.7** (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**.  $\square$

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

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

**Proposition 3.6.** 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 (3.12) 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.,  $b = c$ .  $\blacksquare$

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

*PROOF:* Immediate from prop.3.6.  $\blacksquare$

**Proposition 3.7.** Each of the following algebraic structures is an integral domain:

- $(\mathbb{Z}, +, \cdot)$ : the integers with addition and multiplication,
- $(\mathbb{Q}, +, \cdot)$ : the rational numbers with addition and multiplication,
- $(\mathbb{R}, +, \cdot)$ : the real numbers with addition and multiplication.
- <sup>16</sup>  $(\mathbb{C}, +, \cdot)$ : the complex numbers with addition and multiplication.

<sup>16</sup>Skip this example if you have not learned about complex numbers.

PROOF Will not be given here. ■

We next briefly examine ordered integral domains.

**Definition 3.8** (Ordered integral domains). Let  $(R, \oplus, \odot)$  be an integral domain. Let  $P \subseteq R$  satisfy 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  $R$

We use  $P$  to define on  $R$  an “order relation”  $a < b$  as follows: Let  $a, b \in R$ . We define

$$(3.17) \quad a < b \text{ if and only if } b \ominus a \in P,$$

$$(3.18) \quad a \leq b \text{ if and only if } a < b \text{ or } a = b,$$

$$(3.19) \quad a > b \text{ if and only if } b < a,$$

$$(3.20) \quad a \geq b \text{ if and only if } b \leq a.$$

We say that  $<$  is the **order induced by  $P$** .

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

**Proposition 3.8.** 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 is the subset of all fractions  $\frac{m}{n}$  where both  $m, n$  are non-negative integers and  $n \neq 0$ . The positive cone is the subset of all fractions  $\frac{m}{n}$  where both  $m, n$  are non-negative integers and  $n \neq 0$ .
- 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. <sup>17</sup>

PROOF Will not be given here. ■

### 3.3 Addenda to Ch.3

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 3.9** (Subgroup). Let  $(G, \diamond)$  be a group and  $H \subseteq G$  be nonempty such that

$$(3.21) \quad \text{if } h, h' \in H \text{ then } h \diamond h' \in H,$$

$$(3.22) \quad \text{if } h \in H \text{ then its inverse element } h^{-1} \text{ (in } G!) \text{ belongs to } H.$$

<sup>17</sup> Skip this example if you have not learned about complex numbers.

We call  $(H, \diamond)$  a **subgroup** of  $G$ . We write  $H$  rather than  $(H, \diamond)$  if there is no confusion about the nature of  $\diamond$ .  $\square$

**Proposition 3.9.** *Subgroups are groups.*

**PROOF:**

Let  $(G, \diamond)$  be a group and let  $H$  be a subgroup of  $G$ . We must prove that  $H$  is a monoid (has a neutral element  $e_H$  and satisfies associativity) and that each  $h \in H$  possesses  $h' \in H$  such that  $h \diamond h' = h' \diamond h = e_H$ .  
*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 3.22, to  $H$ . It follows from 3.21 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  $H$ .

We next prove associativity. Let  $h_1, h_2, h_3 \in H$ . Then  $h_1, h_2, h_3 \in H$ . We apply 3.21 four times to obtain that

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

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 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  $h^{-1} \in H$ , 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 a subgroup of  $(\mathbb{R}_{\neq 0}, \cdot)$ ,
- 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 3.10.** Let  $G := \{f : \mathbb{R} \rightarrow \mathbb{R} : f(x) = ax + b \text{ for some } a, b \in \mathbb{R} \text{ where } a \neq 0\}$  be the set of all polynomials of degree 1 with function composition  $g \circ f : x \mapsto g \circ f(x) = g(f(x))$ . Further, let  $H := \{f : \mathbb{R} \rightarrow \mathbb{R} : f(x) = ax \text{ for some nonzero } a \in \mathbb{R}\}$ .

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

**PROOF:**

It was established in prop.3.3 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.32.

**Proposition 3.11.** *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 3.21 and 3.22. 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  $h^{-1} \in H_1 \cap H_2$  and this concludes the proof. ■

## 4 Logic (Skim this!)

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

### 4.1 Statements and Statement Functions

**Note 4.1** (Textual variables). It was mentioned in c of note 2.1 on p.19 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 Joyce 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  $F(\text{Joyce})$  and you will most likely see the expression  $F(\text{Joyce}) =$  "Each morning 'Joyce' writes two pages."  $\square$

**Definition 4.1** (Statements). A **statement** <sup>18</sup> 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**  $\square$

**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).  $\square$

**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!"  $\square$

For the remainder of the entire chapter on logic we define

$$(4.1) \quad \mathcal{S} := \text{the set of all statements}$$

<sup>18</sup> 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.

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

Be sure to understand the material of ch.2.3 (A First Look at Functions) on p.19) 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  $\mathcal{S}$  of all statements. If  $A$  contains  $n$  variables  $x_1, x_2, \dots, x_n$  and if they are **bound**, i.e., assigned to the specific values  $x_1 = x_{10}, x_2 = x_{20}, \dots, x_n = x_{n0}$ , we write  $A(x_{10}, x_{20}, \dots, x_{n0})$  for the resulting statement.

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

There are restrictions for the choice of  $x_1 = x_{10}, x_2 = x_{20}, \dots, x_n = x_{n0}$ : Associated with each variable  $x_j$  in  $A$  is a set  $\mathcal{U}_j$  which we call the **universe of discourse**, in short, **UoD**, for the  $j$ th variable in  $A$ . Each value  $x_{j0}$  ( $j = 1, 2, \dots, n$ ) must be chosen in such a way that  $x_{j0} \in \mathcal{U}_j$ . 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  $\mathcal{S}$  (see (4.1)) just as long as  $x_{j0} \in \mathcal{U}_j$ . In other words we have a function

$$(4.2) \quad A : \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n \rightarrow \mathcal{S}, \quad (x_1, x_2, \dots, x_n) \mapsto A(x_1, x_2, \dots, x_n)$$

in the sense of def. 2.16. with the cartesian product of the UoDs for  $x_1, \dots, x_n$  as domain and  $\mathcal{S}$  as codomain. We call such a function a **statement function**<sup>19</sup> or **predicate**.  $\square$

**Note 4.2** (Relaxed notation for statement functions). You should remember that a statement function is a function in the sense of def.2.16 but 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  $\mathcal{U}_j$  rather than the formal notation

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

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

---

<sup>19</sup> 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 branches 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 (left-most) 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 def.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  $x$ :  $\mathcal{U}_x :=$  all plants and animals in the U.S.,

UoDs for  $y$  and  $z$ :  $\mathcal{U}_y := \mathcal{U}_z :=$  all BU majors in actuarial science.

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

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

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

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

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

*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  $A : \mathbb{R} \rightarrow \mathcal{S}, x \mapsto A(x)$  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 we deal with a dummy variable and we have employed its name consistently in both cases. We have

- a.  $A(\text{Honda}) = "Honda - 4.7$  is an integer" is inadmissible because a car brand is not part of our universe of discourse.
- b. If  $u_0 \in \mathcal{U}$  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  $n \in \mathcal{U}$  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  $n$ .  $\square$

**Example 4.5.** Let  $B(x, y) := "x^2 - y + 2 = 11"$ . Then  $B : \mathbb{R} \times ]1, 100[ \rightarrow \mathcal{S}, (x, y) \mapsto B(x, y)$  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  $B(\cdot, \cdot)$   $\square$

**Example 4.6.** The following are predicates:

- a.  $P := "2x = 27"$  (see example 4.2.b), UoD  $\mathcal{U} := \{x \in \mathbb{R} : x > 10\}$
- b.  $Q := "x^2 + y^2 = 34"$  (example 4.2.c), UoD  $\mathcal{V} := \{(x, y) : x, y \in \mathbb{R} \text{ and } x < y\}$
- c.  $R := "x^2 + y^2 = 34 \text{ and } xy > 100"$ , UoDs are  $\mathcal{W}_x := \mathcal{W}_y := [-50, 25]$ .

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  $30 \notin \mathcal{W}_x$ .

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.  $\square$

**Example 4.7.** Let  $B := "x + 7 = 16 \text{ and } d \text{ is a dog}"$ . Let  $\mathcal{U}_x := \mathbb{N}$  and  $\mathcal{U}_d := \{d : d \text{ is a vegetable or animal}\}$ .

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

Assume for the following that Robby is an animal.

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

## 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**.

<sup>20</sup>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  $C : x \mapsto C(x) := "x + 7 = 16 \text{ and } z \text{ is a dog}"$  and  $D : d \mapsto D(d) := "z + 7 = 16 \text{ and } d \text{ is a dog}"$  and then state that  $z$  is not a variable but a fixed (but unspecified) value.

Here is an overview of the important connectives. <sup>21</sup> Their meaning will be explained subsequently, once we define compound statements and compound statement functions.

<b>negation:</b>	$\neg A$	<b>not</b> $A$
<b>conjunction:</b>	$A \wedge B$	<b>A and</b> $B$
<b>double arrow (biconditional):</b>	$A \leftrightarrow B$	<b>A double arrow</b> $B$
<b>logical equivalence:</b>	$A \Leftrightarrow B$	<b>A if and only if</b> $B$
<b>disjunction (inclusive or):</b>	$A \vee B$	<b>A or</b> $B$
<b>exclusive or:</b>	$A \text{ xor } B$	<b>either A or B, exactly one of A or B</b>
<b>arrow:</b>	$A \rightarrow B$	<b>A arrow B, if A then B</b>
<b>implication:</b>	$A \Rightarrow B$	<b>A implies B, if A then B</b>

**Notations 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 “ $(d+7 = 16) \wedge (F \text{ is a dog})$ ”

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.  $\square$

**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**.  $\square$

**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**.  $\square$

#### 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  $\neg A$ :<sup>22</sup>

A	$\neg A$
F	T
T	F

<sup>21</sup> 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.

<sup>22</sup>The definition of a truth table will be given shortly. See def.4.6 on p.44.

□

**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”  $= A$ .

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:

$A_1 \wedge A_2 \wedge \dots \wedge A_n$  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.  $\square$

**Definition 4.7** (Logically impossible). The statements  $A$  and  $B$  in the truth table of def.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<sup>23</sup> 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

$\square$

**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

$\square$

**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,

<sup>23</sup> 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

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

□

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

**Example 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 <sup>24</sup>  $Q(x) := \neg(U(x) \wedge V(x)) \wedge W(x) \wedge Z(x)$  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  $Q(x) = \text{what-ever} \wedge (W(x) \wedge Z(x))$  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  $x$ . □

<sup>24</sup> 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.

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.  $\square$

**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  $P := \neg(A \wedge \neg A)$  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  $Q := (A \wedge \neg A) \wedge B$  is a contradiction.  $\square$

*PROOF of c and f:*

$P = \neg(A \wedge \neg A)$  (last column) has entries all T, hence  $P$  is a tautology.

$Q = (A \wedge \neg A) \wedge B$  (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** <sup>25</sup> 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.56

<sup>25</sup> [4] Bryant, Kirby Course Notes for MAD 2104 calls this operator the **equivalence operator** but we abstain from that terminology because " $A$  is equivalent to  $B$ " has a different meaning and is written  $A \Leftrightarrow B$ .



#### 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**.  $\square$

**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

$\square$

**Definition 4.12** (Exclusive or). The **exclusive or operator** is represented by the symbol “**xor**”.<sup>26</sup>  $A$  **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**  $B$ :

$A$	$B$	$A$ <b>xor</b> $B$
F	F	F
F	T	T
T	F	T
T	T	F

$\square$

**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**  $y > 0”$

<sup>26</sup> Some documents such as [4] Bryant, Kirby Course Notes for MAD 2104. also use the symbol  $\oplus$ .

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

$A(x)$	$B(y)$	$A(x) \mathbf{xor} B(y)$
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

$A(x)$	$B(y)$	$A(x) \mathbf{xor} B(y)$
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

$A(x)$	$B(x)$	$A(x) \mathbf{xor} B(x)$
F	F	F
T	T	F

This last truth table is remarkable. The truth values for  $A(x) \mathbf{xor} B(x)$  are **false** in each row, hence it is a contradiction as defined in def.4.8 on p.47.  $\square$

**Remark 4.4.** Note that the truth values for  $A \leftrightarrow B$  are the exact opposites of those for  $A \mathbf{xor} B$ :

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

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

**Exercise 4.1.** use that last remark to prove that for any two statements  $A$  and  $B$  the compound statement

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

is a tautology.  $\square$

#### 4.2.5 Arrow and Implication Operators

**Definition 4.13** (Arrow operator). The **arrow operator** <sup>27</sup> 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**.  $\square$

**Definition 4.14** (Implication operator). We say that  $A$  **implies**  $B$  and we write

$$(4.10) \quad A \Rightarrow B$$

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.

<sup>27</sup> [4] Bryant, Kirby Course Notes for MAD 2104 calls this operator the **implication operator** but we abstain from that terminology because " $A$  implies  $B$ " has a different meaning and is written  $A \Rightarrow B$ .

(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:

$A$ implies $B$
if $A$ then $B$
$A$ only if $B$
$B$ if $A$
$B$ whenever $A$
$A$ is sufficient for $B$
$B$ is necessary for $A$

Interpret this as:

The truth of $A$ implies the truth of $B$
if $A$ is true then $B$ is true
$A$ is true only if $B$ is true
$B$ is true if $A$ is true
$B$ is true whenever $A$ is true
The truth of $A$ is sufficient for the truth of $B$
The truth of $B$ is necessary for the truth of $A$

□

**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  $A \rightarrow B$  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  $A \rightarrow C$  yields **false** because only they contain the combination  $A:T, C:F$ . It follows that  $A \rightarrow C$  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 " $\rightarrow$ "). Let  $A, B, C$  be three statements.

Then  $[(A \rightarrow B) \wedge (B \rightarrow C)] \Rightarrow (A \rightarrow C)$ .

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

A	B	C	$A \rightarrow B$	$B \rightarrow C$	$P :=$ $(A \rightarrow B) \wedge (B \rightarrow C)$	$A \rightarrow C$	$P \rightarrow (A \rightarrow C)$
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  $[(A \rightarrow B) \wedge (B \rightarrow C)] \rightarrow (A \rightarrow C)$  contains **true** everywhere and we have proved that this statement is a tautology. ■

**Definition 4.15.** In the context of  $A \rightarrow B$  and  $A \Rightarrow B$  we call  $A$  the **premise** or the **hypothesis**<sup>28</sup> and we call  $B$  the **conclusion**.<sup>29</sup>

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

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.**

- a. The difference between  $A \rightarrow B$  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  $A \rightarrow B$  in case the latter is a tautology as defined in def.4.8 on p.47.
- b. Both  $A \rightarrow B$  and  $A \Rightarrow B$  are interpreted as “if  $A$  then  $B$ ” but we prefer in general to say “ $A$  arrow  $B$ ” for  $A \rightarrow B$  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  $\rightarrow$  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.59. □

**Remark 4.7. What is the connection between the truth tables for  $A \rightarrow B, A \Rightarrow B$  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”<sup>30</sup>

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”

<sup>28</sup> also called the **antecedent**

<sup>29</sup> Another word for conclusion is **consequent** .

<sup>30</sup> A counterexample is the function  $f(x) = |x|$  because it is continuous everywhere but not differentiable at  $x = 0$ .

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

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

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

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

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

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

A	B	$A \rightarrow B$
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”

A	B	$A \rightarrow B$
F	F	T
F	T	T
T	F	??
T	T	T

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

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:  $T \rightarrow F$  is false.

We now have the complete truth table for  $A \rightarrow B$  and it matches the one in def.4.13:

A	B	$A \rightarrow B$
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  $A \rightarrow B$  by demanding that  $A$  and  $B$  be such that  $A \rightarrow B$  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  $A \rightarrow B$  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**  $A \rightarrow B$  and  $A \Rightarrow B$  by means of the truth tables given in def.4.13 and def.4.14 and from there we justified why these operators appropriately model “if  $A$  then  $B$ ”.  $\square$

**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

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

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

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

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

We have seen that some work was involved to show that the “ $A(x) \rightarrow B(y)$ ” 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  $A(x) \rightarrow B(y)$  with an equivalent  $\rightarrow$  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  $\mathcal{U}_x := \mathcal{U}_y := \mathbb{Z}$  be the UoDs for the variables  $x$  and  $y$ .

- Let  $A : \mathcal{U}_x \rightarrow \mathcal{S}$  with  $x \mapsto “x \in \mathbb{N}”$ ,  
 $B : \mathcal{U}_y \rightarrow \mathcal{S}$  with  $y \mapsto “y \in \mathbb{Z}$  and  $y > 0”$ ,  
 $C : \mathcal{U}_x \times \mathcal{U}_y \rightarrow \mathcal{S}$  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

$$(4.12) \quad T : \mathcal{U}_x \times \mathcal{U}_y \rightarrow \mathcal{S} \text{ with } (x, y) \mapsto T(x, y) := [(A(x) \wedge C(x, y)) \rightarrow B(y)] \text{ is a tautology.}$$

Note that

- a. The last arrow in (4.12) is the arrow operator  $\rightarrow$ , not the function assignment operator  $\mapsto$ .
- b. if we can demonstrate that (4.12) is correct then we can replace  $(A(x) \wedge C(x, y)) \rightarrow B(y)$  with  $(A(x) \wedge C(x, y)) \Rightarrow B(y)$ . 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**  $\rightarrow B(y)$  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**:

$A(x)$	$C(x, y)$	$B(y)$	$A(x) \wedge C(x, y)$	$(A(x) \wedge C(x, y)) \rightarrow B(y)$
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 4.12b. 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

$A(x)$	$C(x, y)$	$B(y)$	$A(x) \wedge C(x, y)$	$(A(x) \wedge C(x, y)) \rightarrow B(y)$
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).  $\square$

**Remark 4.8.** Let us compare example 4.16.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 4.16.b allowed us to formulate a truth table in which none of the statements had to explicitly refer to the condition  $x < y$ .

In contrast to this we had to introduce in example 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.  $\square$

We will come back to the subject of proofs in chapter 4.6.1 (Building blocks of mathematical theories) on p.69.

### 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.48.

**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.  $\square$

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:*

$$(4.13) \quad \neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B,$$

$$(4.14) \quad \neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B.$$

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

$$(4.15) \quad \neg(A_1 \wedge A_2 \wedge \dots \wedge A_n) \Leftrightarrow \neg A_1 \vee \neg A_2 \vee \dots \vee \neg A_n,$$

$$(4.16) \quad \neg(A_1 \vee A_2 \vee \dots \vee A_n) \Leftrightarrow \neg A_1 \wedge \neg A_2 \wedge \dots \wedge \neg A_n.$$

*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.  $\blacksquare$

**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  $x$ .  $\square$

#### 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  $B(y) := "y \in \mathbb{Z} \text{ and } y > 0"$ . (see example 4.12 on p. 47).

a. Tautologies:

- $T_0$ ,
- $A_1 := "5 + 7 = 12"$ ,
- $A_2 := "Any \text{ integer is even or odd}"$ ,
- $A_3 := P \vee \neg P$  (Tertium non datur or law of the excluded middle),
- $A_4 := P \vee T_0$ ,
- $A_5 := (P \wedge Q) \vee (P \wedge \neg Q)$ ,
- $A_6 := (P \rightarrow Q) \leftrightarrow (\neg P \vee Q)$  (Implication is logically equivalent to an **or** statement),
- $A_7 := ["x < y" \wedge A(x)] \rightarrow B(y)$  (see 4.16.b on p.54),
- $A_8 := A(x) \leftrightarrow B(x)$  (see 4.16.c).

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

$$(P \rightarrow Q) \Leftrightarrow (\neg P \vee Q), \quad ["x < y" \wedge A(x)] \Rightarrow B(y), \quad A(x) \Leftrightarrow B(x).$$

b. Contradictions:

- $F_0$ ,
- $B_1 := "5 + 7 = 15"$ ,
- $B_2 := "There \text{ are some non-zero numbers } x \text{ such that } x = 2x"$ ,

$B_3 := P \wedge \neg P,$   
 $B_3 := P \wedge F_0,$   
 $B_4 := F_0 \wedge (P \vee \neg P),$   
 $B_5 := [\neg P \vee \neg Q] \wedge [P \wedge Q],$   
 $B_6 := A(x) \text{ xor } B(x)$  (see 4.14.c on p. 49).  $\square$

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:

$P$	$Q$	$P \rightarrow Q$	$\neg P$	$\neg P \vee Q$	$(P \rightarrow Q) \leftrightarrow (\neg P \vee Q)$
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  $R$ .  $\square$

### 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 [4] Bryant, Kirby Course Notes for MAD 2104.

**Theorem 4.4.** Let  $P, Q, R$  be statements.

a. Identity Laws: (4.17)  $P \wedge T_0 \Leftrightarrow P$

(4.18)  $P \vee F_0 \Leftrightarrow P$

b. Domination Laws: (4.19)  $P \vee T_0 \Leftrightarrow T_0$

(4.20)  $P \wedge F_0 \Leftrightarrow F_0$

<i>c. Idempotent Laws:</i>	(4.21)	$P \vee P \Leftrightarrow P$
	(4.22)	$P \wedge P \Leftrightarrow P$
<i>d. Double Negation Law:</i>	(4.23)	$\neg(\neg P) \Leftrightarrow P$
<i>e. Commutative Laws:</i>	(4.24)	$P \vee Q \Leftrightarrow Q \vee P$
	(4.25)	$P \wedge Q \Leftrightarrow Q \wedge P$
<i>f. Associative Laws:</i>	(4.26)	$(P \vee Q) \vee R \Leftrightarrow P \vee (Q \vee R)$
	(4.27)	hence $(P \vee Q) \vee R \Leftrightarrow P \vee Q \vee R$
	(4.28)	$(P \wedge Q) \wedge R \Leftrightarrow P \wedge (Q \wedge R)$
	(4.29)	hence $(P \wedge Q) \wedge R \Leftrightarrow P \wedge Q \wedge R$
<i>g. Distributive Laws:</i>	(4.30)	$P \vee (Q \wedge R) \Leftrightarrow (P \vee Q) \wedge (P \vee R)$
	(4.31)	$P \wedge (Q \vee R) \Leftrightarrow (P \wedge Q) \vee (P \wedge R)$
<i>h. De Morgan's Laws:</i> <sup>31</sup>	(4.32)	$\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$
	(4.33)	$\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$
<i>i. Absorption Laws:</i>	(4.34)	$P \wedge (P \vee Q) \Leftrightarrow P$
	(4.35)	$P \vee (P \wedge Q) \Leftrightarrow P$
<i>j. Implication Law:</i>	(4.36)	$(P \rightarrow Q) \Leftrightarrow (\neg P \vee Q)$
		<i>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.</i>
<i>k. Contrapositive Laws:</i>	(4.37)	$(P \rightarrow Q) \Leftrightarrow (\neg Q \rightarrow \neg P)$
	(4.38)	$(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P)$
<i>l. Tautology:</i>	(4.39)	$(P \vee \neg P) \Leftrightarrow T_0$

---

<sup>31</sup>This is theorem 4.3 (De Morgan's laws for statements).

m. Contradiction: (4.40)  $(P \wedge \neg P) \Leftrightarrow F_0$

n. Equivalence: (4.41)  $(P \rightarrow Q) \wedge (Q \rightarrow P) \Leftrightarrow (P \leftrightarrow Q)$

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

PROOF of **h** (De Morgan's laws):

See theorem 4.3 on p.56.

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  $P \rightarrow Q$  in def.4.13 on p.50 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  $\rightarrow$ ):

We prove (4.37) with the help of the previously given laws **a** through **j**:

$$(P \rightarrow Q) \stackrel{(j)}{\Leftrightarrow} (\neg P \vee Q) \stackrel{(e)}{\Leftrightarrow} (Q \vee \neg P) \stackrel{(d)}{\Leftrightarrow} (\neg(\neg Q) \vee \neg P) \stackrel{(j)}{\Leftrightarrow} (\neg Q \rightarrow \neg P)$$

■

**Example 4.20.** Use the logical equivalences of thm.4.4 to prove that  $\neg(\neg A \wedge (A \wedge B))$  is a tautology.

□

Solution:

$$\begin{aligned} & \neg(\neg A \wedge (A \wedge B)) \\ \Leftrightarrow & \neg(\neg A) \vee \neg(A \wedge B) && \text{De Morgan's Law (4.32)} \\ \Leftrightarrow & A \vee (\neg A \vee \neg B) && \text{De Morgan (4.32) + Double negation (4.23)} \\ \Leftrightarrow & (A \vee \neg A) \vee \neg B && \text{Associative law (4.26)} \\ \Leftrightarrow & T_0 \vee \neg B && \text{Tautology (4.39)} \\ \Leftrightarrow & T_0 && \text{Commutative Law (4.24) + Domination Law (4.19)} \quad \blacksquare \end{aligned}$$

**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  $\neg(A \rightarrow B)$ . We proceed as follows:

$$\neg(A \rightarrow B) \stackrel{(j)}{\Leftrightarrow} \neg(\neg A \vee B) \stackrel{(h)}{\Leftrightarrow} (\neg(\neg A) \wedge \neg B) \stackrel{(d)}{\Leftrightarrow} (A \wedge \neg B)$$

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.  $\square$

#### 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  $A^c$ , 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  $\mathcal{U}$  of which all sets are subsets.

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

$$(4.42) \quad a : \mathcal{U} \rightarrow \mathcal{S}, \quad x \mapsto a(x) =: "x \in A"$$

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

**Example 4.22.** Let  $a \cong A$  and  $b \cong B$ .

We have

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

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.  $x \in A^c$ . Hence  $\neg a \cong A^c$ .

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

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

We expand the table of formulas for statements given in thm 4.4 on p.58 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  $\mathcal{U}$  Let  $p, q, r$  be statement functions and let  $P, Q, R \subseteq \mathcal{U}$  such that  $p \cong P, q \cong Q, r \cong R$ . Then we have the following:

a. Identity:  $(4.43) \quad p \wedge T_0 \Leftrightarrow p \qquad P \cap \mathcal{U} = P$

$(4.44) \quad p \vee F_0 \Leftrightarrow p \qquad P \cup \emptyset = P$

b. Domination:  $(4.45) \quad p \vee T_0 \Leftrightarrow T_0 \qquad P \cup \mathcal{U} = \mathcal{U}$

$(4.46) \quad p \wedge F_0 \Leftrightarrow F_0 \qquad P \cap \emptyset = \emptyset$

*c. Idempotency:*

$$(4.47) \quad p \vee p \Leftrightarrow p \qquad P \cup P = P$$

$$(4.48) \quad p \wedge p \Leftrightarrow p \qquad P \cap P = P$$

*d. Double Negation:*

$$(4.49) \quad \neg(\neg p) \Leftrightarrow p \qquad (P^c)^c = P$$

*e. Commutative:*

$$(4.50) \quad p \vee q \Leftrightarrow q \vee p \qquad P \cup Q = Q \cup P$$

$$(4.51) \quad p \wedge q \Leftrightarrow q \wedge p \qquad P \cap Q = Q \cap P$$

*f. Associative:*

$$(4.52) \quad \begin{aligned} (p \vee q) \vee r & \qquad (P \cup Q) \cup R = P \cup (Q \cup R) \\ \Leftrightarrow p \vee (q \vee r) & \end{aligned}$$

$$(4.53) \quad \begin{aligned} (p \wedge q) \wedge r & \qquad (P \cap Q) \cap R = P \cap (Q \cap R) \\ \Leftrightarrow p \wedge (q \wedge r) & \end{aligned}$$

*g. Distributive:*

$$(4.54) \quad \begin{aligned} p \vee (q \wedge r) & \qquad P \cup (Q \cap R) = (P \cup Q) \cap (P \cup R) \\ \Leftrightarrow (p \vee q) \wedge (p \vee r) & \end{aligned}$$

$$(4.55) \quad \begin{aligned} p \wedge (q \vee r) & \qquad P \cap (Q \cup R) = (P \cap Q) \cup (P \cap R) \\ \Leftrightarrow (p \wedge q) \vee (p \wedge r) & \end{aligned}$$

*h. De Morgan:*

$$(4.56) \quad \neg(p \wedge q) \Leftrightarrow \neg p \vee \neg q \qquad (P \cap Q)^c = P^c \cup Q^c$$

$$(4.57) \quad \neg(p \vee q) \Leftrightarrow \neg p \wedge \neg q \qquad (P \cup Q)^c = P^c \cap Q^c$$

*i. Absorption:*

$$(4.58) \quad p \wedge (p \vee q) \Leftrightarrow p \qquad P \cap (P \cup Q) = P$$

$$(4.59) \quad p \vee (p \wedge q) \Leftrightarrow p \qquad P \cup (P \cap Q) = P$$

*j1. Implication 1:*

$$(4.60) \quad (p \rightarrow q) \Leftrightarrow (\neg p \vee q) \qquad (P \setminus Q)^c = P^c \cup Q$$

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

$$(4.61) \quad p \Rightarrow q$$

*j2. Implication 2:*  $P \setminus Q = \emptyset$ , i.e.,  $P \subseteq Q$

Note that we are not dealing with  $p \rightarrow q$  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:*

$$(4.62) \quad (P \rightarrow Q) \Leftrightarrow (\neg Q \rightarrow \neg P) \quad P^c \cup Q = Q \cup P^c$$

$$(4.63) \quad (P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P) \quad P \subseteq Q \Leftrightarrow Q^c \subseteq P^c$$

*l. Tautology:* (4.64)  $(P \vee \neg P) \Leftrightarrow T_0 \quad P \cup P^c = \mathcal{U}$

*m. Contradiction:* (4.65)  $(P \wedge \neg P) \Leftrightarrow F_0 \quad P \cap P^c = \emptyset$

*n1. Equivalence 1:* (4.66)  $(p \rightarrow q) \wedge (q \rightarrow p) \Leftrightarrow (p \leftrightarrow q)$

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

*n2. Equivalence 2:* (4.67)  $(p \Rightarrow q) \wedge (q \Rightarrow p) \Leftrightarrow (p \Leftrightarrow q)$

$$(P \subseteq Q) \text{ and } (Q \subseteq P) \Leftrightarrow (P = Q)$$

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

PROOF of Equivalence 1:

$$\begin{aligned} (P^c \cup Q) \cap (Q^c \cup P) &= [(P^c \cup Q) \cap Q^c] \cup [(P^c \cup Q) \cap P] \\ &= (P^c \cap Q^c) \cup (Q \cap Q^c) \cup (P^c \cap P) \cup (Q \cap P) \\ &= (P^c \cap Q^c) \cup (Q \cap P) \\ &= \{x : x \text{ neither in } P \text{ nor in } Q \text{ or } x \text{ both in } P, Q\}. \end{aligned}$$

■

## 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 [4] Bryant, Kirby Course Notes for MAD 2104.

### 4.5.1 Quantifiers for One–Variable Statement Functions

**Definition 4.16** (Quantifiers). Let  $A : \mathcal{U} \rightarrow \mathcal{S}$ ,  $x \mapsto A(x)$  be a statement function of a single variable  $x$  with UoD  $\mathcal{U}$  for  $x$ .

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

$$(4.68) \quad \text{“For all } x A(x)\text{”}, \quad \text{written } \forall x A(x).$$

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

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

$$(4.69) \quad \text{“For some } x A(x)\text{”}, \quad \text{written } \exists x A(x).$$

The above is a short for “There exists  $x \in \mathcal{U}$  such that  $A(x)$  is true”.<sup>32</sup> We call the symbol  $\exists$  the **existential quantifier** symbol.

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

$$(4.70) \quad \text{“There exists unique } x \text{ such that } A(x)\text{”}, \quad \text{written } \exists! x A(x).$$

The above is a short for “There exists a unique  $x \in \mathcal{U}$  such that  $A(x)$  is true”.<sup>33</sup> We call the symbol  $\exists!$  the **unique existential quantifier** symbol.  $\square$

**Example 4.23.** Let  $A : [-3, 3] \rightarrow \mathcal{S}$  be the statement function  $x \mapsto “x^2 - 4 = 0”$ .

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

$C = “\text{for all } x \in [-3, 3] \text{ it is true that } x^2 - 4 = 0”$

$D = “\text{there is at least one } x \in [-3, 3] \text{ such that } x^2 - 4 = 0”$

$E = “\text{there is exactly one } x \in [-3, 3] \text{ such that } x^2 - 4 = 0”$

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

**Example 4.24.** Let  $\mathcal{U} := \{ \text{all human beings} \}$  be the UoD for the following three predicates:

$S(x) := “x \text{ is a student at NYU}”$ ,

$C(x) := “x \text{ cheats when taking tests}”$ ,

$H(x) := “x \text{ is honest}”$ ,

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

$A_1 := “\text{All humans are NYU students}”$ ,

$A_2 := “\text{All NYU students cheat on tests}”$ ,

$A_3 := “\text{Any NYU student who cheats on tests is not honest}”$ .

Solution:

$A_1 = \forall x S(x)$ ,

$A_2 = \forall x [S(x) \rightarrow C(x)]$ ,

$A_3 = \forall x [(S(x) \wedge C(x)) \rightarrow \neg H(x)]$ .  $\square$

<sup>32</sup> Equivalently, “ $A(x)$  is true for some  $x \in \mathcal{U}$ ” or “ $A(x)$  is true for at least one  $x \in \mathcal{U}$ ”.

<sup>33</sup> Equivalently, “ $A(x)$  is true for exactly one  $x \in \mathcal{U}$ ”.



**Example 4.25.** We continue example 4.24.

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

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:

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

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

It follows that

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

What a mess! let us drop the “ $(x)$ ” everywhere and the above becomes

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

We have seen in example 4.21 on p.60 that for any two statements  $P$  and  $Q$  the equivalence  $\neg(P \rightarrow Q) \Leftrightarrow (P \wedge \neg Q)$  is true.

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

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

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”.  $\square$

## 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

$$(4.71) \quad B : \mathcal{U}_x \times \mathcal{U}_y \rightarrow \mathcal{S}, \quad x \mapsto B(x, y)$$

It follows that the universes of discourse are  $\mathcal{U}_x$  for  $x$  and  $\mathcal{U}_y$  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  $x \in \mathcal{U}_x$  and for all  $y \in \mathcal{U}_y$  (we have the truth of)  $B(x, y)$ ”,
- b.  $\forall x \exists y B(x, y)$  “for all  $x \in \mathcal{U}_x$  there exists (at least one)  $y \in \mathcal{U}_y$  such that  $B(x, y)$ ”,
- c.  $\exists x \forall y B(x, y)$  “there exists (at least one)  $x \in \mathcal{U}_x$  such that for all  $y \in \mathcal{U}_y$   $B(x, y)$ ”,
- d.  $\exists! x \forall y B(x, y)$  “there exists exactly one  $x \in \mathcal{U}_x$  such that for all  $y \in \mathcal{U}_y$   $B(x, y)$ ”,
- e.  $\exists x \exists y B(x, y)$  “there exists (at least one)  $x \in \mathcal{U}_x$  and (at least one)  $y \in \mathcal{U}_y$  such that  $B(x, y)$ ”.  $\square$

**Example 4.26.** Let  $\mathcal{U}_x := \mathbb{N}$ ,  $\mathcal{U}_y := \mathbb{Z}$  and  $B : \mathcal{U}_x \times \mathcal{U}_y \rightarrow \mathcal{S}$ ,  $(x, y) \mapsto B(x, y) := "x + y = 1"$ . Then

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

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

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

**Proposition 4.1.** Note the following:

$$(4.72) \quad \forall x \forall y B(x, y) \Leftrightarrow \forall y \forall x B(x, y)$$

$$(4.73) \quad \exists x \exists y B(x, y) \Leftrightarrow \exists y \exists x B(x, y)$$

$$(4.74) \quad \forall x \exists y B(x, y) \not\Leftrightarrow \exists y \forall x B(x, y)$$

$$(4.75) \quad \exists y \forall x B(x, y) \Rightarrow \forall x \exists y B(x, y)$$

*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  $y_0 \in \mathcal{U}_y$  such that  $B(x, y_0)$  is true for all  $x \in \mathcal{U}_x$ .

Why does that imply the truth of  $\forall x \exists y B(x, y)$ , i.e., for all  $x \in \mathcal{U}_x$  you can pick some  $y \in \mathcal{U}_y$  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  $x \in \mathcal{U}_x$ . ■

**Remark 4.12.** The last part of the proof of (17.1) 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

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

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  $\mathcal{U}_x := \{ \text{all students at this party} \}$  and  $\mathcal{U}_y := \{ \text{“Linear Algebra”, Discrete Mathematics”, “Multivariable Calculus”, “Ordinary Differential Equations”, “Complex Variables”, “Graph Theory”, “Real Analysis”} \}$ .

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

$$A : \mathcal{U}_x \times \mathcal{U}_y \rightarrow \mathcal{S}, \quad (x, y) \mapsto A(x, y) = “x \text{ studies } y”.$$

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

$$B : \mathcal{U}_y \rightarrow \mathcal{S}, \quad y \mapsto B(y) = “\text{all students at this party study } y”$$

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

$$C : \mathcal{U}_x \rightarrow \mathcal{S}, \quad x \mapsto C(x) = “x \text{ studies exactly one of the courses listed in } \mathcal{U}_y”.$$
  $\square$

### 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

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

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.  $\square$

**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  $f : ]a, b[ \rightarrow \mathbb{R}$  be a function which maps each  $x$  in its domain  $]a, b[$  to a real number  $y = f(x)$ .

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

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

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  $x \in ]a, b[$ .<sup>34</sup>

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  $]a, b[$ .<sup>35</sup>

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

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

<sup>34</sup> See def.11.31 ( $\varepsilon$ - $\delta$  continuity) on p.233.

<sup>35</sup> See def.11.34 (Uniform continuity of functions) on p.241.

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

#### 4.5.4 Quantifiers and Negation (Understand this!)

*Negation of statements involving quantifiers is governed by*

**Theorem 4.6** (De Morgan's laws for quantifiers). *Let  $A$  be a statement function with UoD  $\mathcal{U}$ . Then*

- a.  $\neg(\forall x A(x)) \Leftrightarrow \exists x \neg A(x)$  "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.  $\neg(\exists x A(x)) \Leftrightarrow \forall x \neg A(x)$  "There is **no**  $x$  for which  $A(x)$  is true"  $\Leftrightarrow$  " $A(x)$  is **not** true for all  $x$ "

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

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

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

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  $A : x \mapsto A(x) := [\forall y P(x, y)]$ . We obtain

$$\begin{aligned} [\neg \exists x \forall y P(x, y)] &\Leftrightarrow [\neg \exists x A(x)] \stackrel{\text{(b)}}{\Leftrightarrow} [\forall x \neg A(x)] \Leftrightarrow [\forall x (\neg \forall y P(x, y))] \\ &\stackrel{\text{(a)}}{\Leftrightarrow} [\forall x (\exists y \neg P(x, y))]. \quad \square \end{aligned}$$

**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:

$$\begin{aligned} [\neg \exists x \forall y P(x, y)] &\Leftrightarrow [\neg \exists x (\forall y P(x, y))] \stackrel{\text{(b)}}{\Leftrightarrow} [\forall x \neg (\forall y P(x, y))] \Leftrightarrow [\forall x (\neg \forall y P(x, y))] \\ &\stackrel{\text{(a)}}{\Leftrightarrow} [\forall x (\exists y \neg P(x, y))]. \quad \square \end{aligned}$$

## 4.6 Proofs (Understand this!)

We have informally discussed proofs in examples 4.16 and 4.17 of chapter 4.2.5 (Arrow and Implication Operators) on p.50 and seen in two simple cases how a proof can be done by building a single truth table for an *if... then* statement and showing that it is a tautology. In this chapter we take a deeper look at the concept of “proof”.

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

### 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 [4] Bryant, Kirby Course Notes for MAD 2104.

**Notations 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 or 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**.  $\square$

**Example 4.31.** Let  $A :=$  “all continuous functions are differentiable” (known to be false<sup>36</sup>) 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.  $\square$

#### 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  $a^n + b^n = c^n$ .<sup>37</sup> 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.  $\square$

**Notations 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

<sup>36</sup> see remark 4.7 on p.52 in ch.4.2.5 (Arrow and Implication Operators).

<sup>37</sup> 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 will be discussed in section 4.6.2 (Rules of Inference) on p.72
- b. logical equivalences for statements (see ch.4.2.6 (Biconditional and Logical Equivalence Operators – Part 2) on p.ch.56).

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

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

□

**Notations 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 be 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.39) that what we call a statement, [4] Bryant, Kirby calls a proposition and that we deviate from that approach because mathematics outside logic uses “proposition” to denote a theorem of lesser importance.*

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

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

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

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

**Example 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:<sup>38</sup> (It is postulated that) "if a line segment falling on two line segments makes the interior angles on the same side less than two right angles, the two line segments, if produced indefinitely, meet on that side on which are the angles less than the two right angles".

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

c. Chapters 1 and 2 of [1] Beck/Geoghegan list several axioms which stipulate the existence of a nonempty set called  $\mathbb{Z}$  whose elements are called "integers" which you can "add" and "multiply". Certain algebraic properties such as " $a + b = b + a$ " and " $c \cdot (a + b) = (c \cdot a) + (c \cdot b)$ " 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  $z \in \mathbb{N}$  or  $-z \in \mathbb{N}$  or  $z = 0$ . Finally there is the induction axiom which states that if you create the sequence  $1, 1 + 1, (1 + 1) + 1, \dots$  then you capture all of  $\mathbb{N}$ . This axiom is the basis for the principle of proof by mathematical induction (see thm.2.2 on p. 22).  $\square$

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  $\dots$ " was fair game to define a set. We saw in remark 2.2 (Russell's Antinomy) on p.12 that this led to problems so serious that they caused some of the leading mathematicians of the time to revisit the foundations of mathematics.  $\square$

**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  $z$ . 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.

<sup>38</sup> [https://en.wikipedia.org/wiki/Euclidean\\_geometry#Axioms](https://en.wikipedia.org/wiki/Euclidean_geometry#Axioms)



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!  $\square$

#### 4.6.2 Rules of Inference

**Remark 4.19** (Most important rules of inference). In Notations 4.2 on p.69 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

$$(4.76) \quad \text{the compound statement } [A_1 \wedge A_2 \wedge \dots \wedge A_n] \rightarrow C \text{ is a tautology.}$$

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.  $\square$

**Theorem 4.7.** Let  $P_1, P_2, \dots, P_n$  and  $C$  be statements. Then the statement  $(P_1 \wedge P_2 \wedge \dots \wedge P_n) \rightarrow C$  is a tautology if and only if the following combination of truth values is logically impossible:

$$(4.77) \quad P_j \text{ is } \mathbf{true} \text{ for each } j = 1, 2, \dots, n \text{ and } C \text{ is } \mathbf{false}.$$

PROOF:

Let  $P := (P_1 \wedge P_2 \wedge \dots \wedge P_n)$ . 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  $P \rightarrow C$  is a tautology if and only if the combination of truth values

$$(4.78) \quad P \text{ is } \mathbf{true} \text{ and } C \text{ is } \mathbf{false} \text{ is logically impossible.}$$

In other words, we must prove that  $P \rightarrow C$  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 obvious as row 3 is the only one for which  $P \rightarrow C$  evaluates to **false**.

	$P$	$C$	$P \rightarrow C$
1.	F	F	T
2.	F	T	T
3.	T	F	<b>false</b>
4.	T	T	T

■

**Notations 4.5.** Rules of inference are commonly written in the following form:

	$A_1$
	$A_2$
Your explanations go	$\dots$
into this area	$A_n$
	-----
	$\therefore C$



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

$$\boxed{\begin{array}{c} A_1, A_2, \dots, A_n \\ \hline \therefore C \end{array}}$$

**Theorem 4.8** (The three most important inference rules). *The following lists three inference rules, i.e., those arrow statements are indeed tautologies:*

$$(4.79) \quad \begin{array}{l} \textit{Modus Ponens} \\ \textit{(Law of detachment - the mode that affirms the antecedent (the premise))} \end{array} \quad \begin{array}{c} A \\ A \rightarrow C \\ \hline \therefore C \end{array}$$

$$(4.80) \quad \begin{array}{l} \textit{Modus Tollens} \\ \textit{(The mode that Denies the consequent (the conclusion))} \end{array} \quad \begin{array}{c} \neg C \\ A \rightarrow C \\ \hline \therefore \neg A \end{array}$$

$$(4.81) \quad \begin{array}{l} \textit{Hypothetical syllogism} \end{array} \quad \begin{array}{c} A \rightarrow B \\ B \rightarrow C \\ \hline \therefore A \rightarrow C \end{array}$$

Here is the compact notation:

<i>Modus Ponens</i>	<i>Modus Tollens</i>	<i>Hypothetical syllogism</i>
$A, A \rightarrow C$	$\neg C, A \rightarrow C$	$A \rightarrow B, B \rightarrow C$
$\hline$	$\hline$	$\hline$
$\therefore C$	$\therefore \neg A$	$\therefore A \rightarrow C$

Note that the proof that the hypothetical syllogism is a tautology was given in [thm.4.1](#) on [p.51](#)

PROOF:

■

**Example 4.34.** Here are five more inference rules.

$$(4.82) \quad \begin{array}{l} \textit{Disjunction Introduction} \end{array} \quad \begin{array}{c} A \\ \hline \therefore A \vee B \end{array}$$

$$(4.83) \quad \begin{array}{l} \textit{Conjunction elimination} \end{array} \quad \begin{array}{c} A \wedge B \\ \hline \therefore A \end{array}$$

(4.84)	Disjunctive syllogism	$\frac{A \vee B}{\neg A}$ $\therefore B$
(4.85)	Conjunction introduction	$\frac{A}{B}$ $\therefore A \wedge B$
(4.86)	Constructive dilemma	$\frac{(A \rightarrow B) \wedge (C \rightarrow D)}{A \vee C}$ $\therefore B \vee D$

Compact notation:

<p>Disjunction Introduction</p> $\frac{A}{\therefore A \vee B}$	<p>Conjunction elimination</p> $\frac{A \wedge B}{\therefore A}$	<p>Disjunctive syllogism</p> $\frac{A \vee B, \neg A}{\therefore B}$
<p>Conjunction introduction</p> $\frac{A, B}{\therefore A \wedge B}$	<p>Constructive dilemma</p> $\frac{(A \rightarrow B) \wedge (C \rightarrow D), A \vee C}{\therefore B \vee D}$	

□

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

### 4.6.3 An Example of a Direct Proof

We illustrate in detail a mathematical proof by applying some of 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  $\mathcal{D}$  for the set of all differentiable functions. We now can formulate our theorem.

**Theorem 4.9.** Given the statements

$$\begin{aligned} \mathbf{a}: \quad A &:= "(n \in \mathbb{Z}_{\geq 0}) \wedge (c_0 \in \mathbb{R}) \wedge (c_1 \in \mathbb{R}) \wedge \cdots \wedge (c_n \in \mathbb{R}) \wedge (f(x) = \sum_{j=0}^n c_j x^j)", \\ \mathbf{b}: \quad B &:= "f(x) \in \mathcal{D}", \end{aligned}$$

the following is valid:  $A \Rightarrow B$ .<sup>39</sup>

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.

$$\begin{aligned} \mathbf{a}: \quad Z_j &:= "j \in \mathbb{Z}_{\geq 0}", \\ \mathbf{b}: \quad C_j &:= Z_j \wedge "c_j \in \mathbb{R}", \\ \mathbf{c}: \quad X_j &:= Z_j \wedge "x^j \in \mathcal{D}", \\ \mathbf{d}: \quad D_j &:= Z_j \wedge "c_j x^j \in \mathcal{D}", \\ \mathbf{e}: \quad E &:= Z_n \wedge "f(x) = \sum_{j=0}^n c_j x^j", \\ \mathbf{f}: \quad B &:= "f(x) \in \mathcal{D}" \text{ (repeated for convenient reference)} \end{aligned}$$

We now can write our theorem as

$$(4.87) \quad (Z_n \wedge C_0 \wedge C_1 \wedge \cdots \wedge C_n \wedge E) \rightarrow B.$$

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

$$A_1 := Z_n \wedge "p(x) = x^n", \quad B_1 := X_n.$$

Then Thm-1 states that  $A_1 \Rightarrow B_1$ .<sup>41</sup>

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

We rewrite Thm-2 as an implication. Let

$$\begin{aligned} A_2 &:= "c \in \mathbb{R} \wedge "h(x) \in \mathcal{D} \wedge "g(x) = c \cdot h(x)", \\ B_2 &:= "h(x) \in \mathcal{D}", \end{aligned}$$

Then Thm-2 states that  $A_2 \Rightarrow B_2$ .

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

<sup>39</sup> Note here and for the other theorems the use of  $A_2 \Rightarrow B_2$  instead of  $A_2 \rightarrow B_2$ : We assume that Thm-2 has been proved, i.e.,  $A_2 \rightarrow B_2$  is a tautology.

<sup>40</sup> The expression  $x^j$  in **c** and **d** denotes the function  $x \mapsto x^j$ .

<sup>41</sup> 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  $A_1 \rightarrow B_1$ : Thm-1 has been proved already, i.e., we know that  $A_1 \rightarrow B_1$  is a tautology.

We rewrite Thm-3 as an implication. Let

$$A_3 := "Z_n \wedge "h_1(x) \in \mathcal{D}'' \wedge "h_2(x) \in \mathcal{D}'' \wedge \dots \wedge "h_n(x) \in \mathcal{D}'' \wedge "g(x) = \sum_{j=0}^n h_j(x) "",$$

$$B_3 := "g(x) \in \mathcal{D}'",$$

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

Assertion	Reason
<b>a:</b> $Z_0, Z_1, \dots, Z_n$	evident from $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$
<b>b:</b> $C_0, C_1, \dots, C_n$	part of the premise of $A \rightarrow B$ (see (4.87))
<b>c:</b> $Z_j \rightarrow X_j$ ( $j = 0, 1, \dots, n$ )	Thm-1 with $n := j$
<b>d:</b> $X_j$ ( $j = 0, 1, \dots, n$ )	<b>c</b> and modus ponens
<b>e:</b> $(Z_j \wedge C_j \wedge X_j) \rightarrow D_j$ ( $j = 0, 1, \dots, n$ )	Thm-2 with $c := c_j$ and $h(x) := x^j$
<b>f:</b> $D_j$ ( $j = 0, 1, \dots, n$ )	<b>e</b> and modus ponens
<b>g:</b> $E$	part of the premise of $A \rightarrow B$
<b>h:</b> $(Z_n \wedge D_0 \wedge D_1 \wedge \dots \wedge D_n \wedge E) \rightarrow B$	<b>g</b> and Thm-3 with $h_j(x) := c_j x^j$ and $g(x) := f(x)$
<b>i:</b> $B$	<b>h</b> 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  $Z_j \wedge "c_j x^j \in \mathcal{D}''$  (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.
- 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)

$$\begin{array}{r} P \rightarrow Q \\ Q \\ \hline \therefore P \end{array}$$

(4.89)            Denying the Antecedent  
(indirect proof in the wrong direction)

$$\begin{array}{r} P \rightarrow Q \\ \neg P \\ \hline \therefore \neg Q \end{array}$$

(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: <sup>42</sup>  
“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. ■

### 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

<sup>42</sup> This is example 1.8.3 in ch.3 (Methods of Proofs) of [4] Bryant, Kirby Course Notes for MAD 2104.

### 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 it 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  $P \rightarrow Q$ .

For example, it was mentioned in remark 2.3 (Elements of the empty set and their properties) on p.12 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.74 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 p59). 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  $\Omega$  such that  $A \cap B^c = \emptyset$ . Then  $A \subseteq B$ .

PROOF: We prove the contrapositive instead: If  $A \not\subseteq B$  then  $A \cap B^c \neq \emptyset$ .

So let us assume  $A \not\subseteq B$ . 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  $x \in A \setminus B = A \cap B^c$ , i.e.,  $A \cap B^c \neq \emptyset$ .

We have proved from the negated conclusion  $A \not\subseteq B$  the negated premise  $A \cap B^c \neq \emptyset$ . ■

#### 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:

$$(4.91) \quad m, n \in A \Rightarrow m + n \in A,$$

$$(4.92) \quad m, n \in A \Rightarrow mn \in A,$$

$$(4.93) \quad 0 \notin A,$$

$$(4.94) \quad \text{if } n \in \mathbb{Z} \text{ then either } n \in A \text{ or } -n \in A \text{ or } n = 0.$$

Then  $1 \in A$ .

*Proof by contradiction:* Assume that  $A$  is a set of integers with properties (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

$$(4.95) \quad P_1 \vee P_2 \vee \dots \vee P_n \Leftrightarrow P.$$

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

$$(4.96) \quad (P \Rightarrow Q) \Leftrightarrow [(P_1 \Rightarrow Q) \vee (P_2 \Rightarrow Q) \vee \dots \vee (P_n \Rightarrow Q)].$$

*Proof (outline):* You would do the proof by induction. Prove (4.96) first for  $n = 2$  by expressing  $A \rightarrow B$  as  $\neg A \vee B$  and then building a truth table that compares  $(\neg(P_1 \vee P_2)) \vee Q$  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  $(A \Rightarrow Q) \Leftrightarrow [(P_1 \Rightarrow Q) \vee (P_2 \Rightarrow Q) \vee \dots \vee (P_n \Rightarrow Q)]$  the induction assumption. ■

**Theorem 4.14.** Prove that for any  $x \in \mathbb{R}$  such that  $x \neq 5$  we have

$$(4.97) \quad \frac{x}{x-5} > 0 \Rightarrow [(x < 0) \text{ or } (x > 5)].$$

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

$P := "x/(x - 5) > 0"$ ,<sup>43</sup>  $P_1 := x > 0$  **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). ■

---

<sup>43</sup>  $P := "x/(x - 5) > 0$  **and**  $x \neq 5"$  if you want to be a stickler for precision



## 5 Functions and Relations

We now give an in depth presentation of the material of [ch.2.3](#) (A First Look at Functions).

### 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  $b \in B$ .

Two elements  $(a_1, b_1)$  and  $(a_2, b_2)$  are called **equal** if and only if  $a_1 = a_2$  and  $b_1 = b_2$ . In this case we write  $(a_1, b_1) = (a_2, b_2)$ .

It follows from this definition of equality that the pairs  $(a, b)$  and  $(b, a)$  are different unless  $a = b$ . In other words, the order of  $a$  and  $b$  is important. We express this by saying that the cartesian product consists of **ordered pairs**.

As a shorthand, we abbreviate  $A^2 := A \times A$ .  $\square$

**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:  $(1, 0) \in \mathbb{R}^2$  (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!  $\square$

**Remark 5.1** (Function graphs as subsets of cartesian products). We gave the preliminary definition of a function in [def.2.16](#), [p.20](#) of [ch.2.3](#) (A First Look at Functions).<sup>44</sup> A function

$$f : X \rightarrow Y; \quad y = f(x)$$

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

$$\Gamma_f := \{(x, f(x)) : x \in X\}$$

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.  $\square$

**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.  $\square$

**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

---

<sup>44</sup> The precise definition of a function will be given in [section 5.2](#) on [p.85](#).

- 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  $x_1, x_2 \in X$ .  $\square$

Here are some examples of relations.

**Example 5.2** (Equality as a relation). Given a set  $X$  let  $R := \{(x, x) : x \in X\}$ , i.e.,  $xRy$  if and only if  $x = y$ . This defines a relation on  $X$  which is reflexive, symmetric and transitive.  $\square$

**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.  $\square$

**Example 5.4** (Size of sets as a relation). Let  $X$  be a finite set, i.e., a set which only contains finitely many elements. For  $A \subseteq X$  let  $|A|$  be the number of its elements.<sup>45</sup> Let

$$R := \{(A, B) : A, B \subseteq X \text{ and } |A| = |B|\},$$

i.e.,  $ARB$  if and only if  $A$  and  $B$  possess the same number of elements. This defines a relation on the power set  $2^X$  of  $X$  which is reflexive, symmetric and transitive.  $\square$

**Example 5.5** (Empty relation). Given two sets  $X$  and  $Y$  let  $R := \emptyset$ . This **empty relation** is the only relation on  $(X, Y)$  if  $X$  or  $Y$  is empty.  $\square$

**Example 5.6.** Let  $X := \mathbb{R}^2$  be the  $xy$ -plane. For any point  $\vec{x} = (x_1, x_2)$  in the plane let  $\|\vec{x}\| := \sqrt{x_1^2 + x_2^2}$  be its length<sup>46</sup> and let  $R := \{(\vec{x}, \vec{y}) \in \mathbb{R}^2 \times \mathbb{R}^2 : \|\vec{x}\| = \|\vec{y}\|\}$ . In other words, two points in the plane are related when they have the same length: they are located on a circle which is centered at the origin and has radius  $r = \|\vec{x}\| = \|\vec{y}\|$ . The relation  $R$  is reflexive, symmetric and transitive but not antisymmetric.  $\square$

The relations given in examples 5.2, 5.4, 5.5 and 5.6 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  $X$  which is reflexive, symmetric and transitive. We call such a relation an **equivalence relation** on  $X$ . It is customary to write  $x \sim y$  rather than  $xRy$  (or  $(x, y) \in R$ ) and we say in this case that  $x$  and  $y$  are **equivalent**

Given is an equivalence relation " $\sim$ " on a set  $X$ . For  $x \in X$  let

$$(5.1) \quad [x]_{\sim} := \{y \in X : y \sim x\} = \{\text{all items equivalent to } x\}.$$

We call  $[x]_{\sim}$  the **equivalence class** of  $x$ . If it is clear from the context what equivalence relation is referred to then we simply write  $[x]$  instead of  $[x]_{\sim}$ .  $\square$

**Proposition 5.1** (see [1] B/G prop.6.4). Let " $\sim$ " be an equivalence relation on a nonempty set  $X$  and  $x, y \in X$  Then

- a.  $x \in [x]$ .
- b.  $x \sim y \Leftrightarrow [x] = [y]$ .

<sup>45</sup> For infinite, i.e., not finite sets  $Y$  we define  $|Y| := \infty$ .

<sup>46</sup> See def.10.3 on p.177. of the length or Euclidean norm of a vector in  $n$ -dimensional space.

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). ■

**Proposition 5.2** (see [1] B/G prop.6.5). Let “ $\sim$ ” be an equivalence relation on a nonempty set  $X$  and  $x, y \in X$ . Then either  $[x] = [y]$  or  $[x] \cap [y] = \emptyset$ .

PROOF: It suffices to show that if  $[x] \cap [y] \neq \emptyset$  then  $[x] = [y]$ . So let  $z \in [x] \cap [y]$ . It follows from  $z \in [x]$  that  $z \sim x$  and from  $z \in [y]$  that  $z \sim y$ . Symmetry and transitivity of “ $\sim$ ” yield  $x \sim y$ . It follows from prop.5.1 that  $[x] = [y]$ . ■

**Proposition 5.3** (see [1] B/G prop.6.6 for parts **a** and **b**).

- Let “ $\sim$ ” be an equivalence relation on a nonempty set  $X$  and let  $\mathcal{P}_\sim := \{[x] : x \in X\}$  be the set of all its equivalence classes. Then  $\mathcal{P}_\sim$  is a partition of  $X$ .
- Conversely, let  $\mathcal{P}$  be a partition of  $X$  and define a relation “ $\sim_{\mathcal{P}}$ ” on  $X$  as follows:  $x \sim_{\mathcal{P}} y \Leftrightarrow$  there is  $P \in \mathcal{P}$  such that  $x, y \in P$ . Then  $\sim_{\mathcal{P}}$  is an equivalence relation on  $X$ .
- Let “ $\sim$ ” be an equivalence relation on  $X$ . Let  $\mathcal{P}_\sim$  be the associated partition of its equivalence classes. Let “ $\sim_{\mathcal{P}_\sim}$ ” be the equivalence relation associated with the partition  $\mathcal{P}_\sim$ . Then “ $\sim_{\mathcal{P}_\sim}$ ” = “ $\sim$ ” (i.e., both equivalence relations are equal as subsets of  $X \times X$ ).
- Let  $\mathcal{P}$  be a partition of  $X$ . Let  $\sim_{\mathcal{P}}$  be the associated equivalence relation defined in part **b**.. Let  $\mathcal{P}_{\sim_{\mathcal{P}}}$  be the associated partition of its equivalence classes. Then  $\mathcal{P}_{\sim_{\mathcal{P}}} = \mathcal{P}$ .

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  $\mathcal{P}_\sim$ ! It follows from prop.5.2 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  $[x] \in \mathcal{P}_\sim$ . It follows that  $x \in \bigcup [P : P \in \mathcal{P}_\sim]$  and we obtain that  $\bigcup [P : P \in \mathcal{P}_\sim] = X$ . We have proved that  $\mathcal{P}_\sim$  is a partition of  $X$ .

PROOF of **b**: For the following assume that  $x, y, z \in X$ .

It follows from  $\bigcup [P : P \in \mathcal{P}] = X$  that for each  $x \in X$  there exists  $P \in \mathcal{P}$  such that  $x \in P$ . It further follows from the mutual disjointness of the elements of  $\mathcal{P}$  that there exists exactly one such  $P$  and we are justified to write  $P_x$  for this uniquely defined set  $P$ . In other words, the assignment  $x \mapsto P_x$  defines a function  $P(\cdot) : X \rightarrow \mathcal{P}$ .

Reflexivity:  $x \in P_x$  implies  $x \sim_{\mathcal{P}} x$ .

Symmetry: Let  $x \sim_{\mathcal{P}} y$ . This implies  $P_x = P_y$ , hence  $y \in P_x$ , hence  $y \sim_{\mathcal{P}} x$ .

Transitivity: Let  $x \sim_{\mathcal{P}} y$  and  $y \sim_{\mathcal{P}} z$ .  $x \sim_{\mathcal{P}} y$  implies  $P_x = P_y$  and  $y \sim_{\mathcal{P}} z$  implies  $P_y = P_z$ . It follows that  $P_z = P_x$ , i.e.,  $x \sim_{\mathcal{P}} z$ .

PROOF of **c**:

Let  $x, y \in X$ . Then

$$x \sim y \Leftrightarrow [x] = [y] \Leftrightarrow \text{both } x, y \text{ belong to the same element of } \mathcal{P}_\sim \Leftrightarrow x \sim_{\mathcal{P}_\sim} y.$$

PROOF of *d*:

Let  $P \in \mathcal{P}$  and  $x, y \in X$ . Let  $[x]$  and  $[y]$  be the equivalence classes of  $x$  and  $y$  for " $\sim_{\mathcal{P}}$ ". If  $x \in P$  then

$$y \in P \Leftrightarrow x \sim_{\mathcal{P}} y \Leftrightarrow [x] = [y].$$

It follows that  $P = [x]$ , hence  $P \in \mathcal{P}_{\sim_{\mathcal{P}}}$ . This true for any  $P \in \mathcal{P}$  and it follows that  $\mathcal{P} \subseteq \mathcal{P}_{\sim_{\mathcal{P}}}$ .

Now let  $x \in X$ . If  $y \in X$  then

$$y \in [x] \Leftrightarrow y \sim_{\mathcal{P}} x \Leftrightarrow y \in P_x.$$

It follows that any equivalence class  $[x]$  with respect to " $\sim_{\mathcal{P}}$ " is an element of  $\mathcal{P}$ , hence  $\mathcal{P}_{\sim_{\mathcal{P}}} \subseteq \mathcal{P}$ . We have shown that  $\mathcal{P}_{\sim_{\mathcal{P}}} = \mathcal{P}$  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  $X$ .<sup>47</sup> 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 5.3.** 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  $y \preceq x \Rightarrow y = x$       **antisymmetry**,  
(5.4)  $x \preceq y$  and  $y \preceq z \Rightarrow x \preceq z$       **transitivity**. □

**Remark 5.4** (Partial orderings and reflexivity). Note the following:

**A.** According to the above definition, the following are partial orderings of  $X$ :

1.  $X = \mathbb{R}$  and  $x \preceq y$  if and only if  $x \leq y$ .
2.  $X = 2^\Omega$  for some set  $\Omega$  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  $\Omega$  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  $[x < y \text{ and } y < x]$  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 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  $x \preceq y$ . □

<sup>47</sup> 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** (Inverse Relation). Let  $X$  and  $Y$  be two sets and  $R \subseteq X \times Y$  a relation on  $(X, Y)$ . Let

$$R^{-1} := \{ (y, x) : (x, y) \in R \}.$$

Clearly  $R^{-1}$  is a subset of  $Y \times X$  and hence a relation on  $(Y, X)$ . We call  $R^{-1}$  the **inverse relation** to the relation  $R$ .  $\square$

**Example 5.7.** Let  $R := \{ (x, x^3) : x \in \mathbb{R} \}$ . Then this relation is the graph  $\Gamma_f$  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  $x = f^{-1}(y) = y^{1/3}$ .  $\square$

## 5.2 Functions (Mappings) and Families

### 5.2.1 Some Preliminary Observations about Functions

**Remark 5.5** (A layman's definition of a function). We look at the set  $\mathbb{R}$  of all real numbers <sup>48</sup> 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"):

$$f(0) = \sqrt{4 - 0} = 2, \quad f(2) = f(-2) = \sqrt{4 - 4} = 0, \quad f(2/3) = f(-2/3) = \sqrt{(36 - 4)/9} = \sqrt{30}/3, \dots$$

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.

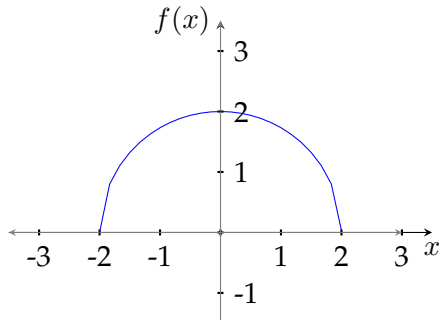
Let us take a closer look at the function  $y = f(x) = \sqrt{4 - x^2}$  and its properties:

- 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., [2] 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  $f(x) = -\pi$ .

---

<sup>48</sup> Real numbers were defined in section "Numbers" on p.15.

- e. On the other hand, there are numbers  $y$  such as  $y = 0$ , which are “hit” more than once by the function:  $f(2) = f(-2) = 0$ .<sup>49</sup>
- 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<sup>50</sup>  $(x, y, z)$  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

$$\Gamma_f := \{(x, f(x)) : x \in D_f\}$$

(see remark 5.1 on p. 81).

□

*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 5.6** (A better definition of a function). We make the following alterations to remark 5.5.

- ▶ We require an upfront specification of the set  $A$  of items that will be allowed as input (arguments) for the function and we require that  $y = f(x)$  makes sense for each  $x \in A$ . Given the function  $y = f(x) = \sqrt{4 - x^2}$  from above this means that  $A$  must be a subset of  $[-2, 2]$ .
- ▶ We require an upfront specification of the set  $B$  of items that will be allowed as output (function values) for the function. This set must be so big that each  $x \in A$  has a function value  $y \in B$ . We do not mind if  $B$  contains redundant  $y$  values. For  $y = f(x) = \sqrt{4 - x^2}$  any superset of the closed interval  $[0, 2]$  will do. We may choose, e.g.,  $B := [0, 2]$  or  $B := [-2, 2\pi]$  or  $B := [0, 4]$  or  $B := \mathbb{R} \cup \{\text{all inhabitants of Chicago}\}$ .

<sup>49</sup>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  $x = \pm\sqrt{4 - y^2}$ .

<sup>50</sup>Skip this example on first reading if you do not know about functions of several variables. You will find information about this in chapter 10 (“Vectors and vector spaces”) on p.174.

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  $\Gamma_f$  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.,

$$(5.5) \quad \Gamma_f := \{(x, f(x)) : x \in A\}.$$

$\Gamma_f$  has the following properties:

- 5a.  $\Gamma_f \subseteq A \times B$ , i.e.,  $\Gamma_f$  is a relation on  $(A, B)$  (see def.5.2 on p.81).
- 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  $\Gamma_f$  and  $\Gamma_g$  do not coincide
6. Conversely, if  $A$  and  $B$  are two nonempty sets, then any relation  $\Gamma$  on  $(A, B)$  which satisfies 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 displeasing 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

*We have seen in remark 5.6 on p. 86 that a function can be thought of equivalently as an assignment rule  $x \mapsto f(x)$  or as a graph. Mathematicians prefer the latter because “assignment rule” is a rather vague*



term (an *undefined term* in the sense of ch. 4.6.1 (Building blocks of mathematical theories) on p.69) whereas “graph” is entirely defined in the language of sets. This chapter starts with the official definition of a function. It then deals with the following concepts: composition of functions, injective, surjective, bijective and inverse functions.

**Definition 5.6** (Mappings (functions)). Given are two arbitrary nonempty sets  $X$  and  $Y$  and a relation  $\Gamma$  on  $(X, Y)$  (see 5.2 on p.81) which satisfies the following:

$$(5.6) \quad \text{for each } x \in X \text{ there exists exactly one } y \in Y \text{ such that } (x, y) \in \Gamma.$$

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 mostly use the words “domain” and “codomain” in this document.

Usually mathematicians simply write  $f$  instead of  $f(\cdot)$ . We mostly follow that convention, but sometimes include the “ $(\cdot)$ ” part to emphasize that a function rather than an “ordinary” element of a set is involved.

Let  $x \in X$ . We write  $f(x)$  for the uniquely determined  $y \in Y$  such that  $(x, y) \in \Gamma$ . We write  $\Gamma_f$  or  $\Gamma(f)$  if we want to stress that  $\Gamma$  is the relation associated with the function  $f = (X, Y, \Gamma)$ . We call  $\Gamma$  the **graph** of the function  $f$ . Clearly

$$(5.7) \quad \Gamma = \Gamma_f = \Gamma(f) = \{(x, f(x)) : x \in X\}.$$

It is customary to write

$$(5.8) \quad f : X \rightarrow Y, \quad x \mapsto f(x)$$

instead of  $f = (X, Y, \Gamma)$  and we henceforth follow that convention. We abbreviate that to  $f : X \rightarrow Y$  if it is clear or irrelevant how to compute  $f(x)$  from  $x$ . We read “ $a \mapsto b$ ” as “ $a$  is assigned to  $b$ ” or “ $a$  maps to  $b$ ” and refer to  $\mapsto$  as the **maps to operator** or **assignment operator**.

Domain elements  $x \in X$  are called **independent variables** or **arguments** and  $f(x) \in Y$  is called the **function value** of  $x$ . The subset

$$(5.9) \quad f(X) := \{y \in Y : y = f(x) \text{ for some } x \in X\} = \{f(x) : x \in X\}$$

of  $Y$  is called the **range** or **image** of the function  $f(\cdot)$ .<sup>51</sup>

We say “ $f$  maps  $X$  into  $Y$ ” and “ $f$  maps the domain value  $x$  to the function value  $f(x)$ ”.  $\square$

We say that two functions  $f = (X, Y, \Gamma)$  and  $f' = (X', Y', \Gamma')$  are **equal equal functions function!equality** if  $X = X'$ ,  $Y = Y'$ , and  $\Gamma = \Gamma'$ . Note that  $X = X'$  follows from  $\Gamma = \Gamma'$  because

$$x \in X \Leftrightarrow (x, y) \in \Gamma \text{ for some (unique) } y \in Y \Leftrightarrow (x, y) \in \Gamma' \text{ for some } y \in Y \Leftrightarrow x \in X'.$$

Figure 5.1 on p.89 illustrates the graph of a function as a subset of  $X \times Y$ .

**Remark:** Note that if  $Y \subsetneq Y'$  and  $f = (X, Y, \Gamma)$  is a function then  $f' = (X, Y', \Gamma)$  also is a function:  $\Gamma$  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  $\Gamma$  as follows:

$$X = \{x : (x, y) \in \Gamma \text{ for some } y\}. \quad \square$$

<sup>51</sup> We distinguish the target (codomain)  $Y$  of  $f(\cdot)$  from its image (range)  $f(X)$  which is a subset of  $Y$ .



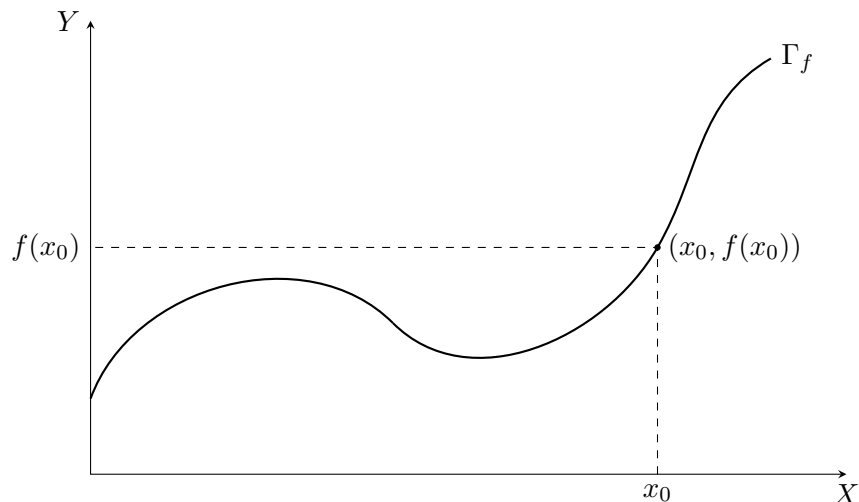


Figure 5.1: Graph of a function.

**Remark 5.7** (Mappings vs. functions). Mathematicians do not always agree 100% on their definitions. The issue of what is called a function and what is called a mapping is subject to debate. Some mathematicians call a mapping a function only if its codomain is a subset of the real numbers<sup>52</sup> 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 part of  $\mathbb{R}$  (see (5.13) on p.99).  $\square$

**Remark 5.8.** 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 non-negative real number its (positive) square root:

$$\begin{aligned} f &: [0, \infty[ \rightarrow \mathbb{R}, & x &\mapsto \sqrt{x}, \\ f &: [0, \infty[ \rightarrow \mathbb{R}, & y &\mapsto \sqrt{y}, \\ f &: [0, \infty[ \rightarrow \mathbb{R}, & f(\gamma) &= \sqrt{\gamma}. \end{aligned}$$

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:

$$\begin{aligned} \varphi &: [0, \infty[ \rightarrow \mathbb{R}, & t &\mapsto \sqrt{t}, \\ A &: [0, \infty[ \rightarrow \mathbb{R}, & x &\mapsto \sqrt{x}, \\ g &: [0, \infty[ \rightarrow \mathbb{R}, & g(A) &= \sqrt{A}. \end{aligned}$$

In contrast, the following three functions all are different from each other and none of them

---

<sup>52</sup> or if the codomain is a subset of the complex numbers, but we won't discuss complex numbers in this document.

equals  $f$  because domain and/or codomain do not match:

$$\begin{aligned} \psi &: ]0, \infty[ \rightarrow \mathbb{R}, & x &\mapsto \sqrt{x} \quad (\text{different domain}). \\ B &: [0, \infty[ \rightarrow ]0, \infty[, & x &\mapsto \sqrt{x} \quad (\text{different codomain}), \\ h &: [0, 1[ \rightarrow [0, 1[, & x &\mapsto \sqrt{x} \quad (\text{different domain and codomain}). \quad \square \end{aligned}$$

**Definition 5.7** (Function composition). Given are three nonempty sets  $X, Y$  and  $Z$  and two functions  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ . Given  $x \in X$  we know the meaning of the expression  $g(f(x))$ :

$y := f(x)$  is the function value of  $x$  for the function  $f$ , i.e., the unique  $y \in Y$  such that  $(x, y) \in \Gamma_f$ .

$z := g(y) = g(f(x))$  is the function value of  $f(x)$  for the function  $g$ , i.e., the unique  $z \in Z$  such that  $(f(x), z) = (f(x), g(f(x))) \in \Gamma_g$ .

The set  $\Gamma := \{(x, g(f(x))) : x \in X\}$  is a relation on  $(X, Z)$  such that

(5.10) for each  $x \in X$  there exists exactly one  $z \in Z$ , namely,  $z = g(f(x))$ , such that  $(x, z) \in \Gamma$ .

It follows that  $\Gamma$  is the graph of a function  $h = (X, Z, \Gamma)$  with function values  $h(x) = g(f(x))$  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.  $\square$

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*

We have a special name for the “do nothing function” which assigns each argument to itself:

**Definition 5.8** (identity mapping). Given any nonempty set  $X$ , we use the symbol  $id_X$  for the **identity** mapping defined as

$$id_X : X \rightarrow X, \quad x \mapsto x.$$

We drop the subscript if it is clear what set is referred to.  $\square$

### 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 5.8.** Let  $\Gamma := \{(x, x^3) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$ . Then  $f = (\mathbb{R}, \mathbb{R}, \Gamma)$  is the function

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto x^3. \quad \square$$

**Example 5.9.** Let  $\Gamma := \{(x, x^2 + 1) : x \in \mathbb{R}\}$ . Then  $g = (\mathbb{R}, \mathbb{R}, \Gamma)$  is the function

$$g : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto x^2 + 1. \quad \square$$

**Example 5.10.** 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

$$h : ]0, \infty[ \rightarrow \mathbb{R}, \quad x \mapsto \ln(x). \quad \square$$

**Example 5.11.** Let  $\Gamma := \{(x, \sqrt{x}) : x \in [0, \infty[ \}$ . Then  $\varphi = ([0, \infty[, \mathbb{R}, \Gamma)$  is the function

$$\varphi : [0, \infty[ \rightarrow \mathbb{R}, \quad x \mapsto \sqrt{x}. \quad \square$$

**Example 5.12.** Let  $\Gamma := \{(x, \sqrt{x}) : x \in [0, \infty[ \}$ . We can consider  $\Gamma$  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

$$\varphi : [0, \infty[ \rightarrow \mathbb{R}, \quad x \mapsto \sqrt{x}.$$

In the second case we obtain a different(!) function  $\psi = ([0, \infty[, [0, \infty[, \Gamma)$ , i.e., the function

$$\psi : [0, \infty[ \rightarrow [0, \infty[, \quad x \mapsto \sqrt{x}. \quad \square$$

*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 5.13.** We define a function which maps two-dimensional vectors to numbers. Let  $A := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ . Let  $\Gamma := \{(x, y), \sqrt{1 - x^2 - y^2}) : (x, y) \in A\}$ . Then  $F = (A, \mathbb{R}, \Gamma)$  is the function

$$F : A \rightarrow \mathbb{R}, \quad (x, y) \mapsto \sqrt{1 - x^2 - y^2}.$$

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.  $\square$

**Example 5.14.** We define a function which maps numbers to two-dimensional vectors (a curve in the plane). Let  $\Gamma := \{(t, (\sin t, \cos t)) : t \in \mathbb{R}\}$ . Then  $G = (\mathbb{R}, \mathbb{R}^2, \Gamma)$  is the function

$$G : \mathbb{R} \rightarrow \mathbb{R}^2, \quad t \mapsto (\sin t, \cos t).$$

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 5.15.** Let  $\Gamma := \{(x, y), (2x - y/3, x/6 + 4y)) : x, y \in \mathbb{R}\}$ . Then  $H = (\mathbb{R}^2, \mathbb{R}^2, \Gamma)$  is the function

$$H : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto (2x - y/3, x/6 + 4y).$$

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 5.16.** As is customary in linear algebra we now think of  $\mathbb{R}^2$  as the collection of column vectors  $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R} \right\}$  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  $2 \times 2$  matrix

$$A := \begin{pmatrix} 2 & -1/3 \\ 1/6 & 4 \end{pmatrix}.$$

We then obtain for any pair of numbers  $\vec{x} = (x, y)^T$ <sup>53</sup> 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  $\Gamma := \left\{ \left( \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 2x - y/3 \\ x/6 + 4y \end{pmatrix} \right) : x, y \in \mathbb{R} \right\}$ . Then  $H = (\mathbb{R}^2, \mathbb{R}^2, \Gamma)$  is the function

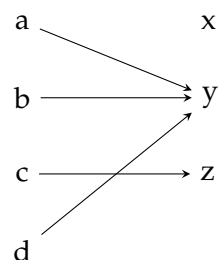
$$H : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \end{pmatrix}.$$

Note that both domain and codomain are the Euclidean plane.  $\square$

*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 5.17.**

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 5.6 (A better definition of a function) on p.86. 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.  $\square$

**Example 5.18.** This example represents a mathematical model for computing probabilities of the outcomes of rolling a fair die and demonstrates that probability can be thought of as a function that maps sets to numbers.

If we roll a die then the outcome will be an integer between 1 and 6, i.e., the “state space” for this random action will be  $X := \{1, 2, 3, 4, 5, 6\}$ . For  $A \subseteq X$  let  $\text{Prob}(A)$  denote the probability that rolling the die results in an outcome  $x \in A$ .

---

<sup>53</sup> Here  $(x, y)^T = \begin{pmatrix} x \\ y \end{pmatrix}$  is the **transpose** of  $(x, y)$ , i.e., the operation that switches rows and columns of any matrix. In particular it transforms a row vector into a column vector and vice versa.

For example  $\text{Prob}(\text{an even number occurs}) = \text{Prob}(\{2, 4, 6\}) = 50\% = 1/2$ . Clearly we have for singletons consisting of a single outcome that

$$\text{Prob}(\{1\}) = \text{Prob}(\{2\}) = \cdots = \text{Prob}(\{6\}) = 1/6 = 16.\bar{6}\%.$$

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

$$\text{Prob}(A) = \text{Prob}(\{x_1\}) + \text{Prob}(\{x_2\}) + \cdots + \text{Prob}(\{x_k\}) = \sum_{j=1}^k \text{Prob}(\{x_j\}).$$

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.,  $\text{Prob}(\emptyset) = 0$ .

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

$$P : 2^X \rightarrow [0, 1], \quad A \mapsto \text{Prob}(A).$$

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 probability 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.  $\square$

**Example 5.19.** In this example we will define a function  $I(\cdot)$  for which both domain and codomain are sets of functions. For the necessary background on antiderivatives see rem.2.11 on p.25.

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

$$\begin{aligned} \mathcal{F} &:= \{f : ]a, b[ \rightarrow \mathbb{R} \text{ such that } f \text{ is continuous on } ]a, b[ \}, \\ \mathcal{G} &:= \{g : ]a, b[ \rightarrow \mathbb{R} \text{ such that } g \text{ is differentiable on } ]a, b[ \text{ and } g(x_0) = c_0 \}. \end{aligned}$$

We have seen in rem.2.11 that for each  $f \in \mathcal{F}$  there exists a unique  $F \in \mathcal{G}$  which is an antiderivative for  $f$ . We now define a function  $I : \mathcal{F} \rightarrow \mathcal{G}$  by specifying its graph as the set

$$\Gamma := \{(f(\cdot), g(\cdot)) : f \in \mathcal{F}, g \in \mathcal{G}, g \text{ is an antiderivative of } f \text{ and } g(x_0) = c_0\}.$$

It follows from e that the function value  $I(f)(x)$  for any  $a < x < b$  is

$$I(f)(x) = \int_{x_0}^x f(u)du - I(f)(x_0) = \int_{x_0}^x f(u)du - c_0. \quad \square$$

## 5.2.4 Injective, Surjective and Bijective functions

**Definition 5.9** (Surjective, injective, bijective). Let  $f : X \rightarrow Y$ . As usual the graph of  $f$  is denoted  $\Gamma_f$ .

**a. Surjectivity:** In general it is not true that  $f(X) = \{f(x) : x \in X\}$  equals the entire codomain  $Y$ , i.e., that

$$(5.12) \quad \text{for each } y \in Y \text{ there exists at least one } x \in X \text{ such that } (x, y) \in \Gamma_f.$$

But if  $f(X) = Y$ , i.e., if (5.12) holds, we call  $f$  **surjective** and we say that  $f$  maps  $X$  **onto**  $Y$ .

**b. Injectivity:** In general it is not true that if  $y \in f(X)$  then  $y = f(x)$  for a unique  $x$ , i.e., that if there is another  $x_1 \in X$  such that also  $y = f(x_1)$  then it follows that  $x_1 = x$ . But if this is the case, i.e., if

$$(5.13) \quad \text{for each } y \in Y \text{ there exists at most one } x \in X \text{ such that } (x, y) \in \Gamma_f.$$

then we call  $f$  **injective**.

We can express (5.13) 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  $f : X \rightarrow Y$  be both injective and surjective. Such a function is called **bijective**. We often write  $f : X \xrightarrow{\sim} Y$  for a bijective function  $f$ .

It follows from (5.12) and (5.13) that  $f$  is bijective if and only if

$$(5.14) \quad \text{for each } y \in Y \text{ there exists exactly one } x \in X \text{ such that } (x, y) \in \Gamma_f.$$

We rewrite (5.14) by employing  $\Gamma_f$ 's inverse relation  $\Gamma_f^{-1} = \{(y, x) : (x, y) \in \Gamma_f\}$  (see def. 5.5 on p.85) and obtain

$$(5.15) \quad \text{for each } y \in Y \text{ there exists exactly one } x \in X \text{ such that } (y, x) \in \Gamma_f^{-1}.$$

But this implies, according to (5.6), that  $\Gamma_f^{-1}$  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

$$(5.16) \quad \Gamma_f^{-1} = \Gamma_g.$$

We call  $g$  the **inverse mapping** or **inverse function** of  $f$  and write  $f^{-1}$  instead of  $g$ .  $\square$

**Remark 5.9.**

a. It follows from (5.16) that

$$(5.17) \quad \Gamma_f^{-1} = \Gamma_{f^{-1}}.$$

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

$$(5.18) \quad \text{if } x \in X, y \in Y \text{ then } f(x) = y \Leftrightarrow x = f^{-1}(y).$$

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:

$$(5.19) \quad \text{Inverse function} \quad \begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow id_X & \downarrow f^{-1} \\ & & X \end{array} \quad \begin{array}{ccc} Y & \xrightarrow{f^{-1}} & X \\ & \searrow id_Y & \downarrow f \\ & & Y \end{array} \quad \square$$

**Theorem 5.1** (Characterization of inverse functions). *Let  $X$  and  $Y$  be nonempty sets and  $f : X \rightarrow Y$ . Then the following are equivalent:*

- a.  $f$  is bijective.
- b. There exists  $g : Y \rightarrow X$  such that both  $g \circ f = id_X$  and  $f \circ g = id_Y$ .

*PROOF of a  $\Rightarrow$  b:* We have seen in part e of remark 5.9 that  $g := f^{-1}$  satisfies b.

*PROOF of b  $\Rightarrow$  a:* We must show that  $f$  is both surjective and injective. First we show that  $f$  is surjective. Let  $y \in Y$ . we must find some  $x \in X$  such that  $f(x) = y$ . Let  $x := g(y)$ . Then

$$f(x) = f(g(y)) = f \circ g(y) = id_Y(y) = y.$$

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

$$x_1 = id_X(x_1) = g \circ f(x_1) = g(f(x_1)) = g(y) = g(f(x_2)) = g \circ f(x_2) = id_X(x_2) = x_2,$$

i.e.,  $x_1 = x_2$ . This proves injectivity of  $f$ . ■

**Remark 5.10.** [Horizontal and vertical line tests] Let  $X$  and  $Y$  be nonempty sets and  $f : X \rightarrow Y$ . 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.88 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. (5.12) on p.93 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. (5.13) on p.94 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 (5.14) on p.94 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  $f^{-1}$ ). □

**Remark 5.11.** Abstract math is about proving theorems and propositions. Functions are very important tools for many proofs. 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 as this simple example shows.

Let  $f : A \rightarrow B$  be the function  $f(x) := x^2$ .

- |                                    |   |
|------------------------------------|---|
| $A = \mathbb{R}, B = \mathbb{R}$ : | $f$ is neither injective nor surjective |
| $A = ] - 2, 3[, B = [0, 9[$ :      | $f$ is surjective but not injective     |
| $A = ]0, 3[, B = [0, 9]$ :         | $f$ is injective but not surjective     |
| $A = ]0, 3[, B = ]0, 9[$ :         | $f$ is bijective                        |

The above demonstrates why domain and codomain are important for the specification of a function.  $\square$

**Proposition 5.4.** Let  $X, Y, Z \neq \emptyset$ . Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ .

- 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 5.7.

PROOF of **c**: Follows from **a** and **b** because bijective = injective + surjective.  $\blacksquare$

**Corollary 5.1.** Let  $X, Y, Z \neq \emptyset$ . Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ .

- 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.5.4.a because bijective functions are injective.

**b** follows from prop.5.4.b because bijective functions are surjective.

**c** follows from prop.5.4.c.  $\blacksquare$

We now examine conditions under which there are functions  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$  such that  $g \circ f = id_X$ , i.e.,

$$(5.20) \quad g(f(x)) = x \text{ for all } x \in X : \quad \begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow id_X & \downarrow g \\ & & X \end{array}$$

**Proposition 5.5.** Let  $X, Y \neq \emptyset$ . Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$  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  $x_0$  and we have proved surjectivity of  $g$ .  $\blacksquare$

**Proposition 5.6.** Let  $X, Y \neq \emptyset$ .

- a. Let  $f : X \rightarrow Y$ . If  $f$  is injective then there exists  $g : Y \rightarrow X$  such that  $g \circ f = id_X$  and any such function  $g$  is necessarily surjective.
- b. Let  $g : Y \rightarrow X$ . If  $g$  is surjective then there exists  $f : X \rightarrow Y$  such that  $g \circ f = id_X$  and any such function  $f$  is necessarily injective.



PROOF of **a**: Let  $Y' := f(X)$  and

$$f' : X \rightarrow Y', \quad x \mapsto f(x),$$

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  $(f')^{-1} : Y' \rightarrow X$  exists.

Let  $x_0$  be an arbitrary, but fixed, element of  $X$ . We define  $g : Y \rightarrow X$  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  $g \circ f(x) = g \circ f'(x) = (f')^{-1}(f'(x)) = x$ . As  $x$  was an arbitrary element of  $X$ , this proves  $g \circ f = id_X$ . We observe that  $g$  is surjective according to prop.5.5b.

PROOF of **b**: For  $x \in X$  let  $Y_x := \{y \in Y : g(y) = x\}$ . It follows from the surjectivity of  $g$  that  $Y_x \neq \emptyset$  and we may select for each  $x \in X$  some  $y_x \in Y_x$ .<sup>54</sup>

Let  $f : X \rightarrow Y$  be the function  $x \mapsto y_x$ . Let  $x \in X$ . 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  $y_x \in Y_x$ . It follows from prop.5.5a that  $f$  is injective. ■

There are special names for functions  $f$  and  $g$  which are related by (5.20).

**Definition 5.10** (Left inverses and right inverses). Let  $X, Y \neq \emptyset$ . Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$  such that  $g \circ f = id_X$ . We call  $g$  a **left inverse** of  $f$  and we call  $f$  a **right inverse** of  $g$ . □

We combine that last definitions with the preceding two proposition and obtain

**Theorem 5.2.** Let  $X, Y \neq \emptyset$ .

- a. Let  $f : X \rightarrow Y$ . Then  $f$  is injective  $\Leftrightarrow f$  has a left inverse (which is necessarily surjective).
- b. Let  $g : Y \rightarrow X$ . Then  $g$  is surjective  $\Leftrightarrow g$  has a right inverse (which is necessarily injective).
- c. An injection  $X \rightarrow Y$  exists  $\Leftrightarrow$  a surjection  $Y \rightarrow X$  exists.

PROOF of **a**  $\Rightarrow$ ): prop.5.6a.

PROOF of **a**  $\Leftarrow$ ): prop.5.5a.

PROOF of **b**  $\Rightarrow$ ): prop.5.6b.

PROOF of **b**  $\Leftarrow$ ): prop.5.5b.

PROOF of **c**  $\Rightarrow$ ): Let  $f : X \rightarrow Y$  be injective. According to part **a** there exists a left inverse  $g : Y \rightarrow X$  and this function is surjective

PROOF of **c**  $\Leftarrow$ ): Let  $g : Y \rightarrow X$  be surjective. According to part **b** there exists a right inverse  $f : X \rightarrow Y$  and this function is injective ■

<sup>54</sup> The ability to do such selections  $x$  regardless of the nature of  $X, Y$  and of the surjective function  $f : X \rightarrow Y$  is not something one can prove. It requires acceptance of the **Axiom of Choice**. See remark 13.1 on p.278 in ch.13 (Applications of Zorn's Lemma).

**Remark 5.12.** Let  $X$  and  $Y$  be two nonempty sets. No assumptions are made concerning how  $X$  and  $Y$  might be related.

- a. Let  $a \in Y$ . Then the function

$$(5.21) \quad f : X \xrightarrow{\sim} \{a\} \times X; \quad x \mapsto (a, x)$$

is bijective because  $f$  has the function  $(a, x) \mapsto x$  as an inverse.

- b. The sets  $\{0\} \times X$  and  $\{1\} \times Y$  are disjoint.  
c. An injection/surjection/bijection  $X \rightarrow Y$  exists if and only if an injection/surjection/bijection  $\{0\} \times X \rightarrow \{1\} \times Y$  exists.  $\square$

### 5.2.5 Binary Operations and Restrictions and Extensions of Functions

**Definition 5.11** (Binary operations). Let  $X$  be a nonempty set. A **binary operation**, on  $X$  is a function

$$(5.22) \quad \diamond : X \times X \longrightarrow X; \quad (x, y) \mapsto x \diamond y := \diamond(x, y).$$

A **unary operation**, on  $X$  is a function

$$(5.23) \quad \bullet : X \longrightarrow X; \quad x \mapsto x^\bullet := \bullet(x).$$

$\square$

**Example 5.20.** Some examples of binary operations are.

- a. Addition on  $X = \mathbb{N}$  or  $X = \mathbb{Z}$  or  $X = \mathbb{Q}$  or  $X = \mathbb{R}$  is a binary operation

$$(5.24) \quad + : X \times X \longrightarrow X; \quad (x, y) \mapsto x + y.$$

- b. Multiplication on  $X = \mathbb{N}$  or  $X = \mathbb{Z}$  or  $X = \mathbb{Q}$  or  $X = \mathbb{R}$  is a binary operation

$$(5.25) \quad \cdot : X \times X \longrightarrow X; \quad (x, y) \mapsto x \cdot y.$$

- c. Let  $X$  be a nonempty set and  $\mathcal{F} := \{ \text{functions } f : X \rightarrow X \}$ . Function composition

$$(5.26) \quad \circ : \mathcal{F} \times \mathcal{F} \longrightarrow \mathcal{F}; \quad (f, g) \mapsto g \circ f$$

where  $g \circ f : X \rightarrow X$  is the function defined by  $x \mapsto g \circ f(x) := g(f(x))$  ( $x \in X$ ).

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

$$(5.27) \quad - : X \rightarrow X; \quad x \mapsto -x.$$

is a unary operation.

e. Reciprocal  $X = \mathbb{Q}_{\neq 0}$  or  $X = \mathbb{R}_{\neq 0}$

$$(5.28) \quad \cdot^{-1} : X \rightarrow X; \quad x \mapsto x^{-1} = 1/x.$$

is a unary operation.

f. Let  $X$  be a nonempty set and  $\mathcal{B} := \{\text{bijective functions } f : X \rightarrow X\}$ . Let

$$(5.29) \quad \cdot^{-1} : \mathcal{B} \rightarrow \mathcal{B}; f(\cdot) \mapsto f^{-1}(\cdot)$$

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  $\mathcal{B}$ .

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 5.12** (Restriction/Extension of a function). Given are three nonempty sets  $A, X$  and  $Y$  such that  $A \subseteq X$ . Let  $f : X \rightarrow Y$  a function with domain  $X$ . We define the **restriction of  $f$  to  $A$**  as the function

$$(5.30) \quad f|_A : A \rightarrow Y \quad \text{defined as} \quad f|_A(x) := f(x) \text{ for all } x \in A.$$

Conversely let  $f : A \rightarrow Y$  and  $\varphi : X \rightarrow Y$  be functions such that  $f = \varphi|_A$ . We then call  $\varphi$  an **extension** of  $f$  to  $X$ .  $\square$

**Example 5.21.** 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  $\varphi_\alpha : \mathbb{R} \rightarrow \mathbb{R}$  defined as  $\varphi_\alpha(x) := 3x^2$  if  $0 \leq x \leq 1$  and  $\alpha x$  otherwise defines a different extension of  $f$  to  $\mathbb{R}$ .  $\square$

**Notations 5.1.** As the only difference between  $f$  and  $f|_A$  is the domain, it is customary to write  $f$  instead of  $f|_A$  to make formulas look simpler if doing so does not give rise to confusions.  $\square$

**Remark 5.13.** The restriction  $f|_A$  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<sup>55</sup> of the extension  $\varphi_\alpha$  of  $f$  in example 5.21 above, only  $\varphi_1(x) = 3x^2$  if  $0 \leq x \leq 1$  and  $x$  otherwise would qualify.  $\square$

## 5.2.6 Operations on Real Valued Functions

**Definition 5.13** (real functions). Let  $X$  be an arbitrary, nonempty set. If the codomain  $Y$  of a mapping

$$f : X \rightarrow Y; \quad x \mapsto f(x)$$

is a subset of  $\mathbb{R}$ , then we call  $f(\cdot)$  a **real function** or **real valued function**.  $\square$

*Remember that this definition does not exclude the case  $Y = \mathbb{R}$  because  $Y \subseteq \mathbb{R}$  is in particular true if both sets are equal.*

<sup>55</sup> Continuity of functions  $y = f(x)$  with real numbers  $x$  and  $y$  will be defined in ch.9.2 (Convergence and Continuity in  $\mathbb{R}$ ). See def.9.12 on p.156. Until then use your knowledge from calculus.

Real functions are a pleasure to work with because, given any fixed argument  $x_0$ , the object  $f(x_0)$  is just an ordinary number. In particular you can add, subtract, multiply and divide real valued functions. Of course, division by zero is not allowed:

**Definition 5.14** (Operations on real valued functions). Let  $X$  be an arbitrary nonempty set. Given are two real valued functions  $f(\cdot), g(\cdot) : X \rightarrow (R)$  and a real number  $\alpha$ . The **sum**  $f + g$ , **difference**  $f - g$ , **product**  $fg$  or  $f \cdot g$ , **quotient**  $f/g$ , and **scalar product**  $\alpha f$  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:

$$(5.31) \quad \begin{aligned} (f + g)(x) &:= f(x) + g(x), \\ (f - g)(x) &:= f(x) - g(x), \\ (fg)(x) &:= f(x)g(x), \\ (f/g)(x) &:= f(x)/g(x) \quad \text{for all } x \in X \text{ where } g(x) \neq 0, \\ (\alpha f)(x) &:= \alpha \cdot g(x). \quad \square \end{aligned}$$

Before we list some basic properties of addition and scalar multiplication of functions (the operations that interest us the most), let us have a quick look at constant functions.

**Definition 5.15** (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:

$$y_0(\cdot) : X \rightarrow Y; \quad x \mapsto y_0.$$

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 real-valued function is the **zero function**  $0(\cdot)$  which maps any  $x \in X$  to the number zero. We usually just write 0 for this function unless doing so would confuse the reader. Note that scalar multiplication  $(\alpha f)(x) = \alpha \cdot g(x)$  is a special case of multiplying two functions  $(gf)(x) = g(x)f(x)$ : Let  $g(x) = \alpha$  for all  $x \in X$  (constant function  $\alpha$ ).  $\square$

One last definition before we get so see some examples:

**Definition 5.16** (Negative function). Let  $X$  be an arbitrary, nonempty set and let  $f : X \rightarrow \mathbb{R}$ . The function

$$-f(\cdot) : X \rightarrow \mathbb{R}; \quad x \mapsto -f(x).$$

is called **negative  $f$**  or **minus  $f$** . We usually write  $-f$  for  $-f(\cdot)$ .  $\square$

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 5.22** (Arithmetic operations on real valued functions).

For simplicity, let  $X := \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ . Let

$$\begin{aligned} f : \mathbb{R}_{\geq 0} &\rightarrow \mathbb{R}; & x &\mapsto (x - 1)(x + 1) \\ g : \mathbb{R}_{\geq 0} &\rightarrow \mathbb{R}; & x &\mapsto (x - 1) \\ h : \mathbb{R}_{\geq 0} &\rightarrow \mathbb{R}; & x &\mapsto (x + 1) \end{aligned}$$

Then

$$\begin{aligned}
 (f+h)(x) &= (x-1)(x+1) + x + 1 = x^2 - 1 + x + 1 = x(x+1) \quad \forall x \in \mathbb{R}_{\geq 0}, \\
 (f-g)(x) &= (x-1)(x+1) - (x-1) = x^2 - 1 - x + 1 = x(x-1) \quad \forall x \in \mathbb{R}_{\geq 0}, \\
 (gh)(x) &= (x-1)(x+1) = f(x) \quad \forall x \in \mathbb{R}_{\geq 0}, \\
 (f/h)(x) &= (x-1)(x+1)/(x+1) = x-1 = g(x) \quad \forall x \in \mathbb{R}_{\geq 0}, \\
 (f/g)(x) &= (x-1)(x+1)/(x-1) = x+1 = h(x) \quad \forall x \in \mathbb{R}_{\geq 0} \setminus \{1\}
 \end{aligned}$$

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.  $\square$

*Many more properties of functions will be discussed later. Now we look at families, sequences and some additional properties of sets.*

### 5.2.7 Sequences, Families and Functions as Families

**Definition 5.17** (Indices). Given is an expression of the form

$$X_i.$$

We say that  $X$  is **indexed by** or **subscripted by** or **tagged by**  $i$ . We call  $i$  the **index** or **subscript** of  $X$  and we call  $X_i$  an **indexed item**.  $\square$

**Remark 5.14.** Both  $X$  and  $i$  can occur in many different ways. Here is a collection of indexed items:

$$x_7, A_\alpha, k_T, \mathfrak{N}_{2/9}, f_x, x_t, h_{\mathcal{A}}, i_{\mathbb{R}}, H_{2\pi}$$

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.,  $2\pi$ ) Others don't even look like numbers (e.g.,  $\alpha, T, x, t, \mathcal{A}$  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: The index  $\mathbb{R}$  of  $i_{\mathbb{R}}$  denotes the set of all real numbers.  $\square$

*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. 11:*

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

We note several things:

- $F$  has the kind of indices that we are familiar with: all of them are positive integers.
- Some of the indices in  $F$  occur multiple times. For example, 3 occurs as an index for  $i_3, o_3, u_3$ .
- All of the indices in  $G$  are unique.
- As in remark 5.14, 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  $\iota \in J$  must be used exactly once. Remember that this automatically takes care of the duplicate indices problem as a set never contains duplicate values (see def.2.1 on p. 11). We also demand that there is a set  $X$  such that each indexed item  $x_\iota$  belongs to  $X$ .

We now are ready to give the definition of a family:

**Definition 5.18** (Indexed families). Let  $J$  and  $X$  be nonempty set and assume that

for each  $\iota \in J$  there exists exactly one indexed item  $x_\iota \in X$ .

Let  $R := \{(\iota, x_\iota) : \iota \in J\}$ . Then  $R$  is a relation on  $(J, X)$  which satisfies (5.6) of the definition of a function

$$F : J \rightarrow X, \quad \iota \mapsto F(\iota) := x_\iota$$

(see def.5.6 on p.88) whose graph  $\Gamma_F$  equals  $R$ .

We write  $(x_\iota)_{\iota \in J}$  for this function if we want to emphasize that we are interested in the collection of indexed elements  $x_\iota$  rather than the function  $F(\cdot)$  or the relation  $R$ . Reasons for this will be given in rem.5.16 on p.103.

$(x_\iota)_{\iota \in J}$  is called an **indexed family** or simply a **family** in  $X$  and  $J$  is called the **index set** of the family. For each  $j \in J$ ,  $x_j$  is called a **member of the family**  $(x_\iota)_{\iota \in J}$ .

$\iota$  is a dummy variable:  $(x_\iota)_{\iota \in J}$  and  $(x_k)_{k \in J}$  describe the same family as long as  $\iota \mapsto x_\iota$  and  $k \mapsto x_k$  describe the same function  $F : J \rightarrow X$ . This should not come as a surprise to you if you recall remark 5.8 on p.89.  $\square$

**Remark 5.15.** The codomain  $X$  does not occur in the notation  $(x_\iota)_{\iota \in J}$ . 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 possible choices for a codomain.

- a. If there is a universal set  $X$  which contains all tagged items of the family then  $X$  is acceptable as the codomain.
- b. If there is no universal set then you can think of

$$X = \bigcup [x_\iota : \iota \in J] := \{x : x = x_{\iota_0} \text{ for some } \iota_0 \in I\}$$

as the codomain. <sup>56</sup>  $\square$

**Definition 5.19** (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  $i \in I$ .  $\square$

Note that this is different from the concept of equality of functions because no demand is made that both families are families in the same set  $X$ .

<sup>56</sup>General unions and intersections will be defined in ch.6.1 (More on set operations). See def.6.1 on p.109.

**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  $(x_i)_i$  instead of  $(x_i)_{i \in J}$ . Even the index outside the right parentheses may be omitted and we write  $(x_i)$  if it is clear that we are talking about families.

For example, a proposition may start as follows: Let  $(A_\alpha)$  and  $(B_\alpha)$  be two families of subsets of  $\Omega$  indexed by the same set. Then .....

It is clear from the formulation that we deal in fact with two families  $(A_\alpha)_{\alpha \in J}$  and  $(B_\alpha)_{\alpha \in J}$ . Nothing is said about  $J$ , probably because the proposition is valid for any index set or because this set was fixed once and for all earlier on for the entire section.  $\square$

**Example 5.23.** 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.  $\square$

**Remark 5.16.** If a family is just some kind of function, why bother with yet another definition? The answer to this is that there are many occasions in which writing something as a collection of indexed items rather than as a function makes things easier to understand. For example, take a peek at theorem 6.1 (De Morgan's Law) on p.111. One of the formulas there states that for any indexed family  $(A_\alpha)_{\alpha \in I}$  of subsets of a universal set  $\Omega$  it is true that

$$\left(\bigcup_{\alpha} A_{\alpha}\right)^c = \bigcap_{\alpha} A_{\alpha}^c.$$

Without the notion of a family you might have to say something like this: Let  $A : I \rightarrow 2^{\Omega}$  be a function which assigns its arguments to subsets of  $\Omega$ . Then

$$\left(\bigcup_{\alpha} A(\alpha)\right)^c = \bigcap_{\alpha} A(\alpha)^c.$$

The additional parentheses around the index  $\alpha$  just add complexity to the formula.  $\square$

**Example 5.24** (Sequences as families). You have worked with special families before: those where  $J = \mathbb{N}$  or  $J = \mathbb{Z}_{\geq 0}$  and  $X$  is a subset of the real numbers. Example:  $x_n := 1/n$ . Here

$$(x_n)_{n \in \mathbb{N}} \text{ corresponds to the indexed collection } 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots \quad \square$$

*Such families are called sequences:*

**Definition 5.20** (Sequences and subsequences). Let  $n_*$  be an integer and let  $J := \{n_*, n_* + 1, n_* + 2, \dots\} = \{k \in \mathbb{Z} : k \geq n_*\}$ . Let  $X$  be an arbitrary nonempty set. An indexed family in  $X$  with index set  $J$  is called a **sequence** in  $X$  with **start index**  $n_*$ .

Let  $I$  be an infinite subset of  $J$ . Because any two different integers  $m$  and  $n$  can be compared, i.e., either  $m < n$  or  $n < m$ ,<sup>57</sup> there are integers

$$(5.32) \quad n_1 < n_2 < n_3 < \dots$$

<sup>57</sup>  $\mathbb{Z}$  is a **linearly ordered set**, also called a **totally ordered set** set in the terminology of def.13.1 (Linear orderings) on p.278.

such that  $I = \{n_j : j \in \mathbb{N}\}$ . An indexed family in  $X$  with index set  $I$  is called a **subsequence** of the original sequence.

Note that, according to this definition, the sequence  $(n_j)_{j \in \mathbb{N}}$  is a subsequence of the full sequence of indices  $(n)_{n \in J}$ .

As for families, the name of the index variable of a sequence does not matter as long as it is applied consistently. It does not matter whether you write  $(x_j)_{j \in J}$  or  $(x_n)_{n \in J}$  or  $(x_\beta)_{\beta \in J}$ .  $\square$

**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. If  $J$  is defined as above then is not unusual to see “ $(x_n)_{n \geq n_*}$ ” instead of “ $(x_n)_{n \in J}$ ”. By default the index set for a sequence is  $J = \mathbb{N} = \{1, 2, 3, 4, \dots\}$  and we can then write  $(x_n)_n$  or just  $(x_n)$
- c. Customary 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.  $\square$

Here are two more examples of sequences.

**Example 5.25** (Oscillating sequence). If  $x_j := (-1)^j$  ( $j \in \mathbb{Z}_{\geq 0}$ ), then

$$x_0 = 1, \quad x_1 = -1, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 1, \quad x_5 = -1, \dots \quad \square$$

**Example 5.26** (Series (summation sequence)). Let  $s_k := 1 + 2^{-1} + 2^{-2} + \dots + 2^{-k}$  ( $k = 1, 2, 3, \dots$ ):

$$s_1 = 1, \quad s_2 = 1 + 1/2 = 2 - 1/2, \quad s_3 = 1 + 1/2 + 1/4 = 2 - 1/4, \quad \dots, \\ s_k = 1 + 1/2 + \dots + 2^{k-1} = 2 - 2^{k-1}; \quad s = 1 + 1/2 + 1/4 + 1/8 + \dots \quad \text{“infinite sum”}.$$

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**.<sup>58</sup>  $\square$

**Remark 5.17.** Having defined the family  $(x_\iota)_{\iota \in J}$  as the function which maps  $\iota \in J$  to  $x_\iota$  means that a family distinguishes any two of its members  $x_\iota$  and  $x_j$  by remembering what their indices are, even if they represent one and the same element of  $X$ : Think of “ $(x_\iota)_{\iota \in J}$ ” as an abbreviation for

$$(5.33) \quad \left( (\iota, x_\iota) \right)_{\iota \in J}.$$

Doing so should also make it much easier to see the equivalence of functions and families: (5.33) looks at its core very much like the graph  $\{(\iota, x_\iota) : \iota \in J\}$  of the function  $\iota \mapsto x_\iota$ .  $\square$

**Remark 5.18** (Families and sequences can contain duplicates). One of the important properties of sets is that they do not contain any duplicates (see def.2.1 (sets) on p.11). On the other hand, remark 5.17 casually mentions that families, and hence sequences as special kinds of families, can contain duplicates. Let us look at this more closely.

<sup>58</sup> The precise definition of a series will be given in ch.11.3 (Function Sequences and Infinite Series) on p.244.



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  $(w_i)_{i \in J}$  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  $\Gamma \subseteq J \times A = J \times B$ <sup>59</sup> are different from each other because two pairs  $(a, b)$  and  $(x, y)$  are equal only if  $x = a$  and  $y = b$ .  $\square$

*In contrast to sets, families and sequences allow us to incorporate duplicates.*

A family  $(x_i)_{i \in J}$  in  $X$  is specified by the function  $F : J \rightarrow X$  which maps  $i \in J$  to  $F(i) = x_i$ . Conversely, let  $X, Y$  be nonempty sets and let  $f : X \rightarrow Y$  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 5.7.** *The following ways of specifying a function  $f : X \rightarrow Y$ ,  $x \mapsto f(x)$  are equivalent:*

- a.  $f = (X, Y, \Gamma)$  is defined by its graph  $\Gamma := \{(x, f(x)) : x \in X\}$ .
- b.  $f$  is defined by the following family in  $Y$ :  $(f(x))_{x \in X}$

*Note that this is one case where we had to explicitly mention the codomain  $Y$  in the specification of the family.*

*There will be a lot more on sequences and series (sequences of sums) in later chapters, but we need to develop more concepts, such as convergence, to continue with this subject. Now let's get back to sets.*

## 5.3 Exercises for Ch.5

### 5.3.1 Exercises for Functions and Relations

#### Exercise 5.1.

- 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  $\mathbb{R}_{< 0}$ ?  $\square$

**Exercise 5.2.** Let  $R \subseteq \mathbb{Z}^2$  be the following relation on  $\mathbb{Z}$ :  $mRn \Leftrightarrow m|n$ .

- 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  $n|m$ .  $\square$

#### Exercise 5.3. Injectivity and Surjectivity

- Let  $f : \mathbb{R} \rightarrow [0, \infty[; \quad x \mapsto x^2$ .
  - Let  $g : [0, \infty[ \rightarrow [0, \infty[; \quad x \mapsto x^2$ .
- 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**.

<sup>59</sup> Be sure to understand that  $J \times A = J \times B$ !

- 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.  $\square$

**Exercise 5.4** (Exercise 5.3 continued). Let  $A \subseteq \mathbb{R}$ .

Part 1.

- Let  $F_1 : A \rightarrow [-2, 20[; \quad x \mapsto x^2$ .
- Let  $F_2 : A \rightarrow [2, 20[; \quad x \mapsto x^2$ .

What choice of  $A$  makes

- a.  $F_1$  surjective?            c.  $F_2$  surjective?  
 b.  $F_1$  injective?            d.  $F_2$  injective?

Part 2.

- Let  $G_1 : A \rightarrow [-2, 20[; \quad x \mapsto \sqrt{x}$ .
- Let  $G_2 : A \rightarrow [2, 20[; \quad x \mapsto \sqrt{x}$ .

What choice of  $A$  makes

- e.  $G_1$  surjective?            g.  $G_2$  surjective?  
 f.  $G_1$  injective?            h.  $G_2$  injective?

Part 3.

- Let  $G_3 : A \rightarrow [-20, 2[; \quad x \mapsto \sqrt{x}$ .
- Let  $G_4 : A \rightarrow [-20, -2[; \quad x \mapsto \sqrt{x}$ .

What choice of  $A$  makes

- i.  $G_3$  surjective?            k.  $G_3$  surjective?  
 j.  $G_4$  injective?            l.  $G_4$  injective?

For the questions above

- Write **impossible** if no choice of  $A \subseteq \mathbb{R}$  exists.
- Write **NAF** for any of  $F_1, F_2, G_1, G_2, G_3, G_4$  which does **not define a function**.  $\square$

**Exercise 5.5.** Find  $f : X \rightarrow Y$  and  $A \subseteq X$  such that  $f(A^c) \neq f(A)^c$ . 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 5.17 on p.92.  $\square$

**Exercise 5.6.**

- a. Prove prop.5.4a: The composition of two injective functions is injective.  
 b. Prove prop.5.4b: The composition of two surjective functions is surjective.  $\square$

**Exercise 5.7.** 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  $f : \{a\} \rightarrow \{b_1, b_2\}$  and  $g : \{b_1, b_2\} \rightarrow \{a\}$  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 5.17 on p.92.  $\square$

**Exercise 5.8.** Prove b of remark 5.12 on p.98: Let  $X$  and  $Y$  be two nonempty sets. Then the sets  $\{0\} \times X$  and  $\{1\} \times Y$  are disjoint.  $\square$

**Exercise 5.9.** Prove c of remark 5.12 on p.98: Let  $X$  and  $Y$  be two nonempty sets. Then an injection/surjection/bijection  $X \rightarrow Y$  exists if and only if an injection/surjection/bijection  $\{0\} \times X \rightarrow \{1\} \times Y$  exists.  $\square$

**Exercise 5.10.** B/G Project 6.9.:

On  $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$  we define the relation  $\sim$  as follows.

$$(5.34) \quad (m_1, n_1) \sim (m_2, n_2) \Leftrightarrow m_1 \cdot n_2 = n_1 \cdot m_2.$$

a. Prove that  $\sim$  defines an equivalence relation on  $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ .

Let

$$(5.35) \quad \Omega := \{[(m, n)] : m, n \in \mathbb{Z} \text{ and } n \neq 0\}$$

be the set of all equivalence classes of  $\sim$ . We define two binary operations  $\oplus$  and  $\otimes$  on  $\Omega$  as follows;

$$(5.36) \quad [(m_1, n_1)] \oplus [(m_2, n_2)] := [(m_1 n_2 + m_2 n_1, n_1 n_2)],$$

$$(5.37) \quad [(m_1, n_1)] \otimes [(m_2, n_2)] := [(m_1 m_2, n_1 n_2)]$$

b. Prove that these binary operations are defined consistently: the right hand sides of (5.36) and (5.37) 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

$$(5.38) \quad [(m_1 n_2 + m_2 n_1, n_1 n_2)] = [(p_1 q_2 + p_2 q_1, q_1 q_2)],$$

$$(5.39) \quad [(m_1 m_2, n_1 n_2)] = [(p_1 p_2, q_1 q_2)].$$

or, equivalently, then

$$(5.40) \quad (m_1 n_2 + m_2 n_1, n_1 n_2) \sim (p_1 q_2 + p_2 q_1, q_1 q_2),$$

$$(5.41) \quad (m_1 m_2, n_1 n_2) \sim (p_1 p_2, q_1 q_2). \quad \square$$

**Exercise 5.11.** What are the graphs  $\Gamma_{f_\star}$  and  $\Gamma_{f^\star}$  of the functions  $f_\star$  and  $f^\star$  of example 5.27 on p.107? Do not use the symbols  $f_\star$  and  $f^\star$  when you write the formulas!

## 5.4 Addenda to Ch.5

**Example 5.27.** Let  $X$  and  $Y$  be nonempty sets and  $f : \mathcal{X} \rightarrow \mathcal{Y}$ . We define two functions  $f_\star$  and  $f^\star$  which are associated with  $f$  and map sets into sets as follows.

- a.  $f_\star : 2^X \rightarrow 2^Y; \quad A \mapsto f_\star(A) := \{f(a) : a \in A\},$   
b.  $f^\star : 2^Y \rightarrow 2^X; \quad B \mapsto f^\star(B) := \{x \in X : f(x) \in B\}.$

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  $X$ . We will encounter those two functions again in ch.7.1 (Direct Images and Indirect Images (Preimages) of a Function), with different notations, as the **direct image**  $f(A)$  (rather than  $f_\star(A)$ ) of  $A \subseteq X$  and the **indirect image** or **preimage**  $f^{-1}(B)$  (rather than  $f^\star(B)$ ) of  $B \subseteq Y$ . See def.7.1 on p.118.

**Example 5.28.** Compare the following to example 5.19.

Let  $a \in \mathbb{R} \cup \{-\infty\}$  and  $b \in \mathbb{R} \cup \{\infty\}$  and let  $X := ]a, b[$  be the open (end points  $a, b$  are excluded) interval of all real numbers between  $a$  and  $b$ . Let  $x_0 \in ]a, b[$  be “fixed but arbitrary”. For example, we could choose  $x_0 := \frac{a+b}{2}$ . Let

$$\begin{aligned}\mathcal{F} &:= \{f : f \text{ is a real valued function with domain } ]a, b[ \}, \\ \mathcal{C} &:= \{f : ]a, b[ \rightarrow \mathbb{R} \text{ such that } f \text{ is continuous on } ]a, b[ \}, \\ \mathcal{D} &:= \{f : ]a, b[ \rightarrow \mathbb{R} \text{ such that } f \text{ is differentiable on } ]a, b[ \}.\end{aligned}$$

Note that  $\mathcal{D} \subseteq \mathcal{C} \subseteq \mathcal{F}$  because differentiable functions are continuous. We define the following equivalence relation on  $\mathcal{D}$ :  $f \sim g \Leftrightarrow f - g = \text{const}$ .<sup>60</sup> Let

$$\mathcal{A} := \{[f] : f \in \mathcal{D}\}$$

be the set of all equivalence classes of differentiable functions on  $]a, b[$ . Then

$$I : \mathcal{C} \rightarrow \mathcal{A}; \quad f \mapsto [I(f)] \quad \text{where } I(f) : ]a, b[ \rightarrow \mathbb{R} \text{ is the function } x \mapsto I(f)(x) := \int_{x_0}^x f(u) du,$$

is a function whose domain  $\mathcal{C}$  is a set of functions and whose codomain  $\mathcal{A}$  is a set of equivalence classes (i.e., sets(!)) of functions.

*The following remark belongs after prop.7.7 on p.124.*

**Remark 5.19.** It follows from prop.7.7 parts **a** and **b** together with thm.5.1 (Characterization of inverse functions) on p.95 that if  $f : X \rightarrow Y$  is a bijection between two nonempty sets  $X$  and  $Y$  then the direct image function  $f : 2^X \rightarrow 2^Y; \quad A \mapsto \{f(a) : a \in A\}$  is a bijection between the two powersets of  $X$  and  $Y$ , and its inverse is the preimage function  $f^{-1} : 2^Y \rightarrow 2^X; \quad B \mapsto \{x \in X : f(x) \in B\}$ .

---

<sup>60</sup> Note that  $f \sim g \Leftrightarrow f - g = \text{const}$  also defines equivalence relations on the supersets  $\mathcal{C}$  and  $\mathcal{F}$ .

## 6 More on Sets (Understand this!)

### 6.1 More on Set Operations (Study this!)

The material in this chapter thematically belongs to ch.2.1 on p.11 but it had to be deferred to this chapter as much of it deals with families of sets, i.e., families  $(A_i)_i$

**Definition 6.1** (Arbitrary unions and intersections). Let  $J$  be a nonempty set and let  $(A_i)_{i \in J}$  be a family of sets. We define

$$(6.1) \quad \bigcup_{i \in I} A_i := \bigcup [A_i : i \in I] := \{x : x \in A_{i_0} \text{ for some } i_0 \in I\},$$

$$(6.2) \quad \bigcap_{i \in I} A_i := \bigcap [A_i : i \in I] := \{x : x \in A_{i_0} \text{ for each } i_0 \in I\}.$$

We call  $\bigcup_{i \in I} A_i$  the **union** and  $\bigcap_{i \in I} A_i$  the **intersection** of the family  $(A_i)_{i \in J}$

It is convenient to allow unions and intersections for the empty index set  $J = \emptyset$ . For intersections this requires the existence of a universal set  $\Omega$ . We define

$$(6.3) \quad \bigcup_{i \in \emptyset} A_i := \emptyset, \quad \bigcap_{i \in \emptyset} A_i := \Omega. \quad \square$$

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

We give some examples of non-finite unions and intersections. For the first one (prop.6.1) We need to review the behavior of the sequence  $x_n = 1/n$ .

**Remark 6.1.** The following definition of the limit of a sequence is known from calculus (see [11] Stewart, J: Single Variable Calculus, 7th edition, ch.11.1) and it will be repeated in ch.9.2 (Convergence and Continuity in  $\mathbb{R}$ ) of this document (see def.9.10 on p.150).

We say that a sequence  $(x_n)$  of real numbers **converges** to  $a \in \mathbb{R}$  as  $n \rightarrow \infty$  if for any  $\varepsilon \in \mathbb{R}$  (no matter how small) there exists a corresponding (possibly extremely large)  $N \in \mathbb{N}$  such that

$$(6.4) \quad |x_j - a| < \varepsilon \text{ for all } j \geq N.$$

We write

$$(6.5) \quad a = \lim_{n \rightarrow \infty} x_n \quad \text{or} \quad x_n \rightarrow a$$

and we call  $a$  the **limit** of the sequence  $(x_n)$ .  $\square$

**Remark 6.2.** It is known from calculus that  $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ . According to (6.4) this implies that for any  $\varepsilon > 0$  there exists  $N \in \mathbb{N}$  such that

$$(6.6) \quad \frac{1}{n} = \left| \frac{1}{n} - 0 \right| < \varepsilon \text{ for all } n \geq N. \quad \square$$

**Proposition 6.1.** For the following note that  $[u, v] = \emptyset$  for  $u > v$  and  $]u, v[ = \emptyset$  for  $u \geq v$  (see (2.14) on p.18). Let  $a, b \in \mathbb{R}$ . Then

$$(6.7) \quad [a, b] = \bigcap_{n \in \mathbb{N}} \left] a - \frac{1}{n}, b + \frac{1}{n} \right[.$$

$$(6.8) \quad ]a, b[ = \bigcup_{n \in \mathbb{N}} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right],$$

**A. PROOF of (6.7)**

**Case 1:** We assume that  $a > b$ . Then  $[a, b] = \emptyset$ , hence (6.7) 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  $\varepsilon > 0$ . According to (6.6) 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

$$\frac{2}{N} < a - b, \quad \text{hence } b + \frac{1}{N} < a - \frac{1}{N}, \quad \text{hence } \left] a - \frac{1}{N}, b + \frac{1}{N} \right[ = \emptyset.$$

**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  $\{a\} \subseteq \bigcap_n ]a - \frac{1}{n}, b + \frac{1}{n}[$ . 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  $\varepsilon := a - x > 0$ . According to (6.6) 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

$$(6.9) \quad a - \frac{1}{n} < x < b + \frac{1}{n} \quad \text{for all } n \in \mathbb{N}.$$

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  $\varepsilon > 0$ . According to (6.6) 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 (6.9) and we have proved that  $x \geq a$ . Demonstrating that  $x \leq b$  is similar. We have proved “ $\supseteq$ ”.

**B. PROOF of (6.8):** This proof is left as exercise 6.1. ■

**Example 6.1.** For any set  $A$  we have  $A = \bigcup_{a \in A} \{a\}$ . According to (6.3) this also is true if  $A = \emptyset$ . □

The following trivial lemma (a lemma is a “proof subroutine” which is not remarkable on its own but very useful as a reference for other proofs) is useful if you need to prove statements of the form  $A \subseteq B$  or  $A = B$  for sets  $A$  and  $B$ . It is a means to simplify the proofs of [1] B/G (Beck/Geoghegan), project 5.12. You must reference this lemma as the “inclusion lemma” when you use it in your homework or exams. Be sure to understand what it means if you choose  $J = \{1, 2\}$  (draw one or two Venn diagrams).

**Lemma 6.1 (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

$$(6.10) \quad U \subseteq \bigcap_{j \in J} X_j \subseteq Y \subseteq \bigcup_{j \in J} Z_j \subseteq W.$$

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  $x \in \bigcap_{j \in J} X_j$ . This proves the first inclusion.
- b. Now let  $x \in \bigcap_{j \in J} X_j$  and  $j_0 \in J$ . Then  $x \in X_{j_0}$  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.,  $x \in \bigcup_{j \in J} Z_j$ . This proves the third inclusion.
- d. Finally, assume  $x \in \bigcup_{j \in J} Z_j$ . 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  $\bigcup_{j \in J} Z_j \subseteq W$ . This finishes the proof of the rightmost inclusion. ■

**Definition 6.2** (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. □

**Definition 6.3** (Partition). Let  $\mathfrak{A} \subseteq 2^\Omega$ . We call  $\mathfrak{A}$  a **partition** or a **partitioning** of  $\Omega$  if

$$\mathbf{a.} \ A \cap B = \emptyset \text{ for any two } A, B \in \mathfrak{A}, \quad \mathbf{b.} \ \Omega = \bigsqcup [A : A \in \mathfrak{A}].$$

We extend this definition to arbitrary families and hence finite collections and sequences of subsets of  $\Omega$ : Let  $J$  be an arbitrary nonempty set, let  $(A_j)_{j \in J}$  be a family of subsets of  $\Omega$ . We call  $(A_j)_{j \in J}$  a partition of  $\Omega$  if it is a mutually disjoint family which satisfies  $\Omega = \bigsqcup [A_j : j \in J]$ , i.e., if  $\mathfrak{A} := \{A_j : j \in J\}$  is a partition of  $\Omega$ .

Note that duplicate nonempty sets cannot occur in a disjoint family of sets because otherwise the condition of mutual disjointness does not hold. □

**Example 6.2.** Here are some examples of partitions.

- a. For any set  $\Omega$  the collection  $\{\{\omega\} : \omega \in \Omega\}$  is a partition of  $\Omega$ .
- b. The empty set is a partition of the empty set and it is its only partition. Do you see that this is a special case of a?
- c. The set of half open intervals  $\{]k, k + 1] : k \in \mathbb{Z}\}$  is a partitioning of  $\mathbb{R}$ .
- d. Given is a strictly increasing sequence  $n_0 = 0 < n_1 < n_2 < \dots$  of non-negative integers. For  $k \in \mathbb{N}$  let  $A_k := \{j \in \mathbb{N} : n_{k-1} < j \leq n_k\}$ . Then the set  $\{A_k : k \in \mathbb{N}\}$  is a partition of  $\mathbb{N}$  (**not** of  $\mathbb{Z}_{\geq 0}$ !) □

**Theorem 6.1** (De Morgan's Law). Let there be a universal set  $\Omega$  (see (2.6) on p.14). Then the following "duality principle" holds for any indexed family  $(A_\alpha)_{\alpha \in I}$  of sets:

$$(6.11) \quad \mathbf{a.} \ \left(\bigcup_{\alpha} A_{\alpha}\right)^c = \bigcap_{\alpha} A_{\alpha}^c \quad \mathbf{b.} \ \left(\bigcap_{\alpha} A_{\alpha}\right)^c = \bigcup_{\alpha} A_{\alpha}^c$$

To put this in words, the complement of an arbitrary union is the intersection of the complements and the complement of an arbitrary intersection is the union of the complements.

Generally speaking the formulas are a consequence of the duality principle for set operations which states that any true statement involving a family of subsets of a universal sets can be converted into its “dual” true statement by replacing all subsets with their complements, all unions with intersections and all intersections with unions.

PROOF of De Morgan’s law, formula a:

1) First we prove that  $(\bigcup_{\alpha} A_{\alpha})^c \subseteq \bigcap_{\alpha} A_{\alpha}^c$ :

Assume that  $x \in (\bigcup_{\alpha} A_{\alpha})^c$ . Then  $x \notin \bigcup_{\alpha} A_{\alpha}$  which is the same as saying that  $x$  does not belong to any of the  $A_{\alpha}$ . That means that  $x$  belongs to each  $A_{\alpha}^c$  and hence also to the intersection  $\bigcap_{\alpha} A_{\alpha}^c$ .

2) Now we prove that  $(\bigcup_{\alpha} A_{\alpha})^c \supseteq \bigcap_{\alpha} A_{\alpha}^c$ : Let  $x \in \bigcap_{\alpha} A_{\alpha}^c$ . Then  $x$  belongs to each of the  $A_{\alpha}^c$  and hence to none of the  $A_{\alpha}$ . Then it also does not belong to the union of all the  $A_{\alpha}$  and must therefore belong to the complement  $(\bigcup_{\alpha} A_{\alpha})^c$ . This completes the proof of formula a. The proof of formula b is very similar and given as exercise 6.4 on p.118. ■

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 6.2** (Distributivity of unions and intersections). Let  $(A_i)_{i \in I}$  be an arbitrary family of sets and let  $B$  be a set. Then

$$(6.12) \quad \bigcup_{i \in I} (B \cap A_i) = B \cap \bigcup_{i \in I} A_i,$$

$$(6.13) \quad \bigcap_{i \in I} (B \cup A_i) = B \cup \bigcap_{i \in I} A_i.$$

PROOF: We only prove (6.12). The proof of (6.13) is left as exercise 7.4.

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  $(\bigcup_i A_i) \cap B$ .

PROOF of “ $\supseteq$ ”: Let  $x \in B \cap \bigcup_i A_i$ . Then  $x \in B$  and  $x \in A_{i^*}$  for some  $i^* \in I$ , hence  $x \in B \cap A_{i^*}$ , hence  $x \in \bigcup_i (B \cap A_i)$ . ■

**Proposition 6.3** (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  $\Omega$ . For  $n \in \mathbb{N}$  let  $B_n := \bigcup_{j=1}^n A_j = A_1 \cup A_2 \cup \dots \cup A_n$

Further, let  $C_1 := A_1 = B_1$  and  $C_{n+1} := A_{n+1} \setminus B_n$  ( $n \in \mathbb{N}$ ). Then

a. The sequence  $(B_j)_j$  is increasing:  $m < n \Rightarrow B_m \subseteq B_n$ ,

b. For each  $n \in \mathbb{N}$ ,  $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^n B_j$ ,

c. The sets  $C_j$  are mutually disjoint and  $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^n B_j = \bigoplus_{j=1}^n C_j$ .

PROOF of a and of b: Left as exercise 6.2 (p.117).



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

$$C_j \cap C_{n+1} \subseteq B_n \cap C_{n+1} = B_n \cap (A_{n+1} \setminus B_n) = B_n \cap (A_{n+1} \cap B_n^c) = A_{n+1} \cap (B_n \cap B_n^c) = \emptyset.$$

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  $x \in \biguplus_{j=1}^n C_j$ . ■

## 6.2 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.

For the following definition see also [1] B/G (Beck/Geoghegan), ch.4.3 (Fishing in a Finite Pool).

**Definition 6.4** (Vectors and finite lists). Let  $n_*, n^* \in \mathbb{Z}$  such that  $n_* \leq n^*$  and let  $J := \{k \in \mathbb{Z} : n_* \leq k \leq n^*\}$ . A **finite list**  $\chi$ , is a family with (finite) index set  $J$ . It is customary to write one of the following:

$$(6.14) \quad \chi = x_{n_*}, x_{n_*+1}, x_{n_*+2}, \dots, x_{n^*-1}, x_{n^*},$$

$$(6.15) \quad \chi = (x_{n_*}, x_{n_*+1}, x_{n_*+2}, \dots, x_{n^*-1}, x_{n^*}),$$

$$(6.16) \quad \chi = (x_k)_{n_*}^{n^*}.$$

If all members  $x_j$  of the list are (real) numbers, we call  $\chi$  a **vector**<sup>61</sup> of dimension  $n^*$ . In this case we always surround the list with parentheses, and often use "arrow notation"  $\vec{x}$  rather than  $\chi$ , i.e., we write

$$(6.17) \quad \vec{x} = (x_1, x_2, x_3, \dots, x_{n-1}, x_n).$$

### Example 6.3.

- a.  $(3.5, -97\pi, 4, \sqrt{8})$  is both a list and a vector of dimension 4. We have  $x_1 = 3.5, x_2 = -97\pi, x_3 = 4, x_4 = \sqrt{8}$ .
2.  $(3k + 2)_{k=-2}^3 = -4, -1, 2, 5, 8, 11 = (-4, -1, 2, 5, 8, 11)$  is both a list and a vector of size/dimension 6.
3. Joe, 5,  $-6.8$ , Dolores is a list of size 4. This is not a vector because not all of its members are numeric.

<sup>61</sup> Vectors can be of a more general nature than just being a finite list of numbers. See ch.10.2 (General Vector Spaces) on p.178 (General Vector Spaces).

**Remark 6.3** (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

$$F : (A \times B) \times C \rightarrow A \times (B \times C), \quad ((a, b), c) \mapsto (a, (b, c))$$

is bijective because it has the mapping

$$G : A \times (B \times C) \rightarrow (A \times B) \times C, \quad (a, (b, c)) \mapsto ((a, b), c)$$

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  $(a, b, c) \mapsto ((a, b), c)$  and  $(a, b, c) \mapsto (a, (b, c))$ .  $\square$

*This remark leads us to the following definition.*

**Definition 6.5** (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 <sup>62</sup>

$$X_1 \times X_2 \times X_3 \times \dots \times X_N := \{(x_1, x_2, \dots, x_N) : x_j \in X_j \text{ for all } 1 \leq j \leq N\}$$

Note that the elements of this set are finite lists in the sense of def.6.4 (Vectors and finite lists) on p.113.

Two elements  $(x_1, x_2, \dots, x_N)$  and  $(y_1, y_2, \dots, y_N)$  of  $X_1 \times X_2 \times X_3 \times \dots \times X_N$  are called **equal** if and only if  $x_j = y_j$  for all  $j$  such that  $1 \leq j \leq N$ . In this case we write  $(x_1, x_2, \dots, x_N) = (y_1, y_2, \dots, y_N)$ .

As a shorthand, we abbreviate  $X^N := \underbrace{X \times X \times \dots \times X}_{N \text{ times}}$ .  $\square$

**Example 6.4** ( $N$ -dimensional coordinates). Here is the most important example of a cartesian product of  $N$  sets. Let  $X_1 = X_2 = \dots = X_N = \mathbb{R}$ . Then

$$\mathbb{R}^N = \{(x_1, x_2, \dots, x_N) : x_j \in \mathbb{R} \text{ for } 1 \leq j \leq N\}$$

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  $N = 3$ .

<sup>62</sup> 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.

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  $N \leq 3$  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  $N$ .<sup>63</sup> You will read more about this in ch.10 about vectors and vector spaces on page 174.

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 10, p.174 on vectors and vector spaces.  $\square$

**Example 6.5** (Parallelepipeds). Let  $a_1 < b_1, a_2 < b_2, a_3 < b_3$  be real numbers. Then

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] = \{(x, y, z) : a_1 \leq x \leq b_1, a_2 \leq y \leq b_2, a_3 \leq z \leq b_3\}$$

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

$$[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_N, b_N] = \{(x_1, x_2, \dots, x_N) : a_j \leq x_j \leq b_j, j \in \mathbb{N}, j \leq N\}$$

is the parallelepiped with sides  $[a_1, b_1], \dots, [a_N, b_N]$ .  $\square$

We now introduce cartesian products of an entire family of sets  $(X_i)_{i \in I}$ .

**Definition 6.6** (Cartesian Product of a family of sets). 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

$$\prod_{i \in I} X_i := \left( \prod_{i \in I} X_i \right) := \{(x_i)_{i \in I} : x_k \in X_k \forall k \in I\}$$

of all families  $(x_i)_{i \in I}$  each of whose members  $x_j$  belongs to the corresponding set  $X_j$ . The “ $\prod$ ” is the greek “upper case” letter “Pi” (whose lower case incarnation “ $\pi$ ” you are probably more familiar with).

Two elements  $(x_i)_{i \in I}$  and  $(y_k)_{k \in I}$  of  $\prod_{i \in I} X_i$  are called **equal** if and only if  $x_j = y_j$  for all  $j \in I$ . In this case we write  $(x_i)_{i \in I} = (y_k)_{k \in I}$ .<sup>64</sup>

If all sets  $X_i$  are equal to one and the same set  $X$ , we also write  $X^I := \prod_{i \in I} X := \prod_{i \in I} X_i$ .  $\square$

<sup>63</sup> See def.6.4 (Vectors and finite lists) on p.113.

<sup>64</sup> In other words, if and only if those two families are equal in the sense of def.5.19 on p.102.

**Remark:**

Note that, because  $I$  is not empty,  $\prod_{i \in I} X_i = \emptyset \Leftrightarrow$  there exists some  $i \in I$  such that  $X_i = \emptyset$ .

**Remark 6.4.** We note that each element  $(y_x)_{x \in X}$  of the cartesian product  $Y^X$  is the function

$$y(\cdot) : X \rightarrow Y, \quad x \mapsto y_x$$

(see def.5.18 (indexed families) and the subsequent remarks concerning the equivalence of functions and families). In other words,

$$(6.18) \quad Y^X = \{f : f \text{ is a function with domain } X \text{ and codomain } Y\}. \quad \square$$

### 6.3 Rings and Algebras of Sets ( $\star$ )

*Note that this chapter is starred, hence optional.*

*You will benefit from examining the proof of prop.6.4 on p.116 and learn how to split a proof which involves 3 or 4 sets can be split into easily dealt with cases.*

**Definition 6.7** (Rings and Algebras of Sets). A subset  $\mathcal{R}$  of  $2^\Omega$  (a set of sets!) is called a **ring of sets** if it is closed with respect to the operations “ $\cup$ ” and “ $\setminus$ ”, i.e.,

$$(6.19) \quad R_1 \cup R_2 \in \mathcal{R} \text{ and } R_1 \setminus R_2 \in \mathcal{R} \text{ whenever } R_1, R_2 \in \mathcal{R}.$$

A subset  $\mathcal{A}$  of  $2^\Omega$  is called an **algebra of sets** if  $\Omega \in \mathcal{A}$  and  $\mathcal{A}$  is a ring of sets.  $\square$

**Proposition 6.4. 1.** Let  $\mathcal{R}$  be a ring of sets and  $A, B \in \mathcal{R}$ . Then  $A \Delta B \in \mathcal{R}$  and  $A \cap B \in \mathcal{R}$ .

2. Let  $\mathcal{R}$  be a ring of sets and  $A, B, C \in \mathcal{R}$ . Then

- a.  $(A \Delta B) \Delta C = A \Delta (B \Delta C)$  (associativity of  $\Delta$ )
- b.  $A \Delta \emptyset = \emptyset \Delta A = A$  (neutral element  $\emptyset$  for  $\Delta$ )
- c.  $A \Delta A = \emptyset$  (inverse element  $\emptyset$  for  $\Delta$ )<sup>65</sup>
- d.  $A \Delta B = B \Delta A$  (commutativity of  $\Delta$ )

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  $\Omega$  for  $\cap$ )
- g.  $A \cap B = B \cap A$  (commutativity of  $\cap$ )

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

- h.  $A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$  (distributivity)

**Solution:**

*PROOF of a.* The proof given here is based on with set membership. It is very tedious and there is a much more elegant proof but it requires knowledge of indicator functions<sup>66</sup> and of base 2 modular arithmetic (see, e.g., [1] B/G (Beck/Geoghegan) ch.6.2).

<sup>65</sup>The inverse element for  $A$  in the sense of def.3.3 on p.30. is  $A$  itself!

<sup>66</sup>Indicator functions will be discussed in ch.7.2 on p.125 and in ch.9.5 on p.167.

By definition  $x \in U \Delta V$  if and only if either  $x \in U$  or  $x \in V$ , i.e.,  
(either)  $[x \in U \text{ and } x \notin V]$  or  $[x \in V \text{ and } x \notin U]$

Hence  $x \in (A \Delta B) \Delta C$  means either  $x \in (A \Delta B)$  or  $x \in C$ , i.e.,  
either  $[x \in A, x \notin B \text{ or } x \in B, x \notin A]$  or  $x \in C$ , i.e., we have one of the following four combinations:

- a.  $x \in A \quad x \notin B \quad x \notin C$
- b.  $x \notin A \quad x \in B \quad x \notin C$
- c.  $x \in A \quad x \in B \quad x \in C$
- d.  $x \notin A \quad x \notin B \quad x \in C$

and  $x \in A \Delta (B \Delta C)$  means either  $x \in A$  or  $x \in (B \Delta C)$ , i.e.,  
either  $x \in A$  or  $[x \in B, x \notin C \text{ or } x \in C, x \notin B]$ , i.e., we have one of the following four combinations:

- 1.  $x \in A \quad x \in B \quad x \in C$
- 2.  $x \in A \quad x \notin B \quad x \notin C$
- 3.  $x \notin A \quad x \in B \quad x \notin C$
- 4.  $x \notin A \quad x \notin B \quad x \in C$

We have a perfect match  $\mathbf{a} \leftrightarrow \mathbf{2}$ ,  $\mathbf{b} \leftrightarrow \mathbf{3}$ ,  $\mathbf{c} \leftrightarrow \mathbf{1}$ ,  $\mathbf{d} \leftrightarrow \mathbf{4}$ . and this completes the proof of **a**.

■

**Remark 6.5** (Rings of sets as rings). A set  $G = (G, \oplus)$  with a “binary operation”  $g_3 = g_1 \oplus g_2$  – i.e., an operation  $\oplus$  which assigns to any two  $g_1, g_2 \in G$  a third element  $g_3 \in G$  – such that  $\oplus$  satisfies **a – d** above with  $\Delta$  playing the role of  $\oplus$  is called an abelian or commutative group (see ch.14 on p.289.)

A set  $R = (R, \oplus, \odot)$  with two binary operations  $g_3 = g_1 \oplus g_2$  and  $g_4 = g_1 \odot g_2$  which satisfies **a – h** above with  $\Delta$  playing the role of  $\oplus$  and  $\cap$  playing the role of  $\odot$  is called a commutative ring with multiplicative unit.  $\oplus$  is customarily referred to as “addition” and  $\odot$  is customarily referred to as “multiplication”. This explains the name “ring of sets” for  $\mathcal{R} = (\mathcal{R}, \Delta, \cap)$ .

We note that rings of sets satisfy axioms 1.1 – 1.4 but not axiom 1.5 (the cancellation axiom) of [1] B/G (Beck/Geoghegan) ch.1.1. This explains the name “ring of sets”.

The name “algebra of sets” for a ring of sets which contains  $\Omega$  stems from the fact that such systems of subsets of  $\Omega$  are “boolean algebras”. □

## 6.4 Exercises for Ch.6

**Exercise 6.1.** Prove (6.8) of prop.6.1 on p.110: Let  $a, b \in \mathbb{R}$ . Then  $]a, b[ = \bigcup_{n \in \mathbb{N}} [a + 1/n, b - 1/n]$ .

Adapt the proof of (6.7) but note that this one is simpler. There are only two cases to worry about:  $a \geq b$  (very easy!) vs  $a < b$ . □

**Exercise 6.2.** Prove **a** and **b** of prop.6.3 (Rewrite unions as disjoint unions) on p.112:

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  $B_n := \bigcup_{j=1}^n A_j = A_1 \cup A_2 \cup \dots \cup A_n$

Further, let  $C_1 := A_1 = B_1$  and  $C_{n+1} := A_{n+1} \setminus B_n$  ( $n \in \mathbb{N}$ ). Then

**a.** The sequence  $(B_j)_j$  is increasing:  $m < n \Rightarrow B_m \subseteq B_n$ ,

b. For each  $n \in \mathbb{N}$ ,  $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^n B_j$ .  $\square$

**Exercise 6.3.** (See example 5.23 on p.103). Let  $X := [0, 2]$ . For  $0 \leq x \leq 2$  let  $A_x := [x, 2x]$ .

a. What is  $\bigcap [A_x : x \in X]$ ?      b. What is  $\bigcup [A_x : x \in X]$ ?  $\square$

**Exercise 6.4.** Prove b of thm.6.1 (De Morgan's Law):

Let  $(A_\alpha)_{\alpha \in I}$  be a family of subsets of a universal set  $\Omega$ . Then  $(\bigcap_{\alpha} A_\alpha)^c = \bigcup_{\alpha} A_\alpha^c$ .  $\square$

**Exercise 6.5.** Prove the second formula of prop.6.2 (Distributivity of unions and intersections): Let  $(A_i)_{i \in I}$  be an arbitrary family of sets and let  $B$  be a set. Then

$$\bigcap_{i \in I} (B \cup A_i) = B \cup \bigcap_{i \in I} A_i. \quad \square$$

**Exercise 6.6.** Let  $(G, \diamond)$  be a group, let  $(H_i)_{i \in J}$  be a family of subgroups of  $G$ , and let  $H := \bigcap_{i \in J} H_i$ .

Then  $H$  is a subgroup of  $G$ .  $\square$

**Exercise 6.7.** Let  $f$  be the function  $f : [-3, 3] \rightarrow \mathbb{R}; \quad x \mapsto x^2$ .

- 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?  $\square$

## 7 Sets and Functions, Direct and Indirect Images (Study this!)

### 7.1 Direct Images and Indirect Images (Preimages) of a Function

**Definition 7.1.** Let  $X, Y$  be two nonempty sets and  $f : X \rightarrow Y$  be an arbitrary function with domain  $X$  and codomain  $Y$ . Let  $A \subseteq X$  and  $B \subseteq Y$ . Let

$$(7.1) \quad f(A) := \{f(x) : x \in A\},$$

$$(7.2) \quad f^{-1}(B) := \{x \in X : f(x) \in B\}.$$

We call  $f(A)$  the **direct image**<sup>67</sup> of  $A$  under  $f$  and we call  $f^{-1}(B)$  the **indirect image** or **preimage** of  $B$  under  $f$ .  $\square$

#### *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 whatsoever. 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.

<sup>67</sup> The range  $f(X)$  of  $f$  (see (5.9) on p.88) is a special case of a direct image.

**Notational conveniences II:**

In measure theory and probability theory the following notation is also very common:

$$\{f \in B\} := f^{-1}(B), \quad \{f = y\} := f^{-1}\{y\}.$$

Let  $a < b \in \mathbb{R}$ . We write  $\{a \leq f \leq b\} := f^{-1}([a, b])$ ,  $\{a < f < b\} := f^{-1}(]a, b[)$ ,

$$\{a \leq f < b\} := f^{-1}([a, b[)$$
,  $\{a < f \leq b\} := f^{-1}(]a, b])$ ,  $\{f \leq b\} := f^{-1}(]-\infty, b])$ , etc.

**Example 7.1 (Direct images).** Let  $f : \mathbb{R} \rightarrow \mathbb{R}; \quad f(x) = x^2$ .

- $f(]-4, -2]) = \{x^2 : x \in ]-4, -2]) = \{x^2 : -4 < x < -2\} = ]4, 16[$ .
- $f([1, 2]) = \{x^2 : x \in [1, 2]) = \{x^2 : 1 \leq x \leq 2\} = [1, 4]$ .
- $f([5, 6]) = \{x^2 : x \in [5, 6]) = \{x^2 : 5 \leq x \leq 6\} = [25, 36]$ .
- $f(]-4, -2[ \cup [1, 2] \cup [5, 6]) = \{x^2 : x \in ]-4, -2[ \text{ or } x \in [1, 2] \text{ or } x \in [5, 6]\} = ]4, 16[ \cup [1, 4] \cup [25, 36] = [1, 16[ \cup [25, 36]$ .  $\square$

**Example 7.2 (Direct images).** Let  $f : \mathbb{R} \rightarrow \mathbb{R}; \quad f(x) = x^2$ .

- $f(]-4, 2]) = \{x^2 : x \in ]-4, 2]) = \{x^2 : -4 < x < 2\} = ]4, 16[$ .
- $f([1, 3]) = \{x^2 : x \in [1, 3]) = \{x^2 : 1 \leq x \leq 3\} = [1, 9]$ .
- $f(]-4, 2[ \cap [1, 3]) = \{x^2 : x \in ]-4, 2[ \text{ and } x \in [1, 3]\} = \{x^2 : 1 \leq x < 2\} = [1, 4[$ .  $\square$

And here are the results for the preimages of the same sets with respect to the same function  $x \mapsto x^2$ .

**Example 7.3 (Preimages).** Let  $f : \mathbb{R} \rightarrow \mathbb{R}; \quad f(x) = x^2$ .

- $f^{-1}(]-4, -2]) = \{x \in \mathbb{R} : x^2 \in ]-4, -2]) = \{-4 < f < -2\} = \emptyset$ .
- $f^{-1}([1, 2]) = \{x \in \mathbb{R} : x^2 \in [1, 2]) = \{1 \leq f \leq 2\} = [-\sqrt{2}, -1] \cup [1, \sqrt{2}]$ .
- $f^{-1}([5, 6]) = \{x \in \mathbb{R} : x^2 \in [5, 6]) = \{5 \leq f \leq 6\} = [-\sqrt{6}, -\sqrt{5}] \cup [\sqrt{5}, \sqrt{6}]$ .
- $f^{-1}(]-4, -2[ \cup [1, 2] \cup [5, 6]) = \{x \in \mathbb{R} : x^2 \in ]-4, -2[ \text{ or } x^2 \in [1, 2] \text{ or } x^2 \in [5, 6]\} = [-\sqrt{2}, -1] \cup [1, \sqrt{2}] \cup [-\sqrt{6}, -\sqrt{5}] \cup [\sqrt{5}, \sqrt{6}]$ .  $\square$

**Example 7.4 (Preimages).** Let  $f : \mathbb{R} \rightarrow \mathbb{R}; \quad f(x) = x^2$ .

- $f^{-1}(]-4, 2]) = \{x \in \mathbb{R} : x^2 \in ]-4, 2]) = \{x \in \mathbb{R} : -4 < x^2 < 2\} = ]-\sqrt{2}, \sqrt{2}[$ .
- $f^{-1}([1, 3]) = \{x \in \mathbb{R} : x^2 \in [1, 3]) = \{x \in \mathbb{R} : 1 \leq x^2 \leq 3\} = [-\sqrt{3}, -1] \cup [1, \sqrt{3}]$ .
- $f^{-1}(]-4, 2[ \cap [1, 3]) = \{x \in \mathbb{R} : x^2 \in ]-4, 2[ \text{ and } x^2 \in [1, 3]\} = \{x \in \mathbb{R} : 1 \leq x^2 < 2\} = ]-\sqrt{2}, -1] \cup [1, \sqrt{2}[$ .  $\square$

Unless stated otherwise,  $X, Y$  and  $f$  are as defined above for the remainder of this chapter:  $f : X \rightarrow Y$  is a function with domain  $X$  and codomain  $Y$ .

**Proposition 7.1.** Some simple properties:

$$(7.3) \quad f(\emptyset) = f^{-1}(\emptyset) = \emptyset$$

$$(7.4) \quad A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2) \quad (\text{monotonicity of } f\{\dots\})$$

$$(7.5) \quad B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2) \quad (\text{monotonicity of } f^{-1}\{\dots\})$$

$$(7.6) \quad x \in X \Rightarrow f(\{x\}) = \{f(x)\}$$

$$(7.7) \quad f(X) = Y \Leftrightarrow f \text{ is surjective}$$

$$(7.8) \quad f^{-1}(Y) = X \quad \text{always!}$$

PROOF: Left as an exercise. ■

**Proposition 7.2** ( $f^{-1}$  is compatible with all basic set ops). In the following we assume that  $J$  is an arbitrary index set, and that  $B \subseteq Y$ ,  $B_j \subseteq Y$  for all  $j$ . Then

$$(7.9) \quad f^{-1}\left(\bigcap_{j \in J} B_j\right) = \bigcap_{j \in J} f^{-1}(B_j)$$

$$(7.10) \quad f^{-1}\left(\bigcup_{j \in J} B_j\right) = \bigcup_{j \in J} f^{-1}(B_j)$$

$$(7.11) \quad f^{-1}(B^c) = f^{-1}(B)^c$$

$$(7.12) \quad f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$$

$$(7.13) \quad f^{-1}(B_1 \Delta B_2) = f^{-1}(B_1) \Delta f^{-1}(B_2)$$

PROOF of (7.9): Let  $x \in X$ . Then

$$(7.14) \quad \begin{aligned} x \in f^{-1}\left(\bigcap_{j \in J} B_j\right) &\Leftrightarrow f(x) \in \bigcap_{j \in J} B_j \quad (\text{def preimage}) \\ &\Leftrightarrow \forall j \, f(x) \in B_j \quad (\text{def } \cap) \\ &\Leftrightarrow \forall j \, x \in f^{-1}(B_j) \quad (\text{def preimage}) \\ &\Leftrightarrow x \in \bigcap_{j \in J} f^{-1}(B_j) \quad (\text{def } \cap) \end{aligned}$$

PROOF of (7.10): Let  $x \in X$ . Then

$$(7.15) \quad \begin{aligned} x \in f^{-1}\left(\bigcup_{j \in J} B_j\right) &\Leftrightarrow f(x) \in \bigcup_{j \in J} B_j \quad (\text{def preimage}) \\ &\Leftrightarrow \exists j_0 : f(x) \in B_{j_0} \quad (\text{def } \cup) \\ &\Leftrightarrow \exists j_0 : x \in f^{-1}(B_{j_0}) \quad (\text{def preimage}) \\ &\Leftrightarrow x \in \bigcup_{j \in J} f^{-1}(B_j) \quad (\text{def } \cup) \end{aligned}$$

PROOF of (7.11): Let  $x \in X$ . Then

$$(7.16) \quad \begin{aligned} x \in f^{-1}(B^c) &\Leftrightarrow f(x) \in B^c \quad (\text{def preimage}) \\ &\Leftrightarrow f(x) \notin B \quad (\text{def } (\cdot)^c) \\ &\Leftrightarrow x \notin f^{-1}(B) \quad (\text{def preimage}) \\ &\Leftrightarrow x \in f^{-1}(B)^c \quad (\cdot)^c \end{aligned}$$

PROOF of (7.12): Let  $x \in X$ . Then

$$(7.17) \quad \begin{aligned} x \in f^{-1}(B_1 \setminus B_2) &\Leftrightarrow x \in f^{-1}(B_1 \cap B_2^c) \quad (\text{def } \setminus) \\ &\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2^c) \quad (\text{see (7.9)}) \\ &\Leftrightarrow x \in f^{-1}(B_1) \cap f^{-1}(B_2)^c \quad (\text{see (7.11)}) \\ &\Leftrightarrow x \in f^{-1}(B_1) \setminus f^{-1}(B_2) \quad (\text{def } \setminus) \end{aligned}$$

PROOF of (7.13): This follows from  $B_1 \Delta B_2 = (B_1 \setminus B_2) \cup (B_2 \setminus B_1)$  and (7.10) and (7.12). ■



**Proposition 7.3** (Properties of the direct image). *In the following we assume that  $J$  is an arbitrary index set, and that  $A \subseteq X$ ,  $A_j \subseteq X$  for all  $j$ . Then*

$$(7.18) \quad f\left(\bigcap_{j \in J} A_j\right) \subseteq \bigcap_{j \in J} f(A_j)$$

$$(7.19) \quad f\left(\bigcup_{j \in J} A_j\right) = \bigcup_{j \in J} f(A_j)$$

*PROOF of (7.18): This follows from the monotonicity of the direct image (see 7.4):*

$$\begin{aligned} \bigcap_{j \in J} A_j \subseteq A_i \quad \forall i \in J &\Rightarrow f\left(\bigcap_{j \in J} A_j\right) \subseteq f(A_i) \quad \forall i \in J \\ &\Rightarrow f\left(\bigcap_{j \in J} A_j\right) \subseteq \bigcap_{i \in J} f(A_i) \quad (\text{def } \cap) \end{aligned}$$

*First proof of (7.19) - "Expert proof":*

$$(7.20) \quad y \in f\left(\bigcup_{j \in J} A_j\right) \Leftrightarrow \exists x \in X : f(x) = y \text{ and } x \in \bigcup_{j \in J} A_j \quad (\text{def } f(A))$$

$$(7.21) \quad \Leftrightarrow \exists x \in X \text{ and } j_0 \in J : f(x) = y \text{ and } x \in A_{j_0} \quad (\text{def } \cup)$$

$$(7.22) \quad \Leftrightarrow \exists x \in X \text{ and } j_0 \in J : f(x) = y \text{ and } f(x) \in f(A_{j_0}) \quad (\text{def } f(A))$$

$$(7.23) \quad \Leftrightarrow \exists j_0 \in J : y \in f(A_{j_0}) \quad (\text{def } f(A))$$

$$(7.24) \quad \Leftrightarrow y \in \bigcup_{j \in J} f(A_j) \quad (\text{def } \cup)$$

*Alternate proof of (7.19) - Proving each inclusion separately. Unless you have a lot of practice reading and writing proofs whose subject is the equality of two sets you should write your proof the following way:*

*A. Proof of " $\subseteq$ ":*

$$(7.25) \quad y \in f\left(\bigcup_{j \in J} A_j\right) \Rightarrow \exists x \in X : f(x) = y \text{ and } x \in \bigcup_{j \in J} A_j \quad (\text{def } f(A))$$

$$(7.26) \quad \Rightarrow \exists j_0 \in J : f(x) = y \text{ and } x \in A_{j_0} \quad (\text{def } \cup)$$

$$(7.27) \quad \Rightarrow y = f(x) \in f(A_{j_0}) \quad (\text{def } f(A))$$

$$(7.28) \quad \Rightarrow y \in \bigcup_{j \in J} f(A_j) \quad (\text{def } \cup)$$

*B. Proof of " $\supseteq$ ":*

*This follows from the monotonicity of  $A \mapsto f(A)$  (see 7.4):*

$$(7.29) \quad A_i \subseteq \bigcup_{j \in J} A_j \quad \forall i \in J \Rightarrow f(A_i) \subseteq f\left(\bigcup_{j \in J} A_j\right) \quad \forall i \in J$$

$$(7.30) \quad \Rightarrow \bigcup_{i \in J} f(A_i) \subseteq f\left(\bigcup_{j \in J} A_j\right) \quad \forall i \in J \quad (\text{def } \cup)$$

■

*The "elementary" proof is barely longer than the first one, but it is so much easier to understand!*

**Remark 7.1.** In general you will not have equality in (7.18). 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,  $\{0\} \subsetneq [0, \infty]$ .  $\square$

**Proposition 7.4** (Direct images and preimages of function composition). *Let  $X, Y, Z$  be an arbitrary, nonempty sets. Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  and let  $W \subseteq Z$ . Then*

$$(7.31) \quad (g \circ f)(U) = g(f(U)) \text{ for all } U \subseteq X.$$

$$(7.32) \quad (g \circ f)^{-1} = f^{-1} \circ g^{-1}, \text{ i.e., } (g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W)) \text{ for all } W \subseteq Z.$$

PROOF of (7.31): Left as exercise ??.

PROOF of (7.32):

a. " $\subseteq$ ": Let  $W \subseteq Z$  and  $x \in (g \circ f)^{-1}(W)$ . Then  $g(f(x)) = (g \circ f)(x) \in W$ , hence  $f(x) \in g^{-1}(W)$ . But then  $x \in f^{-1}(g^{-1}(W))$ . This proves " $\subseteq$ ".

b. " $\supseteq$ ": Let  $W \subseteq Z$ ,  $h := g \circ f$ , and  $x \in f^{-1}(g^{-1}(W))$ . Then  $f(x) \in g^{-1}(W)$ , hence  $h(x) = g(f(x)) \in W$ , hence  $x \in h^{-1}(W) = (g \circ f)^{-1}(W)$ . This proves " $\supseteq$ ".  $\blacksquare$

**Proposition 7.5** (Indirect image and fibers of  $f$ ). *Let  $X, Y$  be nonempty sets and let  $f : X \rightarrow Y$  be a function. We define on the domain  $X$  a relation " $\sim$ " as follows:*

$$(7.33) \quad x_1 \sim x_2 \Leftrightarrow f(x_1) = f(x_2).$$

a. " $\sim$ " is an equivalence relation. Its equivalence classes, which we denote by  $[x]_f$ ,<sup>68</sup> are

$$(7.34) \quad [x]_f = \{a \in X : f(a) = f(x)\} = f^{-1}\{f(x)\}. \quad (x \in X)$$

b. If  $A \subseteq X$  then

$$(7.35) \quad f^{-1}(f(A)) = \bigcup_{a \in A} [a]_f.$$

The proof that " $\sim$ " is an equivalence relation is left as exercise 7.3.

PROOF of (7.34): 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 (7.35):

As  $f(A) = f(\bigcup_{a \in A} \{x\}) = \bigcup_{a \in A} \{f(x)\}$  (see 7.19), it follows that

$$(7.36) \quad f^{-1}(f(A)) = f^{-1}\left(\bigcup_{a \in A} \{f(a)\}\right)$$

$$(7.37) \quad = \bigcup_{a \in A} f^{-1}\{f(a)\} \quad (\text{see 7.10})$$

$$(7.38) \quad = \bigcup_{a \in A} [a]_f \quad (\text{see 7.34})$$

$\blacksquare$

<sup>68</sup>  $[x]_f$  is called the **fiber over**  $f(x)$  of the function  $f$ .

**Corollary 7.1.**

$$(7.39) \quad \text{If } A \subseteq X \text{ then } f^{-1}(f(A)) \supseteq A.$$

PROOF: It follows from  $x \sim x$  for all  $x \in X$  that  $x \in [x]_f$ , i.e.,  $\{x\} \subseteq [x]_f$  for all  $x \in X$ . But then

$$(7.40) \quad A = \bigcup_{x \in A} \{x\} \subseteq \bigcup_{x \in A} [x]_f = f^{-1}(f(A)),$$

where the last equation holds because of (7.35). ■

This next example shows how to work with fibers to prove that certain relations are equivalence relations.

**Example 7.5.** 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.7.5 (Indirect image and fibers of  $f$ ).

For the above examples you do this as follows:

- a.  $F : X \rightarrow \mathbb{R}, \quad x \mapsto |x|$ .
- b.  $G : X \rightarrow \{-1, 1\}, \quad x \mapsto \frac{x}{|x|}$ .
- b.  $H : X \rightarrow \mathbb{R}, \quad (x, y, z) \mapsto z \sin(xy)$ . □

**Proposition 7.6.**

$$(7.41) \quad \text{If } B \subseteq Y \text{ then } f(f^{-1}(B)) = B \cap f(X).$$

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 " $\supseteq$ ":

Let  $y \in f(X) \cap B$ . We must prove that  $y \in f(f^{-1}(B))$ .

It follows from  $y \in f(X)$  that there exists  $x_0 \in X$  such that  $y = f(x_0)$ .

Because  $f(x_0) \in B$ , we conclude that  $x_0 \in f^{-1}(B)$  (def preimage), hence  $y = f(x_0) \in f(f^{-1}(B))$ .

We have shown that if  $y \in f(X)$  and  $y \in B$  then  $y \in f(f^{-1}(B))$ . The proof of " $\supseteq$ " is completed. ■

**Remark 7.2.** Be sure to understand how the assumption  $y \in f(X)$  was used. □

**Corollary 7.2.**

$$(7.42) \quad \text{If } B \subseteq Y \text{ then } f(f^{-1}(B)) \subseteq B.$$

Trivial as  $f(f^{-1}(B)) = B \cap f(X) \subseteq B$ . ■

**Proposition 7.7.**

- a. Let  $A \subseteq X$ . If  $f : X \rightarrow Y$  is injective then  $f^{-1}(f(A)) = A$ .
- b. Let  $B \subseteq Y$ . If  $f : X \rightarrow Y$  is surjective then  $f(f^{-1}(B)) = B$ .
- c. Let  $A \subseteq X$  and  $B \subseteq Y$ . If  $f : X \rightarrow Y$  is injective and if  $B = f(A)$  then  $f^{-1}(B) = A$ .
- d. Let  $A \subseteq X$  and  $B \subseteq Y$ . If  $f : X \rightarrow Y$  is surjective and if  $f^{-1}(B) = A$  then  $B = f(A)$ .
- e. Let  $A \subseteq X$  and  $B \subseteq Y$ . If  $f : X \rightarrow Y$  is bijective then  $B = f(A) \Leftrightarrow f^{-1}(B) = A$ .

PROOF: Left as exercise 7.5.

**Proposition 7.8.** Let  $J$  be an arbitrary nonempty index set and let  $A \subseteq X$ ,  $A_j \subseteq X$  for all  $j$ .

Let  $f : X \rightarrow Y$  be bijective. Then the following all are true:

$$(7.43) \quad f\left(\bigcap_{j \in J} A_j\right) = \bigcap_{j \in J} f(A_j)$$

$$(7.44) \quad f\left(\bigcup_{j \in J} A_j\right) = \bigcup_{j \in J} f(A_j)$$

$$(7.45) \quad f(A^c) = f(A)^c$$

$$(7.46) \quad f(A_1 \setminus A_2) = f(A_1) \setminus f(A_2)$$

$$(7.47) \quad f(A_1 \Delta A_2) = f(A_1) \Delta f(A_2)$$

PROOF: Left as exercise 7.7.

**Proposition 7.9.** Let  $f : X \rightarrow Y$  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  $X = \bigsqcup_j A_j$ . Assume further that none of the  $A_j$  are empty. 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  $f|_{A_j} : A_j \rightarrow Y$  to  $A_j$ . Then  $f|_{A_j}(A_j) = B_j$  and the function

$$f_j : A_j \rightarrow B_j, \quad x \mapsto f_j(x) := f|_{A_j}(x) = f(x)$$

is a bijection.

PROOF: Left as exercise 7.8.

**Corollary 7.3.** Let  $f : X \rightarrow Y$  be bijective. Let  $A \subset X$ ,  $A \neq \emptyset$  (strict inclusion, so  $A^c \neq \emptyset$ ). Then both

$$f_A : A \rightarrow f(A), \quad x \mapsto f(x) \quad \text{and} \quad f_{A^c} : A^c \rightarrow f(A^c), \quad x \mapsto f(x)$$

are bijections.

PROOF: This follows from prop.7.9, applied to  $J = \{1, 2\}$ ,  $A_1 = A$ ,  $A_2 = A^c$ . ■

**Corollary 7.4.** Let  $f : X \rightarrow Y$  be bijective. Let  $a \in X$  and assume that  $X \neq \{a\}$ . Then

$$\tilde{f} : X \setminus \{a\} \rightarrow Y \setminus \{f(a)\}, \quad x \mapsto f(x)$$

also is bijective. <sup>69</sup>

---

<sup>69</sup> This is B/G [1] prop.13.2.

PROOF: This follows from 7.4 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 in ch.8.1 on p.130 when we prove propositions concerning cardinality.

The first proposition shows how to preserve bijectivity if two function values need to be switched around.

**Proposition 7.10.** Let  $X, Y \neq \emptyset$ , let  $f : X \rightarrow Y$  be bijective and let  $x_1, x_2 \in X$ . Let

$$(7.48) \quad g(x) := \begin{cases} f(x_2) & \text{if } x = x_1, \\ f(x_1) & \text{if } x = x_2, \\ f(x) & \text{if } x \neq x_1, x_2. \end{cases}$$

(In other words, we swap two function arguments). Then  $g : X \rightarrow Y$  also is bijective.

PROOF: Left as exercise 7.9.

A more general version of the above shows how to preserve surjectivity if two function values need to be switched around.

**Proposition 7.11.** Let  $X, Y \neq \emptyset$  and assume that  $Y$  contains at least two elements  $y_1$  and  $y_2$ . Let  $f : X \rightarrow Y$  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

$$(7.49) \quad g(x) := \begin{cases} y_2 & \text{if } x \in A_1, \\ y_1 & \text{if } x \in A_2, \\ f(x) & \text{if } x \in B. \end{cases}$$

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  $g : X \rightarrow Y$  also is surjective.

PROOF: Left as exercise 7.10.

**Proposition 7.12.** Let  $X, Y$  be two nonempty sets and let  $f : X \rightarrow Y$  be surjective. Let  $\emptyset \neq B \subsetneq Y$  so that  $Y = B \uplus B^c$  is a partitioning of  $Y$  into two nonempty subsets  $B$  and  $B^c$ . Let  $A := \{f \in B\}$ . Then the restrictions  $f_1 := f|_A : A \rightarrow B$  and  $f_2 := f|_{A^c} : A^c \rightarrow B^c$  of  $f$  to  $A$  and  $A^c$  are surjections.

PROOF: Left as exercise 7.11.

## 7.2 Indicator Functions

Sometimes it is advantageous to think of the subsets of a universal set  $\Omega$  as “binary” functions  $\Omega \rightarrow \{0, 1\}$ .

**Definition 7.2** (indicator function for a set). Let  $\Omega$  be “the” universal set, i.e., we restrict our scope of interest to subsets of  $\Omega$ . Let  $A \subseteq \Omega$ . Let  $1_A : \Omega \rightarrow \{0, 1\}$  be the function defined as

$$(7.50) \quad 1_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

$1_A$  is called the **indicator function** of the set  $A$ .<sup>70</sup> □

<sup>70</sup> Some authors call this **characteristic function** of  $A$  and some choose to write  $\chi_A$  or  $\mathbb{1}_A$  instead of  $1_A$ .

The above association of a subset  $A$  of  $\Omega$  with its indicator function is a bijection:

**Proposition 7.13.** Let  $\mathcal{F}(\Omega, \{0, 1\}) := \{0, 1\}^\Omega$  denote the set of all functions  $f : \Omega \rightarrow \{0, 1\}$ , i.e., all functions  $f$  with domain  $\Omega$  for which the only possible function values  $f(\omega)$  are zero or one. <sup>71</sup>

a. The mapping

$$(7.51) \quad F : 2^\Omega \rightarrow \mathcal{F}(\Omega, \{0, 1\}), \quad \text{defined as } F(A) := 1_A$$

which assigns to each subset of  $\Omega$  its indicator function is injective.

b. Let  $f \in \mathcal{F}(\Omega, \{0, 1\})$ . Further, let  $A := \{f = 1\} = f^{-1}(\{1\}) = \{a \in A : f(a) = 1\}$ . Then  $f = 1_A$ .

c. The function  $F$  above is bijective and its inverse function is

$$(7.52) \quad G : \mathcal{F}(\Omega, \{0, 1\}) \rightarrow 2^\Omega, \quad \text{defined as } G(f) := \{f = 1\}.$$

PROOF of a: This follows from c which will be proved below.

PROOF of b: We have

$$\begin{aligned} f(\omega) = 1 &\Leftrightarrow \omega \in \{f = 1\} \quad (\text{def. of inverse image}) \\ &\Leftrightarrow \omega \in A \quad (\text{because } A = \{f = 1\}) \\ &\Leftrightarrow 1_A(\omega) = 1 \quad (\text{def. of indicator function}). \end{aligned}$$

It follows that  $f(\omega) = 1$  if and only if  $1_A(\omega) = 1$ . As the only other possible function value is 0 we conclude that  $f(\omega) = 0$  if and only if  $1_A(\omega) = 0$ . It follows that  $f(\omega) = 1_A(\omega)$  for all  $\omega \in \Omega$ , i.e.,  $f = 1_A$ . This proves b.

PROOF of c: According to theorem 5.1 on p.95 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

$$(7.53) \quad G \circ F = id_{2^\Omega} \quad \text{and} \quad F \circ G = id_{\mathcal{F}(\Omega, \{0, 1\})}.$$

Let  $A \subseteq \Omega$  and  $A \in \Omega$ . Then

$$G \circ F(A) = G(1_A) = \{1_A = 1\} = \{\omega \in \Omega : 1_A(\omega) = 1\} = \{\omega \in \Omega : \omega \in A\} = A.$$

This proves  $G \circ F = id_{2^\Omega}$ . Let  $\omega \in \Omega$ . Then

$$\begin{aligned} (F \circ G(f))(\omega) &= F(\{f = 1\})(\omega) = 1_{\{f=1\}}(\omega) \\ &= \begin{cases} 1 & \text{iff } \omega \in \{f = 1\}, \\ 0 & \text{iff } \omega \notin \{f = 1\} \end{cases} = \begin{cases} 1 & \text{iff } f(\omega) = 1, \\ 0 & \text{iff } f(\omega) \neq 1 \end{cases} = \begin{cases} 1 & \text{iff } f(\omega) = 1, \\ 0 & \text{iff } f(\omega) = 0 \end{cases} = f(\omega). \end{aligned}$$

The equation next to the last results from the fact that the only possible function values for  $f$  are 0 and 1. It follows that  $F \circ G(f) = f = id_{\mathcal{F}(\Omega, \{0, 1\})}(f)$  for all  $f \in \mathcal{F}(\Omega, \{0, 1\})$ , hence  $F \circ G = id_{\mathcal{F}(\Omega, \{0, 1\})}$ . We have proved (7.53) and hence c. ■

The following definition is a special case of “modular arithmetic”. We will only use it in the context of indicator functions of set differences. For further information we refer to [1] B/G, ch.6.3.

<sup>71</sup> See remark 6.4 on p.116, ch.6.2 (Cartesian Products of More Than Two Sets).

**Definition 7.3** (mod 2 addition). Let  $m, n \in \mathbb{Z}$ . We define

$$(7.54) \quad m + n \pmod{2} := \begin{cases} 0 & \text{if } m + n \text{ is even,} \\ 1 & \text{if } m + n \text{ is odd,} \end{cases}$$

and we call  $m + n \pmod{2}$  the **sum mod 2** of  $m$  and  $n$ .  $\square$

**Proposition 7.14.** Let  $m, n, p \in \mathbb{Z}$ . Then addition mod 2 is associative, i.e.,

$$(7.55) \quad (m + n \pmod{2}) + p \pmod{2} = m + (n + p \pmod{2}) \pmod{2}.$$

The proof is left as exercise 7.12.

**Proposition 7.15.** Let  $A, B, C$  be subsets of  $\Omega$ . Then

$$(7.56) \quad \mathbb{1}_{A \cup B} = \max(\mathbb{1}_A, \mathbb{1}_B),$$

$$(7.57) \quad \mathbb{1}_{A \cap B} = \min(\mathbb{1}_A, \mathbb{1}_B),$$

$$(7.58) \quad \mathbb{1}_{A^c} = 1 - \mathbb{1}_A,$$

$$(7.59) \quad \mathbb{1}_{A \Delta B} = \mathbb{1}_A + \mathbb{1}_B \pmod{2}.$$

PROOF: The proof of the first three equations is left as an exercise.

PROOF of (7.59): This follows easily from the fact that

$$(A \Delta B)^c = \{\omega \in \Omega : \text{either } \omega \in A \cap B \text{ or } \omega \in \text{neither } A \text{ nor } B\} \blacksquare$$

**Proposition 7.16** (Symmetric set differences  $A \Delta B$  are associative). Let  $A, B, C \subseteq \Omega$ . Then

$$(7.60) \quad (A \Delta B) \Delta C = A \Delta (B \Delta C).$$

PROOF: This follows easily from (7.59) and prop.7.14 as follows. Let  $\omega \in \Omega$ . Then

$$\omega \in (A \Delta B) \Delta C \Leftrightarrow \mathbb{1}_{(A \Delta B) \Delta C} = 1 \Leftrightarrow \mathbb{1}_{A \Delta (B \Delta C)} = 1 \Leftrightarrow \omega \in A \Delta (B \Delta C).$$

We obtained the equivalence in the middle from prop.7.14.  $\blacksquare$

### 7.3 Exercises for Ch.7

**Exercise 7.1.** Let  $f : [-2, \infty[ \rightarrow \mathbb{R}; \quad x \mapsto x^2$ . Compute the following.

a.  $f(f^{-1}([-4, 4]))$ ,   b.  $f^{-1}(f([0, 3]))$ .  $\square$

**Exercise 7.2.** Prove prop.7.1 on p.119:

- 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!  $\square$

**Exercise 7.3.** Prove (7.31) of prop.7.4 on p.122: Let  $X, Y, Z$  be an arbitrary, nonempty sets. Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  and let  $W \subseteq Z$ . Then  $(g \circ f)(U) = g(f(U))$  for all  $U \subseteq X$ .  $\square$

**Exercise 7.4.** Let  $X, Y$  be nonempty sets and let  $f : X \rightarrow Y$  be a function. We define on the domain  $X$  a relation “ $\sim$ ” as follows:

$$x_1 \sim x_2 \Leftrightarrow f(x_1) = f(x_2). \quad \square$$

(See prop.7.5 (Indirect image and fibers of  $f$ ) on p.122). Prove that “ $\sim$ ” is an equivalence relation.

**Exercise 7.5.** Prove prop.7.7 on p.124.

**Hint:** The main tools you need are prop.7.5 on p.122, prop.7.6 on p.123, and their corollaries.  $\square$

**Exercise 7.6.** Prove the reverse directions of prop.7.7.a and prop.7.7.b on p.124.

- a. If  $f : X \rightarrow Y$  satisfies  $f^{-1}(f(A)) = A$  for all  $A \subseteq X$  then  $f$  is injective.
- b. If  $f : X \rightarrow Y$  satisfies  $f(f^{-1}(B)) = B$  for all  $B \subseteq Y$  then  $f$  is surjective.

**Exercise 7.7.** Prove prop.7.8 on p.124.

**Hint:** Work with the inverse of  $f$  and apply prop.7.2 on p.120.  $\square$

**Exercise 7.8.** Prove prop.7.9 on p.124.

**Hint:** To prove a, use prop.7.3 on p.121.  $\square$

**Exercise 7.9.** Prove prop.7.10 on p.125: Let  $X, Y \neq \emptyset$ , let  $f : X \rightarrow Y$  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  $g : X \rightarrow Y$  also is bijective.  $\square$

**Exercise 7.10.** Prove prop.7.11 on p.125: Let  $X, Y \neq \emptyset$  and assume that  $Y$  contains at least two elements  $y_1$  and  $y_2$ . Let  $f : X \rightarrow Y$  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  $g : X \rightarrow Y$  also is surjective.  $\square$

**Exercise 7.11.** Prove prop.7.12 on p.125: Let  $X, Y$  be two nonempty sets and let  $f : X \rightarrow Y$  be surjective. Let  $\emptyset \neq B \subsetneq Y$  so that  $Y = B \uplus B^c$  is a partitioning of  $Y$  into two nonempty subsets  $B$  and  $B^c$ . Let  $A := \{f \in B\}$ . Then the restrictions  $f_1 := f|_A : A \rightarrow B$  and  $f_2 := f|_{A^c} : A^c \rightarrow B^c$  of  $f$  to  $A$  and  $A^c$  are surjections.  $\square$



**Exercise 7.12.** Prove prop.7.14 on p.127: Let  $m, n, p \in \mathbb{Z}$ . Then

$$(m + n \pmod 2) + p \pmod 2 = m + (n + p \pmod 2) \pmod 2.$$

**Hint:** There are eight possible combinations of zeroes and ones for the functions

$$(m, n, p) \rightarrow (m + n \pmod 2) + p \pmod 2 \quad \text{and} \quad (m, n, p) \rightarrow m + (n + p \pmod 2) \pmod 2.$$

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  $m + n \pmod 2$ . 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 7.13.** See [1] B/G project 6.8 on p.58 for the following.

Prove that the following are equivalence relations on  $\mathbb{R}^2$ .

- $(x, y) \sim (u, v) \Leftrightarrow \sqrt{x^2 + y^2} = \sqrt{u^2 + v^2}$ .
- $X = \mathbb{R}_{\neq 0} = \{x \in \mathbb{R} : x \neq 0\}$  and  $x \sim y \Leftrightarrow |xy| > 0$ .
- $X = \mathbb{R}^3$  and  $(x, y, z) \sim (u, v, w) \Leftrightarrow z \sin(xy) = w \sin(uv)$ .

Hint: See example 7.5 on p.123.

□

## 8 Some Miscellaneous Topics

Although this chapter only contains a single topic at this time (cardinality), additional topics are planned in the future.

### 8.1 Cardinality - Alternate Approach to Beck/Geoghegan (Study this!)

**Notation:** In this entire chapter on cardinality, if  $n \in \mathbb{N}$ , the symbol  $[n]$  does not denote an equivalence class of any kind but the set  $\{1, 2, \dots, n\}$  of the first  $n$  natural numbers.

**Definition 8.1** (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  $X, Y \neq \emptyset$  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  $f : X \rightarrow Y$  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  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  and  $\mathbf{card}(Y) \neq \mathbf{card}(X)$ , i.e., there is an injective mapping but not a bijection  $f : X \rightarrow Y$ .
- 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  $\mathbf{card}(Y) \leq \mathbf{card}(X)$ .
- e. the **cardinality of  $X$  is greater than the cardinality of  $Y$**  and we write  $\mathbf{card}(X) > \mathbf{card}(Y)$  if  $\mathbf{card}(Y) < \mathbf{card}(X)$ .  $\square$

**Remark 8.1.** Note the following.

- a. The above definition does not specify how  $\mathbf{card}(X)$  itself is defined. This will be done in def.8.4 on p.141.
- b. The above definition includes the cases  $X = \emptyset$  and  $Y = \emptyset$ :  $\mathbf{card}(\emptyset) = \mathbf{card}(\emptyset)$ ,  $\mathbf{card}(\emptyset) \leq \mathbf{card}(Y)$  for any set  $Y$ , and  $\mathbf{card}(\emptyset) \neq \mathbf{card}(Y)$  if  $Y$  is not empty.  $\square$

**Example 8.1.** Let  $A, B$  be two sets such that  $A \subseteq B$ . Then  $\mathbf{card}(A) \leq \mathbf{card}(B)$ .

*PROOF:*

*Case 1:*  $A = \emptyset$ . It then is true by definition that  $\mathbf{card}(\emptyset) \leq \mathbf{card}(B)$  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  $\mathbf{card}(A) \leq \mathbf{card}(B)$ .  $\blacksquare$

**Definition 8.2** (Finite and infinite sets). Let  $n \in \mathbb{N}$ . we say that

- a. A set  $X$  has **size  $n$**  and we write  $|X| = n$  if there is a bijective mapping between  $X$  and the set  $[n] := \{1, 2, \dots, n\}$ . In other words, a set  $X$  of size  $n$  is one whose elements can be enumerated as  $x_1, x_2, \dots, x_n$  and its size simply is the number of elements it contains.
- b. We say that the empty set  $\emptyset$  has size  $|\emptyset| = 0$ .
- c. A set  $X$  is **finite** if there exists  $n \in \mathbb{Z}_{\geq 0}$  such that  $X$  has size  $n$ . 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  $f : \mathbb{N} \xrightarrow{\sim} X$ . This means that all of the elements of  $X$  can be arranged in a sequence

$$X = \{x_1 = f(1), x_2 = f(2), x_3 = f(3), \dots\}.$$

which is infinite, i.e., we rule out the case of sets with finitely many members.  $X$  is called a **countably infinite set**.

- e. We call a set  $X$  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**.  $\square$

**Example 8.2.** Let  $X := \{2n : n \in \mathbb{N}\}$  be the set of all even natural numbers. Then  $X$  is countably infinite because the function

$$f : \mathbb{N} \longrightarrow X; n \mapsto 2n \quad \text{has the function} \quad g : X \longrightarrow \mathbb{N}; k \mapsto \frac{k}{2}$$

as its inverse and hence is bijective.

Note that because the even natural numbers are divisible by 2,  $\frac{k}{2} \in \mathbb{N}$  if  $k \in X$ .  $\square$

**Proposition 8.1.**

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.5.4.c on p.96 which states that the composition of two bijective functions is bijective.

We only need to prove the “ $\Rightarrow$ ” directions because we obtain “ $\Leftarrow$ ” 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  $g : X \xrightarrow{\sim} [n]$ .  $Y \xrightarrow{f^{-1} \circ g} [n]$  is bijective according to prop.5.4.c on p.96. 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  $Y \xrightarrow{f^{-1} \circ g} \mathbb{N}$  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. This proves that  $Y$  is uncountable.  $\blacksquare$

**Theorem 8.1** (B/G thm.13.31). Let  $X$  be a set. Then  $\text{card}(X) < \text{card}(2^X)$ .

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:* We note that the function  $x \mapsto \{x\}$  is an injection from  $X$  into  $2^X$ . It follows that  $\text{card}(X) \leq \text{card}(2^X)$ .

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

**Proposition 8.2.** Let  $X, Y$  be two sets such that  $\text{card}(X) = \text{card}(Y)$ . Then  $\text{card}(2^X) = \text{card}(2^Y)$ .

The proof is left as exercise 8.1. ■

**Proposition 8.3.** Let  $n \in \mathbb{Z}_{\geq 0}$ . Let  $X$  be a set such that  $|X| = n$ . Then its powerset has size  $|2^X| = 2^n$ .

*PROOF:* The proof is done by induction on  $n = |X|$ .

Base case  $n = 0$ :

It follows that  $X$  is empty. We have  $|\emptyset| = 0$  and  $2^\emptyset = \{\emptyset\}$  has size  $2 = 2^1$ .

Induction assumption: Let  $k \in \mathbb{Z}_{\geq 0}$  such that  $|2^Y| = 2^k$  for all sets  $Y$  such that  $|Y| = k$ . (★)

Let  $|X| = k + 1$ . We need to show that  $|2^X| = 2^{k+1}$ . (★★)

By definition  $|X| = k + 1$  means that we have a bijection  $X \xrightarrow{\sim} [k + 1]$ . It follows from prop.8.2 on p.132 that we also have a bijection  $2^X \xrightarrow{\sim} 2^{[k+1]}$ . But then  $|2^X| = |2^{[k+1]}|$ <sup>72</sup> by prop.8.1.e on p.131.

So it is enough to show (★★) for the set  $[k + 1]$  and we assume from now on that  $X = [k + 1]$ . We have seen in prop.7.13 on p.126 that, for any set  $\Omega$ , the function

$$F : 2^\Omega \rightarrow \mathcal{F}(\Omega, \{0, 1\}); \quad F(A) := 1_A$$

which assigns to  $A \subseteq \Omega$  its indicator function  $1_A(\omega) = 1$  if  $\omega \in A$  and zero else is a bijection.

We apply this to  $\Omega = [k + 1]$  and obtain

$$(8.1) \quad |2^{[k+1]}| = |\mathcal{F}([k + 1], \{0, 1\})|.$$

How many  $\{0, 1\}$ -valued functions are there with domain  $[k + 1] = \{j \in \mathbb{N} : j \leq k\}$ ? It is intuitively clear that  $\mathcal{F}([k + 1], \{0, 1\})$  contains exactly twice as many functions as  $\mathcal{F}([k], \{0, 1\})$  because each  $\varphi : [k] \rightarrow \{0, 1\}$  has exactly two extensions  $\varphi^{(0)}, \varphi^{(1)} : [k + 1] \rightarrow \{0, 1\}$  defined as follows:

$$(8.2) \quad \varphi^{(0)}(j) := \varphi(j) \text{ if } j \leq k \text{ and } \varphi^{(0)}(k + 1) := 0;$$

$$(8.3) \quad \varphi^{(1)}(j) := \varphi(j) \text{ if } j \leq k \text{ and } \varphi^{(1)}(k + 1) := 1.$$

In other words,

$$\mathcal{F}([k + 1], \{0, 1\}) = \{ \varphi^{(0)}, \varphi^{(1)} : \varphi \in \mathcal{F}([k], \{0, 1\}) \}.$$

<sup>72</sup> **Careful!** Distinguish between the power set  $2^{[k+1]}$  and the number  $2^{k+1}$  and the set of natural numbers  $[2^{k+1}]$ !

To complete the proof, we will construct a bijection  $G : [2^{k+1}] \rightarrow \mathcal{F}([k+1], \{0, 1\})$ .

It follows from the induction assumption that

$$(8.4) \quad |\mathcal{F}([k], \{0, 1\})| = |[2^k]|, \text{ i.e., there is bijective } F : [2^k] \xrightarrow{\sim} \mathcal{F}([k], \{0, 1\}).$$

We define  $G : [2^{k+1}] \rightarrow \mathcal{F}([k+1], \{0, 1\})$  as follows.

$$(8.5) \quad m \mapsto G(m)(\cdot) := \begin{cases} F(m)^{(0)}(\cdot) & \text{if } m \leq 2^k, \\ F(m - 2^k)^{(1)}(\cdot) & \text{if } 2^k < m \leq 2^{k+1}. \end{cases}$$

The superscripted functions  $F(m)^{(0)}(\cdot)$  and  $F(m - 2^k)^{(1)}(\cdot)$  are defined in (8.2) and (8.3).

We carry through the argument  $(\cdot)$  to remind the reader that  $F(i), G(i), F(m)^{(0)}$ , etc., are functions  $j \mapsto G(i)(j)$ ,  $j \mapsto F(i)(j)$ ,  $j \mapsto F(m)^{(0)}(j)$ , etc., with domain  $[k]$  or  $[k+1]$

*Proof of injectivity of  $G$ .* Let  $m, m' \in \mathbb{N}$ .

**Case 1:** Assume  $1 \leq m < m' \leq 2^k$ .

Let  $\varphi := G(m)|_{[k]}$  and  $\psi := G(m')|_{[k]}$  be the restrictions of the functions  $G(m), G(m') : [k+1] \rightarrow \{0, 1\}$  to the subset  $[k]$  of  $[k+1]$ . We have

$$\varphi = F(m)^{(0)}|_{[k]} = F(m), \quad \psi = F(m')^{(0)}|_{[k]} = F(m').$$

We conclude from  $m \neq m'$  and the injectivity of  $F$  that there exists some  $1 \leq j \leq k$  such that

$$\varphi(j) = F(m)(j) \neq F(m')(j) = \psi(j), \text{ hence } \varphi^{(0)} \neq \psi^{(0)}, \text{ hence } G(m)(\cdot) \neq G(m')(\cdot).$$

**Case 2:** Assume  $1 \leq m \leq 2^k < m' \leq 2^{k+1}$ . We have  $G(m)(k+1) = 0$  and  $G(m')(k+1) = 1$ . It follows that  $G(m)(\cdot) \neq G(m')(\cdot)$ .

**Case 3:** Assume  $2^k < m < m' \leq 2^{k+1}$ .

Let  $\varphi := G(m - 2^k)|_{[k]}$  and  $\psi := G(m' - 2^k)|_{[k]}$  be the restrictions of the functions  $G(m), G(m') : [k+1] \rightarrow \{0, 1\}$  to the subset  $[k]$  of  $[k+1]$ . We have

$$\varphi = F(m - 2^k)^{(1)}|_{[k]} = F(m - 2^k), \quad \psi = F(m' - 2^k)^{(1)}|_{[k]} = F(m' - 2^k).$$

We conclude from  $m - 2^k \neq m' - 2^k$  and the injectivity of  $F$  that there exists some  $1 \leq j \leq k$  such that

$$\varphi(j) = F(m - 2^k)(j) \neq F(m' - 2^k)(j) = \psi(j), \text{ hence } \varphi^{(1)} \neq \psi^{(1)}, \text{ hence } G(m)(\cdot) \neq G(m')(\cdot).$$

Cases 1, 2 and 3 together prove that  $G(m) \neq G(m')$  whenever  $m \neq m'$ , hence  $G$  is injective.

*Proof of surjectivity of  $G$ .*

Let  $f \in \mathcal{F}([k+1], \{0, 1\})$ . We must find  $j \in [k+1]$  such that  $G(j) = f$ .

Let  $\varphi := f|_{[k]}$  be the restriction of  $f$  to  $[k]$ . Then  $\varphi \in \mathcal{F}([k], \{0, 1\})$ . It follows from the surjectivity of  $F$  that there exists  $m \in [2^k]$  such that  $F(m) = \varphi$ .

**Case 1:**  $f(k+1) = 0$ . Then  $f = \varphi^{(0)} = F(m)^{(0)}$ . It follows from (8.5) that  $f = G(m)$ .

**Case 1:**  $f(k+1) = 1$ . Then  $f = \varphi^{(1)} = F(m)^{(1)}$ . It follows from (8.5) that  $f = G(m + 2^k)$ .

Cases 1 and 2 together prove that  $G$  is surjective. We already saw that  $G$  is injective. We have proven the existence of a bijection  $[2^{k+1}] \xrightarrow{\sim} \mathcal{F}([k+1], \{0, 1\})$ , hence  $|[2^{k+1}]| = |\mathcal{F}([k+1], \{0, 1\})|$ . It then follows from (8.1) that  $|2^{k+1}| = |[2^{k+1}]|$ , and this proves  $(\star\star)$ . ■

**Lemma 8.1.** Let  $X$  be an arbitrary set and let  $A$  be a nonempty proper subset of  $X$ , so that  $X = A \uplus A^c$  is a partitioning of  $X$  into two nonempty subsets  $A$  and  $A^c$ . Let  $a \in A$ ,  $a_0 \in A^c$  and  $A' := (A \setminus \{a\}) \uplus \{a_0\}$ . Then the function

$$\varphi : A' \xrightarrow{\sim} A; \quad x \mapsto \begin{cases} x & \text{if } x \neq a_0, \\ a & \text{if } x = a_0 \end{cases}$$

is a bijection.

*PROOF:* The function

$$\psi : A \xrightarrow{\sim} A'; \quad x \mapsto \begin{cases} x & \text{if } x \neq a, \\ a_0 & \text{if } x = a \end{cases}$$

is the inverse of  $\varphi$ . ■

**Proposition 8.4.** 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* 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  $f : \{1\} \rightarrow [2]$  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  $\{2\} \rightarrow [2]$  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  $\Gamma \rightarrow [n-1]$ .

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  $f : A \rightarrow [n]$  exists.

**case 1:**  $n \notin A$ :

Then  $A \subseteq [n-1]$ .

Let  $\Gamma := A \setminus \{f = n\}$ . Because  $f$  is surjective,  $\{f = n\} \neq \emptyset$ , hence  $\Gamma \subsetneq [n-1]$ . It follows from prop.7.12 on p.125 that  $f_\Gamma := f|_\Gamma : \Gamma \rightarrow [n] \setminus \{n\} = [n-1]$  is surjective.

But the induction assumption implies that there is no surjection from the proper subset  $\Gamma$  of  $[n-1]$  onto  $[n-1]$ . We have reached a contradiction.

**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 lemma 8.1 that there is a bijection  $g : A' \xrightarrow{\sim} A$ . Hence  $f \circ g : A' \rightarrow [n]$  is surjective as the composition of a surjection with a bijection (see cor.5.1.b on p.96). But  $A'$  satisfies the conditions of case 1 and we have already proven that such a surjection cannot exist. ■

**Corollary 8.1** (No bijection from  $[m]$  to  $[n]$  exists). B/G Thm.13.4: Let  $m, n \in \mathbb{N}$ . If  $m \neq n$  then there is no bijective  $f : [m] \xrightarrow{\sim} [n]$ .

*PROOF:* We may assume  $m < n$  and can now apply prop. 8.4 with  $A := [m]$ . ■

**Corollary 8.2** (Pigeonhole Principle). B/G Prop.13.5: Let  $m, n \in \mathbb{N}$ . If  $m < n$  then there is no injective  $f : [n] \rightarrow [m]$ .

PROOF: Otherwise, by thm.5.2 on p.97,  $f$  would have a (surjective) left inverse  $g : [m] \rightarrow [n]$  in contradiction to the preceding proposition. ■

**Proposition 8.5** (B/G Prop.13.6, p.122: Subsets of finite sets are finite). Let  $\emptyset \neq B \subseteq A$  and let  $A$  be finite. Then  $B$  is finite.

PROOF: Done by induction on the size  $n$  of  $A$ . We may assume that  $B \subsetneq A$  because bijectivity is reflexive (the identity mapping  $id_A$  maps  $A$  bijectively to itself).

Base case:  $n = 1$  or  $n = 2$ : Proof obvious.

Induction assumption: Assume that all subsets of sets of size less than  $n$  are finite.

Now let  $A$  be a set with size  $|A| = n$ . There is a bijection  $a(\cdot) : [n] \xrightarrow{\sim} A$ . Let  $B \subseteq A$ .

**Case 1:**  $a(n) \in B$ : Let  $B_n := B \setminus \{a(n)\}$  and  $A_n := A \setminus \{a(n)\}$ .

Then the restriction  $a(\cdot)|_{[n-1]}$  of  $a(\cdot)$  to  $[n-1]$  is a bijection  $[n-1] \xrightarrow{\sim} A_n$  according to cor.7.3 on p.124.

As  $|A_n| = n - 1$  and  $B_n \subseteq A_n$  it follows from the induction assumption that  $B_n$  is finite: there exists  $m \in \mathbb{N}$  and a bijection  $b(\cdot) : [m] \xrightarrow{\sim} B_n$ .

We now extend  $b(\cdot)$  to  $[m+1]$  by defining  $b(m+1) := a(n)$ . It follows that this extension remains injective and it is also surjective if we choose as codomain  $B_n \cup \{a(n)\} = B$ .

It follows that  $B$  is finite.

**Case 2:**  $a(n) \notin B$ : We pick an arbitrary  $b \in B$ . Let  $j := a^{-1}(b)$ . Clearly  $j \in [n]$ .

We modify the mapping  $a(\cdot)$  by switching the function values for  $j$  and  $n$ . We obtain another bijection  $f : [n] \xrightarrow{\sim} A$  (see prop.7.10 on p. 125) for which  $f(n) = a(j) = b \in B$ .

We apply to  $f$  what was proved in case 1 and obtain that  $B$  is finite. ■

Our next major result is that subsets of countable sets are countable.

**Lemma 8.2.** Any subset of  $\mathbb{N}$  is countable.

PROOF: Let  $B \subseteq \mathbb{N}$ . Finite sets are countable by definition, so we may assume that  $B$  is infinite. We extract  $j_1 < j_3 < j_2 < \dots$  from  $\mathbb{N}$  as follows.

$$\begin{aligned} j_1 &:= \min\{j \in \mathbb{N} : j \in B\}, \\ j_2 &:= \min\{j \in B : j > j_1\}, \\ &\dots\dots\dots \\ j_{k+1} &:= \min\{j \in B : j > j_k\} \end{aligned}$$

Note that none of the sets  $B_k := \{j \in B : j > j_k\}$  are empty because  $B_k = \emptyset$  would imply  $B \subseteq \{1, 2, \dots, j_k\}$ , i.e.,  $B$  is a subset of the finite set  $[j_k]$ . This is a contradiction to prop.8.5 (Subsets of finite sets are finite) on p.135.

Let  $B' := \{j_k : k \in \mathbb{N}\}$ . Note that  $j_k < j_{k+1}$  for all  $k \in \mathbb{N}$  by construction. Hence

$$(8.6) \quad j_k \geq k \quad \text{for all } k \in \mathbb{N}.$$

In particular,  $B'$  is not bounded above.

We claim that  $B = B'$ . It follows from  $j_k \in B$  for all  $k$  that  $B' \subseteq B$ . We now show that  $B \subseteq B'$ . Otherwise there would be  $n' \in B$  such that  $n' \neq j_k$  for all natural numbers  $k$ . It follows from 8.6 that  $n' \leq j_{n'}$ . Hence there is  $m \leq n'$  such that  $j_{m-1} < n' < j_m$ .

This contradicts  $j_m = \min\{j \in B : j > j_{m-1}\}$ .

It follows from 8.6 that the function

$$\psi : \mathbb{N} \rightarrow B; \quad k \mapsto j_k$$

is injective.  $\psi$  also is surjective because its codomain is  $B' = \{b_k : k \in \mathbb{N}\}$ . Hence  $B$  is countably infinite, hence  $B$  is countable. ■

**Theorem 8.2.** Subsets of countable sets are countable.

PROOF: Let  $X$  be countable and  $A \subseteq X$ .

Before we continue, note that no assumption is made about the nature of  $X$  and  $A$ . We may not deal with natural numbers or any other kind of numbers.  $X$  and hence  $A$  might, e.g., be sets of functions or sets of sets.

We may assume that  $X$  is infinite because otherwise  $A$  is finite, hence countable, as the subset of a finite set. See prop.8.5 on p.135.  $X$  is countably infinite, hence there exists a bijection

$$\phi : \mathbb{N} \xrightarrow{\sim} X; \quad n \mapsto x_n := x(n).$$

The function  $\phi^{-1}|_A : A \rightarrow \phi^{-1}(A), \quad x \mapsto \phi^{-1}(x),$

is injective as a restriction of the bijective, hence injective function  $\phi^{-1}$ .  $\phi^{-1}|_A$  also is surjective because the codomain is  $\phi^{-1}(A) = \phi^{-1}|_A(A)$ . It follows that  $\phi^{-1}|_A$  is bijective.

The diagram to the right shows the role of  $\phi$  and its restriction to  $A$ . The bijectivity of those functions allows us to shift our problem from a subset  $A$  of an abstract set  $X$  to a subset of  $\mathbb{N}$  whose cardinality equals that of  $A$ .

$$\begin{array}{ccc} & A & \xrightarrow{id_A} & X \\ \phi^{-1}|_A \downarrow & & & \downarrow \phi^{-1} \\ & \phi^{-1}(A) & & \mathbb{N} \end{array}$$

According to lemma 8.2,  $\phi^{-1}(A) = \{n \in \mathbb{N} : \phi(x) \in A\}$  is countable as a subset of  $\mathbb{N}$ . It follows from the bijectivity of  $\phi^{-1}|_A$  and prop.8.1.c on p.131 that  $A$  is countable. ■

**Corollary 8.3.** Supersets of uncountable sets are uncountable.

Proof: Assume to the contrary that  $X$  is uncountable,  $Y \supseteq X$ , and  $Y$  is countable. This can be worded as follows:  $Y$  is countable,  $X \subseteq Y$ , and  $X$  is uncountable. We have a contradiction to theorem 8.2. ■

The next proposition is a major stepping stone for proving that countable unions of countable sets are countable.

**Proposition 8.6** (B/G Cor.13.16, p.122).  $\mathbb{N}^2$  is countable.

PROOF:<sup>73</sup> Done by directly specifying a bijection  $F : \mathbb{N}^2 \xrightarrow{\sim} \mathbb{N}$ .

The following definitions and observations will make it easier to understand this proof. Let

$$s_0 := 0; \quad s_n := \sum_{j=1}^n j = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

<sup>73</sup> Understanding this proof is not very important and you will understand the essence of it if you read instead the subsequent remark 8.2.



For the last equality see [1] B/G prop.4.11 on p.37. We note that

$$(8.7) \quad s_{n-1} + n = s_n.$$

Moreover,

$$(8.8) \quad A_n := \{j \in \mathbb{N} : s_{n-1} < j \leq s_n\} = \{s_{n-1} + 1, s_{n-1} + 2, \dots, s_{n-1} + n\} \quad (n \in \mathbb{N})$$

is a partition of  $\mathbb{N}$  (see example 6.2.d on p.111). For  $n \in \mathbb{N} \setminus \{1\} = \{2, 3, 4, \dots\}$  let

$$D_n := \{(i, j) \in \mathbb{N}^2 : i + j = n\}$$

be the set of all pairs of natural numbers whose sum equals  $n$ . Clearly,  $(D_n)_{n \geq 2}$  is a partition of  $\mathbb{N}^2$ .

For  $n \in \mathbb{N}$  let

$$(8.9) \quad f_n : A_n \rightarrow D_{n+1}, \quad s_{n-1} + k \mapsto f_n(s_{n-1} + k) := (n + 1 - k, k) \quad (1 \leq k \leq n).$$

We see from the second equality in (8.8) that the set of argument values  $s_{n-1} + k$  ( $1 \leq k \leq n$ ) coincides with the domain  $A_n$  and it follows that (8.9) indeed defines a function  $A_n \rightarrow \mathbb{N}^2$ .

It is immediate that  $D_{n+1} = \{(n + 1 - k, k) : 1 \leq k \leq n\}$ . We conclude that  $f_n(A_n) = D_{n+1}$  and we have proved surjectivity of  $f_n$ .

To prove injectivity of  $f_n$ , we observe that if  $1 \leq k, k' \leq n$  and  $k \neq k'$  then

$$f_n(s_{n-1} + k) = (n + 1 - k, k) \neq (n + 1 - k', k') = f_n(s_{n-1} + k').$$

We “glue together” the functions  $f_n$  to obtain a function  $f$  with domain  $\bigcup [A_n : n \in \mathbb{N}] = \mathbb{N}$  and codomain  $\bigcup [D_{n+1} : n \in \mathbb{N}] = \mathbb{N}^2$  as follows:

$f(m) := f_n(m)$  for  $m \in A_n$ , i.e.,

$$f(s_{n-1} + k) = f_n(s_{n-1} + k) = (n + 1 - k, k) \quad \text{for } 1 \leq k \leq n.$$

$f$  inherits injectivity from the individual  $f_n$  as the ranges  $f_n(A_n) = D_{n+1}$  are mutually disjoint for different values of  $n$  and  $f$  is surjective because

$$\begin{aligned} f(\mathbb{N}) &= f\left(\bigcup [A_n : n \in \mathbb{N}]\right) = \bigcup [f(A_n) : n \in \mathbb{N}] \\ &= \bigcup [f_n(A_n) : n \in \mathbb{N}] = \bigcup [D_{n+1} : n \in \mathbb{N}] = \mathbb{N}^2. \end{aligned}$$

To summarize, we have proved that  $f$  is a bijective mapping between  $\mathbb{N}$  and  $\mathbb{N}^2$  and this proves that  $\mathbb{N}^2$  is countable. ■

**Remark 8.2.** The following will help to visualize the proof just given. We think of  $\mathbb{N}^2$  as a matrix with “infinitely many rows and columns”

$$(8.10) \quad (1, 1) \ (1, 2) \ (1, 3) \ \dots$$

$$(8.11) \quad (2, 1) \ (2, 2) \ (2, 3) \ \dots$$

$$(8.12) \quad (3, 1) \ (3, 2) \ (3, 3) \ \dots$$

We reorganize this matrix into an ordinary sequence  $(f(j))_{j \in \mathbb{N}}$  as follows:

$$(8.13) \quad f(1) = f_1(1) = (1, 1),$$

$$(8.14) \quad f(2) = f_2(2) = (1, 2), \quad f(3) = f_2(3) = (2, 1),$$

$$(8.15) \quad f(4) = f_3(4) = (1, 3), \quad f(5) = f_3(5) = (2, 2), \quad f(6) = f_3(6) = (3, 1),$$

$$(8.16) \quad f(7) = f_4(7) = (1, 4), \quad f(8) = f_4(8) = (2, 3), \quad f(9) = f_4(9) = (3, 2), \quad f(10) = f_4(10) = (4, 1),$$

$$(8.17) \quad \dots$$

In other words, we traverse first  $D_2$ , then  $D_3$ , then  $D_4$ , ... starting for each  $D_n$  at the upper right  $(1, n - 1)$  and ending at the lower left  $(n - 1, 1)$ .  $\square$

**Theorem 8.3** (B/G prop.13.6: Countable unions of countable sets). *The union of countably many countable sets is countable.*

PROOF: Let the sets

$$A_1, A_2, A_3, \dots \text{ be countable and let } A := \bigcup_{n \in \mathbb{N}} A_n.$$

We assume that at least one of those sets is not empty: otherwise their union is empty, hence finite, hence countable and we are done.

We may assume that the sets  $A_i$  are mutually disjoint, i.e., any two different sets  $A_i, A_j$  have intersection  $A_i \cap A_j = \emptyset$ . Otherwise we replace each  $A_n$  with  $A_n \setminus \bigcup_{j < n} A_j$  (see prop.6.3 on p. [pagerefprop-x:math330-union-as-disjoint-u](#)).

**A.** As each of the nonempty  $A_i$  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 now define the mapping  $f : A \rightarrow \mathbb{N}^2$ ,  $a \mapsto (i_a, j_a)$  as follows: Let  $a \in A$ . As the  $A_j$  are disjoint there is a unique index  $i_a$  such that  $a \in A_{i_a}$  and, as 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, for any  $a \in A$  there exists a unique pair  $(i_a, j_a) \in \mathbb{N}^2$  such that  $a = a_{(i_a, j_a)}$  and the assignment  $a \mapsto (i_a, j_a)$  defines the injective function  $f : A \rightarrow \mathbb{N}^2$  we were looking to create.

We shrink the codomain 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 countable as a subset of the countable set  $\mathbb{N}^2$  and this proves the theorem because any subset of a countable set is countable (see prop.8.2 on p.136).  $\blacksquare$

**Corollary 8.4.** *Let the set  $X$  be uncountable and let  $A \subseteq X$  be countable. Then the complement  $A^c$  of  $A$  is uncountable.*

The proof is left as exercise 8.2.

Here is another corollary to thm.8.3.

**Corollary 8.5.** *The set  $Z$  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.  $\blacksquare$

**Corollary 8.6.** *The rational numbers are countable.*

*PROOF:* Let  $n \in \mathbb{N}$  and  $Q_n := \{\frac{m}{n} : m \in \mathbb{Z}\}$ . Then  $f_n : Q_n \rightarrow \mathbb{Z}$ ,  $\frac{m}{n} \mapsto m$  is a bijection because it has as an inverse the function  $m \mapsto \frac{m}{n}$ . It follows from cor.8.5 that  $Q_n$  is countable. By thm.8.3,  $\mathbb{Q} = \bigcup [Q_n : n \in \mathbb{N}]$  is countable as the union of countably many sets. ■

We now examine the cardinality of cartesian products.

**Theorem 8.4** (Finite Cartesians of countable sets are countable). *The Cartesian product of finitely many countable sets is countable.*

*Proof by induction:* Let  $X := X_1 \times \dots \times X_n$  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  $F_2 : X_2 \rightarrow \{x_1\} \times X_2$ ;  $x_2 \mapsto (x_1, x_2)$  is bijective because it has as an inverse the function  $G_2 : \{x_1\} \times X_2 \rightarrow X_2$ ;  $(x_1, x_2) \mapsto x_2$ . It follows that  $\{x_1\} \times X_2$  is countable.

Hence  $X_1 \times X_2 = \bigcup_{x \in X_1} \{x\} \times X_2$  is countable according to thm.8.3 on p.138. We have proved the base case.

Our induction assumption is that  $X_1 \times \dots \times X_k$  is countable. We must prove that  $X_1 \times \dots \times X_{k+1}$  is countable. We can “identify”

$$(8.18) \quad X_1 \times \dots \times X_{k+1} = (X_1 \times \dots \times X_k) \times X_{k+1}$$

by means of the bijection  $(x_1, \dots, x_n, x_{n+1}) \mapsto ((x_1, \dots, x_n), x_{n+1})$ . According to the induction assumption the set  $X_1 \times \dots \times X_k$  is countable.

The proof for the base case shows that  $X_1 \times \dots \times X_{k+1}$  as the Cartesian product of the two sets  $X_1 \times \dots \times X_k$  and  $X_{k+1}$  is countable. This finishes the proof of the induction step. ■

**Corollary 8.7.** *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. ■

**Theorem 8.5.** *The real numbers are uncountable.*

*Proof by contradiction:* This proof follows the one given in B/G [1] (thm.13.22, p.125).

According to cor.8.3 on p.136 it is enough to prove that the subset  $A$  of all decimals  $x = 0.d_1d_2d_3\dots$  where each  $d_n$  is a digit  $0, 1, 2, \dots, 8$  is uncountable.<sup>74</sup>

Let us assume assume that  $A$  is countable, i.e., there exists a finite or infinite sequence of decimals  $x_n = 0.d_{n,1}d_{n,2}d_{n,3}\dots$  such that  $0 \leq d_{n,j} \leq 8$  for each digit  $d_{n,j}$  as follows: for each  $x \in A$  there exists some  $k = k(x) \in \mathbb{N}$

such that  $x = x_k$ . We will find a contradiction.

For  $n \in \mathbb{N}$  we define the digit  $\tilde{d}_n$  as

$$\tilde{d}_n := \begin{cases} 3 & \text{if } d_{n,n} \neq 3, \\ 4 & \text{if } d_{n,n} = 3. \end{cases}$$

<sup>74</sup> Excluding the digit 9 avoids dealing with numbers which have non-unique decimal representations such as  $1/2 = 0.5\bar{0} = 0.4\bar{9}$ .

Let  $y := 0.\tilde{d}_1\tilde{d}_2\tilde{d}_3\dots$ . Clearly  $y \in A$ . It follows from  $d_{1,1} \neq \tilde{d}_1$  that  $y \neq x_1$ . It follows from  $d_{2,2} \neq \tilde{d}_2$  that  $y \neq x_2$ . You get the idea: Let  $k \in \mathbb{N}$ . It follows from  $d_{k,k} \neq \tilde{d}_k$  that  $y \neq x_k$ . This contradicts the assumption that for each  $x \in A$  there exists some  $k = k(x) \in \mathbb{N}$  such that  $x = x_k$ . ■

**Theorem 8.6.** Let  $a, b \in \mathbb{R}$  such that  $a < b$ . Let  $A$  be one of  $]a, b[, ]a, b], [a, b[, [a, b]$ . Then  $\text{card}(A) = \text{card}(\mathbb{R})$ .

PROOF:

**a.**  $F : \mathbb{R} \rightarrow ]-1, 1[$ ;  $x \mapsto \frac{x}{|x|+1}$  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  $\text{card}(]-1, 1[) = \text{card}(\mathbb{R})$ .

**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  $\text{card}(]-c, c[) = \text{card}(\mathbb{R})$ .

**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  $\text{card}(B) = \text{card}(A)$  because  $B$  is the image of  $A$  under the bijection  $x \mapsto x - \frac{a+b}{2}$ .

**d.** It follows from  $]-c, c[ \subseteq B \subseteq ]-2c, 2c[$  and part **b** that

$$\text{card}(\mathbb{R}) = \text{card}(]-c, c[) \leq \text{card}(B) \leq \text{card}(]-2c, 2c[) = \text{card}(\mathbb{R}).$$

It is a consequence of the Cantor–Schröder–Bernstein Theorem which is formulated and proved later on (ch.8.1.1<sup>75</sup>, thm.8.7 on p.141) that there exists a bijection between  $B$  and  $\mathbb{R}$ . We saw in part **c** that  $\text{card}(B) = \text{card}(A)$ . This proves  $\text{card}(A) = \text{card}(\mathbb{R})$ . ■

**Definition 8.3** (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 8.7** (B/G Prop.13.21, p.125). The set of all algebraic numbers is countable.

PROOF: Let  $P$  be the set of all integer polynomials and  $Z$  the set of zeroes for all such polynomials. Let

$$(8.19) \quad P_n := \{\text{polynomials } p(x) = \sum_{j=0}^n a_j x^j : a_j \in \mathbb{Z} \text{ and } -n \leq a_j \leq n\}.$$

Then  $P_n$  is finite and

$$(8.20) \quad Z_n := \{x \in \mathbb{R} : p(x) = 0 \text{ for some } p \in P_n\}$$

also is finite as a polynomial of degree  $n$  has at most  $n$  zeroes (see [1] B/G (Beck/Geoghegan), prop.6.20).

$Z$  is the countable union of the sets  $Z_n$ . It follows that  $Z$  is countable. ■

**Proposition 8.8.** Let  $k, m, n \in \mathbb{N}$ . Then  $\sqrt[k]{\frac{m}{n}}$  is algebraic.

<sup>75</sup> “Cardinality as a Partial Ordering”. The proof of the Cantor–Schröder–Bernstein Theorem is by no means trivial.

PROOF: Let  $p(x) := nx^k - m$ . Then

$$p(\sqrt[k]{m/n}) = n(\sqrt[k]{m/n})^k - m = \frac{nm}{n} - m = 0. \blacksquare$$

**Proposition 8.9.** Let  $r \in \mathbb{Q}$ . Then  $r$  is algebraic.

PROOF:  $r \in \mathbb{Q}$ . There exist  $m, n \in \mathbb{Z}$  such that  $n \neq 0$  and  $r = \frac{m}{n}$ . Let  $p(x) := nx - m$ . Then

$$p(m/n) = n \cdot \frac{m}{n} - m = 0. \blacksquare$$

Note that if  $m, n > 0$  then the above is a special case of prop.8.8 (let  $k := 1$ ).

Here are some trivial consequences of the fact that  $\mathbb{R}$  is uncountable (see thm. 8.5, p.8.5 and B/G Thm.13.22).

**Proposition 8.10.** All transcendental numbers are uncountable.

PROOF: the uncountable real numbers are the disjoint union of the countable algebraic numbers with the transcendentals. The assertion follows from cor.8.4.  $\blacksquare$

### 8.1.1 Cardinality as a Partial Ordering

We assume in this subchapter that all sets are subsets of a universal set  $\Omega$ . Having such a universal set allows us to declare on its power set  $2^\Omega$  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  $\Omega$ .

We defined in def.8.1 on p.130 the meaning of  $\mathbf{card}(X) = \mathbf{card}(Y)$  and  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  for two sets  $X$  and  $Y$  but we never defined the expression  $\mathbf{card}(X)$  per se. This will be done now.

**Definition 8.4** (Cardinality as an Equivalence Class). Let  $X, Y \subseteq \Omega$ . We say that  $X$  and  $Y$  are equivalent and we write  $X \sim Y$ , if and only if  $\mathbf{card}(X) = \mathbf{card}(Y)$ , 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  $2^\Omega$ . This justifies to define for a set  $X \subseteq \Omega$  its **cardinality** as follows:

$$(8.21) \quad \mathbf{card}(X) := [X] \quad (\text{the equivalence class of } X \text{ w.r.t. “}\sim\text{”}).$$

In other words,

$$(8.22) \quad \mathbf{card}(\emptyset) := \{\emptyset\},$$

$$(8.23) \quad \mathbf{card}(X) := \{Y \subseteq \Omega : \exists \text{ bijection } X \rightarrow Y\} \text{ if } X \neq \emptyset. \quad \square$$

**Proposition 8.11.**  $X \sim Y$  as defined above is an equivalence relation on  $2^\Omega$ .

The proof is left as exercise 8.3.  $\blacksquare$

The following theorem allows us to prove that the relation  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  on  $\mathcal{A} := \{\mathbf{card}(X) : X \subseteq \Omega\}$  is antisymmetric.

**Theorem 8.7** (Cantor–Schröder–Bernstein’s Theorem). Let  $X, Y \subseteq \Omega$  be nonempty. Let there be functions  $f' : X \rightarrow Y$  and  $g' : Y \rightarrow X$  which both are injective. Then there exists a bijection  $X \rightarrow Y$ .

PROOF: The proof given here follows closely that of [1] B/G (Beck/Geoghegan), *Further Topics F: Cardinal Number and Ordinal Number*.

We have no interest in any particulars of the sets  $X$  and  $Y$ . We only are interested in establishing the existence of a bijection  $h : X \rightarrow Y$ . 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 5.12 on p.98).

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

$$(8.24) \quad \sigma : f(X) \uplus g(Y) \rightarrow X \uplus Y; \quad z \mapsto \begin{cases} f^{-1}(z) & \text{if } z \in f(X), \\ g^{-1}(z) & \text{if } z \in g(Y), \end{cases}$$

i.e.,  $\sigma|_{f(X)} = f^{-1}$  and  $\sigma|_{g(Y)} = g^{-1}$ . 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

$$\sigma^2(z) = \sigma(\sigma(z)), \quad \sigma^3(z) = \sigma(\sigma^2(z)), \quad \dots, \sigma^{n+1}(z) = \sigma(\sigma^n(z)), \quad \dots,$$

just as long as  $\sigma^n(z) \in f(X) \uplus g(Y)$ . We further define  $\sigma^0$  for all  $z \in f(X) \uplus g(Y)$  as  $\sigma^0(z) := z$ . We associate with each  $z \in X \uplus Y$  a "score"  $N(z) \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$  as follows.

- a. If  $\sigma^k(z) \in f(X) \uplus g(Y)$  for all  $k \in \mathbb{N}$  then  $N(z) := \infty$ .
- b. If  $\sigma^k(z) \notin f(X) \uplus g(Y)$  for some  $k \in \mathbb{N}$  then  $N(z) := \min\{j \geq 0 : \sigma^j(z) \notin f(X) \uplus g(Y)\}$ .

Note that **b** implies the following: If  $z = \sigma^0(z) \notin f(X) \uplus g(Y)$  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:

$$\begin{aligned} \text{if } x \in g(Y) : & \quad x \xrightarrow{\sigma} f^{-1}(x) \xrightarrow{\sigma} g^{-1}(f^{-1}(x)) \xrightarrow{\sigma} f^{-1}(g^{-1}(f^{-1}(x))) \xrightarrow{\sigma} \dots, \\ \text{if } y \in f(X) : & \quad y \xrightarrow{\sigma} g^{-1}(y) \xrightarrow{\sigma} f^{-1}(g^{-1}(y)) \xrightarrow{\sigma} g^{-1}(f^{-1}(g^{-1}(y))) \xrightarrow{\sigma} \dots \end{aligned}$$

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  $X = X_E \uplus X_O \uplus X_\infty$  and  $Y = Y_E \uplus Y_O \uplus Y_\infty$  of  $X$  and  $Y$ .

Each of the functions  $f, g, \sigma$  changes the score of its argument from odd to even and from even to odd. Hence

$$(8.25) \quad \begin{aligned} f(X_E) &\subseteq Y_O, \quad f^{-1}(Y_O) \subseteq X_E, & f(X_O) &\subseteq Y_E, \quad f^{-1}(Y_E) \subseteq X_O, \\ g(Y_E) &\subseteq X_O, \quad g^{-1}(X_O) \subseteq Y_E, & g(Y_O) &\subseteq X_E, \quad g^{-1}(X_E) \subseteq Y_O, \\ f(X_\infty) &\subseteq Y_\infty, \quad f^{-1}(Y_\infty) \subseteq X_\infty, & g(Y_\infty) &\subseteq X_\infty, \quad g^{-1}(X_\infty) \subseteq Y_\infty, \end{aligned}$$

We define a bijection  $h : X \xrightarrow{\sim} Y$  as follows:

$$h : X \rightarrow Y; \quad x \mapsto \begin{cases} \sigma(x) = g^{-1}(x) & \text{if } x \in X_O \uplus X_\infty, \\ f(x) & \text{if } x \in X_E. \end{cases}$$

Note that  $g^{-1}(x)$  is defined for all  $x \in X_O \uplus X_\infty$  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  $x_1, x_2 \in X_O \uplus X_\infty$ . 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:  $x_1 \in X_O \uplus X_\infty$  and  $x_2 \in X_E$ . It follows from (8.25) 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:  $x_2 \in X_O \uplus X_\infty$  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 1:  $y \in Y_E \uplus Y_\infty$ . It follows from (8.25) that  $g(y) \in X_O \uplus X_\infty$ , hence

$$h(g(y)) = \sigma(g(y)) = g^{-1}(g(y)) = y.$$

Here the second equation follows from (8.25). We have found an item in  $X$  which is mapped by  $h$  to  $y$ , and this proves that  $h$  is surjective.

Case 2:  $y \in Y_O$ . It follows from (8.25) 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. ■

**Corollary 8.8.** The relation  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  on the set  $\mathcal{A} := \{\mathbf{card}(X) : X \subseteq \Omega\}$  is a partial ordering.

*Proof if none of the sets involved is empty:* Reflexivity is obvious, antisymmetry follows from Cantor-Schröder-Bernstein and transitivity follows from prop.5.4.a: The composition of two injective functions is injective.

*Proof if at least one of the sets involved is empty:* Use the fact that if  $\mathbf{card}(A) \leq \mathbf{card}(B)$  and  $B = \emptyset$  then  $A = \emptyset$ .

As an example we prove antisymmetry: If  $X = \emptyset$  and  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  and  $\mathbf{card}(Y) \leq \mathbf{card}(X)$  then the second " $\leq$ " implies  $Y = \emptyset$ , i.e.,  $X = Y$ . On the other hand, if  $Y = \emptyset$  and  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  and  $\mathbf{card}(Y) \leq \mathbf{card}(X)$  then the first " $\leq$ " implies  $X = \emptyset$ , i.e.,  $X = Y$ . ■

**Theorem 8.8.** Let  $X, Y \subseteq \Omega$ . Then  $\mathbf{card}(X) \leq \mathbf{card}(Y)$  or  $\mathbf{card}(Y) \leq \mathbf{card}(X)$ .

In other words, " $\leq$ " is a total ordering on the set of all cardinalities for subsets of  $\Omega$ .

PROOF: The proof will be given in thm.13.2, p.280, of ch.13 (Applications of Zorn's Lemma). ■

## 8.2 Exercises for Ch.8

**Exercise 8.1.** Prove prop.8.2 on p.132:

Let  $X, Y$  be two sets such that  $\mathbf{card}(X) = \mathbf{card}(Y)$ . Then  $\mathbf{card}(2^X) = \mathbf{card}(2^Y)$ . □

**Exercise 8.2.** Prove cor.8.4 on p.138: Let the set  $X$  be uncountable and let  $A \subseteq X$  be countable. Then the complement  $A^c$  of  $A$  is uncountable. □

**Exercise 8.3.** Prove prop.8.11 on p.141: The relation  $X \sim Y$  if and only if  $X = Y$  or both  $X$  and  $Y$  are nonempty and there is a bijective function  $f : X \rightarrow Y$  is an equivalence relation on  $2^\Omega$ . □



### 8.3 Addenda to Ch.8

**Proposition 8.12** (Countability Criterion). *Let  $X \neq \emptyset$ .*

*The following are equivalent:*

- a.**  $X$  is countable.
- b.** There exists an injective  $f : X \rightarrow \mathbb{N}$ .
- c.** There exists a surjective  $g : \mathbb{N} \rightarrow X$ .

*Proof:* It follows from *thm.5.2.c* on *p.97* 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  $\varphi : X \xrightarrow{\sim} [n]$ . Let

$$f : X \rightarrow \mathbb{N}; \quad x \mapsto \varphi(x)$$

be the “same” function as  $\varphi$ , except that we enlarge the codomain to  $\mathbb{N}$ .  $f$  inherits injectivity from  $\varphi$ , and we are done.

**Case 2:** Otherwise, if  $X$  is infinite, i.e., countably infinite, there exists a bijective, hence injective,  $f : X \xrightarrow{\sim} \mathbb{N}$ . 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

$$f' : X \rightarrow f(X); \quad x \mapsto f(x).$$

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  $f'$ . It now follows from *prop.8.1.c* on *p.131* that  $X$  is countable. ■

**Proposition 8.13.** *The function  $F : \mathbb{N}^2 \rightarrow \mathbb{N}; \quad (i, j) \mapsto 2^{i-1}(2j - 1)$  is a bijection.*

**PROOF:** According to *prop.17.14* on *p.315*, for every  $n \in \mathbb{N}$  there exists a unique pair  $i, m \in \mathbb{N}$  such that  $m$  is odd and

$$(8.26) \quad n = 2^{i-1} \cdot m$$

It follows from the Division Algorithm (B/G *Thm.6.13*) for  $n = 2$  that any integer  $k$  can be uniquely written in the form  $k = 2q + r$  with uniquely determined  $q \in \mathbb{Z}$  and  $r = 0$  or  $r = 1$ .  $k$  is odd if and only if  $k$  is not divisible by 2, i.e., if and only if  $r = 1$ , i.e.,  $k = 2q + 1$ .

We replace  $q$  with  $j := q + 2$ , and we obtain  $k = 2j - 1$ .  $j$  is uniquely determined because such is the case with  $q$ .

We claim that if  $k \in \mathbb{N}$  then  $j \in \mathbb{N}$ , i.e.,  $j \geq 1$ . This is true because otherwise it would follow from  $j < 1$  that  $2j \leq 0$ , hence  $k = 2j - 1 < 0$ . This contradicts our assumption that  $k \in \mathbb{N}$ .

We apply all this to  $k := m$  of (8.26) and obtain that for each natural number  $n$  there exist unique  $i, j \in \mathbb{N}$  such that  $n = 2^{i-1} \cdot 2(j - 1)$ , i.e., such that  $F(i, j) = n$ . This proves surjectivity of  $F$ .

It finally follows from the uniqueness of  $i$  and  $j$  for the given  $n$  that there is no different pair  $i', j'$  of natural numbers such that  $n = 2^{i'-1} \cdot 2(j' - 1)$ , i.e., such that  $F(i', j') = n$ . This proves injectivity of  $F$ . ■



## 9 The Completeness of the Real Numbers System (Study this!)

### 9.1 Minima, Maxima, Infima and Suprema

**Definition 9.1** (Upper and lower bounds, maxima and minima). <sup>76</sup> Let  $A \subseteq \mathbb{R}$ . Let  $l, u \in \mathbb{R}$ . We call  $l$  a **lower bound** of  $A$  if  $l \leq a$  for all  $a \in A$ . We call  $u$  an **upper bound** of  $A$  if  $u \geq a$  for all  $a \in A$ .

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

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

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

**Proposition 9.1.** *Let  $A \subseteq \mathbb{R}$ . If  $A$  has a maximum then it is unique. If  $A$  has a minimum then it is unique.*

*Proof for maxima: Let  $u_1$  and  $u_2$  be two maxima of  $A$ : both are upper bounds of  $A$  and both belong to  $A$ . As  $u_1$  is an upper bound, it follows that  $a \leq u_1$  for all  $a \in A$ . Hence  $u_2 \leq u_1$ . As  $u_2$  is an upper bound, it follows that  $u_1 \leq u_2$  and we have equality  $u_1 = u_2$ . The proof for minima is similar.  $\blacksquare$*

**Definition 9.2.** Let  $A \subseteq \mathbb{R}$ . We define

$$(9.1) \quad \begin{aligned} A_{lowb} &:= \{l \in \mathbb{R} : l \text{ is lower bound of } A\} \\ A_{uppb} &:= \{u \in \mathbb{R} : u \text{ is upper bound of } A\}. \end{aligned}$$

$\square$

**Remark 9.1.** Note that  $A$  is bounded above if and only if  $A_{uppb} \neq \emptyset$  and bounded below if and only if  $A_{lowb} \neq \emptyset$ .  $\square$

**Axiom 9.1.** (see [1] B/G axiom 8.52, p.83).

**Completeness axiom for  $\mathbb{R}$ :** Let  $A \subseteq \mathbb{R}$ . If its set of upper bounds  $A_{uppb}$  is not empty then  $A_{uppb}$  has a minimum.

*The above has the status of an axiom due to the fact that the real numbers usually are given axiomatically as an “archimedean ordered field” which satisfies the completeness axiom just stated.*

**Remark 9.2.**  $A_{lowb}$  and/or  $A_{uppb}$  may be empty. Examples are  $A = \mathbb{R}$ ,  $A = \mathbb{R}_{>0}$ ,  $A = \mathbb{R}_{<0}$ .  $\square$

**Definition 9.3.** Let  $A \subseteq \mathbb{R}$ . If  $A_{uppb}$  is not empty then  $\min(A_{uppb})$  exists by axiom 9.1 and it is unique by prop. 9.1. We write  $\sup(A)$  or l.u.b.( $A$ ) for  $\min(A_{uppb})$  and call this element of  $\mathbb{R}$  the **supremum** or **least upper bound** of  $A$ .

We will see in cor.9.1 that, if  $A_{lowb}$  is not empty, then  $\max(A_{lowb})$  exists and is unique by prop. 9.1. We write  $\inf(A)$  or g.l.b.( $A$ ) for  $\max(A_{lowb})$  and call this number the **infimum** or **greatest lower bound** of  $A$ .  $\square$

<sup>76</sup> The definitions are given for  $\mathbb{Z}$  in def.17.2 on p.310.

**Remark:** 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  $P := \mathbb{N}$
- c. Induction axiom: If  $A \subseteq \mathbb{N}$  satisfies **1)**  $1 \in A$ , **2)**  $[n \in A \Rightarrow n + 1 \in A]$ , 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)
- c. Ordered by positive cone  $P := \mathbb{R}_{>0}$
- c. Completeness axiom: If  $A \subseteq \mathbb{R}$  has upper bounds then  $\sup(A)$  exists (as an element of  $\mathbb{R}$ )  $\square$

**Proposition 9.2** (Duality of upper and lower bounds, min and max, inf and sup). Let  $A \subseteq \mathbb{R}$  and  $x \in \mathbb{R}$ . Then the following is true for  $-x$  and  $-A = \{-y : y \in A\}$ :

$$(9.2) \quad \begin{aligned} -x \text{ is a lower bound of } A &\Leftrightarrow x \text{ is an upper bound of } -A, \\ -x = \sup(A) &\Leftrightarrow x = \inf(-A), \\ -x = \max(A) &\Leftrightarrow x = \min(-A). \end{aligned}$$

We switch the roles of  $x$  and  $-x$  and obtain three more equations:

$$(9.3) \quad \begin{aligned} -x \text{ is an upper bound of } A &\Leftrightarrow x \text{ is a lower bound of } -A, \\ -x = \inf(A) &\Leftrightarrow x = \sup(-A), \\ -x = \min(A) &\Leftrightarrow x = \max(-A). \end{aligned}$$

PROOF: A simple consequence of

$$-x \leq y \Leftrightarrow x \geq -y \quad \text{and} \quad -x \geq y \Leftrightarrow x \leq -y. \quad \blacksquare$$

**Corollary 9.1.** Let  $A \subseteq \mathbb{R}$ . If  $A$  has lower bounds then  $\inf(A)$  exists.

PROOF: According to the duality proposition prop.9.2, if  $A$  has lower bounds then  $(-A)$  has upper bounds. It follows from the completeness axiom that  $\sup(-A)$  exists. We apply once more prop.9.2 to prove that  $\inf(A)$  exists:  $\inf(A) = -\sup(-A)$ .

Here are some examples. We define for all three of them  $f(x) := -x$  and  $g(x) := x$ .

**Example 9.1** (Example a: Maximum exists). Let  $X_1 := \{t \in \mathbb{R} : 0 \leq t \leq 1\}$ . For each  $x \in X_1$  we have  $|f(x) - g(x)| = g(x) - f(x) = 2x$  and the biggest possible such difference is  $g(1) - f(1) = 2$ . So  $\max(X_1)$  exists and equals  $\max(X_1) = 2$ .  $\square$

**Example 9.2** (Example b: Supremum is finite). Let  $X_2 := \{t \in \mathbb{R} : 0 \leq t < 1\}$ , i.e., we now exclude the right end point 1 at which the maximum difference was attained. For each  $x \in X$  we have

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

and the biggest possible such difference is certainly bigger than

$$g(0.999999999) - f(0.999999999) = 1.999999998.$$

If you keep adding 5,000 9s to the right of the argument  $x$ , then you get the same amount of 9s inserted into the result  $2x$ , so  $2x$  comes closer than anything you can imagine to the number 2, without actually being allowed to reach it.

The supremum is still considered in a case like this to be 2. This precisely is the difference in behavior between the supremum  $s := \sup(A)$  and the **maximum**  $m := \max(A)$  of a set  $A \subseteq \mathbb{R}$  of real numbers: For the maximum there must actually be at least one element  $a \in A$  so that  $a = \max(A)$ .

For the supremum it is sufficient that there is a sequence  $a_1 \leq a_2 \leq \dots$  which approximates  $s$  from below in the sense that the difference  $s - a_n$  "drops down to zero" as  $n$  approaches infinity. We will not be more exact than this because doing so would require us to delve into the concepts of convergence and contact points.  $\square$

**Example 9.3** (Example c: Supremum is infinite). Let  $X_3 := \mathbb{R}_{\geq 0} = \{t \in \mathbb{R} : 0 \leq t\}$ . For each  $x \in X_3$  we have again  $|f(x) - g(x)| = g(x) - f(x) = 2x$ . But there is no more limit to the right for the values of  $x$ . The difference  $2x$  will exceed all bounds and that means that the only reasonable value for  $\sup\{|f(x) - g(x)| : x \in X_3\}$  is  $+\infty$ .

As in case b above, the max does not exist because there is no  $x_0 \in X_3$  such that  $|f(x_0) - g(x_0)|$  attains the highest possible value among all  $x \in X_3$ .

You should understand that even though  $\sup(A)$  as best approximation of the largest value of  $A \subseteq \mathbb{R}$  is allowed to take the "value"  $+\infty$  or  $-\infty$  this cannot be allowed for  $\max(A)$ .

How so? The infinity values are not real numbers, but, by definition of the maximum, if  $\alpha := \max(A)$  exists, then  $\alpha \in A$ . In particular, the max must be a real number.  $\square$

*That last example motivates the following definition.*

**Definition 9.4** (Supremum and Infimum of unbounded and empty sets). If  $A$  is not bounded from above then we define

$$(9.4) \quad \sup A = \infty$$

If  $A$  is not bounded from below then we define

$$(9.5) \quad \inf A = -\infty$$

Finally we define

$$(9.6) \quad \sup \emptyset = -\infty, \quad \inf \emptyset = +\infty. \quad \square$$

**Proposition 9.3.** Let  $A \subseteq B \subseteq \mathbb{R}$ .

- a. Then  $\inf(A) \geq \inf(B)$  and  $\sup(A) \leq \sup(B)$ .
- b. If both  $A$  and  $B$  have a min then  $\min(A) \geq \min(B)$ . If both  $A$  and  $B$  have a max then  $\max(A) \leq \max(B)$ .

*PROOF:*

**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 **a** for suprema. The proof for infima is similar. First we note that if at least one of the sets is empty or not bounded above then the proof is trivial.<sup>77</sup> We may assume that both  $A$  and  $B$  are not empty

<sup>77</sup> Note that if  $A$  is not bounded above then the same holds for the superset  $B$  and that if  $B$  is empty then  $A$  is empty.

and bounded above.

It follows from  $A \subseteq B$  that any upper bound of  $B$  also is an upper bound of  $A$ .

In particular  $\sup(B) \in A_{\text{uppb}}$ , hence  $\sup(A) = \min(A_{\text{uppb}}) \leq \sup(B)$ . ■

**Definition 9.5** (Translation and dilation of sets of real numbers). Let  $A \subseteq \mathbb{R}$ <sup>78</sup> and  $\alpha, b \in \mathbb{R}$ . We define

$$(9.7) \quad \lambda A + b := \{\lambda a + b : a \in A\}.$$

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

$$(9.8) \quad A + b = \{a + b : a \in A\},$$

$$(9.9) \quad -A = \{-a : a \in A\}. \quad \square$$

**Proposition 9.4** (The sup of a set is 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

$$(9.10) \quad \sup(\lambda A) = \lambda \sup(A).$$

*PROOF:* (9.10) holds for  $\lambda = 0$  and  $\sup(A) < \infty$  because

$$\sup(0A) = \sup(\{0\}) = 0 = 0 \cdot \sup(A).$$

Let  $\lambda > 0$ . If  $B \subseteq \mathbb{R}$ , let  $B_{\text{uppb}} := \{u \in \mathbb{R} : u \text{ is upper bound of } B\}$ . Note that, because  $\lambda > 0$ ,

$$(9.11) \quad u \in A_{\text{uppb}} \Leftrightarrow u \geq a \forall a \in A \Leftrightarrow \lambda u \geq \lambda a \forall a \in A \Leftrightarrow \lambda u \in (\lambda A)_{\text{uppb}}.$$

It follows from  $\sup(A) \in A_{\text{uppb}}$  that  $\lambda \sup(A) \in (\lambda A)_{\text{uppb}}$ , hence  $\lambda \sup(A) \geq \min((\lambda A)_{\text{uppb}}) = \sup(\lambda A)$ .

It remains to show that  $\lambda \sup(A) \leq \sup(\lambda A)$ .

We substitute  $\frac{v}{\lambda}$  for  $u$  in (9.11) and obtain  $\frac{v}{\lambda} \in A_{\text{uppb}} \Leftrightarrow v \in (\lambda A)_{\text{uppb}}$ .

It follows from  $\sup(\lambda A) \in (\lambda A)_{\text{uppb}}$  that  $\frac{\sup(\lambda A)}{\lambda} \in A_{\text{uppb}}$ , hence  $\frac{\sup(\lambda A)}{\lambda} \geq \min(A_{\text{uppb}}) = \sup(A)$ . This proves  $\lambda \sup(A) \leq \sup(\lambda A)$ . ■

**Definition 9.6** (bounded functions). Given is a nonempty set  $X$ . A real-valued function  $f(\cdot)$  with domain  $X$  is called **bounded from above** if there exists a (possibly very large) number  $\gamma_1 > 0$  such that

$$(9.12) \quad f(x) < \gamma_1 \quad \text{for all arguments } x.$$

It is called **bounded from below** if there exists a (possibly very large) number  $\gamma_2 > 0$  such that

$$(9.13) \quad f(x) > -\gamma_2 \quad \text{for all arguments } x.$$

It is called a **bounded function** if it is both bounded from above and below. It is obvious that if you set  $\gamma := \max(\gamma_1, \gamma_2)$  then bounded functions are exactly those that satisfy the inequality

$$(9.14) \quad |f(x)| < \gamma \quad \text{for all arguments } x. \quad \square$$

<sup>78</sup> See also def.17.3 in ch.17.2.3

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 9.7** (supremum and infimum of functions). Let  $X$  be an arbitrary set,  $A \subseteq X$  a subset of  $X$ ,  $f : X \rightarrow \mathbb{R}$  a real valued function on  $X$ . Look at the set  $f(A) = \{f(x) : x \in A\}$ , i.e., the image of  $A$  under  $f(\cdot)$ .

The **supremum of  $f(\cdot)$  on  $A$**  is then defined as

$$(9.15) \quad \sup_A f := \sup_{x \in A} f(x) := \sup f(A)$$

The **infimum of  $f(\cdot)$  on  $A$**  is then defined as

$$(9.16) \quad \inf_A f := \inf_{x \in A} f(x) := \inf f(A). \quad \square$$

**Definition 9.8** (supremum and infimum of families). Let  $(x_i)_{i \in I}$  be an indexed family of real numbers  $x_i$ .

The **supremum of  $(x_i)_{i \in I}$**  is then defined as

$$(9.17) \quad \sup(x_i) := \sup_i(x_i) := \sup(x_i)_i := \sup(x_i)_{i \in I} := \sup_{i \in I} x_i := \sup \{x_i : i \in I\}.$$

The **infimum of  $(x_i)_{i \in I}$**  is then defined as

$$(9.18) \quad \inf(x_i) := \inf_i(x_i) := \inf(x_i)_i := \inf(x_i)_{i \in I} := \inf_{i \in I} x_i := \inf \{x_i : i \in I\}. \quad \square$$

The definition above for families extends to sequences. Let  $I = \{k \in \mathbb{Z} : k \geq k_0 \text{ for some } k_0 \in \mathbb{Z}\}$ .

**Definition 9.9** (supremum and infimum of sequences). Let  $(x_n)_{n \in I}$  be a sequence of real numbers  $x_n$ . The **supremum of  $(x_n)_{n \in I}$**  is then defined as

$$(9.19) \quad \sup(x_n) := \sup(x_n)_{n \in I} := \sup_{n \in I} x_n = \sup \{x_n : n \in I\}$$

The **infimum of  $(x_n)_{n \in I}$**  is then defined as

$$(9.20) \quad \inf(x_n) := \inf(x_n)_{n \in I} := \inf_{n \in I} x_n = \inf \{x_n : n \in I\}. \quad \square$$

We note that the “duality principle” for min and max, sup and inf (9.2 on p.146) 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 9.5.**  $X$  be a nonempty set and  $\varphi, \psi : X \rightarrow \mathbb{R}$  be two real valued functions on  $X$ . Let  $A \subseteq X$ . Then

$$(9.21) \quad \sup\{\varphi(x) + \psi(x) : x \in A\} \leq \sup\{\varphi(y) : y \in A\} + \sup\{\psi(z) : z \in A\},$$

$$(9.22) \quad \inf\{\varphi(x) + \psi(x) : x \in A\} \geq \inf\{\varphi(y) : y \in A\} + \inf\{\psi(z) : z \in A\}.$$

PROOF:

We only prove (9.21). The proof of (9.22) is similar. <sup>79</sup>

Let  $U := \{\varphi(x) + \psi(x) : x \in A\}$ ,  $V := \{\varphi(y) : y \in A\}$ ,  $W := \{\psi(z) : z \in A\}$ . Let  $x \in A$ .

Then  $\sup(V)$  is an upper bound of  $\varphi(x)$  and  $\sup(W)$  is an upper bound of  $\psi(x)$ ,

hence  $\sup(V) + \sup(W) \geq \varphi(x) + \psi(x)$ .

This is true for all  $x \in A$ , hence  $\sup(V) + \sup(W)$  is an upper bound of  $U$ .

It follows that  $\sup(V) + \sup(W)$  dominates the least upper bound  $\sup(U)$  of  $U$  and this proves (9.21).

■

## 9.2 Convergence and Continuity in $\mathbb{R}$

You are familiar with the concepts of convergent sequences and continuous functions whose domain and codomain both are sets of real numbers from calculus. 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.

**Definition 9.10** (convergence of sequences of real numbers). We say that a sequence  $(x_n)$  of real numbers **converges** <sup>80</sup> to  $a \in \mathbb{R}$  for  $n \rightarrow \infty$  if almost all of the  $x_n$  will come arbitrarily close to  $a$  in the following sense:

For any  $\delta \in \mathbb{R}_{>0}$  (no matter how small) there exists (possibly extremely large)  $n_0 \in \mathbb{N}$  such that

$$(9.23) \quad |a - x_j| < \delta \quad \text{for all } j \geq n_0.$$

We write either of

$$(9.24) \quad a = \lim_{n \rightarrow \infty} x_n \quad \text{or} \quad x_n \rightarrow a$$

and we call  $a$  the **limit** of the sequence  $(x_n)$ . □

There is an equivalent way of expressing convergence towards  $a \in \mathbb{R}$ : No matter how small a "neighborhood"  $]a - \delta, a + \delta[$  of  $a$  you choose: at most finitely many of the  $x_n$  will be located outside that neighborhood.

**Example 9.4.** Some simple examples for convergence to a real number:

- a. Let  $x_n := 1/n$  ( $n \in \mathbb{N}$ ). Then  $x_n \rightarrow 0$  as  $n \rightarrow \infty$ .
- 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  $\lim_{n \rightarrow \infty} z_n = x_0$ . □

<sup>79</sup> (9.22) can also be deduced from (9.21) and the fact that  $\inf\{\varphi(u) : u \in A\} = -\sup\{-\varphi(v) : v \in A\}$ .

<sup>80</sup> We will define convergence of a sequence of items more general than real numbers in ch.11.1.4 (see def.11.11 (convergence of sequences in metric spaces) on p.213).

*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. It follows for  $n \geq n_0$  that

$$|x_n - 0| = 1/n \leq 1/n_0 < \delta.$$

*PROOF of b and c:* Left as an exercise. ■

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  $\infty$  and  $y_n$  converges to  $-\infty$  because we think of very big numbers as being very close to  $+\infty$  and very small numbers (i.e., very big ones with a minus sign) as being very close to  $-\infty$ .

**Definition 9.11** (Limit infinity). Given a real number  $K > 0$ , we define

$$(9.25a) \quad N_K(\infty) := \{x \in \mathbb{R} : x > K\}$$

$$(9.25b) \quad N_K(-\infty) := \{x \in \mathbb{R} : x < -K\}$$

We call  $N_K(\infty)$  the  $K$ -neighborhood of  $\infty$  and  $N_K(-\infty)$  the  $K$ -neighborhood of  $-\infty$ . We say that a sequence  $(x_n)$  has limit  $\infty$  and we write either of

$$(9.26) \quad x_n \rightarrow \infty \quad \text{or} \quad \lim_{n \rightarrow \infty} x_n = \infty$$

if the following is true for any (big)  $K$ : There is a (possibly extremely large) integer  $n_0$  such that all  $x_j$  belong to  $N_K(\infty)$  just as long as  $j \geq n_0$ , i.e., if

$$\text{for all } K \in \mathbb{N} \text{ there exists } n_0 \in \mathbb{N} \text{ such that if } j \geq n_0 \text{ then } x_j > K.$$

We say that the sequence  $(x_n)$  has limit  $-\infty$  and we write either of

$$(9.27) \quad x_n \rightarrow -\infty \quad \text{or} \quad \lim_{n \rightarrow \infty} x_n = -\infty$$

if the following is true for any (big)  $K$ : There is a (possibly extremely large) integer  $n_0$  such that all  $x_j$  belong to  $N_K(-\infty)$  just as long as  $j \geq n_0$ . □

*a.* There is an equivalent way of stating that the sequence  $(x_n)$  has limit  $\infty$ : No matter how big a threshold  $K > 0$  you choose: at most finitely many of the  $x_n$  will be located above that threshold.  
*b.*  $x_n \rightarrow -\infty$  can also be expressed as follows: No matter how big a threshold  $K > 0$  you choose: at most finitely many of the  $x_n$  will be located below  $-K$ .

**Remark 9.3.**

The majority of mathematicians agrees that there is no “convergence to  $\infty$ ” or “divergence to  $\infty$ ”. Rather, they will state that a sequence has the limit  $\infty$ . We follow this convention □

**Proposition 9.6** (Subsequences of convergent sequences). Let  $(x_n)_n$  be a sequence of real numbers with limit  $L := \lim_{n \rightarrow \infty} x_n$ . Let  $(x_{n_j})$  be a subsequence. Then  $\lim_{j \rightarrow \infty} x_{n_j} = L$ .

PROOF:

**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  $\lim_{j \rightarrow \infty} x_{n_j} = \infty$ , 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  $\infty$ .

The case  $L = -\infty$  is proved similarly. ■

**Note 9.1** (Notation for limits of monotone sequences). Let  $(x_n)$  be a non-decreasing sequence of real numbers and let  $(y_n)$  be a non-increasing sequence. If  $\xi = \lim_{k \rightarrow \infty} x_k$  (that limit might be  $+\infty$ ) then we write suggestively

$$x_n \nearrow \xi \quad (n \rightarrow \infty)$$

If  $\eta = \lim_{j \rightarrow \infty} y_j$  (that limit might be  $-\infty$ ) then we write suggestively

$$y_j \searrow \eta \quad (j \rightarrow \infty) \quad \square$$

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 9.7.** 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  $n \geq K$ . Let  $L \in \mathbb{R}$ . Then

$$\lim_{n \rightarrow \infty} x_n = L \Leftrightarrow \lim_{n \rightarrow \infty} y_n = L, \quad \lim_{n \rightarrow \infty} x_n = \pm\infty \Leftrightarrow \lim_{n \rightarrow \infty} y_n = \pm\infty.$$

PROOF:

**Case 1:**  $\lim_{n \rightarrow \infty} x_n = L$ .

Let  $\delta > 0$ . Then there exists  $N' \in \mathbb{N}$  such that

$$(9.28) \quad |x_j - L| < \delta \text{ for all } j \geq N'.$$

Let  $N := \max(K, N')$ . It follows from  $N \geq K$ ,  $N \geq N'$   $x_n = y_n$  for all  $n \geq K$  and (9.28) that

$$(9.29) \quad |y_j - L| = |x_j - L| < \delta \text{ for all } j \geq N.$$

This proves that  $y_n$  converges to  $L$ .

**Case 2:**  $\lim_{n \rightarrow \infty} x_n = \infty$ .

Let  $M \in \mathbb{R}$ . It follows from  $x_n \rightarrow \infty$  and def.9.11 that there exists  $N' \in \mathbb{N}$  such that

$$(9.30) \quad x_j > M \text{ for all } j \geq N'.$$

Let  $N := \max(K, N')$ . It follows from  $N \geq K$ ,  $N \geq N'$   $x_n = y_n$  for all  $n \geq K$  and (9.30) that

$$(9.31) \quad y_j = x_j > M \text{ for all } j \geq N.$$

This proves that  $\lim_{n \rightarrow \infty} y_n = \infty$ .

**Case 3:**  $\lim_{n \rightarrow \infty} x_n = -\infty$ .

This is true according to the already proven case 2, applied to the sequences  $(-x_n)_n$  and  $(-y_n)_n$ . ■



**Proposition 9.8.** [See B/G prop.10.16]

Let  $(x_n)_n$  be a sequence of real numbers such that  $\lim_{n \rightarrow \infty} x_n$  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.

PROOF:

**Case 1:**  $\lim_{n \rightarrow \infty} x_n = L \in \mathbb{R}$ .

The proof is left as exercise 9.3.

**Case 2:**  $\lim_{n \rightarrow \infty} x_n = \infty$ .

Let  $M \in \mathbb{R}$ . It follows from  $x_n \rightarrow \infty$  and def.9.11 that there exists  $N \in \mathbb{N}$  such that

$$(9.32) \quad x_j > M \text{ for all } j \geq N.$$

Let  $j \geq N$ . Then  $j + K \geq N$  and it follows from (9.32) that

$$(9.33) \quad y_j = x_{j+K} > M.$$

This proves that  $\lim_{n \rightarrow \infty} y_n = \infty$ .

**Case 3:**  $\lim_{n \rightarrow \infty} x_n = -\infty$ .

This is true according to the already proven case 2, applied to the sequences  $(-x_n)_n$  and  $(-y_n)_n$ . ■

**Proposition 9.9** (convergent  $\Rightarrow$  bounded). Let  $(x_n)_n$  be a sequence in  $\mathbb{R}$  with limit  $a \in \mathbb{R}$ . Then this sequence is bounded.

PROOF: We must prove that there exists  $K \in \mathbb{R}$  such that  $|x_j| \leq K$  for all  $j \in \mathbb{N}$ .

Let  $\delta = 1$  in (9.23). It follows that there is  $n_0 \in \mathbb{N}$  such that  $|x_j - a| < 1$  for all  $j \geq n_0$ . But then

$$|x_j| = |(x_j - a) + a| \leq |x_j - a| + |a| \leq |a| + 1 \text{ for all } j \geq n_0.$$

Let  $K := \max(1, |x_1|, |x_2|, \dots, |x_{n_0-1}|)$ . It follows that  $|x_j| \leq K + |a|$  for all  $j \in \mathbb{N}$ . ■

The following proposition states that the product of a sequence which converges to zero and a bounded sequence converges to zero.

**Proposition 9.10** (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  $\lim_{n \rightarrow \infty} x_n = 0$  and if  $|\alpha_j| \leq \alpha$  for all  $j \in \mathbb{N}$  then

$$(9.34) \quad \lim_{j \rightarrow \infty} (\alpha_j x_j) = 0.$$

PROOF:

**Case 1:**  $\alpha = 0$ . Then  $\alpha_j = 0$  for all  $j \in \mathbb{N}$ , hence  $\alpha_j x_j = 0$ . For any  $\delta > 0$  let  $n_0 = 1$ . Then

$$|\alpha_j x_j - 0| = |\alpha_j x_j| = 0 < \delta \text{ for all } j \in \mathbb{N} \text{ such that } j \geq n_0.$$

This proves convergence  $\alpha_j x_j \rightarrow 0$ .

**Case 2:**  $\alpha \neq 0$ , i.e.,  $|\alpha| > 0$ . Let  $\delta > 0$ . We must show that

$$(9.35) \quad \text{there is } n_0 \in \mathbb{N} \text{ such that } |\alpha_j x_j| < \delta \text{ for all } j \in \mathbb{N} \text{ such that } j \geq n_0.$$

Let  $\varepsilon := \delta/|\alpha|$ . Then  $\varepsilon > 0$  and it follows from  $\lim_{j \rightarrow \infty} x_j = 0$  that

$$(9.36) \quad \text{there is } N \in \mathbb{N} \text{ such that } |x_j| < \varepsilon \text{ for all } j \in \mathbb{N} \text{ such that } j \geq N.$$

But then  $|\alpha_j x_j| = |\alpha_j| \cdot |x_j| < |\alpha| \cdot \varepsilon = \delta$  for all  $j \in \mathbb{N}$  such that  $j \geq N$ . We choose  $n_0 := N$  and (9.35) follows. ■

It is very rare that you need to apply def.9.10 on p.150 to compute a limit. Rather, the previous proposition and the following set of rules are employed.

**Proposition 9.11** (Rules of arithmetic for limits<sup>81</sup>). Let  $(x_n)_n$  and  $(y_n)_n$  be two sequences in  $\mathbb{R}$  and let  $x, y, \alpha \in \mathbb{R}$ .

Let  $\lim_{j \rightarrow \infty} x_j = x$  and  $\lim_{j \rightarrow \infty} y_j = y$ . Then

- a.  $\lim_{j \rightarrow \infty} \alpha = \alpha$ ,
- b.  $\lim_{j \rightarrow \infty} (\alpha \cdot x_j) = \alpha \cdot x$ ,
- c.  $\lim_{j \rightarrow \infty} (x_j + y_j) = x + y$ ,
- d.  $\lim_{j \rightarrow \infty} (x_j \cdot y_j) = x \cdot y$ ,
- e. if  $x \neq 0$  then  $\lim_{j \rightarrow \infty} \frac{1}{x_j} = \frac{1}{x}$ .

PROOF of a: Exercise 9.1.

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

$$(9.37) \quad \text{there is } n_0 \in \mathbb{N} \text{ such that } |\alpha x_j - \alpha x| < \delta \text{ for all } j \in \mathbb{N} \text{ such that } j \geq n_0.$$

Let  $\varepsilon := \delta/|\alpha|$ . Then  $\varepsilon > 0$  and it follows from  $\lim_{j \rightarrow \infty} x_j = x$  that

$$(9.38) \quad \text{there is } N \in \mathbb{N} \text{ such that } |x_j - x| < \varepsilon \text{ for all } j \in \mathbb{N} \text{ such that } j \geq N.$$

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 (9.37) is proved.

PROOF of c:

Let  $\delta > 0$ . It follows from  $\lim_{j \rightarrow \infty} x_j = x$  and  $\lim_{j \rightarrow \infty} y_j = y$  that there exist  $N_1, N_2 \in \mathbb{N}$  such that

$$(9.39) \quad \text{if } j \geq N_1 \text{ then } |x_j - x| < \delta/2 \text{ and if } j \geq N_2 \text{ then } |y_j - y| < \delta/2.$$

It follows from the triangle inequality  $|A + B| \leq |A| + |B|$  (prop.2.3 on p.22) and from (9.39) that

$$(9.40) \quad |(x_j + y_j) - (x + y)| = |(x_j - x) + (y_j - y)| \leq |x_j - x| + |y_j - y| < \delta/2 + \delta/2 = \delta$$

for all  $j \geq \max(N_1, N_2)$ . Let  $n_0 := \max(N_1, N_2)$ . It follows from (9.40) that  $|(x_j + y_j) - (x + y)| < \delta$  for all  $j \geq n_0$ . This proves c.

---

<sup>81</sup> See [1] B/G (Beck/Geoghegan) prop.10.23

PROOF of **d**:

Let  $u_j := (x_j - x)y_j$  and  $v_j := x(y_j - y)$  ( $j \in \mathbb{N}$ ). Then

$$(9.41) \quad x_j y_j - xy = (x_j y_j - x y_j) + (x y_j - xy) = (x_j - x)y_j + x(y_j - y) = u_j + v_j.$$

The convergent sequences  $(y_n)_n$  and  $(x)_n$  (constant sequence!) are bounded by prop.9.9 and it follows from parts **c** and **a** that

$$\begin{aligned} \lim_{j \rightarrow \infty} (x_j - x) &= \lim_{j \rightarrow \infty} x_j - \lim_{j \rightarrow \infty} x = x - x = 0, \\ \lim_{j \rightarrow \infty} (y_j - y) &= \lim_{j \rightarrow \infty} y_j - \lim_{j \rightarrow \infty} y = y - y = 0, \end{aligned}$$

i.e., the sequences  $x_j - x$  and  $y_j - y$  converge to zero.

It now follows from prop.9.10 that  $\lim_{j \rightarrow \infty} u_j = 0$  and  $\lim_{j \rightarrow \infty} v_j = 0$ . According to (9.41)  $x_j y_j = xy + u_j + v_j$  is the limit of three convergent sequences.<sup>82</sup> It follows from part **c** that

$$\lim_{j \rightarrow \infty} x_n y_n = \lim_{j \rightarrow \infty} (xy) + \lim_{j \rightarrow \infty} u_j + \lim_{j \rightarrow \infty} v_j = xy + 0 + 0 = xy.$$

PROOF of **e**:

Because  $\lim_{j \rightarrow \infty} x_n = x$  and  $|x| > 0$  there exists  $N_1 \in \mathbb{N}$  such that  $|x_n - x| \leq |x|/2$  for all  $j \geq N_1$ . Hence

$$(9.42) \quad \begin{aligned} |x| &= |(x - x_n) + x_n| \leq |x - x_n| + |x_n| \leq \frac{|x|}{2} + |x_n| \\ \Rightarrow \frac{|x|}{2} &\leq |x_n| \Rightarrow |x| |x_n| \geq \frac{x^2}{2} \Rightarrow \frac{1}{|x x_n|} \leq \frac{2}{x^2}. \end{aligned}$$

Let  $z_n := (x x_n)^{-1}$  and  $K := \max(2/x^2, |z_1|, |z_2|, \dots, |z_{N_1}|)$ . It follows from (9.42) that the sequence  $(z_n)_n$  is bounded by  $K$ . It follows from part **a** that  $\lim_{j \rightarrow \infty} x = x$  and from part **c** that

$$\lim_{j \rightarrow \infty} (x_n - x) = \lim_{j \rightarrow \infty} (x_n) - x = 0.$$

It now follows from prop.9.10 that

$$(9.43) \quad \lim_{j \rightarrow \infty} \left( \frac{1}{x_n} - \frac{1}{x} \right) = \lim_{j \rightarrow \infty} \frac{1}{x x_n} \cdot (x - x_n) = \lim_{j \rightarrow \infty} z_n (x - x_n) = 0.$$

Let  $\delta > 0$ . On account of (9.43) there exists  $n_0 \in \mathbb{N}$  such that

$$(9.44) \quad \left| \frac{1}{x_n} - \frac{1}{x} \right| = \left| \left( \frac{1}{x_n} - \frac{1}{x} \right) - 0 \right| < \delta \text{ for all } j \geq n_0.$$

This proves convergence of  $1/x_n$  to  $1/x$ . ■

### Proposition 9.12.

- Let  $x_n$  be a sequence of real numbers that is non-decreasing, i.e.,  $x_n \leq x_{n+1}$  for all  $n$  (see def. 18.1 on p.329), and bounded above. Then  $\lim_{n \rightarrow \infty} x_n$  exists and coincides with  $\sup\{x_n : n \in \mathbb{N}\}$
- Further, if  $y_n$  is a sequence of real numbers that is non-increasing, i.e.,  $y_n \geq y_{n+1}$  for all  $n$ , and bounded below, the analogous result is that  $\lim_{n \rightarrow \infty} y_n$  exists and coincides with  $\inf\{y_n : n \in \mathbb{N}\}$ .

<sup>82</sup> The constant sequence  $(xy)$  has limit  $xy$  according to part **a**

*PROOF of a:* Let  $x := \sup_{n \in \mathbb{N}} x_n$ . This is an upper bound of the sequence, hence  $x_j \leq x$  for all  $j \in \mathbb{N}$ . Let  $\varepsilon > 0$ .  $x - \frac{\varepsilon}{2}$  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 non-decreasing, it follows for all  $j \geq N$  that

$$x - \varepsilon < x - \frac{\varepsilon}{2} \leq x_N \leq j \leq x, \text{ hence } |x_j - x| = x - x_j < \varepsilon \text{ for all } j \geq N.$$

This proves that  $x = \sup_{n \in \mathbb{N}} x_n = \lim_{j \rightarrow \infty} x_j$ .

The proof of **b** is similar (or you apply **a** to the sequence  $x_n := -y_n$ ). ■

**Proposition 9.13** (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

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

*PROOF:* Let  $x_0 := \lim_{n \rightarrow \infty} x_n$  and  $y_0 := \lim_{n \rightarrow \infty} y_n$ . We will show that  $\lim_{n \rightarrow \infty} (y_n - x_n) \geq 0$ . This suffices to prove the proposition because, according to prop. 9.11.c,

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

We abbreviate  $z_n := y_n - x_n$  and  $z_0 := \lim_{n \rightarrow \infty} z_n$ . We assume to the contrary that  $z_0 < 0$ . Let  $\varepsilon := -\frac{1}{2}z_0$ . Then  $\varepsilon > 0$ . It follows from the definition of limits that there exists  $N \in \mathbb{N}$  such that

$$|z_j - z_0| < \varepsilon, \text{ i.e., } |z_j + 2\varepsilon| < \varepsilon, \text{ i.e., } -\varepsilon < z_j + 2\varepsilon < \varepsilon, \text{ i.e., } -3\varepsilon < z_j < -\varepsilon < 0. \quad (\star)$$

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  $j$ . ■

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.11.2.1 on p.233. Let  $A \subseteq \mathbb{R}$ . Informally speaking, a continuous function  $f : A \rightarrow \mathbb{R}$  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  $\mathbb{R}$ ). Let  $A \subseteq \mathbb{R}$ ,  $x_0 \in A$ , and let  $f : A \rightarrow \mathbb{R}$  be a real-valued function with domain  $A$ . We say that  $f$  is **continuous at  $x_0$** <sup>83</sup> and we write

$$(9.45) \quad \lim_{x \rightarrow x_0} f(x) = f(x_0)$$

if the following is true for **any** sequence  $(x_n)$  with values in  $A$ :

$$(9.46) \quad \text{if } x_n \rightarrow x_0 \text{ then } f(x_n) \rightarrow f(x_0).$$

In other words, the following must be true for any sequence  $(x_n)$  in  $A$  and  $x_0 \in A$ :

---

<sup>83</sup> We call such a function **sequence continuous** in def.11.30 (Sequence continuity) on p.233 where continuity is generalized to metric spaces.

$$(9.47) \quad \lim_{n \rightarrow \infty} x_n = x_0 \Rightarrow \lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right) = f(x_0).$$

We say that  $f$  is **continuous** if  $f$  is continuous at  $x_0$  for all  $x_0 \in A$ .  $\square$

**Remark 9.4.** 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  $x_n$ .  $\square$

**Proposition 9.14** (Rules of arithmetic for continuous real-valued functions with domain in  $\mathbb{R}$ ). *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  $x_0$ . Hence, setting  $\alpha := -1$ ,  $-f(\cdot) : x \mapsto -f(x)$  is continuous at  $x_0$ .*
- c. *The sum  $f + g(\cdot) : x \mapsto f(x) + g(x)$  is continuous at  $x_0$ .*
- d. *Any linear combination <sup>84</sup>  $\sum_{j=0}^n a_j f_j(\cdot) : x \mapsto \sum_{j=0}^n a_j f_j(x)$  is continuous in  $x_0$ .*

*PROOF:* This proposition will be generalized in prop.11.26 on p.238 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  $n \in \mathbb{N}$  such that  $\lim_{n \rightarrow \infty} x_n = x_0$ . We must show  $\lim_{n \rightarrow \infty} f(x_n)g(x_n) = f(x_0)g(x_0)$ , and this follows from part d of prop.9.11 (Rules of arithmetic for limits) on p.154.  $\blacksquare$

**Theorem 9.1.** *Let  $A \subseteq \mathbb{R}$ ,  $x_0 \in A$ , and let  $f : A \rightarrow \mathbb{R}$  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:*

$$(9.48) \quad f(\{x \in A : |x - x_0| < \delta\}) \subseteq \{y \in \mathbb{R} : |y - f(x_0)| < \varepsilon\},$$

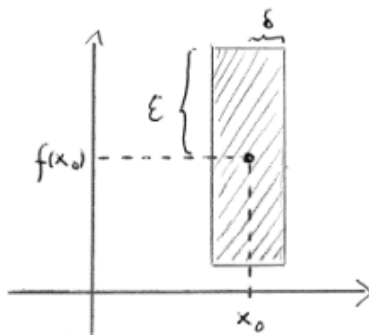
$$(9.49) \quad |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon \text{ for all } x \in A.$$

*PROOF:* A generalized version of this theorem will be proved in a later chapter. <sup>85</sup>  $\blacksquare$

<sup>84</sup>See def.10.6 (linear combinations) on p.183

<sup>85</sup> See thm.11.13 on p.234.

Figure 9.1:  $\varepsilon$ - $\delta$  continuity: Any  $(x, f(x))$  such that  $x_0 - \delta < x < x_0 + \delta$  must be contained in the shaded rectangle.



### 9.3 Geometric Series

The following chapter provides some basic facts about geometric series. More advanced material on series of real numbers is deferred to ch.11.3.2 (Infinite Series) on p.248 because it needs the concept of Cauchy-sequences.<sup>86</sup>

**Definition 9.13** (Numeric Sequences and Series). A sequence  $(a_j)$  is called a **numeric sequence** if each  $a_j$  is a real number. For any such sequence, we can build another sequence  $(s_n)$  as follows:

$$(9.50) \quad s_1 := a_1; \quad s_2 := a_1 + a_2; \quad s_3 := a_1 + a_2 + a_3; \cdots \quad s_n := \sum_{k=1}^n a_k$$

We write this more compactly as

$$(9.51) \quad a_1 + a_2 + a_3 + \cdots = \sum a_k$$

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

$$(9.52) \quad \sum_{k=1}^{\infty} a_k = s$$

if this is true for the associated sequence of partial sums (9.50). We then also say that the **series has limit**  $s$ .

We say that the **series has limit**  $\pm\infty$  if this true for the associated partial sums. In this case we write

$$(9.53) \quad \sum_{k=1}^{\infty} a_k = \pm\infty. \quad \square$$

<sup>86</sup> See ch.11.1.10 (Completeness in Metric Spaces) on p.226

**Proposition 9.15** (Limits of Geometric Series). *A. Let  $|q| < 1$ . Then  $\lim_{j \rightarrow \infty} q^n = 0$ . Proof: AoP end-of-ch.10.4 (limits)*

*B. Equations for geom. series:*

$$\sum_{j=0}^n q^j = \frac{1 - q^{n+1}}{1 - q}$$

$$\sum_{j=0}^{\infty} q^j = \frac{1}{1 - q}$$

PROOF:

Later!

Later!

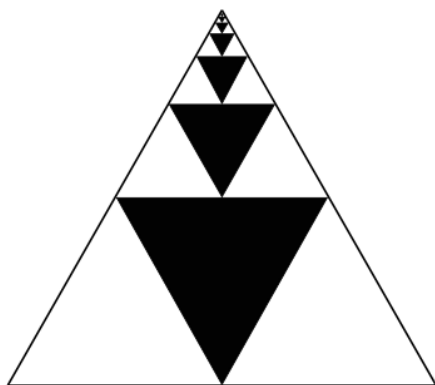
Later! ■

**Example 9.5** (Cover page of the B/G book). Examine the geometric series with  $q = \frac{1}{4}$ :

$$\sum_{j=0}^n (1/4)^j = \frac{4}{3} = 1 + \sum_{j=1}^n (1/4)^j, \quad \text{i.e., } \sum_{j=1}^n (1/4)^j = \frac{1}{3}$$

It has the following geometric meaning:

- 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  $j + 1$ .
- 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 [1] B/G: Art of Proof. □

## 9.4 Limit Inferior and Limit Superior

**Definition 9.14** (Tail sets of a sequence). Let  $(x_k)_{k \in \mathbb{N}}$  be a sequence in  $\mathbb{R}$ . Let

$$(9.54) \quad T_n := \{x_j : j \in \mathbb{N} \text{ and } j \geq n\} = \{x_n, x_{n+1}, x_{n+2}, x_{n+3}, \dots\}$$

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  $(x_k)_k$ .  $\square$

**Remark 9.5.** 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.9.3 on p.147 and prop.9.12 that

$$\begin{aligned}\beta_n &:= \sup(T_n) \text{ is non-increasing, hence } \lim_{n \rightarrow \infty} \beta_n = \inf_n \beta_n; \\ \alpha_n &:= \inf(T_n) \text{ is non-decreasing, hence } \lim_{n \rightarrow \infty} \alpha_n = \sup_n \alpha_n.\end{aligned}$$

These limits can also be expressed as follows.

$$(9.55) \quad \begin{aligned}\lim_{n \rightarrow \infty} (\sup\{x_j : j \in \mathbb{N}, j \geq n\}) &= \lim_{n \rightarrow \infty} (\sup(T_n)) = \inf(\{\sup(T_n) : n \in \mathbb{N}\}), \\ \lim_{n \rightarrow \infty} (\inf\{x_j : j \in \mathbb{N}, j \geq n\}) &= \lim_{n \rightarrow \infty} (\inf(T_n)) = \sup(\{\inf(T_n) : n \in \mathbb{N}\}).\end{aligned}$$

An expression like  $\sup\{x_j : j \in \mathbb{N}, j \geq n\}$  can be written more compactly as  $\sup_{j \in \mathbb{N}, j \geq n} \{x_j\}$ . 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  $\sup_{j \geq n} \{x_j\}$ . This can also be written as  $\sup_{j \geq n} (x_j)$  or  $\sup_{j \geq n} x_j$ .

In other words, (9.55) becomes

$$(9.56) \quad \begin{aligned}\inf_{n \in \mathbb{N}} (\sup_{j \geq n} x_j) &= \inf(\{\sup(T_n) : n \in \mathbb{N}\}) = \lim_{n \rightarrow \infty} (\sup(T_n)) = \lim_{n \rightarrow \infty} (\sup_{j \geq n} x_j), \\ \sup_{n \in \mathbb{N}} (\inf_{j \geq n} x_j) &= \sup(\{\inf(T_n) : n \in \mathbb{N}\}) = \lim_{n \rightarrow \infty} (\inf(T_n)) = \lim_{n \rightarrow \infty} (\inf_{j \geq n} x_j). \quad \square\end{aligned}$$

The above leads us to the following definition:

**Definition 9.15.** Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathbb{R}$  and let  $T_n = \{x_j : j \in \mathbb{N}, j \geq n\}$  be the tail set for  $x_n$ . Assume that  $T_n$  is bounded above for some  $n_0 \in \mathbb{N}$  (and hence for all  $n \geq n_0$ ). We call

$$\limsup_{n \rightarrow \infty} x_j := \lim_{n \rightarrow \infty} (\sup_{j \geq n} x_j) = \inf_{n \in \mathbb{N}} (\sup_{j \geq n} x_j) = \inf_{n \in \mathbb{N}} (\sup(T_n))$$

the **lim sup** or **limit superior** of the sequence  $(x_n)$ .

If, for each  $n$ ,  $T_n$  is not bounded above then we say  $\limsup_{n \rightarrow \infty} x_j = \infty$ .

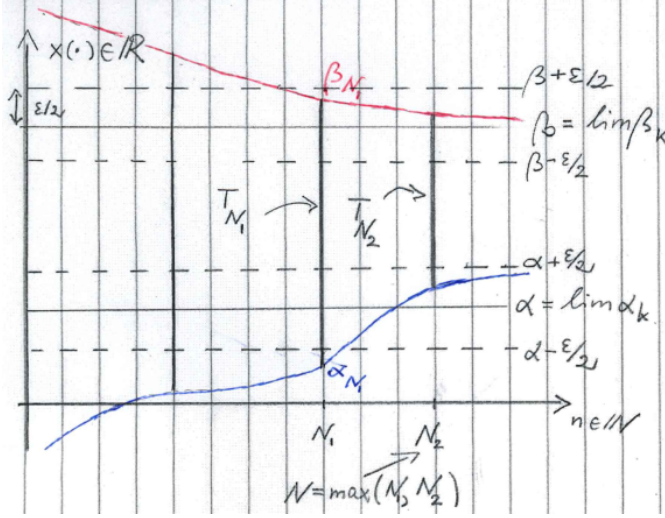
Assume that  $T_n$  is bounded below for some  $n_0$  (and hence for all  $n \geq n_0$ ). We call

$$\liminf_{n \rightarrow \infty} x_j := \lim_{n \rightarrow \infty} (\inf_{j \geq n} x_j) = \sup_{n \in \mathbb{N}} (\inf_{j \geq n} x_j) = \sup_{n \in \mathbb{N}} (\inf(T_n))$$

the **lim inf** or **limit inferior** of the sequence  $(x_n)$ .

If, for each  $n$ ,  $T_n$  is not bounded below then we say  $\liminf_{n \rightarrow \infty} x_j = -\infty$ .  $\square$





Given  $\varepsilon > 0$

There is  $N_1 = N_1(\varepsilon)$ :

$$k \geq N_1 \Rightarrow |x_k - \beta| \leq \frac{\varepsilon}{2};$$

There is  $N_2 = N_2(\varepsilon)$ :

$$k \geq N_2 \Rightarrow |x_k - \alpha| \leq \frac{\varepsilon}{2};$$

$$k \geq N = \max(N_1, N_2) \Rightarrow \text{both } |x_k - \beta| \leq \frac{\varepsilon}{2}, |x_k - \alpha| \leq \frac{\varepsilon}{2};$$

Hence only finitely many  $x_k$  above  $\beta + \varepsilon$  or below  $\alpha - \varepsilon$ ;

Hence we have subsequences  $x_{m_i}, x_{n_j}$ :

$$\lim_{i \rightarrow \infty} x_{m_i} = \beta, \lim_{j \rightarrow \infty} x_{n_j} = \alpha$$

**Proposition 9.16.** Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathbb{R}$  which is bounded above with tail sets  $T_n$ .

A. Let

$$(9.57) \quad \begin{aligned} \mathcal{U} &:= \{y \in \mathbb{R} : T_n \cap [y, \infty[ \neq \emptyset \text{ for all } n \in \mathbb{N}\}, \\ \mathcal{U}_1 &:= \{y \in \mathbb{R} : \text{for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{Z}_{\geq 0} \text{ such that } x_{n+k} \geq y\}, \\ \mathcal{U}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \text{ such that } x_{n_j} \geq y \text{ for all } j \in \mathbb{N}\}, \\ \mathcal{U}_3 &:= \{y \in \mathbb{R} : x_n \geq y \text{ for infinitely many } n \in \mathbb{N}\}. \end{aligned}$$

Then  $\mathcal{U} = \mathcal{U}_1 = \mathcal{U}_2 = \mathcal{U}_3$ .

B. There exists  $z = z(\mathcal{U}) \in \mathbb{R}$  such that  $\mathcal{U}$  is either an interval  $] -\infty, z]$  or an interval  $] -\infty, z[$ .

C. Let  $u := \sup(\mathcal{U})$ . Then  $u = z = z(\mathcal{U})$  as defined in part B. Further,  $u$  is the only real number such that

$$C1. \quad (9.58) \quad u - \varepsilon \in \mathcal{U} \quad \text{and} \quad u + \varepsilon \notin \mathcal{U} \quad \text{for all } \varepsilon > 0.$$

C2. There exists a subsequence  $(n_j)_{j \in \mathbb{N}}$  of integers such that  $u = \lim_{j \rightarrow \infty} x_{n_j}$  and  $u$  is the largest real number for which such a subsequence exists.

PROOF of A:

A.1 -  $\mathcal{U} = \mathcal{U}_1$ : 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 -  $\mathcal{U}_1 \subseteq \mathcal{U}_2$ :

Let  $y \in \mathcal{U}_1$  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  $y \in \mathcal{U}_1$  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 -  $\mathcal{U}_2 \subseteq \mathcal{U}_3$ : This is trivial: Let  $y \in \mathcal{U}_2$ . 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  $\mathcal{U}_3$ .

A.4 -  $\mathcal{U}_3 \subseteq \mathcal{U}$ : Let  $y \in \mathcal{U}_3$ . 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  $y \in \mathcal{U}$

We have shown the following sequence of inclusions:

$$\mathcal{U} = \mathcal{U}_1 \subseteq \mathcal{U}_2 \subseteq \mathcal{U}_3 \subseteq \mathcal{U}$$

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  $y_2 \in \mathcal{U}$ .

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.,  $y_1 \in \mathcal{U}$ .

We conclude that  $\mathcal{U}$  is an interval of the form  $] - \infty, z]$  or  $] - \infty, z[$  for some  $z \in \mathbb{R}$ .

PROOF of C: Let  $z = z(\mathcal{U})$  as defined in part B and  $u := \sup(\mathcal{U})$ .

PROOF of C.1 - (9.58) part 1,  $u - \varepsilon \in \mathcal{U}$ :

As  $u - \varepsilon$  is smaller than the least upper bound  $u$  of  $\mathcal{U}$ ,  $u - \varepsilon$  is not an upper bound of  $\mathcal{U}$ . Hence there is  $y > u - \varepsilon$  such that  $y \in \mathcal{U}$ . It follows from part B that  $u - \varepsilon \in \mathcal{U}$ . ✓

PROOF of C.1 - (9.58) part 2,  $u + \varepsilon \notin \mathcal{U}$ :

This is trivial as  $u + \varepsilon > u = \sup(\mathcal{U})$  implies that  $y \leq u < u + \varepsilon$  for all  $y \in \mathcal{U}$ .

But then  $y \neq u$  for all  $y \in \mathcal{U}$ , i.e.,  $u \notin \mathcal{U}$ . This proves  $u + \varepsilon \notin \mathcal{U}$ .

PROOF of C.2: We construct by induction a sequence  $n_1 < n_2 < \dots$  of natural numbers such that

$$(9.59) \quad u - 1/j \leq x_{n_j} \leq u + 1/j.$$

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  $u - 1 \in \mathcal{U}_3$ , 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 (9.59) and this proves the base case.

Induction step: Let us now assume that  $n_1 < n_2 < \dots < n_k$  satisfying (9.59) 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  $u - 1/(k + 1) \in \mathcal{U}_3$ , 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 (9.59) and this finishes the proof by induction.

We now show that  $\lim_{j \rightarrow \infty} x_{n_j} = u$ . Given  $\varepsilon > 0$  there is  $N = N(\varepsilon)$  such that  $1/N < \varepsilon$ . It follows from (9.59) that  $|x_{n_j} - u| \leq 1/j < 1/N < \varepsilon$  for all  $j \geq N$  and this proves that  $x_{n_j} \rightarrow u$  as  $j \rightarrow \infty$ .

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  $x_{n_j} \rightarrow w$  as  $j \rightarrow \infty$ .

Let  $\varepsilon := (w - u)/2$ . According to (9.58),  $u + \varepsilon \notin \mathcal{U}$ . But then, by definition of  $\mathcal{U}$ , 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  $\lim_{j \rightarrow \infty} x_{n_j} = w$ . ■

**Corollary 9.2.** As in prop.9.16, let  $u := \sup(\mathcal{U})$ . Then  $\mathcal{U} = ] - \infty, u]$  or  $\mathcal{U} = ] - \infty, u[$ .

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  $\mathcal{U} = \mathcal{U}_3$  and parts B and C of prop.9.16. ■

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. 9.16 is a direct consequence of the duality of upper/lower bounds, min/max, inf/sup proposition prop.9.2, p.146.

**Proposition 9.17.** Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathbb{R}$  which is bounded below with tail sets  $T_n$ .

A. Let

$$(9.60) \quad \begin{aligned} \mathcal{L} &:= \{y \in \mathbb{R} : T_n \cap ] - \infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\}, \\ \mathcal{L}_1 &:= \{y \in \mathbb{R} : \text{for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{Z}_{\geq 0} \text{ such that } x_{n+k} \leq y\}, \\ \mathcal{L}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \text{ such that } x_{n_j} \leq y \text{ for all } j \in \mathbb{N}\}, \\ \mathcal{L}_3 &:= \{y \in \mathbb{R} : x_n \leq y \text{ for infinitely many } n \in \mathbb{N}\}. \end{aligned}$$

Then  $\mathcal{L} = \mathcal{L}_1 = \mathcal{L}_2 = \mathcal{L}_3$ .

B. There exists  $z = z(\mathcal{L}) \in \mathbb{R}$  such that  $\mathcal{L}$  is either an interval  $[z, \infty[$  or an interval  $]z, \infty[$ .

C. Let  $l := \inf(\mathcal{L})$ . Then  $l = z = z(\mathcal{L})$  as defined in part B. Further,  $l$  is the only real number such that

$$C1. \quad (9.61) \quad l + \varepsilon \in \mathcal{L} \quad \text{and} \quad l - \varepsilon \notin \mathcal{L}$$

C2. There exists a subsequence  $(n_j)_{j \in \mathbb{N}}$  of integers such that  $l = \lim_{j \rightarrow \infty} x_{n_j}$  and  $l$  is the smallest real number for which such a subsequence exists.

PROOF: Let  $y_n = -x_n$  and apply prop.9.16. ■

**Proposition 9.18.** Let  $(x_n)$  be a bounded sequence of real numbers. As in prop. 9.16 and prop 9.17, let

$$(9.62) \quad \begin{aligned} u &= \sup(\mathcal{U}) = \sup\{y \in \mathbb{R} : T_n \cap [y, \infty[ \neq \emptyset \text{ for all } n \in \mathbb{N}\}, \\ l &= \inf(\mathcal{L}) = \inf\{y \in \mathbb{R} : T_n \cap ] - \infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\}, \end{aligned}$$

Then  $u = \limsup_{n \rightarrow \infty} x_j$  and  $l = \liminf_{n \rightarrow \infty} x_j$ .

Proof that  $u = \limsup_{n \rightarrow \infty} x_j$ : Let

$$(9.63) \quad \beta_n := \sup_{j \geq n} x_j, \quad \beta := \inf_n \beta_n = \limsup_{n \rightarrow \infty} x_n.$$

We will prove that  $\beta$  has the properties listed in prop.9.16.C that uniquely characterize  $u$ : For any  $\varepsilon > 0$ , we have

$$\beta - \varepsilon \in \mathcal{U} \quad \text{and} \quad \beta + \varepsilon \notin \mathcal{U}$$

Another way of saying this is that

$$(9.64) \quad b \in \mathcal{U} \text{ for } b < \beta \quad \text{and} \quad a \notin \mathcal{U} \text{ for } a > \beta.$$

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  $\beta_n$ :  $\beta_{n_0} < a$  for some  $n_0 \in \mathbb{N}$ .

As the  $\beta_n$  are not increasing in  $n$ , this implies strict inequality  $\beta_j < a$  for all  $j \geq n_0$ . By definition,  $\beta_j$  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  $a \notin \mathcal{U}$ .

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  $b \in \mathcal{U}$ .

We put everything together and see that  $\beta$  has the properties listed in (9.64). This finishes the proof that  $u = \limsup_{n \rightarrow \infty} x_j$ . The proof that  $l = \liminf_{n \rightarrow \infty} x_j$  follows again by applying what has already been proved to the sequence  $(-x_n)$ . ■

We have collected everything to prove

**Theorem 9.2** (Characterization of limsup and liminf). Let  $(x_n)_{n \in \mathbb{N}}$  be a bounded sequence in  $\mathbb{R}$ . Then

- a1.**  $\limsup_{n \rightarrow \infty} x_n$  is the largest of all real numbers  $x$  for which  $n_1 < n_2 < \dots \in \mathbb{N}$  can be found such that  $x = \lim_{j \rightarrow \infty} x_{n_j}$ .
- a2.**  $\limsup_{n \rightarrow \infty} x_n$  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.**  $\liminf_{n \rightarrow \infty} x_n$  is the smallest of all real numbers  $x$  for which  $n_1 < n_2 < \dots \in \mathbb{N}$  can be found such that  $x = \lim_{j \rightarrow \infty} x_{n_j}$ .
- b2.**  $\liminf_{n \rightarrow \infty} x_n$  is the only real number  $l$  such that, for all  $\varepsilon > 0$ , the following is true:  
 $x_n < l - \varepsilon$  for at most finitely many  $n$  and  $x_n < l + \varepsilon$  for infinitely many  $n$ .

**PROOF:** We know from prop.9.18 on p.163 that  $\limsup_{n \rightarrow \infty} x_n$  is the unique number  $u$  described in part C of prop.9.16, p.161:

$$u - \varepsilon \in \mathcal{U} \quad \text{and} \quad u + \varepsilon \notin \mathcal{U} \quad \text{for all } \varepsilon > 0$$

and  $u$  is the largest real number for which there exists a subsequence  $(n_j)_{j \in \mathbb{N}}$  of integers such that  $u = \lim_{j \rightarrow \infty} x_{n_j}$ .

$u - \varepsilon \in \mathcal{U} = \mathcal{U}_3$  (see part A of prop.9.18) means that there are infinitely many  $n$  such that  $x_n \geq u - \varepsilon$  and  $u + \varepsilon \notin \mathcal{U} = \mathcal{U}_3$  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.9.18 that  $\liminf_{n \rightarrow \infty} x_n$  is the unique number  $l$  described in part C of prop.9.17, p.163:  $l + \varepsilon \in \mathcal{L}$  and  $l - \varepsilon \notin \mathcal{L}$  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  $u = \lim_{j \rightarrow \infty} x_{n_j}$ .

$l + \varepsilon \in \mathcal{L} = \mathcal{L}_3$  (see part A of prop.9.18) means that there are infinitely many  $n$  such that  $x_n \leq l + \varepsilon$  and  $l - \varepsilon \notin \mathcal{L} = \mathcal{L}_3$  means that there are at most finitely many  $n$  such that  $x_n \leq l - \varepsilon$ . This proves **b1** and **b2**. ■

PROOF of thm.9.2 without the use of prop.9.18, prop.9.16 and the dual propositions for the liminf.

Step 1: Let  $\varepsilon > 0$ . It follows from  $\beta_n = \sup(T_n) = \sup\{x_j : j \geq n\}$  and  $\beta_n \searrow \beta = \limsup_n x_n$  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  $\beta_N$  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  $\beta$  satisfies the first half of **a1** of thm.9.2.

Step 2: We create subsequences  $(x_{n_j})_j$  and  $(\beta_{n_j})_j$  such that

$$(9.65) \quad \beta_{n_j} \geq x_{n_j} > \beta_{n_j} - 1/j$$

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  $n_1 < n_2 < \dots < n_k$  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

$$(9.66) \quad x_{n_k+i} > \beta_{n_{k+1}} - \frac{1}{k+1}.$$

Let  $n_{k+1} := n_k + i$ . The sequence  $\beta_n$  non-increasing (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 (9.66) 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 (9.66) we have

$$(9.67) \quad \beta_{n_{k+1}} \geq x_{n_{k+1}} > \beta_{n_{k+1}} - \frac{1}{k+1}.$$

It follows that  $x_{n_{k+1}}$  satisfies (9.65) 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  $\lim_k \beta_{n_k} = \beta$ ,  $\lim_k \beta_{n_k} - 1/k = \lim_k \beta_{n_k} - \lim_k 1/k = \beta - 0 = \beta$  and  $x_{n_j}$  is "sandwiched" between two sequences which both converge to the same limit  $\beta$ .

No subsequence of  $(x_n)$  can converge to a number  $u$  bigger than  $\beta$ : Let  $\varepsilon := 1/2(u - \beta)$ . 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.9.2.

Step 4. We still must prove the missing half of thm.9.2.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 (9.65) and  $\beta_{n_j} \geq \beta$  and  $1/j < \varepsilon$  that  $x_{n_j} > \beta - \varepsilon$ . This proves the missing half of thm.9.2.a2.

Uniqueness of  $\beta$ : 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.9.2.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.9.2.a2. We have proved that  $\limsup_n x_n$  is uniquely determined by the inequalities of thm.9.2.a2 and we have shown both **a1** and **a2** of thm.9.2.

Parts **b1** and **b2** of thm.9.2 follow now easily from applying parts **a1** and **a2** to the sequence  $y_n := -x_n$ .

■

**Theorem 9.3** (Characterization of limits via limsup and liminf). Let  $(x_n)_{n \in \mathbb{N}}$  be a bounded sequence in  $\mathbb{R}$ . Then  $(x_n)$  converges to a real number if and only if liminf and limsup for that sequence coincide and we have

$$(9.68) \quad \lim_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n.$$

PROOF of “ $\Rightarrow$ ”: Let  $L := \lim_{n \rightarrow \infty} x_n$ . It follows from prop.9.2 (Characterization of limsup and liminf) parts a1 and b1 on p.164 that subsequences can be found which converge to  $\liminf_{n \rightarrow \infty} x_n$  and to  $\limsup_{n \rightarrow \infty} x_n$ . We also know from prop.9.6 (Subsequences of convergent sequences) on p.151 that any convergent subsequence has limit  $L$ . This proves  $\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n = L$ .

PROOF of “ $\Leftarrow$ ”: <sup>87</sup> Let  $L := \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n$ . 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.9.2 (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  $L = \lim_{n \rightarrow \infty} x_n$ . ■

**Proposition 9.19.** 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

$$\liminf_{n \rightarrow \infty} x_n \leq \liminf_{n \rightarrow \infty} x'_n \quad \text{and} \quad \limsup_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} x'_n.$$

PROOF:

We only 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

$$-\liminf_{n \rightarrow \infty} x'_n = \limsup_{n \rightarrow \infty} (-x'_n) \leq \limsup_{n \rightarrow \infty} (-x_n) = -\liminf_{n \rightarrow \infty} x_n,$$

hence  $\liminf_{n \rightarrow \infty} x_n \leq \liminf_{n \rightarrow \infty} x'_n$ . 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.9.2 on p.163 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

$$x'_{n_j} \geq x_{n_j} > u - \varepsilon = u' + \varepsilon.$$

<sup>87</sup> Here is a shorter proof of “ $\Leftarrow$ ”. The drawback: It makes use of prop.9.16 and prop.9.17. Let  $L := \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n$ . Let  $\varepsilon > 0$ . We know from (9.58), p.161 and (9.61), p.163 that  $L + \varepsilon/2 \notin \mathcal{U}$  and  $L - \varepsilon/2 \notin \mathcal{L}$ . 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  $L = \lim_{n \rightarrow \infty} x_n$ . ■



We employ cor.9.2 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  $\sup = \infty$ , 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  $\sup = \infty$ , 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.9.19.

**Corollary 9.3.** 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

$$\limsup_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} y_n \quad \text{and} \quad \liminf_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} y_n.$$

PROOF: Immediate from prop.9.19. ■

Here is the second corollary to prop.9.19.

**Corollary 9.4.** 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

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

PROOF: Immediate from prop.9.19. ■

## 9.5 Sequences of Sets and Indicator functions and their liminf and limsup

Let  $\Omega$  be a nonempty set and let  $f_n : \Omega \rightarrow \mathbb{R}$  be a sequence of real-valued functions. Let  $\omega \in \Omega$ . Then  $(f_n(\omega))_{n \in \mathbb{N}}$  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  $\omega$ .

**Example 9.6.** The following are examples of sequences of real-valued functions.

- a.  $f_n : [0, 1] \rightarrow \mathbb{R}; x \mapsto x^n$  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  $\varphi_n : \Omega \rightarrow \mathbb{R}; \varphi_n(x, y) := \sqrt{x^2 + y^2}$  is a sequence of real-valued functions.
- c. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  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  $h_n : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto D_n f(x)$  is a sequence of real-valued functions.  $\square$

**Definition 9.16** (limsup and liminf of a sequence of real valued functions). Let  $\Omega$  be a nonempty set and let  $f_n : \Omega \rightarrow \mathbb{R}$  be a sequence of real-valued functions such that  $f_n(\omega)$  is bounded for all  $\omega \in \Omega$ .  
88

We define

$$(9.70) \quad \liminf_{n \rightarrow \infty} f_n : \Omega \rightarrow \mathbb{R} \quad \text{as follows: } \omega \mapsto \liminf_{n \rightarrow \infty} f_n(\omega),$$

$$(9.71) \quad \limsup_{n \rightarrow \infty} f_n : \Omega \rightarrow \mathbb{R} \quad \text{as follows: } \omega \mapsto \limsup_{n \rightarrow \infty} f_n(\omega). \quad \square$$

**Remark 9.6.** We recall from thm.9.2 (Characterization of limsup and liminf) on p.164 that

$$(9.72) \quad \liminf_{n \rightarrow \infty} f_n(\omega) = \inf\{\alpha \in \mathbb{R} : \lim_{j \rightarrow \infty} f_{n_j}(\omega) = \alpha \text{ for some subsequence } n_1 < n_2 < \dots\},$$

$$(9.73) \quad \limsup_{n \rightarrow \infty} f_n(\omega) = \sup\{\beta \in \mathbb{R} : \lim_{j \rightarrow \infty} f_{n_j}(\omega) = \beta \text{ for some subsequence } n_1 < n_2 < \dots\}. \quad \square$$

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.7.13 on p.126 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 9.20** (liminf and limsup of  $\{0, 1\}$ -functions). Let  $\Omega \neq \emptyset$  and  $f_n : \Omega \rightarrow \{0, 1\}$ . Let  $\omega \in \Omega$ . Then either  $\liminf_n f_n(\omega) = 1$  or  $\liminf_n f_n(\omega) = 0$  and either  $\limsup_n f_n(\omega) = 1$  or  $\limsup_n f_n(\omega) = 0$ . Further

$$(9.74) \quad \liminf_{n \rightarrow \infty} f_n(\omega) = 1 \Leftrightarrow f_n(\omega) = 1 \text{ except for at most finitely many } n \in \mathbb{N}$$

$$(9.75) \quad \limsup_{n \rightarrow \infty} f_n(\omega) = 1 \Leftrightarrow f_n(\omega) = 1 \text{ for infinitely many } n \in \mathbb{N}$$

*PROOF:* It follows from (9.72), (9.73) and  $0 \leq f_n(\omega) \leq 1$  that  $0 \leq \liminf_n f_n(\omega) \leq \limsup_n f_n(\omega) \leq 1$ .

<sup>88</sup> In more advanced texts you will find the following

**Definition 9.17** (Extended real valued functions). The set  $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$  is called the **extended real numbers line**. A mapping  $F$  whose codomain is a subset of  $\bar{\mathbb{R}}$  is called an **extended real valued function**.  $\square$

The above allows to define the functions  $\liminf_n f_n$  and  $\limsup_n f_n$  even if there are arguments  $\omega$  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:

$$(9.69) \quad 0 \cdot \pm\infty = \pm\infty \cdot 0 = 0$$



We conclude from (9.72) 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) (9.74) but also that either  $\liminf_n f_n(\omega) = 1$  or  $\liminf_n f_n(\omega) = 0$

We conclude from (9.73) 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) (9.75) but also that either  $\limsup_n f_n(\omega) = 1$  or  $\limsup_n f_n(\omega) = 0$ . ■

We now look at indicator functions  $1_{A_n}$  of a sequence of sets  $A_n \subseteq \Omega$ . For such a sequence we define

$$(9.76) \quad A_\star := \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j, \quad A^\star := \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j.$$

**Proposition 9.21.** *Let  $\omega \in \Omega$ . Then*

$$(9.77) \quad \omega \in A_\star \Leftrightarrow \omega \in A_n \text{ for all except at most finitely many } n \in \mathbb{N}.$$

$$(9.78) \quad \omega \in A^\star \Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N},$$

*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  $A_{n_k} \supseteq \bigcap_{j \geq n_k} A_j \supseteq \bigcap_{j \geq k} A_j$  that there

is no  $k \in \mathbb{N}$  such that  $\omega \in \bigcap_{j \geq k} A_j$ .

But then  $\omega \notin \bigcup_k \bigcap_{j \geq k} A_j = A_\star$  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  $\omega \in \bigcap_{n \geq N} A_n \subseteq \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} A_n = A_\star$  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  $\omega \in \bigcup_{j \geq n} A_j$  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  $1 \leq n_1 < n_2 < \dots$  such that  $\omega \in A_{n_j}$  for all  $j \in \mathbb{N}$ . Let  $k \in \mathbb{N}$ .

Then  $n_k \geq k$ , 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 9.22** (liminf and limsup of indicator functions).

$$(9.79) \quad 1_{A_\star} = \liminf_{n \rightarrow \infty} 1_{A_n} \quad \text{and} \quad 1_{A^\star} = \limsup_{n \rightarrow \infty} 1_{A_n}$$

*PROOF:* Let  $\omega \in \Omega$ . Then

$$(9.80) \quad \begin{aligned} 1_{A_\star}(\omega) = 1 &\Leftrightarrow \omega \in A_\star \Leftrightarrow \omega \in A_n \text{ for all except at most finitely many } n \in \mathbb{N} \\ &\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for all except at most finitely many } n \in \mathbb{N} \\ &\Leftrightarrow \liminf_n 1_{A_n}(\omega) = 1 \end{aligned}$$

The second equivalence follows from prop.9.21 and the last equivalence follows from prop.9.20 and this proves the first equation. Similarly we have

$$(9.81) \quad \begin{aligned} 1_{A^\star}(\omega) = 1 &\Leftrightarrow \omega \in A^\star \Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N} \\ &\Leftrightarrow 1_{A_n}(\omega) = 1 \text{ for infinitely many } n \in \mathbb{N} \\ &\Leftrightarrow \limsup_n 1_{A_n}(\omega) = 1 \end{aligned}$$

Again the second equivalence follows from prop.9.21 and the last equivalence follows from prop.9.20. ■

This last proposition is the reason for the following definition.

**Definition 9.18** (limsup and liminf of a sequence of sets). Let  $\Omega$  be a nonempty set and let  $A_n \subseteq \Omega$  ( $n \in \mathbb{N}$ ). We define

$$(9.82) \quad \liminf_{n \rightarrow \infty} A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j,$$

$$(9.83) \quad \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j.$$

We call  $\liminf_{n \rightarrow \infty} A_n$  the **limit inferior** and  $\limsup_{n \rightarrow \infty} A_n$  the **limit superior** of the sequence  $A_n$ .

We note that  $\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$  if and only if the functions  $\liminf_{n \rightarrow \infty} 1_{A_n}$  and  $\limsup_{n \rightarrow \infty} 1_{A_n}$  coincide (prop. 9.22) which is true if and only if the sequence  $1_{A_n}(\omega)$  has a limit for all  $\omega \in \Omega$  (thm.9.3 on p.166). In this case we define

$$(9.84) \quad \lim_{n \rightarrow \infty} A_n := \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$$

and we call this set the **limit** of the sequence  $A_n$ . □

**Note 9.2** (Notation for limits of monotone sequences of sets). Let  $(A_n)$  be a non-decreasing sequence of sets, i.e.,  $A_1 \subseteq A_2 \subseteq \dots$  and let  $A := \bigcup_n A_n$ . Further let  $B_n$  be a non-increasing sequence of sets, i.e.,  $B_1 \supseteq B_2 \supseteq \dots$  and let  $B := \bigcap_n B_n$ . We write suggestively<sup>89</sup>

$$A_n \nearrow A \quad (n \rightarrow \infty), \quad B_n \searrow B \quad (n \rightarrow \infty). \quad \square$$

---

<sup>89</sup> See note 9.1 on p.152.

**Example 9.7.** Let  $A_n \subseteq \Omega$ .

$$(9.85) \quad \text{a. If } A_n \nearrow \text{ then } \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} A_n.$$

$$(9.86) \quad \text{b. If } A_n \searrow \text{ then } \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} A_n. \quad \square$$

**Note 9.3** (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  $\liminf_{n \rightarrow \infty} x_n = \sup_n (\inf(T_n))$  with  $\liminf_{n \rightarrow \infty} T_n = \bigcup_n (\bigcap_{k \geq n} T_k)$ .
- b. Do not confuse  $\limsup_{n \rightarrow \infty} x_n = \inf_n (\sup(T_n))$  with  $\limsup_{n \rightarrow \infty} T_n = \bigcap_n (\bigcup_{k \geq n} T_k)$ .

Those two concepts are very different:  $\liminf_n x_n$  ( $\limsup_n x_n$ ) 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 9.7 and the fact the the sequence of tailsets  $T_n$  is always non-increasing.  $\square$

## 9.6 Sequences that Enumerate Parts of $\mathbb{Q}$

We informally defined the real numbers in ch.2.2 (Numbers) on p.15 as the set of all decimals, i.e., all numbers  $x$  which can be written as

$$(9.87) \quad x = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad \text{where } d_j \text{ is a digit, i.e., } d_j = 0, 1, 2, \dots, 9,$$

$$(9.88) \quad \text{i.e., } x = \lim_{k \rightarrow \infty} x_k \quad \text{where } x_k = m + \sum_{j=1}^k d_j 10^{-j}.$$

Each  $x_k$  is a (finite) sum of fractions, hence  $x_k \in \mathbb{Q}$ .

We proved in cor.8.6 on p.139 that  $\mathbb{Q}$  and hence all of its subsets can be sequenced: If  $A \subseteq \mathbb{Q}$  there is a sequence  $(q_n)_n$  of fractions such that  $A = \{q_n : n \subseteq \mathbb{N}\}$ . We apply this to  $A := \mathbb{Q}$  as follows.

Let  $x \in \mathbb{R}$  have the representation (9.88). 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 (9.88) that  $q_{n_k} \rightarrow x$  as  $k \rightarrow \infty$ . In other words, we have proved the following

**Theorem 9.4** (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  $x = \lim_{k \rightarrow \infty} q_{n_k}$  ■.

**Remark 9.7.**

- 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.9.4: 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  $x_k = m + \sum_{j=1}^k d_j 10^{-j}$  which converge to  $x$  belong to  $A_{\mathbb{Q}}$ . (Why? Especially, why also for  $x = \pm\sqrt{2}$ ?)  $\square$

## 9.7 Exercises for Ch.9

**Exercise 9.1.** Prove example 9.4 part c: Let  $z_n := x_0$  for some  $x_0 \in \mathbb{R}$  ( $n \in \mathbb{N}$ ). Then  $\lim_{n \rightarrow \infty} z_n = x_0$ .

If that is too abstract, try to prove the special case **b** first.  $\square$

**Exercise 9.2.** Prove the assertions of example 9.7 on p.171.  $\square$

**Exercise 9.3.** Prove prop.9.8 on p.153 for the case that  $\lim_{n \rightarrow \infty} x_n$  exists in  $\mathbb{R}$ :

Let  $(x_n)_n$  be a sequence of real numbers such that  $\lim_{n \rightarrow \infty} x_n$  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.  $\square$

**Exercise 9.4.** 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  $f_1 : A_1 \rightarrow \mathbb{R}$  and  $f_2 : A_2 \rightarrow \mathbb{R}$  be two continuous, real-valued functions such that  $f_1(c) = f_2(c)$ . Prove that the “spliced” function

$$(9.89) \quad f(x) := \begin{cases} f_1(x) & \text{for } x \in A_1, \\ f_2(x) & \text{for } x \in A_2 \end{cases}$$

is continuous on  $A$ .  $\square$

**Exercise 9.5.** Let  $a, b, c, d \in \mathbb{R}$  such that  $a < b$  and  $c < d$ . Let  $f : ]a, b[ \rightarrow ]c, d[$  be bijective and strictly monotone, i.e., strictly increasing or decreasing. Prove that both  $f$  and  $f^{-1}$  are continuous.

Hint: Use thm.9.1 on p.157.

**Exercise 9.6.** Prove cor.9.3 on p.167:

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

$$\limsup_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} y_n \quad \text{and} \quad \liminf_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} y_n. \quad \square$$

**Exercise 9.7.** Prove cor.9.4 on p.167:

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

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

**Exercise 9.8.** 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 def.9.14 (Tail sets of a sequence) on p.159! **Hint:** What is  $\alpha_n$  and  $\beta_n$ ?

*Very similar to the previous exercise, but slightly more difficult:*

**Exercise 9.9.** Let  $x_n := (-1)^{\lfloor n + \frac{1}{n} \rfloor}$  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 def.9.14 (Tail sets of a sequence) on p.159! **Hint:** Compute  $\alpha_n$  and  $\beta_n$  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  $\alpha_n$  and  $\beta_n$ ?

## 10 Vectors and Vector spaces (Understand this!)

### 10.1 $\mathbb{R}^n$ : Euclidean Space

Most if not all of the material of this chapter with the exception of ch.10.2.2 (normed Vector Spaces) on p.188 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.

#### 10.1.1 $n$ -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” .<sup>90</sup>

**Definition 10.1** ( $n$ -dimensional vectors). Let  $n \in \mathbb{N}$ . An  $n$ -dimensional vector is a finite, ordered collection  $\vec{v} = (x_1, x_2, x_3, \dots, x_n)$  of real numbers  $x_1, x_2, x_3, \dots, x_n$ , i.e., an element of  $\mathbb{R}^n$ .<sup>91</sup>  $n$  is called the **dimension** of the vector  $\vec{v}$ .  $\square$

Here are some examples of vectors:

**Example 10.1** (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  $(\sqrt{2}, -1.5) \in \mathbb{R}^2$ .  $\square$

**Example 10.2** (Three-dimensional vectors).  $\vec{v}_t = (3 - t, 15, \sqrt{5t^2 + \frac{22}{7}}) \in \mathbb{R}^3$  with coordinates  $x = 3 - t$ ,  $y = 15$  and  $z = \sqrt{5t^2 + \frac{22}{7}}$  is an example of a parametrized vector (parametrized by  $t$ ). Each specific value of  $t$  defines an element of  $\mathbb{R}^3$ , e.g.,  $\vec{v}_{-2} = (5, 15, \sqrt{20 + \frac{22}{7}})$ . note that

$$F : \mathbb{R} \rightarrow \mathbb{R}^3 \quad t \mapsto F(t) = \vec{v}_t$$

defines a function from  $\mathbb{R}$  into  $\mathbb{R}^3$  in the sense of definition ( 5.6 ) on p.88. Each argument  $s$  has assigned to it one and only one argument  $\vec{v}_s = (3 - s, 15, \sqrt{5s^2 + \frac{22}{7}}) \in \mathbb{R}^3$ .

Or, is it rather that we have three functions

$$\begin{aligned} x(\cdot) : \mathbb{R} &\rightarrow \mathbb{R} & t &\rightarrow x(t) = 3 - t, \\ y(\cdot) : \mathbb{R} &\rightarrow \mathbb{R} & t &\rightarrow y(t) = 15, \\ z(\cdot) : \mathbb{R} &\rightarrow \mathbb{R} & t &\rightarrow z(t) = \sqrt{5t^2 + \frac{22}{7}}, \end{aligned}$$

and  $t \rightarrow \vec{v}_t = (x(t), y(t), z(t))$  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  $\vec{v}_t$ .  $\square$

**Example 10.3** (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  $\vec{w}_2 = 5.7a$ .  $\square$

<sup>90</sup> The definition of an abstract vector space will be given in def.10.4 on p.179.

<sup>91</sup> See def.6.5 (Cartesian Product of three or more sets) on p.114) concerning  $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ times}} .$

**Example 10.4** (Vectors as functions). An  $n$ -dimensional vector  $\vec{x} = (x_1, x_2, x_3, \dots, x_n)$  can be interpreted as a real valued function

$$(10.1) \quad \begin{aligned} f_{\vec{x}}(\cdot) : \{1, 2, 3, \dots, n\} &\longrightarrow \mathbb{R} & m &\mapsto x_m, \text{ i.e.,} \\ f_{\vec{x}}(1) = x_1, f_{\vec{x}}(2) = x_2, \dots, f_{\vec{x}}(n) = x_n, \end{aligned}$$

This can be done in reverse. Any real valued function  $f(\cdot) : \{1, 2, 3, \dots, n\} \longrightarrow \mathbb{R}$  can be associated with the vector  $\vec{v}_{f(\cdot)}$  that lists the function values  $f(j)$ :

$$(10.2) \quad \vec{v}_{f(\cdot)} := (f(1), f(2), f(3), \dots, f(n)) \in \mathbb{R}^n. \quad \square$$

### 10.1.2 Addition and Scalar Multiplication for $n$ -Dimensional Vectors

**Definition 10.2** (Addition and scalar multiplication in  $\mathbb{R}^n$ ). 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  $\alpha$ .

We define the **sum**  $\vec{x} + \vec{y}$  of  $\vec{x}$  and  $\vec{y}$  as the vector  $\vec{z}$  with the components

$$(10.3) \quad z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_n = x_n + y_n;$$

We define the **scalar product**  $\alpha\vec{x}$  of  $\alpha$  and  $\vec{x}$  as the vector  $\vec{w}$  with the components

$$(10.4) \quad w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_n = \alpha x_n. \quad \square$$

Figure 10.1 below describes vector addition.

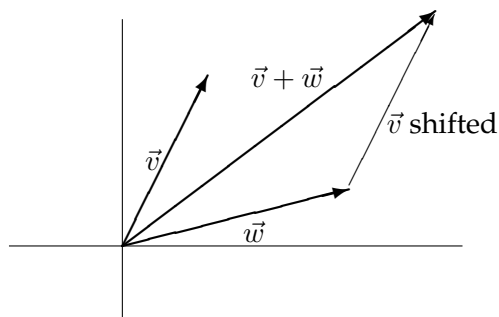


Figure 10.1: Adding two vectors.

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.

### 10.1.3 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 10.5** (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 10.6** (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 10.2). How long is that arrow?

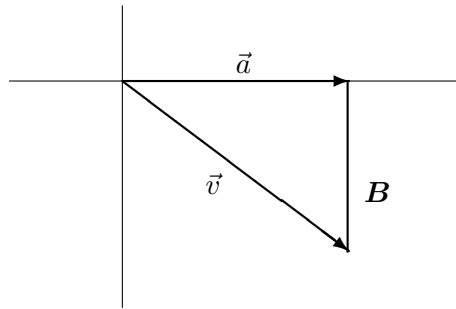


Figure 10.2: Length of a 2-dimensional vectors.

Think of it as the hypotenuse 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  $\mathbf{B}$  between  $(4, 0)$  and  $(4, -3)$ . Note that  $\mathbf{B}$  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  $\text{length-of}(\mathbf{B}) = 3$ . Pythagoras tells us that

$$\|\vec{v}\|_2^2 = \|\vec{a}\|_2^2 + (\text{length-of-}\mathbf{B})^2$$

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,

$$(10.5) \quad \|(x, y)\|_2 = \sqrt{x^2 + y^2} \quad \square$$

**Example 10.7** (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  $B$  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  $B$  is not a vector because it does not start at the origin! It should be clear that  $\text{length-of}(B) = |z| = 12$ .

We connect the origin  $(0, 0, 0)$  with the point  $(4, -3, 0)$  in the  $xy$ -plane (the endpoint of  $B$ ).

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  $\vec{v}$ .<sup>92</sup> 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**) hypotenuse  $\vec{v} = (x, y, z)$  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 hypotenuse 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  $\|(x, y, z)\|_2^2 = x^2 + y^2 + z^2$ , hence

$$(10.6) \quad \|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2} \quad \square$$

*The previous examples show how to extend the concept of “length” to vector spaces of any finite dimension:*

**Definition 10.3** (Euclidean norm). Let  $n \in \mathbb{N}$  and  $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  be an  $n$ -dimension vector. The **Euclidean norm**  $\|\vec{v}\|_2$  of  $\vec{v}$  is defined as follows:

$$(10.7) \quad \|\vec{v}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\sum_{j=1}^n x_j^2}. \quad \square$$

*This definition is important enough to write the special cases for  $n = 1, 2, 3$  where  $\|\vec{v}\|_2$  coincides with the length of  $\vec{v}$ :*

$$(10.8) \quad \begin{aligned} 1\text{-dimensional} : & \quad \|(x)\|_2 = \sqrt{x^2} = |x| \\ 2\text{-dimensional} : & \quad \|(x, y)\|_2 = \sqrt{x^2 + y^2} \\ 3\text{-dimensional} : & \quad \|(x, y, z)\|_2 = \sqrt{x^2 + y^2 + z^2} \end{aligned}$$

---

<sup>92</sup> You will learn in the chapter on vector spaces that the vector  $\vec{a} = (4, -3)$  is the projection on the  $xy$ -coordinates  $\pi_{1,2}(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^2; (x, y, z) \mapsto (x, y)$  of the vector  $\vec{v} = (4, -3, 12)$ . See Example 10.19) on p.185.

**Proposition 10.1** (Properties of the Euclidean norm). Let  $n \in \mathbb{N}$ . Then the Euclidean norm, viewed as a function

$$\|\cdot\|_2 : \mathbb{R}^n \rightarrow \mathbb{R}; \quad \vec{v} = (x_1, x_2, \dots, x_n) \mapsto \|\vec{v}\|_2 = \sqrt{\sum_{j=1}^n x_j^2}$$

has the following three properties:

$$\begin{array}{ll} (10.9a) & \|\vec{v}\|_2 \geq 0 \quad \forall \vec{v} \in \mathbb{R}^n \quad \text{and} \quad \|\vec{v}\|_2 = 0 \Leftrightarrow \vec{v} = \mathbf{0} & \text{positive definiteness} \\ (10.9b) & \|\alpha\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R} & \text{absolute homogeneity} \\ (10.9c) & \|\vec{v} + \vec{w}\|_2 \leq \|\vec{v}\|_2 + \|\vec{w}\|_2 \quad \forall \vec{v}, \vec{w} \in \mathbb{R}^n & \text{triangle inequality} \end{array}$$

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, non-negative. If  $\mathbf{0}$  is the zero vector with coordinates  $x_1 = x_2 = \dots = x_n = 0$  then obviously  $\|\mathbf{0}\|_2 = \sqrt{0 + \dots + 0} = 0$ . Conversely, let  $\vec{v} = (x_1, x_2, \dots, x_n)$  be a vector in  $\mathbb{R}^n$  such that  $\|\vec{v}\|_2 = 0$ . This means that  $\sqrt{\sum_{j=1}^n x_j^2} = 0$  which is only possible if everyone of the non-negative  $x_j$  is zero. In other words,  $\vec{v}$  must be the zero vector  $\mathbf{0}$ .

**b.** Let  $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$ . Then

$$\begin{aligned} \|\alpha\vec{v}\|_2 &= \sqrt{\sum_{j=1}^n (\alpha x_j)^2} = \sqrt{\sum_{j=1}^n \alpha^2 x_j^2} = \sqrt{\alpha^2 \sum_{j=1}^n x_j^2} = \sqrt{\alpha^2} \sqrt{\sum_{j=1}^n x_j^2} \\ &= \sqrt{\alpha^2} \|\vec{v}\|_2 = |\alpha| \cdot \|\vec{v}\|_2 \end{aligned}$$

because it is true that  $\sqrt{\alpha^2} = |\alpha|$  for any real number  $\alpha$  (see assumption 2.1 on p.18).

**c.** The proof will only be given for  $n = 1, 2, 3$ .

**$n = 1$**  : (10.9.c) simply is the triangle inequality for real numbers (see (17.15) on 316) and we are done.

**$n = 2, 3$**  : Look back at the picture about addition of vectors in the plane or in space (see p.175). 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 (10.9.c). An exact proof will be given for arbitrary  $n \in \mathbb{N}$  as a consequence of the so-called Cauchy-Schwartz inequality (cor.10.1). The inequality itself is stated and proved in prop.10.12 on p.189 in the section which discusses inner products (dot products) on vector spaces.

■

## 10.2 General Vector Spaces

### 10.2.1 Vector spaces: Definition and Examples

Part of this follows [3] Brin, Matthew and Marchesi, Gerald: Linear Algebra, a text for Math 304, Spring 2016.

Mathematicians are very fond of looking at 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 10.3 on p.174). We also saw that  $n$ -dimensional vectors can be thought of as real valued functions with domain  $X = \{1, 2, 3, \dots, n\}$ . (see 10.4 on p.175). 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 (10.1.2) on p.175). Given are two  $n$ -dimensional vectors

$$\vec{x} = (x_1, x_2, \dots, x_n) \text{ and } \vec{y} = (y_1, y_2, \dots, y_n) \text{ and a real number } \alpha.$$

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  $\alpha$  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 10.8** (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  $\lambda$ :

$$100a = 100(-3, 1/5) = (-300, 20). \quad \square$$

In the last example I have avoided using the notation " $\vec{x}$ " with the cute little arrows on top for vectors. I did that on purpose, because this notation is not all that popular in Math, even for  $n$ -dimensional vectors, and definitely not for abstract vectors as elements of a vector space. Here now is the definition of a vector space, taken almost word for word from the book "Introductory Real Analysis" (Kolmogorov/Fomin [8]). This definition is quite lengthy because a set needs to satisfy many rules to be a vector space.

**Definition 10.4** (Vector spaces (linear spaces)). A nonempty set  $V$  of elements  $x, y, z, \dots$  is called a **vector space** or **linear space** if it satisfies the following:

**A.** Any two elements  $x, y \in V$  uniquely determine a third element  $x + y \in V$ , called the **sum** of  $x$  and  $y$  with the following properties:

1.  $x + y = y + x$  (**commutativity**);

2.  $(x + y) + z = x + (y + z)$  (**associativity**);
3. There exists an element  $0 \in V$ , called the **zero element**, or **zero vector**, or **null vector**, with the property that  $x + 0 = x$  for each  $x \in V$ ;
4. For every  $x \in V$ , there exists an element  $-x \in V$ , called the **negative** of  $x$ , with the property that  $x + (-x) = 0$  for each  $x \in V$ . When adding negatives, then there is a convenient short form. We write  $x - y$  as an abbreviation for  $x + (-y)$ ;

**B.** Any real number  $\alpha$  and element  $x \in V$  together uniquely determine an element  $\alpha x \in V$  (sometimes also written  $\alpha \cdot x$ ), called the **scalar product** of  $\alpha$  and  $x$ . It has the following properties:

1.  $\alpha(\beta x) = (\alpha\beta)x$ ;
2.  $1x = x$ ;

**C.** The operations of addition and scalar multiplication obey the two **distributive laws**

1.  $(\alpha + \beta)x = \alpha x + \beta x$ ;
2.  $\alpha(x + y) = \alpha x + \alpha y$ ;

The elements of a vector space are called **vectors**.  $\square$

**Definition 10.5** (Subspaces of vector spaces). Let  $V$  be a vector space and let  $A \subseteq V$  be a nonempty subset of  $V$  with the following property: For any  $x, y \in A$  and  $\alpha \in \mathbb{R}$  the sum  $x + y$  and the scalar product  $\alpha x$  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**.  $\square$

**Remark 10.1** (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.12). 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.  $\square$

**Proposition 10.2** (Subspaces are vector spaces). *A subspace of a vector space is a vector space, i.e., it satisfies all requirements of definition (10.4).*

*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  $\lambda x$  and  $\lambda y$  belong to  $A$  because each is the scalar product of  $\lambda$  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  $V$ .  $\blacksquare$

**Remark 10.2** (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.  $\square$

*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 10.9** (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.  $\square$

**Example 10.10** (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 (10.2) on p.175.  $\square$

*The following example should be thought of as the **definition** of the very important function spaces  $\mathcal{F}(X, \mathbb{R})$ ,  $\mathcal{B}(X, \mathbb{R})$ ,  $\mathcal{C}(X, \mathbb{R})$ .*

**Example 10.11** (Vector spaces of real-valued functions). Let  $X$  be an arbitrary, nonempty set. Then

$$(10.10) \quad \mathcal{F}(X, \mathbb{R}) := \{f(\cdot) : f(\cdot) \text{ is a real valued function on } X\}$$

denotes the set of all real valued functions with domain  $X$ <sup>93</sup> and

$$\mathcal{B}(X, \mathbb{R}) := \{g(\cdot) : g(\cdot) \text{ is a bounded real valued function on } X\}$$

denotes the subset of all bounded real valued functions with domain  $X$ .

Let  $A \subseteq \mathbb{R}$ . Then

$$\mathcal{C}(A, \mathbb{R}) := \{\psi(\cdot) : \psi(\cdot) \text{ is a continuous real valued function on } A\}$$

denotes the set of all real-valued continuous functions with domain  $A$ .<sup>94</sup>

We list separately the case  $X = [a, b]$  where  $a, b \in \mathbb{R}$  such that  $a < b$ . Then

$$\mathcal{C}([a, b], \mathbb{R}) := \{h(\cdot) : h(\cdot) \text{ is a continuous real valued function for } a \leq x \leq b\}$$

denotes the set of all continuous 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 (5.14) on p.100, 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.

<sup>93</sup> Note that  $\mathcal{F}(X, \mathbb{R}) = \mathbb{R}^X$  (see remark 6.4, p.116 which follows def.6.6 of the Cartesian Product of a family of sets.)

<sup>94</sup> Continuity for such functions was discussed in ch.9.2 on p.150.

**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  $f, g \in \mathcal{B}(X, \mathbb{R})$  a third item  $f + g \in \mathcal{B}(X, \mathbb{R})$  and “ $\cdot$ ” associates with any  $f \in \mathcal{B}(X, \mathbb{R})$  and  $\alpha \in \mathbb{R}$  a third item  $\alpha \cdot f \in \mathcal{B}(X, \mathbb{R})$ .

**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  $f, g \in \mathcal{C}([a, b], \mathbb{R})$  a third item  $f + g \in \mathcal{C}([a, b], \mathbb{R})$  and “ $\cdot$ ” associates with any  $f \in \mathcal{C}([a, b], \mathbb{R})$  and  $\alpha \in \mathbb{R}$  an item  $\alpha \cdot f \in \mathcal{C}([a, b], \mathbb{R})$ .

It follows from the above that all three function sets are vector spaces and also that **1)**  $\mathcal{B}(X, \mathbb{R})$  is a subspace of  $\mathcal{F}(X, \mathbb{R})$ , **2)**  $\mathcal{C}(X, \mathbb{R})$  is a subspace of  $\mathcal{F}(X, \mathbb{R})$ .

We will see in ch.12 (Compactness) on p.263 that continuous functions defined on a closed interval are bounded. It follows that

$$\mathcal{C}([a, b], \mathbb{R}) \subseteq \mathcal{B}([a, b], \mathbb{R}) \subseteq \mathcal{F}([a, b], \mathbb{R}).$$

We deduce from this that **3)**  $\mathcal{C}([a, b], \mathbb{R})$  also is a subspace of  $\mathcal{B}([a, b], \mathbb{R})$ .

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  $b = 1$ .  $\square$

*Here are some more examples.*

**Example 10.12** (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  $\alpha$  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 (10.5) on p.180).  $\square$

*A proof for the following is omitted even though it is not difficult:*

**Example 10.13** (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:  $V_\alpha$  is the straight line through the origin in the  $xy$ -plane with slope  $\alpha$ .  $\square$

**Example 10.14** (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 10.15** (All finite-dimensional vectors). Let

$$\mathcal{G} := \bigcup_{n \in \mathbb{N}} \mathbb{R}^n = \mathbb{R}^1 \cup \mathbb{R}^2 \cup \dots \cap \mathbb{R}^n \cup \dots$$

be the set of all vectors of finite (but unspecified) dimension.

We can define addition for any two elements  $\vec{x}, \vec{y} \in \mathfrak{S}$  as follows: If  $\vec{x}$  and  $\vec{y}$  both happen to have the same dimension  $n$  then we add them as usual: the sum will be  $x_1 + y_1, x_2 + y_2, \dots, x_n + y_n$ . If not, then one of them, say  $\vec{x}$  will have dimension  $m$  smaller than the dimension  $n$  of  $\vec{y}$ . We 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 10.16** (All sequences of real numbers). Let  $\mathbb{R}^{\mathbb{N}} = \prod_{j \in \mathbb{N}} \mathbb{R}$  (see (6.6) on p.115). 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  $\alpha$ .

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  $j > n$ .  $\square$

**Definition 10.6** (linear combinations). Let  $V$  be a vector space and let  $x_1, x_2, x_3, \dots, x_n \in V$  be a finite number of vectors in  $V$ . Let  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$ . We call the finite sum

$$(10.11) \quad \sum_{j=0}^n \alpha_j x_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_n x_n$$

a **linear combination** of the vectors  $x_j$ . The multipliers  $\alpha_1, \alpha_2, \dots$  are called **scalars**.  $\square$

*In other words, linear combinations are sums of scalar multiples of vectors. The expression in (10.11) always is an element of  $V$ , no matter how big  $n \in \mathbb{N}$  was chosen:*

**Proposition 10.3** (Vector spaces are closed w.r.t. linear combinations). *Let  $V$  be a vector space and let  $x_1, x_2, x_3, \dots, x_n \in V$  be a finite number of vectors in  $V$ . Let  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$ . Then the linear combination  $\sum_{j=0}^n \alpha_j x_j$  also belongs to  $V$ . Note that this is also true for subspaces because those are vector spaces, too.*

*PROOF: Trivial.*  $\blacksquare$

**Proposition 10.4.** *Let  $V$  be a vector space and let  $(W_i)_{i \in I}$  be a family of subspaces of  $V$ . Let  $W := \bigcap [W_i : i \in I]$ . 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  $x \in W$ .*  $\blacksquare$

**Definition 10.7** (Linear span). Let  $V$  be a vector space and  $A \subseteq V$ . Then the set

$$(10.12) \quad \text{span}(A) = \left\{ \sum_{j=1}^k \alpha_j x_j : k \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in A (1 \leq j \leq k) \right\}.$$

of all linear combinations of vectors in  $A$  is called the **span** or **linear span** of  $A$ .  $\square$



**Proposition 10.5.** Let  $V$  be a vector space and  $A \subseteq V$ . Then  $\text{span}(A)$  is a subspace of  $V$ .

*PROOF:* Let  $y_j \in \text{span}(A)$  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 10.1.** Let  $V$  be a vector space and  $A \subseteq V$ .

Let  $\mathfrak{W} := \{W \subseteq V : W \supseteq A \text{ and } W \text{ is a subspace of } V\}$ . Then  $\text{span}(A) = \bigcap [W : W \in \mathfrak{W}]$ .

*PROOF:* Clearly,  $\text{span}(A) \supseteq A$ . It follows from prop.10.5 that  $\text{span}(A) \in \mathfrak{W}$ , hence  $\text{span}(A) \supseteq \bigcap [W : W \in \mathfrak{W}]$ .

On the other hand, Any subspace  $W$  of  $V$  that contains  $A$  also contains all its linear combinations, hence  $\text{span}(A) \subseteq W$  for all  $W \in \mathfrak{W}$ . But then  $\text{span}(A) \subseteq \bigcap [W : W \in \mathfrak{W}]$ .  $\blacksquare$

**Remark 10.3** (Linear  $\text{span}(A) =$  subspace generated by  $A$ ). Let  $V$  be a vector space and  $A \subseteq V$ . Theorem 10.1 justifies to call  $\text{span}(A) :=$  the **subspace generated by  $A$** .  $\square$

**Definition 10.8** (linear mappings). Let  $V_1, V_2$  be two vector spaces.

Let  $f(\cdot) : V_1 \rightarrow V_2$  be a function with the following properties:

$$(10.13a) \quad f(x + y) = f(x) + f(y) \quad \forall x, y \in V_1 \quad \textbf{additivity}$$

$$(10.13b) \quad f(\alpha x) = \alpha f(x) \quad \forall x \in V_1, \forall \alpha \in \mathbb{R} \quad \textbf{homogeneity}$$

Then we call  $f(\cdot)$  a **linear function** or **linear mapping**.  $\square$

**Note 10.1** (Note on homogeneity). We encountered “absolute homogeneity” when examining the properties of the Euclidean norm ((10.9) on p.178). That is not the same concept as homogeneity for linear functions because you had to take the absolute value  $|\alpha|$  instead of  $\alpha$ .  $\square$

**Remark 10.4** (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.

Mathematicians express this by saying that linear mappings **preserve** or are **compatible with** linear combinations.  $\square$

The proof of item C in the previous remark is given in the next proposition.

**Proposition 10.6** (Linear mappings preserve linear combinations). Let  $V_1, V_2$  be two vector spaces.

Let  $f(\cdot) : V_1 \rightarrow V_2$  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.,

$$(10.14) \quad f\left(\sum_{j=0}^n \lambda_j x_j\right) = \sum_{j=0}^n \lambda_j f(x_j).$$



*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*

$$f\left(\sum_{j=0}^k \lambda_j x_j\right) = \sum_{j=0}^k \lambda_j f(x_j) \quad \text{for all } 1 \leq k < n.$$

We use it in the second equation ( $k = 2$ ) and the third equation ( $k = n - 1$ ) of the following:

$$f\left(\sum_{j=0}^n \lambda_j x_j\right) = f\left(\sum_{j=0}^{n-1} \lambda_j x_j + \lambda_n x_n\right) = f\left(\sum_{j=0}^{n-1} \lambda_j x_j\right) + f(\lambda_n x_n) = \sum_{j=0}^{n-1} \lambda_j f(x_j) + f(\lambda_n x_n) = \sum_{j=0}^n \lambda_j f(x_j)$$

■

Here are some examples of linear mappings.

**Example 10.17** (Projection on the first coordinate). Let  $n \in \mathbb{N}$ . The map

$$\pi_1(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R} \quad (x_1, x_2, \dots, x_n) \mapsto x_1$$

is called the **projection** on the first coordinate or the **first coordinate function**. □

**Example 10.18** (Projections on any coordinate). More generally, let  $n \in \mathbb{N}$  and  $1 \leq j \leq n$ .

$$\pi_j(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R} \quad (x_1, x_2, \dots, x_n) \mapsto x_j$$

is called the **projection** on the  $j$ th 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 10.19** (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  $1 \leq i_1 < i_2 < \dots < i_m \leq n$ . The map

$$(10.15) \quad \pi_{i_1, i_2, \dots, i_m}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (x_1, x_2, \dots, x_n) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_m})$$

is called the **projection** on the coordinates  $i_1, i_2, \dots, i_m$ .<sup>95</sup> □

**Example 10.20.** Let  $x_0 \in A$ . The mapping

$$(10.16) \quad \varepsilon_{x_0} : \mathcal{F}(A, \mathbb{R}) \rightarrow \mathbb{R}; \quad f(\cdot) \mapsto f(x_0)$$

which assigns to any real valued function on  $A$  its value at the specific point  $x_0$  is linear because if

$$h(\cdot) = \sum_{j=0}^n a_j f_j(\cdot) \quad \text{then}$$

$$\varepsilon_{x_0}\left(\sum_{j=0}^n a_j f_j\right) = \varepsilon_{x_0}(h) = h(x_0) = \sum_{j=0}^n a_j f_j(x_0) = \sum_{j=0}^n a_j \varepsilon_{x_0}(f_j).$$

$\varepsilon_{x_0}(\cdot)$  is called the **abstract integral** with respect to point mass at  $x_0$ . □

<sup>95</sup> You previously encountered an example where we made use of the projection

$$\pi_{1,2}(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \quad (x, y, z) \mapsto (x, y).$$

This was in the course of computing the length of a 3-dimensional vector (see (10.5) on p.176).

**Lemma 10.1** ( $F \circ \text{span} = \text{span} \circ F$ ). [3] Brin/Marchesi Linear Algebra, general lemma 4.1.7: Let  $V, W$  be two vector spaces and  $F : V \rightarrow W$  a linear mapping from  $V$  to  $W$ . Let  $A \subseteq V$ . Then

$$(10.17) \quad F(\text{span}(A)) = \text{span}(F(A)).$$

*Proof:* See Brin/Marchesi Linear Algebra, general lemma 4.1.7. ■

**Definition 10.9** (Linear dependence and independence). Let  $V$  be a vector space and  $A \subseteq V$

**a.**  $A$  is called **linearly dependent** if the following is true: There exist distinct vectors  $x_1, x_2, \dots, x_k \in A$  and scalars  $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$  ( $k \in \mathbb{N}$ ) such that not all scalars  $\alpha_j$  are

zero ( $1 \leq j \leq k$ ) and  $\sum_{j=1}^k \alpha_j x_j = 0$ .

**b.**  $A$  is called **linearly independent** if  $A$  is not linearly dependent, i.e., if the following is true: Let  $x_1, x_2, \dots, x_k \in A$  and  $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$  ( $k \in \mathbb{N}$ ).

If  $\sum_{j=1}^k \alpha_j x_j = 0$  then  $\alpha_j = 0$  for all  $1 \leq j \leq k$ . □

**Definition 10.10** (Basis of a vector space). Let  $V$  be a vector space and  $B \subseteq V$ .  $B$  is called a **basis** of  $V$  if **a.**  $B$  is linearly independent and **b.**  $\text{span}(B) = V$ . □

**Lemma 10.2.** Let  $V$  be a vector space and  $A \subseteq V$  linearly independent. If  $\text{span}(A) \subsetneq V$  and  $y \in \text{span}(A)^c$  then  $A' := A \cup \{y\}$  is linearly independent.

*Proof:* 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

$$(10.18) \quad \sum_{j=1}^k \alpha_j x_j = 0$$

We must show that each  $\alpha_j$  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  $\alpha_j$  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

$$(10.19) \quad x_{j_0} = \sum_{j \neq j_0} \frac{-\alpha_j}{\alpha_{j_0}} x_j$$

would be a linear combination of elements of  $A$ , contrary to the assumption that  $x_{j_0} = y \in \text{span}(A)^c$ .

We deduce from (10.18) and  $\alpha_{j_0} = 0$  that

$$(10.20) \quad \sum_{j \neq j_0} \alpha_j x_j = 0.$$

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 10.2.** 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., [3] Brin/Marchesi Linear Algebra. ■

This last theorem gives rise to the following definition.

**Definition 10.11** (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 10.7.** For  $a \in \mathbb{R}$  define  $f_a(\cdot) \in \mathcal{B}(\mathbb{R}, \mathbb{R})$  as follows.

$$f_a(x) := \begin{cases} 0 & \text{if } x \neq a, \\ 1 & \text{if } x = a. \end{cases}$$

Then  $\mathcal{A} := \{f_a : a \in \mathbb{R}\}$  is a linearly independent subset of  $\mathcal{B}(\mathbb{R}, \mathbb{R})$ .

PROOF: left as exercise 10.1 on p.202. ■

**Proposition 10.8.** 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  $\tilde{U} = \text{span}(U \cup \{x_0\})$ .

The proof is left as exercise 10.2.

**Proposition 10.9.** Let  $V$  and  $V'$  be two vector spaces and let  $U$  be a proper (linear) subspace of  $V$ , i.e.,  $U \subsetneq V$ . Let  $x_0 \in U^c$ ,  $y_0 \in V'$ . Let  $f := U \rightarrow V'$  be a linear function from  $U$  into  $V'$ . Let  $\alpha \in \mathbb{R}$ . Then

$$(10.21) \quad g : U \uplus \{x_0\} \rightarrow V'; \quad g(x) := \begin{cases} f(x) & \text{if } x \in U, \\ y_0 & \text{if } x = x_0, \end{cases}$$

uniquely extends to a linear function  $\tilde{f} : \text{span}(U \uplus \{x_0\}) \rightarrow V'$  as follows:

$$(10.22) \quad \tilde{f}(x) := \begin{cases} f(x) & \text{if } x \in U, \\ \alpha y_0 & \text{if } x = \alpha x_0, \end{cases}$$

PROOF:

Let  $\tilde{U} := \text{span}(U \uplus \{x_0\})$ . It follows from prop.10.8 on p.187 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 (10.22) is in fact defined on all of  $\tilde{U}$ . Clearly  $\tilde{f}$  coincides with  $g$  on  $U \uplus \{x_0\}$ , hence  $\tilde{f}$  extends  $g$  from  $U \uplus \{x_0\}$  to  $\tilde{U}$ . The proof of linearity is left as exercise 10.3.

It remains to show the uniqueness of  $\tilde{f}$ . So let  $h : \tilde{U} \rightarrow V'$  linear such that  $h(x) = \tilde{f}(x)$  for all  $x \in U \uplus \{x_0\}$ . 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  $h|_{U \uplus \{x_0\}} = \tilde{f}|_{U \uplus \{x_0\}}$ , the second and fourth equalities from the linearity of  $h$  and  $\tilde{f}$ . This proves uniqueness of  $\tilde{f}$ . ■

## 10.2.2 Normed Vector Spaces (Study this!)

Definition 10.3 on p.177 in ch.10.1.3 (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.10.1 on p. 177 “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 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 [13]).

**Definition 10.12** (Inner products). Let  $V$  be a vector space with a function

$$\bullet(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}; \quad (x, y) \mapsto x \bullet y := \bullet(x, y)$$

which satisfies the following properties:

- |   |                              |
|---|------------------------------|
| (10.23a) $x \bullet x \geq 0 \quad \forall x \in V$ and $x \bullet x = 0 \Leftrightarrow x = 0$                       | <b>positive definiteness</b> |
| (10.23b) $x \bullet y = y \bullet x \quad \forall x, y \in V$   | <b>symmetry</b>              |
| (10.23c) $(x + y) \bullet z = x \bullet z + y \bullet z \quad \forall x, y, z \in V$                                  | <b>additivity</b>            |
| (10.23d) $(\lambda x) \bullet y = \lambda(x \bullet y) \quad \forall x, y \in V \quad \forall \lambda \in \mathbb{R}$ | <b>homogeneity</b>           |

We call such a function an **inner product**.<sup>96</sup>  $\square$

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 bilinear in the following sense:

**Definition 10.13** (Bilinearity). Let  $V$  be a vector space with a function

$$F(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}; \quad (x, y) \mapsto F(x, y).$$

$F(\cdot, \cdot)$  is called **bilinear** if it is linear in each component, i.e., the mappings

$$\begin{aligned} F_1 : V &\rightarrow \mathbb{R}; & x &\mapsto F(x, y) \\ F_2 : V &\rightarrow \mathbb{R}; & y &\mapsto F(x, y) \end{aligned}$$

are both linear.  $\square$

<sup>96</sup> also called **dot product**, e.g., in [3] Brin/Marchesi Linear Algebra, ch.6, Orthogonality.

**Proposition 10.10** (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*

$$(10.24a) \quad (a + b) \bullet (x + y) = a \bullet x + b \bullet x + a \bullet y + b \bullet y$$

$$(10.24b) \quad (x + y) \bullet (x + y) = x \bullet x + 2(x \bullet y) + y \bullet y$$

$$(10.24c) \quad (x - y) \bullet (x - y) = x \bullet x - 2(x \bullet y) + y \bullet y$$

*PROOF of a:*

$$\begin{aligned} (a + b) \bullet (x + y) &= (a + b) \bullet x + (a + b) \bullet y \\ &= a \bullet x + b \bullet x + a \bullet y + b \bullet y. \end{aligned}$$

*We used linearity in the second argument for the first equality and linearity in the first argument for the second equality.*

*PROOF of b and c: Left as an exercise.*

*The following is the most important example of an inner product.*

**Proposition 10.11** (Inner product on  $\mathbb{R}^n$ ). *Let  $n \in \mathbb{N}$ . Then the real-valued function*

$$(10.25) \quad (\vec{x}, \vec{y}) \mapsto x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_{j=1}^n x_jy_j \quad (\vec{x} = (x_1, \dots, x_n), \vec{y} = (y_1, \dots, y_n))$$

*is an inner product on  $\mathbb{R}^n \times \mathbb{R}^n$ .*

*PROOF:*

a. For  $\vec{x} = \vec{y}$  we obtain  $\vec{x} \bullet \vec{x} = \sum_{j=1}^n x_j^2$  and positive definiteness of the inner product follows from

$$\sum_{j=1}^n x_j^2 = 0 \Leftrightarrow x_j^2 = 0 \forall j \Leftrightarrow x_j = 0 \forall j.$$

b. Symmetry is clear because  $x_jy_j = y_jx_j$ .

c. Let  $\vec{z} = (z_1, \dots, z_n)$ . 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 10.12** (Cauchy–Schwartz inequality for inner products). *Let  $V$  be a vector space with an inner product*

$$\bullet(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}; \quad (x, y) \mapsto x \bullet y := \bullet(x, y)$$

*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 (10.10) on p.189.*

*This means  $2(x \bullet y) \leq 2$ , i.e.,  $x \bullet y \leq 1 = (x \bullet x)(y \bullet y)$  where the last equality is true because we had assumed  $x \bullet x = y \bullet y = 1$ . The Cauchy–Schwartz inequality is thus true under that special assumption.*

**Step 2: General case:** We do not assume anymore that  $x \bullet x = y \bullet y = 1$ . If  $x$  or  $y$  is zero then the Cauchy–Schwartz inequality is trivially true because, say if  $x = 0$  then the left hand side becomes

$$(x \bullet y)^2 = (0x \bullet y)^2 = 0(x \bullet y)^2 = 0$$

whereas the right hand side is, as the product of two non–negative numbers  $x \bullet x$  and  $y \bullet y$ , non–negative.

So we can assume that  $x$  and  $y$  are not zero. On account of the positive definiteness we have  $x \bullet x > 0$  and  $y \bullet y > 0$ . This allows us to define  $u := x/\sqrt{x \bullet x}$  and  $v := y/\sqrt{y \bullet y}$ . But then

$$\begin{aligned} u \bullet u &= (x \bullet x)/\sqrt{x \bullet x}^2 = 1 \\ v \bullet v &= (y \bullet y)/\sqrt{y \bullet y}^2 = 1. \end{aligned}$$

We have already seen in step 1 that  $u \bullet v \leq 1$ . 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 proved. ■

**Definition 10.14** (sup–norm of bounded real valued functions). Let  $X$  be an arbitrary, nonempty set. Let  $f : X \rightarrow \mathbb{R}$  be a bounded real valued function on  $X$ , i.e., there exists a (possibly very large) number  $K \geq 0$  such that  $|f(x)| \leq K$  for all  $x \in X$ .<sup>97</sup> Let

$$(10.26) \quad \|f\|_\infty := \sup\{|f(x)| : x \in X\}$$

We call  $\|f\|_\infty$  the **supremum norm** or **sup–norm** of the function  $f$ . □

**Proposition 10.13** (Properties of the sup norm). Let  $X$  be an arbitrary, nonempty set. Let

$$\mathcal{B}(X, \mathbb{R}) := \{h(\cdot) : h(\cdot) \text{ is a bounded real valued function on } X\}$$

(see example 10.11 on p. 181). Then the function

$$\|\cdot\|_\infty : \mathcal{B}(X, \mathbb{R}) \rightarrow \mathbb{R}_+, \quad h \mapsto \|h\|_\infty = \sup\{|h(x)| : x \in X\}$$

which assigns to a bounded function on  $X$  its sup–norm satisfies the following:

---

<sup>97</sup> see def.9.6 (bounded functions) on p.148

$$(10.27a) \quad \|f\|_\infty \geq 0 \quad \forall f \in \mathcal{B}(X, \mathbb{R}) \quad \text{and} \quad \|f\|_\infty = 0 \Leftrightarrow f(\cdot) = 0$$

$$(10.27b) \quad \|\alpha f(\cdot)\|_\infty = |\alpha| \cdot \|f(\cdot)\|_\infty \quad \forall f \in \mathcal{B}(X, \mathbb{R}), \forall \alpha \in \mathbb{R}$$

$$(10.27c) \quad \|f(\cdot) + g(\cdot)\|_\infty \leq \|f(\cdot)\|_\infty + \|g(\cdot)\|_\infty \quad \forall f, g \in \mathcal{B}(X, \mathbb{R})$$

*positive definiteness*  
*absolute homogeneity*  
*triangle inequality*

The proof is left as exercise 10.4.

**Note 10.2.** We previously discussed the Euclidean norm

$$(10.28) \quad \|\vec{x}\|_2 = \sqrt{\sum_{j=1}^n x_j^2}$$

for  $n$ -dimensional vectors  $\vec{x} = (x_1, x_2, \dots, x_n)$ . You saw in (10.1) on p.177 that it satisfies positive definiteness, absolute homogeneity and the triangle inequality, just like the sup-norm.<sup>98</sup> 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.  $\square$

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.

**Definition 10.15** (Normed vector spaces). Let  $V$  be a vector space. A **norm** on  $V$  is a real valued function

$$\|\cdot\| : V \rightarrow \mathbb{R} \quad x \mapsto \|x\|$$

with the following three properties:

$$(10.29a) \quad \|x\| \geq 0 \quad \forall x \in V \quad \text{and} \quad \|x\| = 0 \Leftrightarrow x = 0$$

$$(10.29b) \quad \|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in V, \forall \alpha \in \mathbb{R}$$

$$(10.29c) \quad \|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$$

**positive definiteness**  
**absolute homogeneity**  
**triangle inequality**

We call  $V$  a **normed vector space** and we write  $(V, \|\cdot\|)$  instead of  $V$  when we wish to emphasize what norm on  $V$  we are dealing with.  $\square$

**Proposition 10.14.** Let  $(V, \|\cdot\|)$  be a normed vector space and let  $\gamma > 0$ . Let  $p : V \rightarrow \mathbb{R}$  be defined as  $p(x) := \gamma\|x\|$ . Then  $p$  also is a norm.

The proof is left as exercise 10.6.  $\blacksquare$

<sup>98</sup> 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.10.1 (Inner products define norms) on p.193.

**Definition 10.16** ( $p$ -norms for  $\mathbb{R}^n$ ). Let  $p \geq 1$ . It will be proved in prop.10.17 on p.196 that the function

$$(10.30) \quad \vec{x} \mapsto \|\vec{x}\|_p := \left( \sum_{j=1}^n |x_j|^p \right)^{1/p}$$

is a norm on  $\mathbb{R}^n$ ). This norm is called the  **$p$ -norm**.

The Euclidean norm is a  $p$ -norm; it is the 2-norm.  $\square$

**Remark 10.5.** We have seen that a vector space can be endowed with more than one norm.

- a. We have seen in prop.10.14 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  $\mathbb{R}^n$ .  $\square$

The following theorem shows that an inner product can be associated in a natural fashion with a norm.

**Theorem 10.3** (Inner products define norms). Let  $V$  be a vector space with an inner product

$$\bullet(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}; \quad (x, y) \mapsto x \bullet y$$

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

$$\begin{aligned} \|x + y\|_{\bullet}^2 &= (x + y) \bullet (x + y) \\ &= x \bullet x + 2(x \bullet y) + y \bullet y \\ &\leq x \bullet x + 2|x \bullet y| + y \bullet y \\ &\leq x \bullet x + 2\sqrt{x \bullet x} \sqrt{y \bullet y} + y \bullet y \\ &= \|x\|_{\bullet}^2 + 2\|x\|_{\bullet} \|y\|_{\bullet} + \|y\|_{\bullet}^2 \\ &= (\|x\|_{\bullet} + \|y\|_{\bullet})^2. \end{aligned}$$

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  $\alpha$ . 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  $\|x + y\|_{\bullet}^2 \leq (\|x\|_{\bullet} + \|y\|_{\bullet})^2$  and obtain  $\|x + y\|_{\bullet} \leq \|x\|_{\bullet} + \|y\|_{\bullet}$ , the triangle inequality we set out to prove.  $\blacksquare$



**Definition 10.17** (Norm for an inner product). Let  $V$  be a vector space with an inner product

$$\bullet(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}; \quad (x, y) \mapsto x \bullet y$$

Then

$$(10.31) \quad \|\cdot\|_{\bullet} : x \mapsto \|x\|_{\bullet} := \sqrt{(x \bullet x)}$$

is called the **norm associated with the inner product**  $\bullet(\cdot, \cdot)$ .  $\square$

It was stated in prop.10.1 on p. 177 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 10.1.** The Euclidean norm in  $\mathbb{R}^n$  defined as  $\|(x_1, x_2, \dots, x_n)\|_2 = \sqrt{\sum_{j=1}^n x_j^2}$  (see def. 10.3 on p.177) is a norm.

PROOF: This follows from the fact that

$$\vec{x} \bullet \vec{y} = \sum_{j=1}^n x_j y_j \quad \text{where } \vec{x} = (x_1, \dots, x_n) \text{ and } \vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$$

defines an inner product on  $\mathbb{R}^n \times \mathbb{R}^n$  (see prop.10.11) for which  $\|(x_1, x_2, \dots, x_n)\|_2$  is the associated norm.  $\blacksquare$

We now look at an inner product on the vector space  $\mathcal{C}([a, b], \mathbb{R})$  of all continuous real-valued functions on the interval  $[a, b]$  which was defined in example 10.11 (Vector spaces of real valued functions) on p.181. We use the terminology of [11] Stewart, J: Single Variable Calculus) for the following.

**Definition 10.18.** Let  $a, b \in \mathbb{R}$ ,  $a < b$  and assume that  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable functions. (See example 5.19 on p.93.)

- 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.  $\square$
- 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  $y = b$ .  $\square$

**Example 10.21.** Let  $f : [-1, 1]; x \mapsto 4x^3$ . The antiderivative (see example 5.19 on p.93) 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 =  $\int_{(-1)}^1 4|x^3| dx = \int_{(-1)}^0 (-4x^3) dx + \int_0^1 4x^3 dx$   
 $= -x^4 \Big|_{-1}^0 + -x^4 \Big|_0^1 = (0 - (-1)) + (1 - 0) = 2. \quad \square$

Let  $a, b \in \mathbb{R}$  such that  $a < b$ . We remember from example 5.19 on p.93) that continuous functions are integrable. This allows us to compare for  $f \in \mathcal{C}([a, b], \mathbb{R})$  the expressions

$$(10.32) \quad \|f\|_\infty = \sup\{|f(x)| : x \in X\}, \quad \int_a^b |f(x)| dx, \quad \text{and} \quad \int_a^b (f(x))^2 dx.$$

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  $\mathcal{C}([a, b], \mathbb{R})$  whose associated norm is  $f \mapsto \int_a^b (f(x))^2 dx$ . We will discuss that now. In preparation we prove the following proposition.

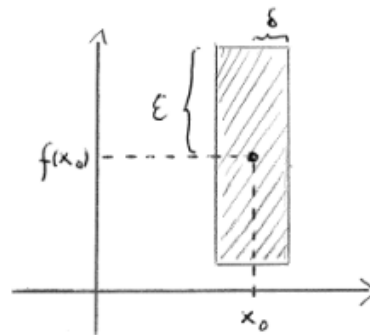
**Proposition 10.15.** Let  $a, b \in \mathbb{R}$  such that  $a < b$ . and let  $f : [a, b] \rightarrow [0, \infty[$  be continuous. Then  $\int_a^b f(x) dx = 0$  only if  $f(x) = 0$  for all  $x \in ]a, b[$ .  $\square$

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.9.1 on p.157 some  $\delta > 0$  such that

$$(10.33) \quad |f(x_0) - f(x)| < \varepsilon, \text{ hence } f(x) > f(x_0) - \varepsilon = \frac{f(x_0)}{2} = \varepsilon \text{ for all } x_0 - \delta < x < x_0 + \delta.$$

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  $g : [a, b] \rightarrow \mathbb{R}$  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 (10.33) that  $f \geq g$ , hence  $\int_a^b f(x) dx \geq \int_a^b g(x) dx = (2\delta)\varepsilon > 0$ .  $\blacksquare$

**Proposition 10.16.** Let  $a, b \in \mathbb{R}$  such that  $a < b$ . Then the mapping

$$(10.34) \quad (f, g) \mapsto f \bullet g := \int_a^b f(x)g(x)dx$$

defines an inner product on  $f \in \mathcal{C}([a, b], \mathbb{R})$ .  $\square$

*PROOF:* We must prove positive definiteness, symmetry, and linearity in the left argument. In the following let  $f, g, h \in \mathcal{C}([a, b], \mathbb{R})$  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.10.15.

**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:

(10.35)

$$(f + g) \bullet h = \int_a^b (f(x) + g(x))h(x)dx = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx = f \bullet h + g \bullet h,$$

$$(10.36) \quad (\lambda f) \bullet g = \int_a^b \lambda f(x)g(x)dx = \lambda \int_a^b f(x)g(x)dx = \lambda(f \bullet g). \quad \blacksquare$$

According to def.10.17 (norm for an inner product) and thm.10.3 (inner products define norms) we now define the norm associated with  $f \bullet g = \int_a^b f(x)g(x)dx$ .

**Definition 10.19** ( $L_2$ -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  $\mathcal{C}([a, b], \mathbb{R})$  of all continuous functions  $[a, b] \rightarrow \mathbb{R}$ :

$$(10.37) \quad f \bullet g := \int_a^b f(x)g(x)dx.$$

The associated norm

$$(10.38) \quad \|\cdot\|_{L^2} : f \mapsto \|f\|_{\bullet} = \sqrt{\int_a^b f^2(x)dx}$$

is called the  $L^2$ -norm. of  $f$ .  $\square$

We saw in def.10.16 that the Euclidean norm is the  $p$ -norm  $\|\vec{x}\|_p = \left(\sum_{j=1}^n x_j^p\right)^{1/p}$  for the special case  $p = 2$ . There is an analogue for the  $L^2$  norm.

**Definition 10.20** ( $L^p$ -norms for  $\mathbb{R}^n$ ). Let  $a, b \in \mathbb{R}$  such that  $a < b$  and  $p \geq 1$ . It will be shown in prop.10.18 (The  $L^p$ -norm is a norm) on p.196 that

$$(10.39) \quad f \mapsto \|f\|_{L^p} := \left( \int_a^b |f(x)|^p \right)^{1/p}$$

is a norm on  $\mathcal{C}([a, b], \mathbb{R})$ . This norm is called the  $L^p$ -norm of  $f$ . .  $\square$

### 10.2.3 The Inequalities of Young, Hoelder, and Minkowski (★)

*Note that this chapter is starred, hence optional.*

**Proposition 10.17** (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 non-negative, hence  $\|\vec{x}\|_p = \sqrt[p]{\sum_{j=1}^n |x_j|^p} \geq 0$ . 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  $(\|\vec{x}\|_p)^{1/p} \geq |x_{j_0}|^p > 0$ .

*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.10.7 below). That  $\|\cdot\|_2$  satisfies the triangle inequality (i.e.,  $p = 2$ ) also follows independently from cor.10.1 on p.193.  $\blacksquare$

**Proposition 10.18** (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  $\mathcal{C}([a, b], \mathbb{R})$ .

PROOF:

*a. Positive definiteness:*

Follows from prop.10.15 on p.194 and the fact that  $x \mapsto |f(x)|^p$  is a non-negative and continuous function.

*b. Absolute homogeneity:*

If  $\lambda \in \mathbb{R}$  then

$$\begin{aligned} \|(\lambda f)\|_{L^p} &= \left( \int_a^b (|\lambda| |f(x)|)^p dx \right)^{1/p} \\ &= \left( |\lambda|^p \int_a^b |f(x)|^p dx \right)^{1/p} = |\lambda| \left( \int_a^b |f(x)|^p dx \right)^{1/p} = |\lambda| \|f\|_{L^p}. \end{aligned}$$

c. Triangle inequality for  $p = 1$ :

It follows from  $|f(x) + g(x)| \leq |f(x)| + |g(x)|$  for all  $x$  that

$$\begin{aligned} \|f + g\|_{L^1} &= \int_a^b |f(x) + g(x)| dx \leq \int_a^b (|f(x)| + |g(x)|) dx \\ &= \int_a^b |f(x)| dx + \int_a^b |g(x)| dx = \|\vec{x}\|_1 + \|\vec{y}\|_1. \end{aligned}$$

d. Triangle inequality for  $p > 1$ :

This is Minkowski's inequality for  $L^p$ -norms (thm.10.5 below). That  $\|\cdot\|_{L^2}$  satisfies the triangle inequality (i.e.,  $p = 2$ ) also follows independently from cor.10.1 on p.193. ■

We were referring to Minkowski'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 allow us to prove those inequalities.

**Proposition 10.19** (Young's Inequality). Let  $a, b > 0$  and let  $p, q > 1$  be conjugate indices, i.e.,

$$(10.40) \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Then the following holds **Young's inequality**:

$$(10.41) \quad ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

PROOF:

**Step 1:** We show that  $q - 1 = \frac{1}{p-1}$ :

$$(10.42) \quad \begin{aligned} \frac{1}{p} + \frac{1}{q} = 1 &\Rightarrow q + p = pq \Rightarrow q(1 - p) = -p \\ &\Rightarrow q = \frac{p}{p-1} \Rightarrow q - 1 = \frac{p - (p-1)}{p-1} = \frac{1}{p-1}. \end{aligned}$$

**Step 2:** The functions

$$\varphi : [0, \infty[ \rightarrow [0, \infty[; \quad x \mapsto x^{p-1} \quad \text{and} \quad \psi : [0, \infty[ \rightarrow [0, \infty[; \quad y \mapsto y^{q-1}$$

are inverse to each other because we have

$$\psi(\varphi(x)) = \psi(x^{p-1}) = (x^{p-1})^{q-1} \stackrel{(*)}{=} (x^{p-1})^{1/(p-1)} = x$$

(\*) follows from step 1). We further have

$$\varphi(\psi(y)) = \varphi(y^{q-1}) = (y^{q-1})^{p-1} \stackrel{(**)}{=} (y^{q-1})^{1/(q-1)} = y$$

(\*\*) 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  $\varphi(0) = 0 = \psi(0)$ .

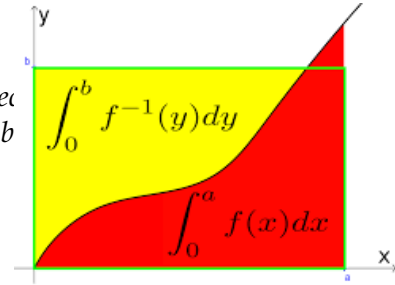
**Step 3:** Let  $f : [0, \infty[ \rightarrow [0, \infty[$  be a continuous and strictly increasing (hence invertible) function such that  $f(0) = 0$ . Then the following is true for any two real numbers  $a, b > 0$ :

$$(10.43) \quad ab \leq \int_0^a f(x) dx + \int_0^b f^{-1}(y) dy.$$

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  $b < f(a)$ : The rectangle  $ab$  is covered by the areas determined by the two integrals, but not all of the area of  $\int_0^a f(x)dx$  is covered by the 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 (10.42)). We integrate and obtain

$$\int_0^a f(x)dx = \int_0^a x^{p-1} = \frac{x^p}{p} \Big|_0^a = \frac{a^p}{p}, \quad \int_0^b f^{-1}(y)dy = \int_0^b y^{q-1} = \frac{y^q}{q} \Big|_0^b = \frac{b^q}{q}.$$

Young's inequality (10.41) now follows from (10.43). ■

**Theorem 10.4** (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.,

$$(10.44) \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Then **Hoelder's inequality** is true:

$$(10.45) \quad \|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}, \quad \text{i.e.,} \quad \int_a^b |f(x)g(x)|dx \leq \left( \int_a^b |f(x)|^p dx \right)^{1/p} \left( \int_a^b |g(x)|^q dx \right)^{1/q}.$$

**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.10.15 on p.194). Likewise,  $\|g\|_{L^q} = 0$  implies  $g(x) = 0$  for all  $x$ . In either case,  $\int_a^b f(x)g(x) = 0$  and (10.45) 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

$$A := \|f\|_{L^p}, \quad a_x := \frac{|f(x)|}{A}, \quad B := \|g\|_{L^q}, \quad b_x := \frac{|g(x)|}{B}.$$

It follows from Young's inequality (10.41) that  $a_x b_x \leq \frac{a_x^p}{p} + \frac{b_x^q}{q}$ . We integrate both sides of that inequality  $\int_a^b \dots dx$  and obtain from the monotonicity of the integral (see example 5.19 on p.93) that

$$(10.46) \quad \int_a^b a_x b_x dx \leq \int_a^b \left( \frac{a_x^p}{p} + \frac{b_x^q}{q} \right) dx.$$

i.e.,

$$(10.47) \quad \begin{aligned} \frac{1}{AB} \int_a^b |f(x)g(x)| dx &\leq \int_a^b \left( \frac{|f(x)|^p}{pA^p} + \frac{|g(x)|^q}{qB^q} \right) dx \\ &= \frac{1}{pA^p} \int_a^b |f(x)|^p dx + \frac{1}{qB^q} \int_a^b |g(x)|^q dx. \end{aligned}$$

We use

$$(10.48) \quad \int_a^b |f(x)|^p dx = (\|f\|_{L^p})^p, \quad \int_a^b |g(x)|^q dx = (\|g\|_{L^q})^q$$

in (10.47) 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 10.5** (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:*

$$(10.49) \quad \|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}, \quad \text{i.e.,}$$

$$(10.50) \quad \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/p} \leq \left( \int_a^b |f(x)|^p dx \right)^{1/p} + \left( \int_a^b |g(x)|^p dx \right)^{1/p}.$$

*PROOF:* This follows for  $p = 1$  from part c of the proof of prop.10.18. We may assume that  $p > 1$ . Let  $q$  be the conjugate index to  $p$ , i.e.,

$$(10.51) \quad \frac{1}{p} + \frac{1}{q} = 1, \quad \text{hence } (p-1)q = p$$

(see (10.42)). 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

$$(10.52) \quad \int_a^b |f(x) + g(x)|^p dx \leq \int_a^b |f(x)| |f(x) + g(x)|^{p-1} dx + \int_a^b |g(x)| |f(x) + g(x)|^{p-1} dx.$$

We apply Hoelder's inequality to the first of the two integrals on the right hand side of (10.52) and obtain

$$(10.53) \quad \begin{aligned} \int_a^b (|f(x)|) (|f(x) + g(x)|^{p-1}) dx &\leq \left( \int_a^b (|f(x)|)^p dx \right)^{1/p} \left( \int_a^b (|f(x) + g(x)|^{p-1})^q dx \right)^{1/q} \\ &= \left( \int_a^b |f(x)|^p dx \right)^{1/p} \left( \int_a^b |f(x) + g(x)|^{(p-1)q} dx \right)^{1/q} \\ &= \left( \int_a^b |f(x)|^p dx \right)^{1/p} \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/q}. \end{aligned}$$

The last equality results from  $(p-1)q = p$  (see (10.51)). Similarly, we obtain from the second integral on the right hand side of (10.51) the following:

$$(10.54) \quad \int_a^b (|g(x)|) (|f(x) + g(x)|^{p-1}) dx \leq \left( \int_a^b |g(x)|^p dx \right)^{1/p} \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/q}.$$

We apply (10.53) and (10.54) to (10.52) and obtain

$$(10.55) \quad \int_a^b |f(x) + g(x)|^p dx \leq \left( \int_a^b |f(x)|^p dx \right)^{1/p} \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/q} \\ + \left( \int_a^b |g(x)|^p dx \right)^{1/p} \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/q}.$$

Minkowski's inequality (10.50) 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 (10.55) by  $(\int_a^b |f(x) + g(x)|^p dx)^{1/q}$ . We obtain

$$(10.56) \quad \left( \int_a^b |f(x) + g(x)|^p dx \right)^{1-1/q} \leq \left( \int_a^b |f(x)|^p dx \right)^{1/p} + \left( \int_a^b |g(x)|^p dx \right)^{1/p}.$$

Note that  $1 - \frac{1}{q} = \frac{1}{p}$  because  $\frac{1}{q} + \frac{1}{p} = 1$ , and (10.56) reads

$$\left( \int_a^b |f(x) + g(x)|^p dx \right)^{1/p} \leq \left( \int_a^b |f(x)|^p dx \right)^{1/p} + \left( \int_a^b |g(x)|^p dx \right)^{1/p}. \blacksquare$$

**Theorem 10.6** (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.,

$$(10.57) \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Then **Hoelder's inequality** in  $\mathbb{R}^n$  is true:

$$(10.58) \quad \sum_{j=1}^n |x_j y_j| \leq \|\vec{x}\|_p \|\vec{y}\|_q, \quad \text{i.e.,} \quad \sum_{j=1}^n |x_j y_j| \leq \left( \sum_{j=1}^n |x_j|^p \right)^{1/p} \left( \sum_{j=1}^n |y_j|^q \right)^{1/q}.$$

*PROOF:* Let  $\vec{x}, \vec{y} \in \mathbb{R}^n$ . If  $\vec{x} = 0$  or  $\vec{y} = 0$  then  $\sum_{j=1}^n |x_j y_j| = 0$  and (10.58) 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.10.17 (positive definiteness of  $\|\cdot\|_p$  for all  $p$ ) on p.196 that  $\|\vec{x}\|_p > 0$  and  $\|\vec{y}\|_q > 0$ . For some fixed index  $1 \leq j \leq n$  let

$$A := \|\vec{x}\|_p, \quad a_j := \frac{|x_j|}{A}, \quad B := \|\vec{y}\|_q, \quad b_x := \frac{|y_j|}{B}.$$

It follows from Young's inequality (10.41) that

$$a_j b_j \leq \frac{a_j^p}{p} + \frac{b_j^q}{q}.$$



We take sums  $\sum_{j=1}^n \dots$  of both sides of that inequality and obtain from the monotonicity of summation

$$(10.59) \quad \sum_{j=1}^n a_j b_j \leq \sum_{j=1}^n \left( \frac{(a_j)^p}{p} + \frac{(b_j)^q}{q} \right),$$

i.e.,

$$(10.60) \quad \frac{1}{AB} \sum_{j=1}^n |x_j y_j| \leq \sum_{j=1}^n \left( \frac{|a_j|^p}{pA^p} + \frac{|b_j|^q}{qB^q} \right) = \frac{1}{pA^p} \sum_{j=1}^n |a_j|^p + \frac{1}{qB^q} \sum_{j=1}^n |b_j|^q.$$

But

$$(10.61) \quad \sum_{j=1}^n |x_j|^p = (\|f\|_p)^p, \quad \sum_{j=1}^n |y_j|^q = (\|g\|_q)^q.$$

It follows from (10.60) 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

$$\sum_{j=1}^n |x_j y_j| \leq AB = \|\vec{x}\|_p \|\vec{x}\|_q. \quad \blacksquare$$

**Theorem 10.7** (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:

$$(10.62) \quad \|\vec{x} + \vec{y}\|_p \leq \|\vec{x}\|_p + \|\vec{y}\|_p, \quad \text{i.e.,}$$

$$(10.63) \quad \left( \sum_j |x_j + y_j|^p \right)^{1/p} \leq \left( \sum_j |f(x)|^p \right)^{1/p} + \left( \sum_j |g(x)|^p \right)^{1/p}.$$

*PROOF:* This follows for  $p = 1$  from part **c** of the proof of prop.10.18. We hence may assume that  $p > 1$ . Let  $q$  be the conjugate index to  $p$ , i.e.,

$$(10.64) \quad \frac{1}{p} + \frac{1}{q} = 1, \quad \text{hence } (p-1)q = p$$

(see (10.42)). 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

$$(10.65) \quad \sum_j |x_j + y_j|^p \leq \sum_j |x_j| |x_j + y_j|^{p-1} + \sum_j |y_j| |x_j + y_j|^{p-1}.$$

Hoelder's inequality applied to the first of the two integrals on the right hand side of (10.65) yields

$$\begin{aligned}
 \sum_j (|x_j|) (|x_j + y_j|^{p-1}) &\leq \left( \sum_j (|x_j|^p) \right)^{1/p} \left( \sum_j (|x_j + y_j|^{p-1})^q \right)^{1/q} \\
 (10.66) \qquad \qquad \qquad &= \left( \sum_j |x_j|^p \right)^{1/p} \left( \sum_j |x_j + y_j|^{(p-1)q} \right)^{1/q} \\
 &= \left( \sum_j |x_j|^p \right)^{1/p} \left( \sum_j |x_j + y_j|^p \right)^{1/q}.
 \end{aligned}$$

The last equality results from  $(p-1)q = p$  (see (10.64)). Similarly, we obtain from the second integral on the right hand side of (10.65) the following:

$$(10.67) \qquad \sum_j (|y_j|) (|x_j + y_j|^{p-1}) \leq \left( \sum_j |y_j|^p \right)^{1/p} \left( \sum_j |x_j + y_j|^p \right)^{1/q}.$$

We apply (10.66) and (10.67) to (10.65) and obtain

$$\begin{aligned}
 \sum_j |x_j + y_j|^p &\leq \left( \sum_j |x_j|^p \right)^{1/p} \left( \sum_j |x_j + y_j|^p \right)^{1/q} \\
 (10.68) \qquad \qquad \qquad &+ \left( \sum_j |y_j|^p \right)^{1/p} \left( \sum_j |x_j + y_j|^p \right)^{1/q}.
 \end{aligned}$$

Minkowski's inequality (10.63) 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 (10.68) by  $(\sum_j |x_j + y_j|^p)^{1/q}$  and obtain

$$(10.69) \qquad \left( \sum_j |x_j + y_j|^p \right)^{1-1/q} \leq \left( \sum_j |x_j|^p \right)^{1/p} + \left( \sum_j |y_j|^p \right)^{1/p}.$$

Note that  $1 - \frac{1}{q} = \frac{1}{p}$  because  $\frac{1}{q} + \frac{1}{p} = 1$ , and (10.69) reads

$$\left( \sum_j |x_j + y_j|^p \right)^{1/p} \leq \left( \sum_j |x_j|^p \right)^{1/p} + \left( \sum_j |y_j|^p \right)^{1/p}. \quad \blacksquare$$

### 10.3 Exercises for Ch.10

**Exercise 10.1.** Prove prop.10.7 on p.187:

For  $a \in \mathbb{R}$  define  $f_a(\cdot) \in \mathcal{B}(\mathbb{R}, \mathbb{R})$  as follows.

$$f_a(x) := \begin{cases} 0 & \text{if } x \neq a, \\ 1 & \text{if } x = a. \end{cases}$$

Then  $\mathcal{A} := \{f_a : a \in \mathbb{R}\}$  is a linearly independent subset of  $\mathcal{B}(\mathbb{R}, \mathbb{R})$ .  $\square$

**Exercise 10.2.** Prove prop.10.8 on p.187:

Let  $V$  be a vector space and let  $U$  be a (linear) subspace of  $V$ . Let  $x_0 \in V$  and  $\tilde{U} := \text{span}(U \cup \{x_0\})$ . Then  $\tilde{U} = \{u + \lambda x_0 : u \in U \text{ and } \lambda \in \mathbb{R}\}$ .  $\square$

**Exercise 10.3.** Prove prop.6.1 on p.110:

Let  $V$  and  $V'$  be two vector spaces and let  $U$  be a proper (linear) subspace of  $V$ , i.e.,  $U \subsetneq V$ . Let  $x_0 \in U^c$ ,  $y_0 \in V'$ . Let  $f := U \rightarrow V'$  be a linear function from  $U$  into  $V'$  and let  $\alpha \in \mathbb{R}$ . We defined in prop.10.9 on p.187 the function

$$\tilde{f} : \text{span}(U \cup \{x_0\}) \rightarrow V'; \quad \tilde{f}(x) := \begin{cases} f(x) & \text{if } x \in U, \\ \alpha y_0 & \text{if } x = \alpha x_0. \end{cases}$$

It was stated that this extension of  $f$  is linear. Prove it.  $\square$

**Exercise 10.4.** Prove prop.10.13 on p.190: Let  $X$  be an arbitrary, nonempty set. Then the function

$$\|\cdot\|_\infty : \mathcal{B}(X, \mathbb{R}) \rightarrow \mathbb{R}_+, \quad h \rightarrow \|h\|_\infty = \sup\{|h(x)| : x \in X\} \text{ defines a norm. } \square$$

**Exercise 10.5.** Prove prop.10.10 (Algebraic properties of the inner product) on p.189:

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.  $(x - y) \bullet (x - y) = x \bullet x - 2(x \bullet y) + y \bullet y \quad \square$

**Exercise 10.6.** Prove prop.10.14 on p.191:

Let  $(V, \|\cdot\|)$  be a normed vector space and let  $\gamma > 0$ . Let  $p : V \rightarrow \mathbb{R}$  be defined as  $p(x) := \gamma\|x\|$ . Then  $p$  also is a norm.  $\square$

**Exercise 10.7.** Prove that the  $p$ -norm (see def.10.16 on p.192) is a norm on  $\mathbb{R}^n$  for the special case  $p = 1$ :

$$\|\vec{x}\|_1 = \sum_{j=1}^n |x_j| \quad \square$$

*In chapter 11 on the topology of real numbers (p. 204) 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 Metric Spaces

There is a branch of Mathematics, called topology, which deals with the concept of closeness. The concept of limits of a sequence  $(x_n)_n$  is based on closeness: The points of the sequence must get “arbitrarily close” to its limit as  $n \rightarrow \infty$ . Continuity of functions also can be phrased in terms of closeness: They map arbitrarily close elements of the domain to arbitrarily close elements of the codomain. In the most general setting Topology is about neighborhoods of a point without having the concept of measuring the distance of two points. We mostly won't deal with such a level of generality in this document. Instead we'll 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.9.2 (Convergence and Continuity in  $\mathbb{R}$ ) and ch.10.2.2 (Normed Vector Spaces).

### 11.1 The Topology 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 non-negative, i.e.,

$$\text{dist}(x, y) = |y - x| = |x - y|$$

More importantly, you must forget what you learned 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 or 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

$$\text{dist}(12.3, 15) = |12.3 - 15| = 2.7.$$

In section 10.1.3 on p.176 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.

#### 11.1.1 Definition and Examples of Metric Spaces

**Definition 11.1** (Metric spaces). Let  $X$  be an arbitrary, nonempty set. A **metric** on  $X$  is a real valued function of two arguments

$$d(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}, \quad (x, y) \mapsto d(x, y)$$

with the following three properties: <sup>99</sup>

$$(11.1a) \quad d(x, y) \geq 0 \quad \forall x, y \in X \quad \text{and} \quad d(x, y) = 0 \Leftrightarrow x = y \quad \text{positive definiteness}$$

$$(11.1b) \quad d(x, y) = d(y, x) \quad \forall x, y \in X \quad \text{symmetry}$$

$$(11.1c) \quad d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in X \quad \text{triangle inequality}$$

Let  $x, y \in X$ . We say that  $x$  and  $y$  are  $\varepsilon$ -**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 11.1** (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 11.2.** 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 satisfied indeed the definition of a metric. <sup>100</sup> □

*The triangle inequality generalizes to more than two terms.*

**Proposition 11.1.** Let  $(X, d)$  be a metric space. Let  $n \in \mathbb{N}$  and  $x_1, x_2, \dots, x_n \in X$ . Then

$$(11.2) \quad d(x_1, x_n) \leq \sum_{j=1}^{n-1} d(x_j, x_{j+1}) = d(x_1, x_2) + d(x_2, x_3) + \dots + d(x_{n-1}, x_n).$$

<sup>99</sup> If you forgot the meaning of  $X \times X$ , it's time to review the definition of a cartesian product (def.5.1 in ch.5.1 on p.81.

<sup>100</sup> see def.11.3 on p.207 and prop.11.2 directly thereafter.

The proof is left as exercise 11.1.

Before we give some examples of metric spaces, here is a theorem that tells you that a vector space with a norm (see def.10.15 on p.191), becomes a metric space as follows:

**Theorem 11.1** (Norms define metric spaces). Let  $(V, \|\cdot\|)$  be a normed vector space. Then the function

$$(11.3) \quad d_{\|\cdot\|}(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}_{\geq 0}; \quad (x, y) \mapsto d_{\|\cdot\|}(x, y) := \|y - x\|$$

defines a metric space  $(V, d_{\|\cdot\|})$ .

The proof is left as exercise 11.2.

**Definition 11.2** (Metric induced by a norm). We say that the metric  $d_{\|\cdot\|}(\cdot, \cdot)$  defined by (11.3) is **induced by the norm**  $\|\cdot\|$ . We also say that  $d_{\|\cdot\|}(\cdot, \cdot)$  is **derived from the norm**  $\|\cdot\|$  or that  $d_{\|\cdot\|}(\cdot, \cdot)$  is **associated with the norm**  $\|\cdot\|$ .<sup>101</sup>  $\square$

Here are some examples of metric spaces.

**Example 11.1** ( $\mathbb{R}$  with  $d_{|\cdot|}(a, b) = |b - a|$ ). According to thm.11.1  $(\mathbb{R}, d_{|\cdot|})$  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.17.15 on p.316):

$$\begin{aligned} d_{|\cdot|}(x, z) &= |x - z| = |(x - y) - (z - y)| \\ &\leq |x - y| + |z - y| = d_{|\cdot|}(x, y) + d_{|\cdot|}(z, y) = d_{|\cdot|}(x, y) + d_{|\cdot|}(y, z). \quad \square \end{aligned}$$

**Example 11.2** (bounded real valued functions with  $d_{\|\cdot\|_{\infty}} f, g = \sup\text{-norm of } g(\cdot) - f(\cdot)$ ).

$$(11.4) \quad d_{\|\cdot\|_{\infty}}(f, g) = \|g - f\|_{\infty} = \sup\{|g(x) - f(x)| : x \in X\}$$

is a metric on the set  $\mathcal{B}(X, \mathbb{R})$  of all bounded real valued functions on  $X$ . This follows from thm.11.1 and prop.10.13 on p. 190, according to which  $(\mathcal{B}(X, \mathbb{R}), \|\cdot\|_{\infty})$  is a normed vector space.  $\square$

**Example 11.3** (continuous real valued functions on  $[a, b]$  with  $d_{\|\cdot\|_{L^2}}(f, g) = \|g - f\|_{L^2}$ ). We will see in ch.11.1.2 on p.207 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 metric

$$(11.5) \quad d_{\|\cdot\|_{L^2}}(f, g) = \|g - f\|_{L^2} = \sqrt{\int_a^b (g(x) - f(x))^2 dx}.$$

(see def.10.19 on p.195).  $\square$

<sup>101</sup> Compare this to def.10.17 (Norm for an inner product) on p.193.

**Example 11.4** ( $\mathbb{R}^N$  with the Euclidean metric).

$$d_{\|\cdot\|_2}(\vec{x}, \vec{y}) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_N - x_N)^2} = \sqrt{\sum_{j=1}^N (y_j - x_j)^2}$$

This follows from the fact that the Euclidean norm is a norm on the vector space  $\mathbb{R}^N$  (see cor.10.1 on p.193.)  $\square$

*Just in case you think that all metrics are derived from norms, here is a counterexample.*

**Definition 11.3** (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  $X$ .  $\square$

*The above definition makes sense because of the following proposition.*

**Proposition 11.2.** *The discrete metric satisfies the properties of a metric.*

*PROOF:* Obviously the function is non-negative and it is zero if and only if  $x = y$ . Symmetry is obvious too.

*The triangle inequality  $d(x, z) \leq d(x, y) + d(y, z)$  is clear in the special case  $x = z$ . (Why?)*

*So let us assume  $x \neq z$ . But then  $x \neq y$  or  $y \neq z$  or both must be true. (Why?) That means that*

$$d(x, z) = 1 \leq d(x, y) + d(y, z)$$

*and this proves the triangle inequality. ■*

### 11.1.2 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 11.4** (Maximal displacement distance between real valued functions). Let  $X$  be an arbitrary, nonempty set and let  $f(\cdot), g(\cdot) : X \rightarrow \mathbb{R}$  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

$$(11.6) \quad d_\infty(f, g) := \|f(\cdot) - g(\cdot)\|_\infty = \sup\{|f(x) - g(x)| : x \in X\},$$

*i.e., as the metric induced by the sup-norm on the set  $\mathcal{B}(X, \mathbb{R})$  of all bounded real valued function on  $X$ .  $\square$*

**Remark 11.3.** We have previously encountered formula (11.6) in example 11.2 on p.206. We will see in prop.11.29 on p.247 of ch.11.3.1 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  $h(\cdot)$ .  $\square$

**Remark 11.4.** The following picture 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.

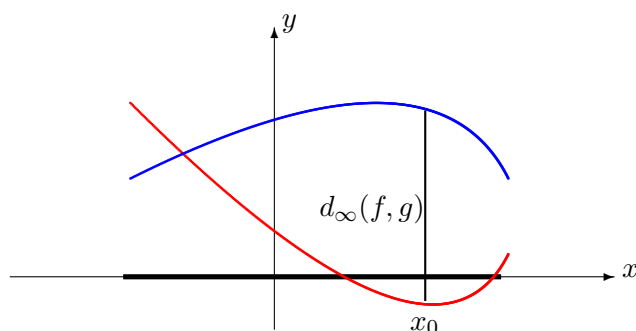


Figure 11.1: Distance of two real valued functions.

This figure allows you to visualize for a given  $\delta > 0$  and  $f : X \rightarrow \mathbb{R}$  the “ $\delta$ -neighborhood” of  $f(\cdot)$  defined as

$$(11.7) \quad N_\delta(f) := \{g : X \rightarrow \mathbb{R} : d(\infty f, g) < \delta\} = \{g(\cdot) : X \rightarrow \mathbb{R} : \sup_{x \in X} |f(x) - g(x)| < \delta\},$$

i.e., the set of all functions  $g(\cdot)$  with distance less than  $\delta$  from  $f(\cdot)$ .

You draw the graph of  $f(\cdot) + \delta$  (the graph of  $f(\cdot)$  shifted up by the amount of  $\delta$ ) and the graph of  $f(\cdot) - \delta$  (the graph of  $f(\cdot)$  shifted down by the amount of  $\delta$ ). Any function  $g(\cdot)$  which stays completely inside this band, without actually touching it, belongs to the  $\delta$ -neighborhood of  $f(\cdot)$ .



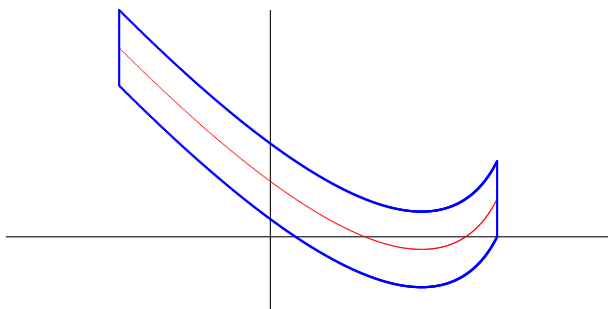


Figure 11.2:  $\delta$ -neighborhood of a real valued function.

In other words, assuming that the domain  $A$  is a single, connected chunk and not a collection of several separate intervals, the  $\delta$ -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  $\delta$ .  $\square$

**Definition 11.5** (Mean distances between real valued functions). Let  $a, b \in \mathbb{R}$  such that  $a < b$  and let  $f(\cdot), g(\cdot) : X \rightarrow \mathbb{R}$  be two continuous real valued functions on  $X$ . We define the **mean square distance** between  $f(\cdot)$  and  $g(\cdot)$  on  $[a, b]$  as

$$(11.8) \quad d_{L^2}(f, g) := d_{\|\cdot\|_{L^2}(f, g)} = \|g - f\|_{L^2} = \int_a^b (g(x) - f(x))^2 dx,$$

i.e., as the metric induced by the  $L^2$ -norm on the set  $\mathcal{C}_{\mathcal{B}}([a, b], \mathbb{R})$  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

$$(11.9) \quad d_{L^1}(f, g) := d_{\|\cdot\|_{L^1}(f, g)} = \|g - f\|_{L^1} = \int_a^b |g(x) - f(x)| dx,$$

i.e., as the metric induced by the  $L^1$ -norm on the set  $\mathcal{C}_{\mathcal{B}}([a, b], \mathbb{R})$ .

$\square$

**Remark 11.5.** We saw in def.10.18, example 10.21 and def.10.19 on pp.193 that both

$$(11.10) \quad d_{L^1}(f, g) := d_{\|\cdot\|_{L^1}(f, g)} = \|g - f\|_{L^1} = \int_a^b |g(x) - f(x)| dx,$$

$$(11.11) \quad d_{L^2}(f, g) := d_{\|\cdot\|_{L^2}(f, g)} = \|g - f\|_{L^2} = \int_a^b (g(x) - f(x))^2 dx,$$

are often better suitable than the distance derived from the sup-norm to measure the distance of two functions. One of the drawbacks from a teaching perspective is that there is no picture like figure 11.2 to visualize the set of all functions with an  $L^1$ -distance or  $L^2$ -distance from a given function.

### 11.1.3 Neighborhoods and Open Sets

A. Given a point  $x_0 \in \mathbb{R}$  (a real number) and  $\varepsilon > 0$ , we can look at

$$(11.12) \quad \begin{aligned} N_\varepsilon(x_0) &= (x_0 - \varepsilon, x_0 + \varepsilon) = \{x \in \mathbb{R} : x_0 - \varepsilon < x < x_0 + \varepsilon\} \\ &= \{x \in \mathbb{R} : d(x, x_0) = |x - x_0| < \varepsilon\} \end{aligned}$$

which is the set of all real numbers  $x$  with a distance to  $x_0$  of strictly less than a number  $\varepsilon$  (the open interval with end points  $x_0 - \varepsilon$  and  $x_0 + \varepsilon$ ). (see example (11.1) on p.206).

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

$$(11.13) \quad \begin{aligned} N_\varepsilon(\vec{x}_0) &= \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\} \\ &= \{\vec{x} \in \mathbb{R}^2 : d_{\|\cdot\|_2}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{x}_0\|_2 < \varepsilon\} \end{aligned}$$

which is the set of all points in the plane with a distance to  $\vec{x}_0$  of strictly less than a number  $\varepsilon$  (the open disc around  $\vec{x}_0$  with radius  $\varepsilon$  from which the points on the boundary (those with distance equal to  $\varepsilon$ ) 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

$$(11.14) \quad \begin{aligned} N_\varepsilon(\vec{x}_0) &= \{(x, y, z) \in \mathbb{R}^3 : (\vec{x} - \vec{x}_0)^2 + (\vec{y} - \vec{y}_0)^2 + (\vec{z} - \vec{z}_0)^2 < \varepsilon^2\} \\ &= \{\vec{x} \in \mathbb{R}^3 : d_{\|\cdot\|_2}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{x}_0\|_2 < \varepsilon\} \end{aligned}$$

which is the set of all points in space with a distance to  $\vec{x}_0$  of strictly less than a number  $\varepsilon$  (the open ball around  $\vec{x}_0$  with radius  $\varepsilon$  from which the points on the boundary (those with distance equal to  $\varepsilon$ ) are excluded).

D. Given a normed vector space  $(V, \|\cdot\|)$  and a vector  $x_0 \in V$ , we can look at

$$(11.15) \quad N_\varepsilon(x_0) = \{x \in V : \|x - x_0\| < \varepsilon\}$$

which is the set of all vectors in  $V$  with a distance to  $x_0$  of strictly less than a number  $\varepsilon$  (the open set around  $x_0$  with "radius"  $\varepsilon$  from which the points on the boundary (those with distance equal to  $\varepsilon$ ) are excluded).

E. Given a bounded real valued function  $f \in \mathcal{B}(X, \mathbb{R})$ , we can look at the sets  $N_\varepsilon(f)$  ( $\varepsilon > 0$ ) defined in (11.7) on p.208, i.e., the set of all functions  $g(\cdot)$  with distance less than  $\varepsilon$  from  $f(\cdot)$ .

F. Given is a closed interval  $[a, b]$  ( $a, b \in \mathbb{R}$ ). For a continuous (hence bounded) real valued function  $f \in \mathcal{B}([a, b], \mathbb{R})$ , we can look at the sets

$$(11.16) \quad N_\varepsilon(f) = \{g \in \mathcal{B}([a, b], \mathbb{R}) : \|g - f\|_{L^2} < \varepsilon\},$$

i.e., the set of all functions  $g(\cdot)$  such that  $\sqrt{\int_a^b (g(x) - f(x))^2 dx} < \varepsilon$  (see def.10.19 on p.195)

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 11.6** ( $\varepsilon$ -Neighborhood). Given a metric space  $(X, d)$  and an element  $x_0 \in X$ , let

$$(11.17) \quad N_\varepsilon(x_0) = \{x \in X : d(x, x_0) < \varepsilon\}$$

be the set of all elements of  $X$  with a distance to  $x_0$  of strictly less than the number  $\varepsilon$  (the open set around  $x_0$  with "radius"  $\varepsilon$  from which the points on the boundary (those with distance equal to  $\varepsilon$ ) are excluded). We call  $N_\varepsilon(x_0)$  the  $\varepsilon$ -**neighborhood** of  $x_0$ .  $\square$

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  $\delta$ -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. 11.4 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 11.7** (Interior points in metric spaces). Given is a metric space  $(X, d)$ .

An element  $a \in A \subseteq X$  is called an **interior point** of  $A$  if we can find some  $\varepsilon > 0$ , however small it may be, so that  $N_\varepsilon(a) \subseteq A$ .  $\square$

**Definition 11.8** (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**.  $\square$

**Proposition 11.3.** Let  $(X, d)$  be a metric space. Let  $x, y \in X$  and  $\varepsilon > 0$  such that  $y \in N_\varepsilon(x)$ .

$$\text{If } \delta > 0 \text{ Then } N_\delta(y) \subseteq N_{\delta+\varepsilon}(x)$$

*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  $N_\delta(y) \subseteq N_{\delta+\varepsilon}(x)$ .  $\blacksquare$

**Proposition 11.4.**  $N_\varepsilon(x_0)$  is an open set

It is worth while to examine the following proof<sup>102</sup> 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,

$$(11.18) \quad \varepsilon - d(a, x_0) = 2\delta$$

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 (11.18).

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.  $\blacksquare$

**Proposition 11.5** (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:*

left as exercise 11.3 on p.260.  $\blacksquare$

<sup>102</sup> A shorter proof can be given if the previous proposition is used.

**Definition 11.9** (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  $x_0$ .  $\square$

**Remark 11.6** (Open neighborhoods are the important ones). You will see that the important neighborhoods are the small ones, not the big ones. The definition above says that for any neighborhood  $A_x$  of a point  $x \in X$  you can find an **open neighborhood**  $U_x$  of  $x$  such that  $U_x \subseteq A_x$ .

Because of this there are many propositions and theorems where you may assume that a neighborhood you deal with is open.  $\square$

**Definition 11.10** (boundary points). Let  $A$  be a subset of the metric space  $(X, d)$ .  $x \in X$  is called a **boundary point** of  $A$  if any neighborhood of  $x$  intersects both  $A$  and  $A^c$ . We write  $\partial A$  for the set of all boundary points of  $A$  and call this set the **boundary** of  $A$ .  $\square$

**Theorem 11.2** (Metric spaces are topological spaces). *The following is true about open sets of a metric space  $(X, d)$ :*

(11.19a) An arbitrary union  $\bigcup_{i \in I} U_i$  of open sets  $U_i$  is open.

(11.19b) A finite intersection  $U_1 \cap U_2 \cap \dots \cap U_n$  ( $n \in \mathbb{N}$ ) of open sets is open.

(11.19c) 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.

*PROOF of b:* Let  $x \in U := U_1 \cap U_2 \cap \dots \cap U_n$ . 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

$$\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n\}$$

Then  $\varepsilon > 0$  and <sup>103</sup>

$$N_\varepsilon(x) \subseteq N_{\varepsilon_j}(x) \subseteq U_j \quad (1 \leq j \leq n), \quad \text{hence} \quad N_\varepsilon(x) \subseteq \bigcap_{j=1}^n U_j.$$

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

<sup>103</sup> by the way, this is the exact spot where the proof breaks down if you deal with an infinite intersection of open sets: the minimum would have to be replaced by an infimum and there is no guarantee that it would be strictly larger than zero.

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.11.1.5 on p.214.

### 11.1.4 Convergence

You have already encountered the precise definition of the convergence of sequences of real numbers in ch.9.2. It is only a small step to generalize this concept to all metric spaces and therefore also to all normed vector spaces.

**Definition 11.11** (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  $n \rightarrow \infty$  if almost all of the  $x_n$  will come arbitrarily close to  $a$  in the following sense:

Let  $\delta$  be an arbitrarily small positive real number. Then there is a (possibly extremely large) integer  $n_0$  such that all  $x_j$  belong to  $N_\delta(a)$  just as long as  $j \geq n_0$ . To say this another way: Given any number  $\delta > 0$ , however small, you can find an integer  $n_0$  such that

$$(11.20) \quad d(a, x_j) < \delta \text{ for all } j \geq n_0$$

We write either of

$$(11.21) \quad a = \lim_{n \rightarrow \infty} x_n \quad \text{or} \quad x_n \rightarrow a$$

and we call  $a$  the **limit** of the sequence  $(x_n)$

There is an equivalent way of expressing convergence towards  $a$ : No matter how small a neighborhood of  $a$  you choose: at most finitely many of the  $x_n$  will be located outside that neighborhood.

□

**Theorem 11.3** (Limits in metric spaces are uniquely determined). Let  $(X, d)$  be a metric space . Let  $(x_n)_n$  be a convergent sequence in  $X$  Then its limit is uniquely determined.

*PROOF:* Otherwise there would be two different points  $L_1, L_2 \in X$  such that both  $\lim_{n \rightarrow \infty} x_n = L_1$  and  $\lim_{n \rightarrow \infty} x_n = L_2$  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  $n \geq \max(N_1, N_2)$ ,<sup>104</sup>

$$d(L_1, L_2) \leq d(L_1, x_n) + d(x_n, L_2) < 2\varepsilon = d(L_1, L_2)$$

and we have reached a contradiction. ■

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.

<sup>104</sup> You could have used  $N_1 + N_2$  instead. Do you see why?

**Proposition 11.6.** <sup>105</sup> 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  $\lim_{n \rightarrow \infty} (y_n)_n = L$ .

PROOF:

Left as an exercise. ■

**Proposition 11.7.** <sup>106</sup> 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

$$\lim_{n \rightarrow \infty} x_n = L \Leftrightarrow \lim_{n \rightarrow \infty} y_n = L.$$

PROOF:

Left as an exercise. ■

**Remark 11.7.**

The following allows you to prove convergence of  $x_n$  to  $L$  in  $(X, d)$  by utilizing what you know about convergence in  $(\mathbb{R}, d_1 \cdot | \cdot |)$ .

$$\lim_{n \rightarrow \infty} x_n = L \Leftrightarrow \lim_{n \rightarrow \infty} d(x_n, L) = 0. \quad \square$$

**Remark 11.8** (Opposite of convergence). Given a metric space  $(X, d)$ , what is the opposite of  $\lim_{k \rightarrow \infty} x_k = L$ ?

Beware! It is NOT the statement that  $\lim_{k \rightarrow \infty} x_k \neq L$  because such a statement would mislead you to believe that such a limit exists, it just happens not to coincide with  $L$

The correct answer: There exists some  $\varepsilon > 0$  such that for **all**  $N \in \mathbb{N}$  there exists some natural number  $j = j(N)$  such that  $j \geq N$  and  $d(x_j, L) \geq \varepsilon$ . □

It is easy to prove from the above remark the following:

**Proposition 11.8** (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  $\varepsilon$ -neighborhood of  $L$ .

The proof is left as exercise 11.23. ■

### 11.1.5 Abstract Topological spaces

Theorem 11.2 on p.212 gives us a way of defining neighborhoods for sets which do not have a metric.

<sup>105</sup> See prop.9.7 on p.152.

<sup>106</sup> See prop.9.8 on p.153.

**Definition 11.12** (Abstract topological spaces). Let  $X$  be an arbitrary nonempty set and let  $\mathfrak{U}$  be a set of subsets<sup>107</sup> of  $X$  whose members satisfy the properties a, b and c of (11.19) on p.212:

(11.22a) An arbitrary union  $\bigcup_{i \in I} U_i$  of sets  $U_i \in \mathfrak{U}$  belongs to  $\mathfrak{U}$ ,

(11.22b)  $U_1, U_2, \dots, U_n \in \mathfrak{U}$  ( $n \in \mathbb{N}$ )  $\Rightarrow U_1 \cap U_2 \cap \dots \cap U_n \in \mathfrak{U}$ ,

(11.22c)  $X \in \mathfrak{U}$  and  $\emptyset \in \mathfrak{U}$ .

Then  $(X, \mathfrak{U})$  is called a **topological space**. The members of  $\mathfrak{U}$  are called **open sets**. of  $(X, \mathfrak{U})$  and the collection  $\mathfrak{U}$  of open sets is called the **topology** of  $X$ .  $\square$

**Definition 11.13** (Topology induced by a metric). Let  $(X, d)$  be a metric space and let  $\mathfrak{U}_d$  be the set of open subsets of  $(X, d)$ , i.e., all sets  $U \subseteq X$  which consist of interior points only: for each  $x \in U$  there exist  $\varepsilon > 0$  such that

$$N_\varepsilon(x) = \{y \in X : d(x, y) < \varepsilon\} \subseteq U$$

(see (11.7) on p.211). We have seen in theorem (11.2) that those open sets satisfy the conditions of the previous definition. In other words,  $(X, \mathfrak{U}_d)$  defines a topological space. We say that its topology is **induced by the metric**  $d(\cdot, \cdot)$  or that it is **generated by the metric**  $d(\cdot, \cdot)$ . If there is no confusion about which metric we are talking about, we also simply speak about the **metric topology**.

Let  $V$  be a vector space with a norm  $\|\cdot\|$ . Remember that any norm defines a metric  $d_{\|\cdot\|}(\cdot, \cdot)$  via  $d_{\|\cdot\|}(x, y) = \|x - y\|$  (see (11.1) on p.206). Obviously, this norm defines open sets

$$\mathfrak{U}_{\|\cdot\|} := \mathfrak{U}_{d_{\|\cdot\|}}$$

on  $V$  by means of this metric. We say that this topology is **induced by the norm**  $\|\cdot\|$  or that it is **generated by the norm**  $\|\cdot\|$ . If there is no confusion about which norm we are talking about, we also simply speak about the **norm topology** of  $V$ .  $\square$

**Example 11.5** (Discrete topology). Let  $X$  be nonempty. Def.11.3 on p.207 gave the discrete metric as the function

$$d(x, y) = \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y. \end{cases}$$

The associated topology is

$$\mathfrak{U}_d = 2^X = \{A : A \subseteq X\}.$$

Note that the discrete metric defines the biggest possible topology on  $X$ , i.e., the biggest possible collection of subsets of  $X$  whose members satisfy properties a, b, c of definition 11.12 on p.214. We call this topology the **discrete topology** of  $X$ .  $\square$

**Example 11.6** (Indiscrete topology). Here is an example of a topology which is not generated by a metric. Let  $X$  be an arbitrary nonempty set and define  $\mathfrak{U} := \{\emptyset, X\}$ . Then  $(X, \mathfrak{U})$  is a topological space. This is trivial because any intersection of members of  $\mathfrak{U}$  is either  $\emptyset$  (if at least one member is  $\emptyset$ ) or  $X$  (if all members are  $X$ ). Conversely, any union of members of  $\mathfrak{U}$  is either  $\emptyset$  (if all members are  $\emptyset$ ) or  $X$  (if at least one member is  $X$ ).

The topology  $\{\emptyset, X\}$  is called the **indiscrete topology** of  $X$ . It is the smallest possible topology on  $X$ .  $\square$

<sup>107</sup> We encountered subsets of  $2^X$  with special properties previously when looking at rings of sets in def.6.7 (Rings and Algebras of Sets) on p.116.

**Definition 11.14** (Neighborhoods and interior points in topological spaces). Let  $x \in X$  and  $A \subseteq X$ . It is not assumed that  $A$  be open.  $A$  is called a **neighborhood** of  $x$  and  $x$  is called an **inner point** or **interior point** of  $A$  if there exists an open set  $U$  such that

$$(11.23) \quad x \in U \subseteq A.$$

$x$  is called an **exterior point** of  $A$  if there exists an open set  $U'$  such that

$$(11.24) \quad x \in U' \subseteq A^c,$$

i.e.,  $x$  is an inner point of  $A^c$ .

We call  $ext(A) := \{ \text{all exterior points of } A \}$  the **open exterior** of  $A$ . <sup>108</sup>  $\square$

**Definition 11.15** (Interior of a set in a topological space). Let  $(X, \mathcal{U})$  be a topological space and let  $A \subseteq X$ . The union of all open subsets of  $A$

$$(11.25) \quad A^\circ := \bigcup [U \in \mathcal{U} : U \subseteq A]$$

is called the **interior** of  $A$ . An alternate notation for  $A^\circ$  is  $int(A)$ .  $\square$

*It follows from def.11.12 (abstract topological spaces) on p.214 that  $A^\circ$  is an open set which is, as a union of subsets of  $A$ , also a subset of  $A$ . Because  $A^\circ$  is the union of all such sets, we conclude that*

*The interior  $A^\circ$  of  $A$  is the largest of all open subsets of  $A$ .*

**Remark 11.9.** Let  $(X, \mathcal{U})$  be a topological space and  $A \subseteq X$ . Note that the open exterior of  $A$  is

$$(11.26) \quad ext(A) = \overline{A^c}^\circ. \quad \square$$

**Remark 11.10.** Let  $(X, \mathcal{U})$  be a topological space.

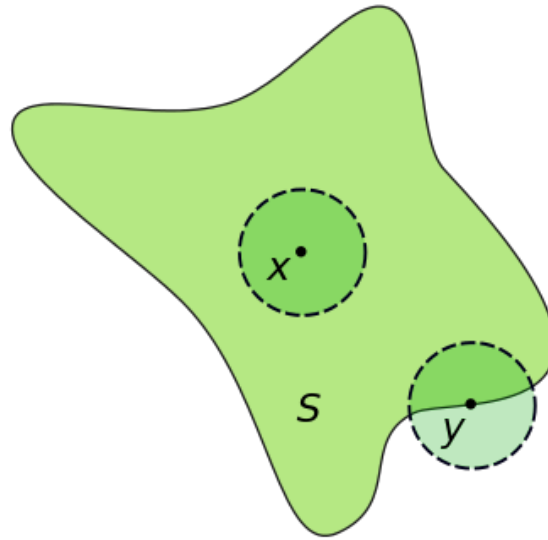
$$(11.27) \quad \text{If } A \subseteq B \subseteq X \text{ then } A^\circ \subseteq B^\circ. \quad \square$$

**Definition 11.16** (boundary points in topological spaces). Let  $(X, \mathcal{U})$  be a topological space and  $A \subseteq X$ . Then  $x \in X$  is called a **boundary point** of  $A$  if any neighborhood of  $x$  intersects both  $A$  and  $A^c$ . We write  $\partial A$  for the set of all boundary points of  $A$  and call this set the **boundary** of  $A$ .

<sup>108</sup> Source: [https://en.wikipedia.org/wiki/Interior\\_\(topology\)](https://en.wikipedia.org/wiki/Interior_(topology))



Figure 11.3: Inner points v.s. boundary points.  
 Source: [https://en.wikipedia.org/wiki/Interior\\_\(topology\)](https://en.wikipedia.org/wiki/Interior_(topology))



We note that this definition is exactly the same as that given for metric spaces (compare def.11.10 on p.212).  $\square$

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 11.11** (Hierarchy of topological spaces). We have seen the following:

- a.  $\mathbb{R}^N$  is an inner product space (see (10.11) on p.189).
- b. All inner product spaces are normed spaces (see (10.3) on p.192).
- c. All normed spaces are metric spaces (see (11.1) on p.206).
- d. All metric spaces are topological spaces. (see (11.12) on p.214 and (11.13) on p.215).

### 11.1.6 Bases and Neighborhood Bases

**Definition 11.17** (Base of the topology). Let  $(X, \mathcal{U})$  be a topological space.

A subset  $\mathfrak{B} \subseteq \mathcal{U}$  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  $\mathfrak{B}$ :

$$(11.28) \quad U = \bigcup_{i \in I} B_i \quad (B_i \in \mathfrak{B} \text{ for all } i \in I)$$

where  $I$  is a suitable index set which of course will in general depend on  $U$ .  $\square$

We note that, because  $X$  itself is open, (11.28) 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 11.18** (Neighborhood base). Let  $(X, \mathfrak{U})$  be a topological space.

The set of subsets of  $X$

$$(11.29) \quad \mathfrak{N}(x) := \{A \subseteq X : A \text{ is a neighborhood of } x\}$$

is called the **neighborhood system of  $x$**

Given a point  $x \in X$ , any subset  $\mathfrak{B} := \mathfrak{B}(x) \subseteq \mathfrak{N}(x)$  of the neighborhood system of  $x$  is called a **neighborhood base of  $x$**  if it satisfies the following condition: For any  $A \in \mathfrak{N}(x)$  you can find a  $B \in \mathfrak{B}(x)$  such that  $B \subseteq A$ .  $\square$

*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 11.19** (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.  $\square$

**Proposition 11.9** ( $\varepsilon$ -neighborhoods are a base of the topology). Let  $(X, d)$  be a metric space. Then the set  $\mathcal{B}_1 := \{N_\varepsilon(x) : x \in X, \varepsilon > 0\}$  is a base for the topology of  $(X, d)$  (see 11.17 on p.217) and the same is true for the “thinner” set  $\mathcal{B}_2 := \{N_{1/n}(x) : x \in X, n \in \mathbb{N}\}$ .

*PROOF:* To show that  $\mathcal{B}_1$  (resp.,  $\mathcal{B}_2$ ) 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  $\mathcal{B}_1$  (resp.,  $\mathcal{B}_2$ ). We prove this for  $\mathcal{B}_2$ .

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  $U \subseteq \bigcup [N_{1/n(x)}(x) : x \in U] \subseteq U$ .

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 6.1 on p.110).

We obtain  $U = \bigcup [N_{1/n(x)}(x) : x \in U]$  and we have managed to represent our open  $U$  as a union of elements of  $\mathcal{B}_2$ . This proves that  $\mathcal{B}_2$  is a base for the topology of  $(X, d)$ .

As  $\mathcal{B}_2 \subseteq \mathcal{B}_1$  it follows that  $\mathcal{B}_1$  also is such a base.  $\blacksquare$

**Theorem 11.4** (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

$$(11.30) \quad \mathfrak{B}(x) := \{N_{1/n}(x) : n \in \mathbb{N}\}.$$

Then  $\mathfrak{B}(x)$  is a neighborhood base of  $x$  because, by def.11.7 and def.11.9 (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  $\frac{1}{n} < \varepsilon$ .  $\blacksquare$

**Proposition 11.10.** 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  $(X, \mathfrak{U}_d)$ .

The proof is left as exercise 11.12.

**Definition 11.20** (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  $\mathfrak{U}$ .  $\square$

**Theorem 11.5** (Euclidean space  $\mathbb{R}^n$  is second countable). *Let*

$$(11.31) \quad \mathfrak{B} := \{ N_{1/j}(\vec{q}) : \vec{q} \in \mathbb{Q}^n, j \in \mathbb{N} \}.$$

Here  $\mathbb{Q}^n = \{ \vec{q} = (q_1, \dots, q_n) : q_j \in \mathbb{Q}, 1 \leq j \leq n \}$  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.11.3 on p.211, applied to  $\delta = \varepsilon = \frac{1}{n_{\vec{x}}}$ , that*

$$N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}}) \subseteq N_{2/n_{\vec{x}}}(\vec{x}) \subseteq U \text{ for all } \vec{x} \in U.$$

$$\text{Hence } U = \bigcup_{\vec{x} \in U} \{ \vec{x} \} \subseteq \bigcup [ N_{1/n_{\vec{x}}}(\vec{q}_{\vec{x}}) : \vec{x} \in U ] \subseteq U.$$

*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.8.7 on p. 139 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.  $\blacksquare$*

### 11.1.7 Metric and Topological Subspaces

**Definition 11.21** (Metric subspaces). Given is a metric space  $(X, d)$  and a nonempty  $A \subseteq (X, d)$ . Let  $d|_{A \times A} : A \times A \rightarrow \mathbb{R}_{\geq 0}$  be the restriction  $d|_{A \times A}(x, y) := d(x, y)$  ( $x, y \in A$ ) of the metric  $d$  to  $A \times A$  (see def.5.12 on p.99). It is trivial to verify that  $(A, d|_{A \times A})$  is a metric space in the sense of def.11.1 on p.204. We call  $(A, d|_{A \times A})$  a **metric subspace** of  $(X, d)$  and we call  $d|_{A \times A}$  the **metric induced by  $d$**  or the **metric inherited from  $(X, d)$** .  $\square$

**Remark 11.12.**



Metric subspaces come with their own collections of open and closed sets, neighborhoods,  $\varepsilon$ -neighborhoods, convergent sequences, ...

You must watch out when looking at statements and their proofs whether those concepts refer to the entire space  $(X, d)$  or to the subspace  $(A, d|_{A \times A})$ .

$\square$

#### Notations 11.1.

- a) Because the only difference between  $d$  and  $d_{A \times A}$  is the domain, it is customary to write  $d$  instead of  $d|_{A \times A}$  to make formulas look simpler if doing so does not give rise to confusion.
- b) We often shorten “open in  $(A, d|_{A \times A})$ ” to “open in  $A$ ”, “closed in  $(A, d|_{A \times A})$ ” to “closed in  $A$ ”, “convergent in  $(A, d|_{A \times A})$ ” to “convergent in  $A$ ”, .....  $\square$

**Definition 11.22** (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  $(A, d|_{A \times A})$  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  $\varepsilon$ -neighborhood of  $a$  (in  $(X, d)$ ). We define

$$N_\varepsilon^A(a) := N_\varepsilon(a) \cap A,$$

i.e.,  $N_\varepsilon^A(a)$  is defined as the trace of  $N_\varepsilon(a)$  in  $A$ .  $\square$

**Proposition 11.11** (Open sets in metric subspaces as traces of open sets in  $X$ ). Let  $(X, d)$  be a metric space and  $A \subseteq X$  a nonempty subset of  $X$ .

**a.** Let  $\varepsilon > 0$  and  $a \in A$ . Then

$$(11.32) \quad N_\varepsilon^A(a) = \{x \in A : d|_{A \times A}(x, a) < \varepsilon\},$$

i.e.,  $N_\varepsilon^A(a)$  is the  $\varepsilon$ -Neighborhood of  $a$  as defined in def.11.6 on p.210.

It then follows from

$$(11.33) \quad N_\varepsilon^A(a) = N_\varepsilon(a) \cap A.$$

that each  $\varepsilon$ -neighborhood in the subspace  $A$  is the trace of an  $\varepsilon$ -neighborhood in  $X$ .

**b.** More generally, a set  $U \subseteq A$  is open in  $A$  if and only if there is an open  $V \subseteq X$  such that

$$(11.34) \quad U = V \cap A,$$

i.e.,  $U$  is the trace of a set  $V$  which is open in  $X$ .

*PROOF of a:* First we prove (11.32). As  $d|_{A \times A}$  is the restriction of  $d$  to  $A \times A$  it follows that

$$\begin{aligned} N_\varepsilon^A(a) &= N_\varepsilon(a) \cap A = \{x \in X : d(x, a) < \varepsilon\} \cap A \\ &= \{x \in A : d(x, a) < \varepsilon\} = \{x \in A : d|_{A \times A}(x, a) < \varepsilon\}. \end{aligned}$$

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  $(A, d|_{A \times A})$ .

Because  $a \in V$  and  $V$  is open in  $X$ , there is  $\varepsilon > 0$  such that  $N_\varepsilon(a) \subseteq V$ . It follows that

$N_\varepsilon^A(a) = N_\varepsilon(a) \cap A \subseteq V \cap A = U$  and  $N_\varepsilon^A(a)$  is open in  $A$ , hence  $a$  is interior point of  $U$  with respect to the subspace  $(A, d|_{A \times A})$ .

Finally we prove that if  $U \subseteq A$  is open in  $A$  there is  $V \subseteq X$  open in  $X$  such that  $U = V \cap A$ :

We can write  $U = \bigcup [N_{\varepsilon(a)}^A(a) : a \in U]$  for suitable  $\varepsilon(a) > 0$  (see the proof of prop.11.9 on p.218).

Let  $V := \bigcup [N_{\varepsilon(a)}(a) : a \in U]$ . Then  $V$  is open in  $(X, d)$  as union of the open sets  $N_{\varepsilon(a)}(a)$ , hence

$$\begin{aligned} V \cap A &= A \cap \bigcup [N_{\varepsilon(x)}(x) : x \in U] = \bigcup [N_{\varepsilon(x)}(x) \cap A : x \in U] \\ &= \bigcup [N_{\varepsilon(x)}^A(x) : x \in U] = U \end{aligned}$$

(the second equality follows from prop.6.2 on p.112). This finishes the proof.  $\blacksquare$

The last proposition justifies to define subspaces of abstract topological spaces as follows.

**Definition 11.23** (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  $\mathfrak{U}_A$ , i.e.,

$$\mathfrak{U}_A = \{V \cap A : V \in \mathfrak{U}\}.$$

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 **topology induced by the metric  $d$**  or the **topology inherited from  $(X, d)$** .  $\square$

**Proposition 11.12** (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 the definition def.11.12 on p.214 of an abstract topological space.*

*PROOF: Left as an exercise.*

**Remark 11.13** (Convergence does not extend to subspaces). Let  $A \subseteq (X, d)$  and  $a_n \in A$ .

Let  $a_n \in A$  for all  $n \in \mathbb{N}$ . Note that convergence of the sequence  $(a_n)$  in the space  $(X, d)$  (i.e., there exists  $x \in X$  such that  $x = \lim_{n \rightarrow \infty} a_n$ ), does NOT imply convergence of the sequence in the subspace  $(A, d|_{A \times A})$ . Rather, we have the following **dichotomy**:

- a.  $x \in A$ : Then  $a_n$  converges in the subspace  $(A, d|_{A \times A})$  (and also in  $(X, d)$ ).
- b.  $x \in A^c$ : Then  $a_n$  converges in  $(X, d)$  but not in  $(A, d|_{A \times A})$ .  $\square$

### 11.1.8 Bounded Sets and Bounded Functions

**Definition 11.24** (bounded sets). Given is a subset  $A$  of a metric space  $(X, d)$ . The **diameter** of  $A$  is defined as

$$(11.35) \quad \text{diam}(\emptyset) := 0, \quad \text{diam}(A) := \sup\{d(x, y) : x, y \in A\} \text{ if } A \neq \emptyset.$$

We call  $A$  a **bounded set** if  $\text{diam}(A) < \infty$ .  $\square$

**Proposition 11.13.** *Given is a metric space  $(X, d)$  and a nonempty subset  $A$ . The following are equivalent:*

$$(11.36) \quad \mathbf{a.} \text{ diam}(A) < \infty, \quad \text{i.e., } A \text{ is bounded.}$$

$$(11.37) \quad \mathbf{b.} \text{ There exists } \gamma > 0 \text{ and } x_0 \in X \text{ such that } A \subseteq N_\gamma(x_0).$$

$$(11.38) \quad \mathbf{c.} \text{ For all } x \in X \text{ there exists } \gamma > 0 \text{ such that } A \subseteq N_\gamma(x).$$

*PROOF of “b  $\Rightarrow$  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  $\Rightarrow$  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 “ $c \Rightarrow a$ ”: 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 “ $a \Rightarrow c$ ”: Given  $x \in X$ , pick an arbitrary  $x_0 \in A$  and let  $\gamma := d(x, x_0) + \text{diam}(A)$ . 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  $A \subseteq N_\gamma(x)$ . ■

### 11.1.9 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 11.25** (Contact points). Given is a topological space  $(X, \mathfrak{U})$ .

Let  $A \subseteq X$  and  $x \in X$  ( $x$  may or may not belong to  $A$ ).  $x$  is called a **contact point**<sup>109</sup> of  $A$  if

$$(11.39) \quad A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x. \quad \square$$

The following theorem shows that we can characterize contact points of subsets of metric spaces by means of sequences.

**Theorem 11.6** (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 “ $\Rightarrow$ ”: Let  $x \in X$  be such that  $N \cap A \neq \emptyset$  for any neighborhood  $N$  of  $x$ . Let  $x_n \in N_{1/n}(x) \cap A$ . Such  $x_n$  exists because the neighborhood  $N_{1/n}(x)$  has nonempty intersection with  $A$ .

Given  $\varepsilon > 0$ , let  $N \in \mathbb{N}$  be chosen such that  $\frac{1}{N} < \varepsilon$ . 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  $x_n \rightarrow x$ .

PROOF of “ $\Leftarrow$ ” Let  $x \in X$  and assume there is  $(x_n)_{n \in \mathbb{N}}$  such that  $x_n \in A$  for all  $n$  and  $x_n \rightarrow x$ .

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  $x_n \rightarrow x$  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  $U_x \cap A \neq \emptyset$ . ■

**Note 11.1.** 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  $A$ .

<sup>109</sup> German: Berührungspunkt - see [12] Von Querenburg, p.21

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  $]0, 1[$ .  $\square$

**Note 11.2** (Contact points vs Limit points). Besides contact points there also is the concept of a limit point. Here is the definition (see [9] Munkres, a standard book on topology):

Given is a metric space  $(X, d)$ . Let  $A \subseteq X$  and  $x_0 \in X$ .  $x_0$  is called a **limit point** or **cluster point** or **point of accumulation** of  $A$  if any neighborhood  $U$  of  $x_0$  intersects  $A$  in at least one point other than  $x_0$ . This definition excludes “isolated points”<sup>110</sup> of  $A$  from being limit points of  $A$ .  $\square$

**Definition 11.26** (Closed sets). Given is a metric space  $(X, d)$  and a subset  $A \subseteq X$ . We call

$$(11.40) \quad \bar{A} := \{x \in X : x \text{ is a contact point of } A\}$$

the **closure** of  $A$ . A set that contains all its contact points is called a **closed set**.  $\square$

**Remark 11.14.** It follows from note 11.1.a that  $A \subseteq \bar{A}$ .  $\square$

**Proposition 11.14.** *The complement of an open set is closed.*

*PROOF of 11.14:* Let  $A$  be an open set. Each point  $a \in A$  is an interior point which can be surrounded by a  $\delta$ -neighborhood  $N_\delta(a)$  which, for small enough  $\delta$ , will be entirely contained within  $A$ .

Assume  $x \in X$  is a contact point of  $A^c$ . We want to prove that  $A^c$  is a closed set, so we must show that  $x \in A^c$ .

We assume to the contrary that  $x \in A$ .  $A$  is open, so  $x$  is an interior point of  $A$ . It follows that there is a neighborhood  $N_\delta(x)$  surrounding  $x$  which is entirely contained in  $A$ , hence  $N_\delta(x) \cap A^c = \emptyset$ .

We also assumed that  $x$  is a contact point of  $A^c$ . This implies that  $N_\delta(x) \cap A^c \neq \emptyset$ . We have reached a contradiction.  $\blacksquare$

**Proposition 11.15.** *The complement of a closed set is open.*

*PROOF:* We will give two proofs of the above. The first one makes use of the criterion for contact points (theorem 11.6) and works with sequences. The second proof is based on the definition of contact points and works with neighborhoods and interior points.

*a. First proof of prop.11.15:*

Let  $A \subseteq (X, d)$  be closed. If  $A^c$  is not open then there must be some  $b \in A^c$  which is not an interior point of  $A^c$ .

We show that this assumption leads to a contradiction.

Because  $b$  is not an interior point of  $A^c$ , there is no  $\delta$ -neighborhood, for whatever small  $\delta$ , that entirely belongs to  $A^c$ . So, for each  $j \in \mathbb{N}$ , there is an  $x_j \in N_{1/j}(b)$  which does not belong to  $A^c$ , i.e.,  $x_j \in A$ .

<sup>110</sup>  $a \in A \subseteq (X, d)$  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  $A$ .



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  $b \in A^c$  and we have a contradiction. ■

**b.** Alternate proof of prop.11.15, based on neighborhoods and interior points:

Let  $A$  be closed set and  $b \in A^c$ .

The closed set  $A$  contains all its contact points, so  $b \notin A$  implies that  $b$  is not a contact point of  $A$ : according to def.11.25 there exists some neighborhood  $V$  of  $b$  such that  $V \cap A = \emptyset$ , i.e.,  $V \subseteq A^c$ .

We have proved that an arbitrary  $b \in A^c$  is an interior point of  $A^c$ , i.e., the complement of the closed set  $A$  is open. ■

**Theorem 11.7** (Open iff complement is closed). Let  $(X, d)$  be a metric space and  $A \subseteq X$ . Then  $A$  is open if and only if  $A^c$  is closed.

*PROOF:* Immediate from prop.11.14 and prop.11.15 ■

**Remark 11.15.** Note the following.

**a.** We have seen that def.11.25 for contact points and hence def.11.26 for closed sets are entirely based on the concept of neighborhood which itself is entirely based on that of open sets. It follows that those two definitions make perfect sense not only in metric spaces but, more generally, in abstract topological spaces  $(X, \mathcal{U})$  which are characterized by the set  $\mathcal{U}$  of all open subsets of  $X$  (see def.11.12 on p.214).

**b.** Moreover the proof for prop.11.14 (complements of open sets are closed) and the first proof for prop.11.15 (complements of closed sets are open) are based on those definitions and do not employ specific properties of metric spaces either; theorem 11.7 also works for abstract topological spaces.

**c.** Matter of fact, 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 proved from those definitions that closed sets are exactly those that contain all their contact points. □

**Definition 11.27** (Contact points, closed sets and boundary points in topological spaces).<sup>111</sup> Given is an abstract topological space  $(X, \mathcal{U})$ .

Let  $A \subseteq X$  and  $x \in X$  ( $x$  may or may not belong to  $A$ ).  $x$  is called a **contact point** of  $A$  if

$$A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x.$$

We call

$$\bar{A} := \{x \in X : x \text{ is a contact point of } A\}$$

the **closure** of  $A$ . A set that contains all its contact points is called a **closed set**.

A point  $x \in X$  is called a **boundary point** of  $A$  if any neighborhood of  $x$  intersects both  $A$  and  $A^c$ . As for metric spaces, we write  $\partial A$  for the set of all boundary points of  $A$  and call this set the **boundary** of  $A$ .

□

---

<sup>111</sup> see def.11.25, def.11.26 and def.11.10 for the corresponding definitions in metric spaces.



**Remark 11.16.** We note that  $A \subseteq \bar{A}$ : Let  $a \in A$  and let  $V_a$  be a neighborhood of  $a$ . Because  $a \in V_a$ , we obtain  $a \in V_a \cap A$ , hence  $V_a \cap A \neq \emptyset$ , hence  $a \in \bar{A}$ .

It follows that  $A$  is closed if and only if  $A = \bar{A}$  (which justifies the name “closure of  $A$ ” for  $\bar{A}$ .)  
□

**Proposition 11.16.** Let  $(X, \mathfrak{A})$  be a topological space and  $A \subseteq B \subseteq X$ . Then  $\bar{A} \subseteq \bar{B}$ .

The proof is left as exercise 11.16.

**Proposition 11.17.** Let  $(X, \mathfrak{A})$  be a topological space.

The closed sets of  $X$  satisfy the following property:

- (11.41)     **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 (6.1) on p.111). Observe that  $X$  plays the role of a universal set because all members  $U$  of  $\mathfrak{A}$  and their complements  $U^c$  are subsets of  $X$ .

PROOF of **a**: Let  $(C_\alpha)$  be an arbitrary family of closed sets. Then  $U_\alpha := C_\alpha^c$  is an open set for each  $\alpha$ . Observe that  $C_\alpha^c = U_\alpha$  because the complement of the complement of any set gives you back that set. Let  $C := \bigcap_\alpha C_\alpha$ . Then

$$C^c = \left( \bigcap_\alpha C_\alpha \right)^c = \bigcup_\alpha C_\alpha^c = \bigcup_\alpha U_\alpha.$$

In other words  $C^c$  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  $U_j := C_j^c$  is an open set for each  $j$ . Let  $C := \bigcup_{1 \leq j \leq n} C_j$ .

Then

$$C^c = \left( \bigcup_j C_j \right)^c = \bigcap_j C_j^c = \bigcap_j U_j$$

Hence,  $C^c$  is the intersection of finitely many open sets. This shows that  $C^c$  is open, i.e.,  $C$  is closed. We have proved **b**.

PROOF of **c**: Trivial because

$$X^c = \emptyset \quad \text{and} \quad \emptyset^c = X. \quad \blacksquare$$

**Proposition 11.18.** Let  $(X, \mathfrak{A})$  be a topological space and  $A \subseteq X$ . Then

$$(11.42) \quad \partial A = \bar{A} \cap \overline{A^c},$$

i.e.,  $x \in X$  is a boundary point of  $A$  if and only if  $x$  is a contact point of both  $A$  and  $A^c$ .

PROOF: Left as an exercise.

**Proposition 11.19** (Minimality of the closure of a set). Let  $(X, \mathfrak{U})$  be a topological space and  $A \subseteq X$ . Then

$$(11.43) \quad \bar{A} = \bigcap \left[ C \supseteq A : C \text{ is closed} \right].$$

The closure  $\bar{A}$  of  $A$  is the smallest of all closed supersets of  $A$ .

*PROOF:* Let  $\mathfrak{C} := \{C \supseteq A : C \text{ is closed}\}$  and let  $F := \bigcap \mathfrak{C}$ . We need to show that  $\bar{A} = F$ .

It follows from prop.11.17.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.11.16 on p.225). ■

**Proposition 11.20** (Closure of a set as a hull operator<sup>112</sup>). Let  $(X, \mathfrak{U})$  be a topological space. We can think of the closure of sets as a function  $\bar{\cdot} : 2^X \rightarrow 2^X$ ;  $A \mapsto \bar{A}$ . This function has the following properties:

$$\text{a. } \bar{\emptyset} = \emptyset, \quad \text{b. } A \subseteq \bar{A}, \quad \text{c. } \bar{\bar{A}} = \bar{A}, \quad \text{d. } \overline{A \cup B} = \bar{A} \cup \bar{B}.$$

*PROOF:* a follows from (11.41).c and b follows from remark 11.14.

The proof of c and d is left as exercise 11.18.

### 11.1.10 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 11.20 (page 213)

**Definition 11.28** (Cauchy sequences). Given is a metric space  $(X, d)$ .

A sequence  $(x_n)$  in  $X$  is called a **Cauchy sequence**<sup>113</sup> or, in short, it is Cauchy if it has the following property: Given any whatever small number  $\varepsilon > 0$ , you can find a (possibly very large) number  $n_0$  such that

$$(11.44) \quad d(x_i, x_j) < \varepsilon \quad \text{for all } i, j \geq n_0$$

This is called the **Cauchy criterion for convergence** of a sequence. □

<sup>112</sup> This proposition states that the closure is a so-called **closure operator** which is defined to be a function

$$cl : 2^X \rightarrow 2^X; \quad A \mapsto cl(A) := \bar{A}$$

on some abstract, nonempty set  $X$  (which need not be a topological space) such that the following are satisfied:

$$\text{a. } cl(\emptyset) = \emptyset, \quad \text{b. } A \subseteq cl(A), \quad \text{c. } cl(cl(A)) = cl(A), \quad \text{d. } cl(A \cup B) = cl(A) \cup cl(B).$$

It can be shown that if we define

$$\mathfrak{U} := \{A^c : cl(A) = A\}$$

then  $(X, \mathfrak{U})$  satisfies the properties of a topological space.

<sup>113</sup> so named after the great french mathematician Augustin-Louis Cauchy (1789–1857) who contributed massively to the most fundamental ideas of Calculus.

**Example 11.7** (Cauchy criterion for real numbers). In  $\mathbb{R}$  we have  $d(x, y) = |x - y|$  and the Cauchy criterion requires for any given  $\varepsilon > 0$  the existence of  $n_0 \in \mathbb{N}$  such that

$$(11.45) \quad |x_i - x_j| < \varepsilon \quad \text{for all } i, j \geq n_0. \quad \square$$

**Proposition 11.21.** *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  $(x_n)_n$ . ■

The following theorem of the completeness of the set of all real numbers<sup>114</sup> states that any Cauchy sequence converges to a real number. This is a big deal: To show that a sequence has a finite limit you need not provide the actual value of that limit. All you must show is that this sequence satisfies the Cauchy criterion. One can say that this preoccupation with proving existence rather than computing the actual value is one of the major points which distinguish Mathematics from applied Physics and the engineering disciplines.

Here is the formal definition of a complete set in a metric space.

**Definition 11.29** (Completeness in metric spaces). Given is a metric space  $(X, d)$ . A subset  $A \subseteq X$  is called **complete** if any Cauchy sequence  $(x_n)$  with elements in  $A$  converges to an element of  $A$ .  $\square$

**Remark 11.17.**

- a. In particular,  $X$  itself is complete iff any Cauchy sequence in  $X$  converges.
- b.  $A$  is complete as a subset of  $(X, d)$  iff the subspace  $((A, d)|_{A \times A})$  is complete "in itself".  $\square$

**Theorem 11.8** (Completeness of the real numbers). *The following is true for the real numbers with the metric  $d(a, b) = |b - a|$  but will in general be false for arbitrary metric spaces: Let  $(x_n)$  be a Cauchy sequence in  $\mathbb{R}$ . then there exists a real number  $L$  such that  $L = \lim_{n \rightarrow \infty} x_n$ .*

*PROOF:* It follows from prop.11.21 that  $x_n$  is bounded, hence  $(x_n)_n$  possesses finite liminf and limsup.<sup>115</sup> We now show that  $\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n$ .

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

$$0 \leq \beta_N - \alpha_N = |\beta_N - \alpha_N| \leq |\beta_N - x_j| + |x_j - x_i| + |x_i - \alpha_N| \leq 3\varepsilon.$$

<sup>114</sup> Remember the completeness axiom for  $\mathbb{R}$  (axiom 9.1 on p.145) 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.9.3 (Characterization of limits via limsup and liminf) on p.166, a theorem which will be used in this chapter to prove the completeness of  $\mathbb{R}$  as a metric space.

<sup>115</sup> See ch.9.1 (Minima, Maxima, Infima and Suprema).

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

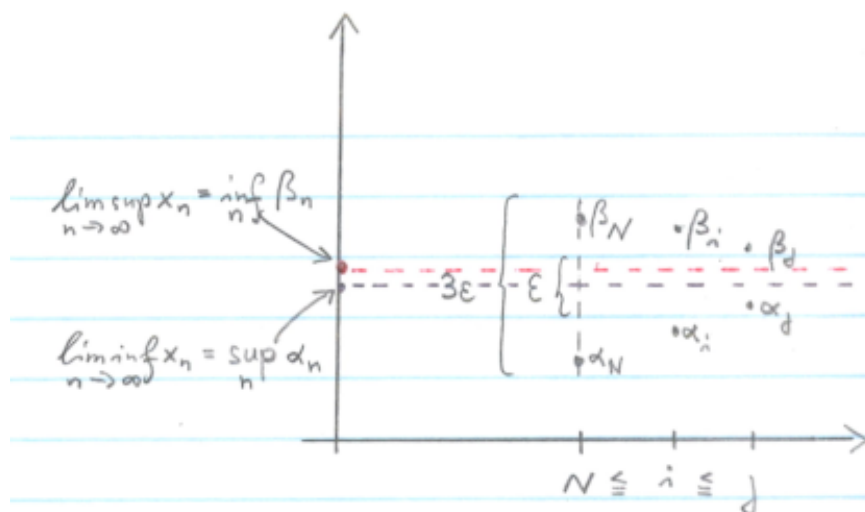
$$\inf \beta_k - \sup \alpha_k \leq \beta_N - \alpha_N \leq 3\varepsilon.$$

But then

$$0 \leq \limsup_{k \rightarrow \infty} x_k - \liminf_{k \rightarrow \infty} x_k = \inf_k \beta_k - \sup_k \alpha_k \leq 3\varepsilon.$$

$\varepsilon > 0$  was arbitrary, hence  $\limsup_{k \rightarrow \infty} x_k = \liminf_{k \rightarrow \infty} x_k$ .

Figure 11.4:  $\liminf = \limsup$  for Cauchy sequences



Part 3: It follows from theorem 9.3 on p.166 that the sequence  $(x_n)_n$  converges to  $L := \limsup_{k \rightarrow \infty} x_k$  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 11.9** (Completeness of  $\mathbb{R}^N$ ). Let  $(\vec{x}_n)$  be a Cauchy sequence in  $\mathbb{R}^N$ . Then there exists a vector  $\vec{a} \in \mathbb{R}^N$  such that  $\vec{a} = \lim_{n \rightarrow \infty} \vec{x}_n$ .

*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

$$a_1, a_2, a_3, \dots, a_N \quad \text{such that} \quad a_k = \lim_{n \rightarrow \infty} x_{n,k} \quad (1 \leq k \leq N).$$

For a given number  $\varepsilon$  we can find natural numbers  $n_{0,1}, n_{0,2}, \dots, n_{0,N}$  such that

$$|x_{n,k} - a_k| < \frac{\varepsilon}{N} \quad \text{for all } n \geq n_{0,j} \text{ and for all } 1 \leq k \leq N.$$

Let  $n^* := \max(n_{0,1}, n_{0,2}, \dots, n_{0,N})$  and  $\vec{a} := (a_1, a_2, \dots, a_n)$ . It follows that

$$d(\vec{x}_n - \vec{a}) = \sqrt{\sum_{k=1}^N |x_{n,k} - a_k|^2} \leq \sqrt{N \cdot \left(\frac{\varepsilon}{N}\right)^2} = \frac{\varepsilon}{\sqrt{N}} \leq \varepsilon \quad \text{for all } n \geq n^*. \blacksquare$$

**Example 11.8** (Approximation of decimals). The following illustrates Cauchy sequences and completeness in  $\mathbb{R}$ . Take any real number  $x \geq 0$  and write it as a decimal:

$$x = m + \sum_{j=1}^{\infty} d_j \cdot 10^{-j} \quad (m \in \mathbb{Z}_{>0}, d_j \in \{0, 1, 2, \dots, 9\})$$

As was explained in (??) on (p.??), anything that can be written as a decimal number is a real number. Let's say,  $x$  starts out on the left as

$$x = 258.1408926584207531 \dots$$

We define as  $x_k$  the leftmost part of  $x$ , truncated  $k$  digits after the decimal points:

$$x_1 = 258.1, \quad x_2 = 258.14, \quad x_3 = 258.140, \quad x_4 = 258.1408, \quad x_5 = 258.14089, \quad \dots$$

We further define  $y_k$  the leftmost part of  $x$ , truncated  $k$  digits after the decimal points, but the rightmost digit incremented by 1 (there might be a carry-over to the left when you add 1 to 9)

$$y_1 = 258.2, \quad y_2 = 258.15, \quad y_3 = 258.141, \quad y_4 = 258.1409, \quad y_5 = 258.14090, \quad \dots$$

The sequence  $(x_n)$  is non-decreasing:  $x_{n+1} \geq x_n$  for all  $n$  and the sequence  $(y_n)$  is non-increasing:  $y_{n+1} \leq y_n$  for all  $n$ . We have the sandwiching property:  $x_n \leq x \leq y_n$  for all  $n$ . Both sequences are Cauchy because

$$d(x_{n+i}, x_{n+j}) = |x_{n+i} - x_{n+j}| \leq 10^{-n} \rightarrow 0 \quad (n \rightarrow \infty),$$

$$d(y_{n+i}, y_{n+j}) = |y_{n+i} - y_{n+j}| \leq 10^{-n} \rightarrow 0 \quad (n \rightarrow \infty).$$

It follows that both sequences have limits. It is obvious that  $x = \lim_{n \rightarrow \infty} x_n = \lim_{m \rightarrow \infty} y_m$ .

What just has been illustrated is that there a natural way to construct for a given  $x \in \mathbb{R}$  Cauchy sequences that converge towards  $x$ . The completeness principle states that the reverse is true: For any Cauchy sequence there is an element  $x$  towards which the sequence converges.  $\square$

*This proposition gives an important example of a metric space which is not complete.*

**Proposition 11.22.** *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  $x = m + \sum_{j=1}^{\infty} d_j 10^{-j}$  where each  $d_j \in \mathbb{Z}_{\geq 0}$  is a digit, i.e.,  $0 \leq d_j \leq 9$ . Let

$$s_n = m + \sum_{j=1}^n d_j 10^{-j} \text{ Then}$$

$$(11.46) \quad |x - s_n| = x - s_n = \sum_{j=n+1}^{\infty} d_j 10^{-j} \leq 9 \cdot \sum_{j=n+1}^{\infty} 10^{-j} = 10^{-n}$$

and it follows, not surprisingly, that  $x = \lim_{n \rightarrow \infty} x_n$ .

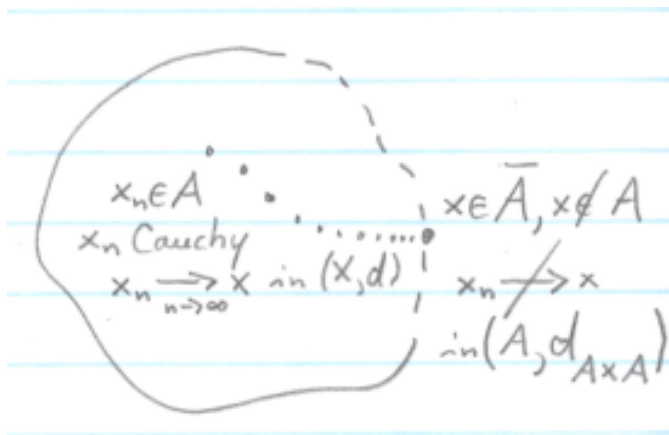
We know from thm.11.12 (Convergent sequences are Cauchy) on p.231 that the sequence  $s_n$  is Cauchy in  $(\mathbb{R}, d_{|\cdot|})$ . But  $s_n \in \mathbb{Q}$  for all  $n$  and the distance  $d_{|\cdot|}(s_n, s_m) = |s_n - s_m|$  is the same in  $(\mathbb{R}, d_{|\cdot|})$  and  $(\mathbb{Q}, d_{|\cdot|})$ .

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. ■

We won't really talk about completeness in general until the chapter on compact spaces. Here is a very simple fact about completeness.

**Theorem 11.10** (Complete sets are closed). Any complete subset of a metric space is closed.

Figure 11.5: complete  $\Rightarrow$  closed



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 def.11.25 on p.222: 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  $A \cap N_{1/m}(a) \neq \emptyset$  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  $N > \frac{1}{2\varepsilon}$ . 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  $\rho$ -balls  $N_\rho(a)$  and  $N_\rho(b)$  where  $\rho$  is one half the distance between  $a$  and  $b$  (see the proof of thm.11.3 on p.213).

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.11.10.

**Theorem 11.11** (Closed subsets of a complete space are complete). Let  $(X, d)$  be a complete metric space and let  $A \subseteq X$  be closed. Then  $A$  is complete, i.e., the metric subspace  $(A, d|_{A \times A})$  is complete.

*PROOF:* Let  $(x_n)_n$  be Cauchy in  $A$ . We must show that there is  $a \in A$  such that  $x_n \rightarrow a$ . 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  $x_n \rightarrow x$ . All  $x_n$  belong to  $A$ . According to thm.11.6 (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. ■

**Theorem 11.12** (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  $x_n \rightarrow L$ . Let  $\varepsilon > 0$ . There exists  $N \in \mathbb{N}$  such that

$$(11.47) \quad k \geq N \Rightarrow d(x_k, L) < \varepsilon/2.$$

It follows from (11.47) 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 (11.44) of the definition of a Cauchy sequence (def. 11.28 on p.226). ■

**Proposition 11.23.** Let  $(x_n)_n$  be a Cauchy sequence in a metric space  $(X, d)$  such that some subsequence  $x_{n_j}$  converges to a limit  $L$ . Then

- a.** ANY subsequence of  $(x_n)_n$  converges to  $L$ .
- b.**  $(x_n)_n$  is a convergent sequence.
- c.** Any subsequence  $x_{n_j}$  of a convergent sequence  $(x_n)_n$  converges to the limit of  $(x_n)_n$ .

*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  $y_j \rightarrow L$  implies that there is  $N \in \mathbb{N}$  such that

$$(11.48) \quad d(y_j, L) < \varepsilon/2 \text{ for all } j \geq N.$$

Because  $(x_j)$  is Cauchy there also exists  $N' \in \mathbb{N}$  such that

$$(11.49) \quad d(x_i, x_j) < \varepsilon/2 \text{ for all } i, j \geq N'.$$

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 (11.49) that  $d(x_j, y_K) = d(x_j, x_{n_K}) < \varepsilon/2$  and it follows from (11.48) that  $d(y_K, L) < \varepsilon/2$ . We conclude that  $d(x_j, L) < \varepsilon$  for all  $j \geq K$  and this proves convergence  $x_j \rightarrow L$ .

*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 c.:* This is trivial, too: The convergent sequence  $(x_n)_n$  is Cauchy, and the assumption now follows from part a. ■



## 11.2 Continuity (Study this!)

### 11.2.1 Definition and Characterizations of Continuous Functions

We have briefly discussed in ch.9.2 on p.150. 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.

**Definition 11.30** (Sequence continuity). Given are two metric spaces  $(X, d_1)$  and  $(Y, d_2)$ . Let  $A \subseteq X$ ,  $x_0 \in A$  and let  $f : A \rightarrow Y$  be a mapping from  $A$  to  $Y$ . We say that  $f$  is **sequence continuous at  $x_0$**  and we write

$$(11.50) \quad \lim_{x \rightarrow x_0} f(x) = f(x_0)$$

if the following is true for any sequence  $(x_n)$  with values in  $A$ :

$$(11.51) \quad \text{if } x_n \rightarrow x_0 \text{ then } f(x_n) \rightarrow f(x_0).$$

In other words, the following must be true for any sequence  $(x_n)$  in  $A$  and  $x_0 \in A$ :

$$(11.52) \quad \lim_{n \rightarrow \infty} x_n = x_0 \Rightarrow \lim_{n \rightarrow \infty} f(x_n) = f(\lim_{n \rightarrow \infty} x_n) = f(x_0).$$

We say that  $f$  is **sequence continuous** if  $f$  is sequence continuous at  $x_0$  for all  $x_0 \in A$ .  $\square$

**Remark 11.18.** Important points to notice:

- a) It is not enough for the above to be true for some sequences that converge to  $x_0$ . Rather, it must be true for all such sequences!
- b) We restrict our universe to the domain  $A$  of  $f$ :  $x_0$  and the entire sequence  $(x_n)_{n \in \mathbb{N}}$  must belong to  $A$  because there must be function values for all  $x$ -values. In other words,  $f$  is continuous at  $x_0 \in A$  if and only if  $f$  is continuous at  $x_0$  in the metric subspace  $(A, d|_{A \times A})$ .  $\square$

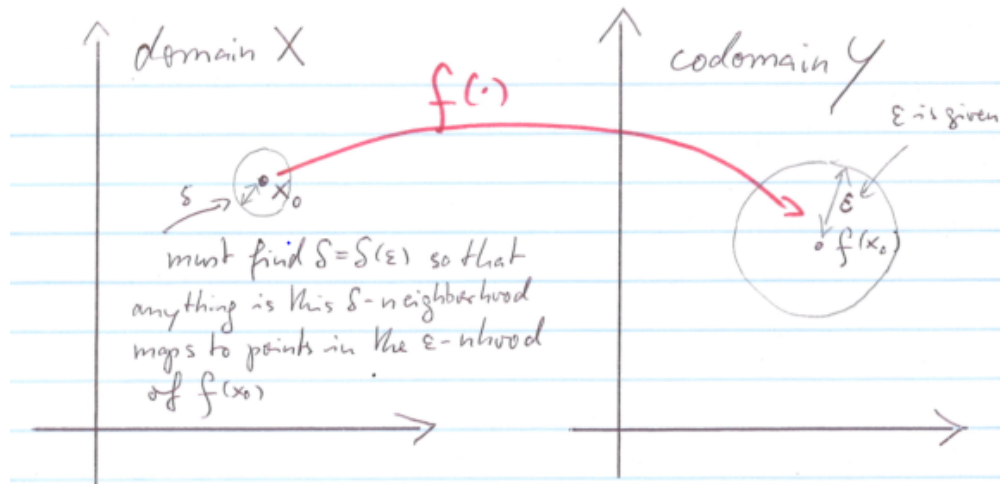
**Definition 11.31** ( $\varepsilon$ - $\delta$  continuity). Given are two metric spaces  $(X, d_1)$  and  $(Y, d_2)$ . Let  $A \subseteq X$ ,  $x_0 \in A$  and let  $f(\cdot) : A \rightarrow Y$  be a mapping from  $A$  to  $Y$ . We say that  $f(\cdot)$  is  $\varepsilon$ - $\delta$  **continuous at  $x_0$**  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:

$$(11.53) \quad f(N_\delta(x_0) \cap A) \subseteq N_\varepsilon(f(x_0)),$$

$$(11.54) \quad d_1(x, x_0) < \delta \Rightarrow d_2(f(x), f(x_0)) < \varepsilon \text{ for all } x \in A.$$

We say that  $f(\cdot)$  is  $\varepsilon$ - $\delta$  **continuous** if  $f(\cdot)$  is  $\varepsilon$ - $\delta$  continuous at  $a$  for all  $a \in A$ .  $\square$

Figure 11.6:  $\varepsilon$ - $\delta$  continuity



**Remark 11.19.** We recall from thm.11.32 on p.220 that

$$(11.55) \quad N_\delta \cap A = N_\delta^A(a) = \{x \in A : d|_{A \times A}(x, a) < \delta\}.$$

Hence, (11.53) states that  $f$  is  $\varepsilon$ - $\delta$  continuous at  $x_0$  if and only if  $f(N_\delta^A(x_0)) \subseteq N_\varepsilon(f(x_0))$ .  $\square$

**Theorem 11.13 (Continuity criterion).** Let  $(X, d_1)$  and  $(Y, d_2)$  be two metric spaces. Let  $A \subseteq X$ ,  $x_0 \in A$  and let  $f(\cdot) : A \rightarrow Y$  be a mapping from  $A$  to  $Y$ . Then  $f$  is sequence continuous at  $x_0$  if and only if  $f$  is  $\varepsilon$ - $\delta$  continuous at  $x_0$ .

In particular  $f$  is sequence continuous if and only if  $f$  is  $\varepsilon$ - $\delta$  continuous.

**a)  $\Rightarrow$ :** Proof that sequence continuity implies  $\varepsilon$ - $\delta$ -continuity:

We assume to the contrary that there exists some function  $f$  which is sequence continuous but not  $\varepsilon$ - $\delta$ -continuous at  $x_0$ , i.e., there exists some  $\varepsilon > 0$  such that neither (11.53) nor the equivalent (11.54) is true for any  $\delta > 0$ .

**a.1.** In other words, No matter how small  $\delta$  is chosen, 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

$$(11.56) \quad \text{there exists some } x_m \in N_{1/m}(x_0) \cap A \text{ such that } d_2(f(x_m), f(x_0)) \geq \varepsilon.$$

**a.2.** We now show that the sequence  $(x_m)_{m \in \mathbb{N}}$  converges to  $x_0$ : Let  $\gamma > 0$ . There exists  $N = N(\gamma) \in \mathbb{N}$  so big that  $N > 1/\gamma$ , i.e.,  $1/N < \gamma$ . As  $x_m \in N_{1/m}(x_0)$ , we obtain for all  $m \geq N$  that

$$d_1(x_m, x_0) < 1/m \leq 1/N < \gamma.$$

This proves that  $x_m \rightarrow x_0$ .

**a.3.** Clearly, the sequence  $(f(x_m))_{m \in \mathbb{N}}$  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 (11.56) 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)  $\Leftarrow$ :** Proof that  $\varepsilon$ - $\delta$ -continuity implies sequence continuity:

Let  $x_n \rightarrow x_0$ . Let  $y_n := f(x_n)$  and  $y_0 := f(x_0)$ . We must prove that  $y_n \rightarrow y_0$  as  $n \rightarrow \infty$ .

**b.1.** Let  $\varepsilon > 0$ . We can find  $\delta > 0$  such that (11.53) and hence (11.54) is satisfied. We assumed that  $x_n \rightarrow x_0$ . Hence 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 (11.54) that  $d_2(y_n, y) = d_2(f(x_n), f(x_0)) < \varepsilon$  for all  $n \geq N$ . In other words,  $y_n \rightarrow y_0$  as  $n \rightarrow \infty$  and the proof of “ $\Leftarrow$ ” is finished.

Finally, the equivalence  $f$  is  $\varepsilon$ - $\delta$  continuous  $\Leftrightarrow f$  is  $\varepsilon$ - $\delta$  continuous. ■

**Definition (Continuity in metric spaces):**

From now on we can use the terms “ $\varepsilon$ - $\delta$  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  $x_0$** .

We saw in the  $\varepsilon$ - $\delta$  continuity definition of a function with metric spaces for both domain and codomain and the subsequent remark 11.19 that continuity of  $f : (A, d_1|_{A \times A}) \rightarrow (Y, d_2)$  in  $x_0 \in A$  was equivalent to demanding that for any  $\varepsilon$ -neighborhood of  $f(x_0)$  there is a  $\delta$ -neighborhood of  $x_0$  such that

$$f(N_\delta^A(x_0)) \subseteq N_\varepsilon(f(x_0)).$$

The fact that any neighborhood of a point  $z$  in a metric space contains a  $\gamma$ -neighborhood of  $z$  for suitably small  $\gamma$ , is at the basis of the following theorem.

**Theorem 11.14** (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  $f(\cdot) : A \rightarrow Y$  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

$$(11.57) \quad f(U_{x_0} \cap A) \subseteq V_{f(x_0)}.$$

Equivalently, this can be stated in terms of the subspace  $(A, d_1|_{A \times A})$  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  $(A, d_1|_{A \times A})$  such that

$$(11.58) \quad f(U_{x_0}^A) \subseteq V_{f(x_0)}.$$

**a)  $\Rightarrow$ :** Assume that  $f$  is continuous, i.e.,  $\varepsilon$ - $\delta$  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  $N_\varepsilon(f(x_0)) \subseteq V_{f(x_0)}$ .  $\varepsilon$ - $\delta$  continuity at  $a$  implies the existence of  $\delta > 0$  such that  $f(N_\delta(x_0) \cap A) \subseteq V_{f(x_0)}$ .

This proves both (11.57) (choose  $U_{x_0} := N_\delta(x_0)$ ) and (11.58) (choose  $U_{x_0}^A := N_\delta(x_0) \cap A$ ).

**b)  $\Leftarrow$ :** Assume that (11.57) 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

$$(11.59) \quad f(N_\delta(x_0) \cap A) \subseteq N_\varepsilon(f(x_0)).$$

$N_\varepsilon(f(x_0))$  is a neighborhood of  $f(x_0)$ . It follows from (11.57) that there exists a neighborhood  $U_{x_0}$  of  $x_0$  such that

$$(11.60) \quad f(U_{x_0} \cap A) \subseteq N_\varepsilon(f(x_0)).$$

$x_0$  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

$$(11.61) \quad f(N_\delta(x_0) \cap A) \subseteq f(U_{x_0} \cap A) \subseteq N_\varepsilon(f(x_0)).$$

The second inclusion relation follows from (11.60). We have proved the existence of  $\delta > 0$  such that (11.59) is satisfied. This finishes the proof of “ $\Leftarrow$ ”. ■

The last theorem allows us to define continuity for functions between abstract topological spaces.

**Definition 11.32** (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  $f(\cdot) : A \rightarrow Y$  be a mapping from  $A$  to  $Y$ . We say that  $f(\cdot)$  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

$$(11.62) \quad f(U_{x_0} \cap A) \subseteq V_{f(x_0)}.$$

Equivalently, this 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 exists a neighborhood  $U_{x_0}^A$  of  $x_0$  in  $(A, \mathfrak{U}_{1_A})$  such that

$$(11.63) \quad f(U_{x_0}^A) \subseteq V_{f(x_0)}.$$

We say that  $f(\cdot)$  is **continuous** if  $f(\cdot)$  is continuous at  $a$  for all  $a \in A$ . □

[1] 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 (11.53) for all  $x_0 \in X$ .

**Proposition 11.24** (“ $f^{-1}(\text{open}) = \text{open}$ ” continuity). Let  $(X, \mathfrak{U})$  and  $(Y, \mathfrak{V})$  be two topological spaces and let  $f(\cdot) : X \rightarrow Y$  be a mapping from  $X$  to  $Y$ . Then  $f(\cdot)$  is continuous if and only if the following is true: If  $V$  is open in  $Y$  then the inverse image  $f^{-1}(V)$  is open in  $X$ . <sup>116</sup>

*PROOF of “ $\Rightarrow$ ”:* Let  $V$  be an open set in  $Y$ . Let  $U := f^{-1}(V)$ ,  $a \in U$  and  $b := f(a)$ . Then  $b \in V$  by the definition of inverse images.  $b$  is inner point of  $V$  because  $V$  is open. According to def.11.32 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.7.1 on p.123 that

$$U_a \subseteq f^{-1}(f(U_a)) \subseteq f^{-1}(V) = U.$$

It follows that the arbitrarily chosen  $a \in U$  is an interior point of  $U$  and this proves that  $U$  is open.

*PROOF of “ $\Leftarrow$ ”:* We now assume that all inverse images of open sets in  $Y$  are open in  $X$ .

Let  $a \in X$ ,  $b = f(a)$ , and 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

$$(11.64) \quad f(U_a) \subseteq V_b$$

<sup>116</sup> This is easily extended to  $f : A \rightarrow Y$  ( $\emptyset \neq A \subseteq X$ ) by demanding that  $f^{-1}(V)$  is open in  $(A, \mathfrak{U}_A)$ .

Let  $U_a := f^{-1}(V_b)$ . 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.7.6 on p.123 that

$$f(U) = f(f^{-1}(V_b)) = V_b \cap f(X) \subseteq V_b.$$

We have proved (11.64) ■

**Remark:**

Note the following for the previous proposition 11.24.

- a. In the interest of simplicity  $f$  was defined on all of  $X$  and not more generally on some arbitrary subset  $A$  of  $X$ .  
Note that the general case of  $f : A \rightarrow Y$  is covered by replacing  $(X, \mathfrak{U})$  with  $(A, \mathfrak{U}_A)$ , i.e., we deal with  $f : (A, \mathfrak{U}_A) \rightarrow (Y, \mathfrak{V})$ .
- b. Also note that the above proposition only addresses “global” continuity of  $f$  for **all**  $x \in X$  and there is no local version which handles continuity at a specific  $x_0$ .

**Remark 11.20** (continuity for real valued functions of real numbers). Let  $(X, d_1) = (Y, d_2) = \mathbb{R}$ . In this case equation (11.54) on p.233 becomes

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon.$$

See thm.9.1 on p.157. □

**Proposition 11.25** (continuity of the identity mapping). Let  $X$  be nonempty and let

$$id_X : X \rightarrow X; \quad x \mapsto x$$

be the identity function on  $X$ . Then  $id_X$  is continuous in the following two cases:

- a.  $X$  is a metric space  $(X, d)$ ,
- b.  $X$  is a topological space  $(X, \mathfrak{U})$ .

**PROOF of b:** 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.

**PROOF of a:** We associate with  $d(\cdot, \cdot)$  its metric topology  $\mathfrak{U}_d = \{U \subseteq X : U \text{ open in } (X, d)\}$  (see def.11.13 (Topology induced by a metric) on p.215). We apply part **b** which we have already proved to  $(X, \mathfrak{U}_d)$  and conclude that  $id_X$  is continuous.

**Alternate proof of a:** Here is a direct proof of **a** which works with the metric  $d$ .

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 (11.54) of the  $\varepsilon$ - $\delta$  characterization of continuity and it follows that  $id_X$  is continuous at  $x$ .<sup>117</sup>  $x$  was an arbitrary point in  $X$  and it follows that the identity is continuous. ■

Which one of the two proofs of **a** do you find more appealing?

<sup>117</sup> Actually, we have proved a very strong form of continuity. Generally speaking,  $\delta = \delta(\varepsilon, x_0)$  is tailored not only to the given  $\varepsilon$ , but also to the particular argument  $x_0$  at which continuity needs to be verified. We were able to find  $\delta$  which does not depend on the argument  $x_0$  but only on  $\varepsilon$ . We will learn later that this makes  $id_X$  **uniformly continuous** on its domain  $X$ . See def.11.34 (Uniform continuity of functions) on p.241.

### 11.2.2 Continuity of Constants and Sums and Products

For all the following, unless stated differently, let  $(X, d)$  be a metric space and  $A \subseteq X$ ,  $A \neq \emptyset$ . Let

$$f : A \rightarrow \mathbb{R}, \quad g : A \rightarrow \mathbb{R}$$

be two real valued functions which both are continuous in a point  $x_0 \in A$ . Moreover, let  $a, b \in \mathbb{R}$ . You can think of any fixed number  $a$  as a function

$$a(\cdot) : A \rightarrow \mathbb{R}; x \mapsto a.$$

In other words, the function  $a(\cdot)$  assigns to each  $x \in X$  one and the same value  $a$ . We called such a function a constant function (see (5.15) on p.100).

**Proposition 11.26** (Rules of arithmetic for continuous real-valued functions). *Given is a metric space  $(X, d)$ . Let the functions*

$$f(\cdot), g(\cdot), f_1(\cdot), f_2(\cdot), f_3(\cdot), \dots, f_n(\cdot) : A \rightarrow \mathbb{R}$$

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,  $af(\cdot) : x \mapsto a \cdot f(x)$  is continuous at  $x_0$  and, using  $-1$  as a constant,  $-f(\cdot) : x \mapsto -f(x)$  is continuous at  $x_0$ .
- c. The sum  $f + g(\cdot) : x \mapsto f(x) + g(x)$  is continuous at  $x_0$ .
- d. Any linear combination <sup>118</sup>  $\sum_{j=0}^n a_j f_j(\cdot) : x \mapsto \sum_{j=0}^n a_j f_j(x)$  is continuous in  $x_0$ .

*PROOF of a:* Let  $\varepsilon > 0$ . We do not even have to look for a suitable  $\delta$  to restrict the distance between two arguments  $x$  and  $x_0$  because it is always true that  $|a(x) - a(x_0)| = |a - a| = 0 < \varepsilon$ . This proves **a**.

*PROOF of b:* Let  $x_n \in A$  for all  $n \in \mathbb{N}$  such that  $x_n \rightarrow x_0$  as  $n \rightarrow \infty$ . All we need to show is  $f(x_n)g(x_n) \rightarrow f(x_0)g(x_0)$ . This follows from prop.9.11 (Rules of arithmetic for limits) on p.154 and we have proved **b**.

*PROOF of c:* Let  $x_n \in A$  for all  $n \in \mathbb{N}$  such that  $x_n \rightarrow x_0$  as  $n \rightarrow \infty$ . We must show  $f(x_n) + g(x_n) \rightarrow f(x_0) + g(x_0)$ . This also follows from prop.9.11 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

$$\sum_{j=0}^{n+1} a_j f_j(x) = \left( \sum_{j=0}^n a_j f_j(x) \right) + a_{n+1} f_{n+1}(x) = I + II.$$

The left term “ $I$ ” is continuous by the induction assumption and the entire sum  $I + II$  then is continuous as the sum of two continuous functions (proved in **c**). This proves **d**. ■

**Example 11.9** (Vector spaces of continuous real valued functions). Let  $(X, d)$  be a metric space. Then

$$(11.65) \quad \mathcal{C}(X, \mathbb{R}) := \{f(\cdot) : f(\cdot) \text{ is a continuous real valued function on } X\}$$

<sup>118</sup>See def.10.6 (linear combinations) on p.183

of all real continuous functions on  $X$  is a vector space. Note that we have seen this before in example 10.11 (Vector spaces of real valued functions) on p.181 for the special case of  $X \subseteq (\mathbb{R}, d_{|\cdot|})$ .

The sup-norm

$$\|f(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\}$$

(see (10.14) on p.190) is **not a real valued function** on all of  $\mathcal{C}(X, \mathbb{R})$  because  $\|f(\cdot)\|_\infty = +\infty$  for any unbounded  $f(\cdot) \in \mathcal{C}(X, \mathbb{R})$ . To avoid complications from dealing with infinity, we often restrict the scope to the subspace

$$\mathcal{C}_B(X, \mathbb{R}) := \{h : h \text{ is a bounded continuous real valued function on } X\}$$

(see prop.10.13 on p. 190) of the normed vector space  $\mathcal{B}(X, \mathbb{R})$  of all bounded real valued functions on  $X$ . On this subspace the sup-norm truly is a real valued function in the sense that  $\|f(\cdot)\|_\infty < \infty$ .  $\square$

### 11.2.3 Continuity of Polynomials (Understand this!)

**Definition 11.33** (Polynomials). Anything that has to do with polynomials takes place in  $\mathbb{R}$  and not on a metric space.

Let  $A$  be subset of the real numbers and let  $p(\cdot) : A \rightarrow \mathbb{R}$  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

$$(11.66) \quad p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n = \sum_{j=0}^n a_jx^j.$$

In other words, polynomials are linear combinations of the **monomials**  $x \rightarrow x^k$  ( $k \in (N)_0$ ).  $\square$

**Proposition 11.27** (All polynomials are continuous).

*The proof is left as exercise 11.24.*

**Proposition 11.28** (Vector space property of polynomials). *Sums and scalar products of polynomials are polynomials.*

*PROOF of a. Additivity:*

Let

$$p_1(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx_1^n = \sum_{j=0}^{n_1} a_jx^j$$

and

$$p_2(x) = b_0 + b_1x + b_2x^2 + \dots + b_nx_2^n = \sum_{j=0}^{n_2} b_jx^j$$

be two polynomials. We may assume that  $n_1 \leq n_2$ . Let  $a_{n_1+1} = a_{n_1+2} = \dots = a_{n_2} = 0$ . Then  $p_1(x) =$



$\sum_{j=0}^{n_2} a_j x^j$ , hence

$$\begin{aligned} p_1(x) + p_2(x) &= \sum_{j=0}^{n_2} a_j x^j + \sum_{j=0}^{n_2} b_j x^j \\ &= \sum_{j=0}^{n_2} (a_j + b_j) x^j \\ &= \sum_{j=0}^{n_2} c_j x^j \quad (c_j := a_j + b_j) \end{aligned}$$

This proves that the function  $p_1(\cdot) + p_2(\cdot)$  is of the form (11.66) and we have shown that it is a polynomial. The proof for the sum of more than two polynomials is easily done by induction. See thm.2.2 on p.22.

PROOF of **b**. Scalar product:

Let  $p(x) = \sum_{j=0}^n a_j x^j$  be a polynomial. Let  $\lambda \in \mathbb{R}$ . Then

$$(\lambda p)(x) = \lambda p(x) = \lambda \sum_{j=0}^n a_j x^j = \sum_{j=0}^n \lambda a_j x^j = \sum_{j=0}^n c_j x^j \quad (c_j := \lambda a_j)$$

This proves that the function  $\lambda p(\cdot)$  is of the form (11.66). ■

Polynomials may not always be given in their **normalized form** (11.66) on p.239. Here is an example:

$$\begin{aligned} p(x) &= a_0 x^0 (1-x)^n + a_1 x^1 (1-x)^{n-1} + a_2 x^2 (1-x)^{n-2} + \dots + a_{n-1} x^{n-1} (1-x)^1 + a_n x^n \\ &= \sum_{k=0}^n a_k x^k (1-x)^{n-k} \end{aligned}$$

is a linear combination of monomials and hence a polynomial. All you need to do is “multiply out” the  $x^k (1-x)^{n-k}$  terms and then regroup the resulting mess. The so called **Bernstein polynomials**<sup>119</sup>

$$p(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}$$

are of that form.

**Example 11.10** (Vector space of polynomials with sup-norm). Let  $A \subseteq \mathbb{R}$ . It follows from (11.28) and (11.27) that the set

$$\{p(\cdot) : p(\cdot) \text{ is a polynomial on } A\}$$

of all polynomials on an arbitrary nonempty subset  $A$  of the real numbers is a subspace of the vector space  $\mathcal{C}(A, \mathbb{R})$ . (see example (11.9) on p.238. If  $A$  is not bounded, the sup-norm

$$\|p\|_\infty = \sup\{|p(x)| : x \in A\}$$

is **not a real valued function** on the set of all polynomials on  $A$  as its value may be  $\infty$ . 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  $A$ . □

<sup>119</sup> Here  $f(\cdot)$  is a function, not necessarily continuous, on the unit interval  $[0, 1]$ . The binomial coefficient  $\binom{n}{k}$  is defined as  $\frac{n!}{k!(n-k)!}$  where  $0! = 1$  and  $n! = 1 \cdot 2 \cdot 3 \cdots n$  for  $n \in \mathbb{N}$  (see ch.4 of [1] B/G Art of Proof)



## 11.2.4 Uniform Continuity

It will be proved in theorem 12.12 (Uniform continuity on sequence compact spaces) on p.276 <sup>120</sup> that continuous real valued functions on the compact set  $[0, 1]$  are uniformly continuous in the sense of the following definition. <sup>121</sup>

**Definition 11.34** (Uniform continuity of functions). Let  $(X, d_1), (Y, d_2)$  be metric spaces and let  $A$  be a subset of  $X$ . A function

$$f(\cdot) : A \rightarrow Y \quad \text{is called **uniformly continuous**}$$

if for any  $\varepsilon > 0$  there exists a (possibly very small)  $\delta > 0$  such that

$$(11.67) \quad d_2(f(x) - f(y)) < \varepsilon \quad \text{for any } x, y \in A \text{ such that } d_1(x, y) < \delta. \quad \square$$

**Remark 11.21** (Uniform continuity vs. continuity). Note the following:

**a.** Condition (11.67) for uniform continuity looks very close to the  $\varepsilon$ - $\delta$  characterization of ordinary continuity (11.54) on p.233. 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  $\varepsilon$  and  $x_0$  according to which you must find a suitable  $\delta$ . In other words, for plain continuity

$$\delta = \delta(\varepsilon, x_0).$$

In the case of uniform continuity all you get is  $\varepsilon$ . You must come up with a suitable  $\delta$  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 neighborhoods, interior points and contact points, subspaces and continuity. Uniform continuity is not a concept that can be defined without a metric. <sup>122</sup>

□

**Example 11.11** (Uniform continuity of the identity mapping). Let us have another look at proposition(11.25) where we proved the continuity of the identity mapping on a metric space. We chose  $\delta = \varepsilon$  no matter what value of  $x$  we were dealing with and it follows that the identity mapping is always uniformly continuous. □

**Remark 11.22.** Now that you have learned the definitions for both continuity and uniform continuity, have another look at example 4.28, p.67 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. □

<sup>120</sup> see chapter 12.6 (Continuous Functions and Compact Spaces) on p.274

<sup>121</sup> For the special case of  $(X, d) = (\mathbb{R}, d_{|\cdot|})$  where  $d_{|\cdot|}(x, y) = |y - x|$ , see [1] Beck/Geoghegan, Appendix A.3, "Uniform continuity".

<sup>122</sup> Well, that's 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.

### 11.2.5 Continuity of Linear Functions (Understand this!)

**Lemma 11.1.** Let  $f : (V, \|\cdot\|) \rightarrow (W, |\cdot|)$  be a linear function between two normed vector spaces. Let

$$\begin{aligned} a &:= \sup\{ |f(x)| : x \in V, \|x\| = 1\}, \\ b &:= \sup\{ |f(x)| : x \in V, \|x\| \leq 1\}, \\ c &:= \sup\left\{ \frac{|f(x)|}{\|x\|} : x \in V, x \neq 0 \right\}. \end{aligned}$$

Then  $a = b = c$ .

*PROOF:* We introduce the following three sets for this proof:

$$\begin{aligned} A &:= \{ |f(x)| : x \in V, \|x\| = 1\}, \\ B &:= \{ |f(x)| : x \in V, \|x\| \leq 1\}, \\ C &:= \left\{ \frac{|f(x)|}{\|x\|} : x \in V, x \neq 0 \right\}. \end{aligned}$$

*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

$$|f(y)| = |f(x/\|x\|)| = (1/\|x\|) |f(x)| \geq |f(x)|.$$

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

$$|f(x)|/\|x\| = |f(x)/\|x\|| = |f(x/\|x\|)| = |f(y)|.$$

It follows that the sup over the bigger set  $C$  does not exceed the sup over  $A$ , hence  $c = b$ . ■

**Definition 11.35** (norm of linear functions). Let  $f : (V, \|\cdot\|) \rightarrow (W, |\cdot|)$  be a linear function between two normed vector spaces. We denote the quantity  $a = b = c$  from lemma 11.1 by  $\|f\|$ , i.e.,

$$\begin{aligned} \|f\| &= \sup\{ |f(x)| : x \in V, \|x\| = 1\} \\ &= \sup\{ |f(x)| : x \in V, \|x\| \leq 1\} \\ (11.68) \quad &= \sup\left\{ \frac{|f(x)|}{\|x\|} : x \in V, x \neq 0 \right\}. \end{aligned}$$

$\|f\|$  is called the **norm of  $f$** . <sup>123</sup>

We note that  $\|f\|$  need not be finite. □

<sup>123</sup> 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!**

**Theorem 11.15** (Continuity criterion for linear functions). Let  $f : (V, \|\cdot\|) \rightarrow (W, \|\cdot\|)$  be a linear function between two normed vector spaces. Then the following are equivalent.

- A.  $f$  is continuous at  $x = 0$ ,
- B.  $f$  is continuous in all points of  $V$ ,
- C.  $f$  is uniformly continuous on  $V$ ,
- D.  $\|f\| < \infty$ .

Moreover, we then have

$$(11.69) \quad \|f(x)\| \leq \|f\| \cdot \|x\| \quad \text{for all } x \in V.$$

PROOF: Clearly we have  $C \Rightarrow B \Rightarrow A$ . We now show  $A \Rightarrow D$ .

It follows from the continuity of  $f$  at 0 that there exists  $\delta > 0$  such that

$$(11.70) \quad \text{if } z \in V \text{ and } \|z\| < \delta \text{ then } \|f(z)\| = \|f(z) - f(0)\| < 1.$$

Let  $x \in V$  such that  $\|x\| \leq 1$ . Then  $\|\delta/2 \cdot x\| \leq \delta/2 < \delta$ , hence, according to (11.70),

$$\delta/2 \cdot \|f(x)\| = \|f(\delta/2 \cdot x)\| < 1, \quad \text{hence } \|f(x)\| < 2/\delta.$$

Because this last inequality is true for all  $x \in V$  with norm bounded by 1, it follows that

$$\|f\| = \sup\{\|f(x)\| : x \in V, \|x\| \leq 1\} < 2/\delta < \infty.$$

We have proved that  $A \Rightarrow D$ .

We finally show  $D \Rightarrow C$  and we do this in two steps.

First we show  $D \Rightarrow (11.69)$ . 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 (11.68).

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  $\|f(x) - f(y)\| < \varepsilon$ , then  $f$  is indeed uniformly continuous and the proof is done. We show this as follows.

$$\|f(x) - f(y)\| = \|f(x - y)\| \stackrel{(11.69)}{\leq} \|f\| \cdot \|x - y\| < \|f\| \cdot \delta = \|f\| \cdot \varepsilon/\|f\| = \varepsilon. \quad \blacksquare$$

**Note 11.3** ( $\|f\|$  is a norm). Let

$$(11.71) \quad \mathcal{C}_{\text{lin}}(V, W) := \mathcal{C}_{\text{lin}}((V, \|\cdot\|), (W, \|\cdot\|)) := \{f : V \rightarrow W : f \text{ is linear and continuous}\}.$$

Then  $\mathcal{C}_{\text{lin}}(V, W)$  is a vector space and

$$(11.72) \quad f \mapsto \|f\| = \sup\{\|f(x)\| : \|x\| = 1\}$$

defines a norm on  $\mathcal{C}_{\text{lin}}(V, W)$ .  $\square$

In all of the proof that is given now let  $A := \{x \in V : \|x\| = 1\}$ .

A. Proof that  $\mathcal{C}_{\text{lin}}(V, W)$  is a vector space.

Let  $f, g \in \mathcal{C}_{\text{lin}}(V, W)$ . We need to show that  $f+g \in \mathcal{C}_{\text{lin}}(V, W)$ , i.e.,  $f+g$  is both linear and continuous. Linearity is immediate. We now show continuity.

Let  $x \in A$ . Then

$$(11.73) \quad |f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\| < \infty.$$

The first inequality holds because the norm  $|f(x)|$  satisfies the triangle inequality for norms. The second follows from (11.68) on p.242, and the finiteness of  $\|f\| + \|g\|$  is, according to the continuity criterion for linear functions (thm.11.15 on p.243), equivalent to the continuity of both  $f$  and  $g$ .

We still must show that if  $f \in \mathcal{C}_{lin}(V, W)$  and  $\lambda \in \mathbb{R}$  then  $\lambda f : x \mapsto \lambda f(x) \in \mathcal{C}_{lin}(V, W)$ , i.e., we must show that this function is linear and continuous. Again, linearity is immediate. To show continuity we proceed as follows.

Let  $x \in A$ .  $\lambda \cdot \mathbf{1}$  is absolutely homogeneous. Hence

$$(11.74) \quad |\lambda f(x)| = |(\lambda \mathbf{1})f(x)| = |\lambda| \cdot |f(x)|.$$

It follows from prop.9.4 (positive homogeneity of the sup) on p.148 that

$$(11.75) \quad \|\lambda f\| = \sup\{| \lambda f(x) | : \|x\| = 1\} = \sup\{|\lambda| |f(x)| : \|x\| = 1\}$$

$$(11.76) \quad = |\lambda| \cdot \sup\{|f(x)| : \|x\| = 1\} = |\lambda| \cdot \|f\| < \infty.$$

This proves that  $\lambda f$  is continuous.

**B.** Proof that  $\|f\|$  is a norm on  $\mathcal{C}_{lin}(V, W)$ .

Because (11.73) is valid for all  $x \in A$ , we obtain

$$(11.77) \quad \|f + g\| = \sup\{|f(x) + g(x)| : x \in A\} \leq \|f\| + \|g\|.$$

This proves the triangle inequality.

Likewise, we obtain from the validity of (11.75) for all  $x \in A$ ,

$$(11.78) \quad \|\lambda f\| = \sup\{|\lambda| |f(x)| : x \in A\} = |\lambda| \sup\{|f(x)| : x \in A\} = |\lambda| \|f\|.$$

This proves absolute homogeneity.

Finally we show positive definiteness. Clearly  $\|f\|$  is non-negative as the sup of non-negative numbers  $|f(x)|$ . Assume that  $\|f\| > 0$ . Then  $\delta := \frac{1}{2}\|f\| > 0$  and there exists  $x_0 \in A$  such that

$$(11.79) \quad \sup\{|f(x)| : x \in A\} - |f(x_0)| < \delta, \text{ i.e., } \|f\| - |f(x_0)| < \delta, \text{ hence } |f(x_0)| > \delta.$$

Positive definiteness of  $\mathbf{1} \cdot \mathbf{1}$  implies that  $f(x_0) \neq 0$  and hence  $f \neq 0$ . We have proved positive definiteness of  $\|\cdot\|$ . ■

## 11.3 Function Sequences and Infinite Series

### 11.3.1 Convergence of Function Sequences (Study this!)

**Notation Alert:**

This chapter makes heavy use of the notation  $f(\cdot)$  instead of  $f$  for a function  $X \rightarrow \mathbb{R}$  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 grouping of a finite number  $n$  of real numbers. Matter of fact, any such vector  $(x_1, x_2, x_3, \dots, x_n)$  can be interpreted as a real valued function (remember: a real valued function is one which maps its arguments into  $\mathbb{R}$ )

$$(11.80) \quad f(\cdot) : \{1, 2, 3, \dots, n\} \rightarrow \mathbb{R} \quad j \mapsto x_j$$

(see (10.4) on p.175).

Next come sequences  $(x_j)_{j \in \mathbb{N}}$  which can be interpreted as real valued functions

$$(11.81) \quad g(\cdot) : \mathbb{N} \rightarrow \mathbb{R} \quad j \mapsto x_j.$$

Finally we deal with any kind of real valued function

$$(11.82) \quad h(\cdot) : X \rightarrow \mathbb{R} \quad x \mapsto h(x)$$

as the most general case.

Now we add more complexity by not just dealing with one or two or three real valued functions but with an entire sequence of functions

$$(11.83) \quad f_n(\cdot) : X \rightarrow \mathbb{R} \quad x \mapsto f_n(x)$$

For any fixed argument  $x_0$  we have a sequence  $f_1(x_0), f_2(x_0), f_3(x_0), \dots$  of real numbers which we can examine for convergence. This sequence may converge for some or all arguments  $x_0 \in X$  to some  $L = L(x_0) \in \mathbb{R}$ . It's time now for some definitions.

**Definition 11.36** (Pointwise convergence of function sequences). Let  $X$  be a nonempty set,  $(Y, d)$  a metric space and let  $f_n(\cdot) : X \rightarrow Y$  and  $f(\cdot) : X \rightarrow Y$  be functions on  $X$  ( $n \in \mathbb{N}$ ). Let  $A \subseteq X$  be a subset of  $X$ . We say that  $f_n(\cdot)$  **converges pointwise** or, simply, **converges** to  $f(\cdot)$  on  $A$  and we write  $f_n(\cdot) \rightarrow f(\cdot)$  on  $A$  if

$$(11.84) \quad f_n(x) \rightarrow f(x) \quad \text{for all } x \in A. \quad \square$$

We omit the phrase "on  $A$ " if it is clear how  $A$  is defined, in particular if  $A = X$ .

**Definition 11.37** (Uniform convergence of function sequences). Let  $X$  be a nonempty set,  $(Y, d)$  a metric space and let  $f_n(\cdot) : X \rightarrow Y$  and  $f(\cdot) : X \rightarrow Y$  be functions on  $X$  ( $n \in \mathbb{N}$ ). Let  $A \subseteq X$  be a subset of  $X$ . We say that  $f_n(\cdot)$  **converges uniformly** to  $f(\cdot)$  on  $A$  and we write <sup>124</sup>

$$(11.85) \quad f_n(\cdot) \xrightarrow{uc} f(\cdot) \text{ on } A$$

if the following is true: For each  $\varepsilon > 0$  (no matter how small) there exists a (possibly huge) number  $n_0$  which can be chosen once and for all, independently of the specific argument  $x$ , such that

$$(11.86) \quad d(f_n(x), f(x)) < \varepsilon \quad \text{for all } x \in A \quad \text{and } n \geq n_0.$$

We omit the phrase "on  $A$ " if it is clear how  $A$  is defined, in particular if  $A = X$ .  $\square$

<sup>124</sup> I must confess that " $f_n(\cdot) \xrightarrow{uc} f(\cdot)$ " is a notation that I coined myself because it is not as tedious as writing " $f_n(\cdot) \rightarrow f(\cdot)$  uniformly"

**Remark 11.23** (Uniform convergence implies pointwise convergence). Take another look at definition def.11.11 (convergence of sequences in metric spaces) on p.213. Note that (11.86) implies, for any given  $x \in A$ , ordinary convergence  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ . The reason is that the number  $n_0 = n_0(\varepsilon)$  chosen in (11.86) will also satisfy (11.20) of that definition for  $x_n = f_n(x)$  and  $a = f(x)$ .

In other words, uniform convergence implies pointwise convergence. But what is the difference between pointwise and uniform convergence? The difference is that, for pointwise convergence, the number  $n_0$  will depend on both  $\varepsilon$  and  $x$ :  $n_0 = n_0(\varepsilon, x)$ . In the case of uniform convergence, the number  $n_0$  will still depend on  $\varepsilon$  but can be chosen independently of the argument  $x \in A$ .  $\square$

**Example 11.12** (a. Constant sequence of functions). Let  $X$  be a set and let  $f : X \rightarrow \mathbb{R}$  be a real valued function on  $X$  which may or may not be continuous anywhere. Define a sequence of functions

$$f_n : X \rightarrow \mathbb{R} \quad (n \in \mathbb{N}) \quad \text{as} \quad f_1 = f_2 = \cdots = f$$

i.e.,

$$f_1(x) = f_2(x) = \cdots = f(x) \quad \forall n \in \mathbb{N}, \quad \forall x \in X.$$

In other words, we are looking at a constant sequence of functions (not to be confused with a sequence of constant functions – seriously!).

Then  $f_n(\cdot) \xrightarrow{uc} f(\cdot)$ .  $\square$

*PROOF of the example: This is trivial. No matter how small an  $\varepsilon$  and  $n_0$  we choose and no matter what argument  $x \in X$  we are looking at, we have*

$$|f_n(x) - f(x)| = 0 < \varepsilon \quad \text{for all } x \in A \quad \text{and } n > n_0 \quad \blacksquare$$

**Example 11.13** (b. Pointwise but not uniformly convergent sequence of functions). Let  $X = [0, 1]$ , i.e.,  $X$  is the closed unit interval  $\{x \in \mathbb{R} : 0 \leq x \leq 1\}$ . Let the functions  $f_n$  be defined as follows on  $X$ :

$$f_n(x) = \begin{cases} n^2 x & \text{for } 0 \leq x \leq \frac{1}{n} \\ \frac{1}{x} & \text{for } \frac{1}{n} \leq x \leq 1 \end{cases}$$

Note that both pieces 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 “ $n^2 x$  definition” gives the same value  $n = n^2 \frac{1}{n}$ . We do not give a formal proof that  $f_n(\cdot)$  is continuous at every point of  $[0, 1]$  but you should draw a picture of  $f_n$  to see that the two graphs flow into each other at the “splicing point”  $1/n$ . Instead you are asked in exercise 11.26 on p.262 to give such a proof.

Now we define the function  $f(\cdot) : [0, 1] \rightarrow \mathbb{R}$  as

$$f(x) = \begin{cases} \frac{1}{x} & \text{for } 0 < x \leq 1 \\ 0 & \text{for } x = 0 \end{cases}$$

Then the functions  $f_n(\cdot)$  converge pointwise but not uniformly to  $f(\cdot)$  on the entire unit interval.  $\square$

*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 zeroes 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 prop.9.7 on p.152. We have thus proved pointwise convergence.

*Proof that there is no uniform convergence:*

To prove that (11.86) 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

$$|f_{2N}(x_N) - f_N(x_N)| = 3.$$

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  $|f_j(x_N) - f(x_N)| > \varepsilon$ : we chose

$$\varepsilon = 2, \quad j = 2N, \quad x_N = \frac{1}{N^2}$$

We have proved that convergence is pointwise but not uniform. ■

**Proposition 11.29** (Uniform convergence is  $\|\cdot\|_\infty$  convergence). Let  $X$  be a nonempty set and  $\mathcal{B}(X, \mathbb{R})$  the set of all bounded real valued functions on  $X$ . We remember that this set is a vector space with the norm  $\|f\|_\infty = \sup\{|g(x) - f(x)| : x \in X\}$  and it is a metric space with the corresponding metric

$$d_{\|\cdot\|_\infty}(f, g) = \sup\{|g(x) - f(x)| : x \in X\}$$

(see example 11.2 on p.206). The following is true:

$$f_n(\cdot) \xrightarrow{uc} f(\cdot) \quad \text{if and only if} \quad f_n(\cdot) \xrightarrow{\|\cdot\|_\infty} f(\cdot).$$

In other words, the sequence  $f_n$  converges to  $f$  uniformly if and only if it converges to  $f$  in the metric space  $(\mathcal{B}(X, \mathbb{R}), d_{\|\cdot\|_\infty}(\cdot, \cdot))$ .

PROOF of “ $\Rightarrow$ ”: Assume that  $f_n(\cdot) \xrightarrow{uc} f(\cdot)$ . Let  $\varepsilon > 0$ . According to def.11.37 (Uniform convergence of function sequences) on p.245, 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 def.11.37 is  $\mathbb{R}$ , so  $d(f_n(x), f(x))$  becomes  $|f_n(x) - f(x)|$ . We obtain

$$\|f_n - f\|_\infty = \sup\{|f_n(x) - f(x)| : x \in X\} \leq \varepsilon/2 \quad \text{for all } n \geq n_0,$$

i.e.,  $d_{\|\cdot\|_\infty}(f_n, f) < \varepsilon$  for all  $n \geq n_0$ . It follows that  $f_n(\cdot) \xrightarrow{\|\cdot\|_\infty} f(\cdot)$ .

PROOF of “ $\Leftarrow$ ”: Assume that  $f_n \xrightarrow{\|\cdot\|_\infty} f$ , i.e.,  $\lim_{n \rightarrow \infty} f_n = f$  in the metric space  $(\mathcal{B}(X, \mathbb{R}), d_{\|\cdot\|_\infty})$ .

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  $f_n(\cdot) \xrightarrow{uc} f(\cdot)$ . ■

The last proposition justifies the next definition.

**Definition 11.38** (Norm and metric of uniform convergence). We also call the sup-norm on  $\mathcal{B}(X, \mathbb{R})$  the **norm of uniform convergence** on  $X$  and its associated metric  $d_{\|\cdot\|_\infty}(\cdot, \cdot)$  the **metric of uniform convergence** on  $X$ . □

**Theorem 11.16** (Uniform limits of continuous functions are continuous). Let  $(X, d_1)$  and  $(Y, d_2)$  be metric spaces and let  $f_n(\cdot) : X \rightarrow Y$  and  $f(\cdot) : X \rightarrow Y$  be functions on  $X$  ( $n \in \mathbb{N}$ ). Let  $x_0 \in X$  and let  $V \subseteq X$  be a neighborhood of  $x_0$ . Assume **a**) that the functions  $f_n(\cdot)$  are continuous at  $x_0$  for all  $n$  and **b**) that  $f_n(\cdot) \xrightarrow{uc} f(\cdot)$  on  $V$ . Then  $f$  is continuous at  $x_0$ .

PROOF: Let  $\varepsilon > 0$ .

**A.** Uniform convergence  $f_n(\cdot) \xrightarrow{uc} f(\cdot)$  on  $V$  guarantees the existence of some  $N = N(\varepsilon)$  such that

$$d_2(f_n(x), f(x)) < \frac{\varepsilon}{3} \quad \text{for all } x \in V \text{ and } n \geq N.$$

In particular, for  $n = N$ ,

$$(11.87) \quad d_2(f_N(x), f(x)) < \frac{\varepsilon}{3} \quad \text{for all } x \in V.$$

**B.** All functions  $f_n$  and in particular  $f_N$  are continuous in  $V$ . There is  $\tilde{\delta} > 0$  such that

$$(11.88) \quad d_2(f_N(x), f_N(x_0)) < \frac{\varepsilon}{3} \quad \text{for all } x \in N_{\tilde{\delta}}(x_0).$$

**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  $\delta$  be the smaller of  $\tilde{\delta}$  and  $\hat{\delta}$ .

Then (11.87) and (11.88) 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. ■

### 11.3.2 Infinite Series

We start by repeating the definition of a sequence given in section 5.2 on p.85: A **sequence**  $(x_j)$  is nothing but a family of things  $x_j$  which are indexed by integers, usually the natural numbers or the non-negative integers. We make throughout this entire document the following



**Assumption 11.1** (indices of sequences).

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.  $\square$

**Proposition 11.30** (Convergence criteria for series). *A series  $s := \sum a_k$  of real numbers converges if and only if for all  $\varepsilon > 0$  there exists  $n_0 \in \mathbb{N}$  such that one of the following is true:*

$$(11.89a) \quad \left| \sum_{k=n}^{\infty} a_k \right| < \varepsilon \quad \text{for all } n \geq n_0$$

$$(11.89b) \quad \left| \sum_{k=n}^m a_k \right| < \varepsilon \quad \text{for all } m, n \geq n_0$$

*PROOF:* Write

$$(11.90) \quad s = \sum_{k=1}^{\infty} a_k = \sum_{k=1}^n a_k + \sum_{k=n+1}^{\infty} a_k = s_n + \sum_{k=n+1}^{\infty} a_k$$

Remember the convergence criteria for numeric sequences. Convergence of a sequence  $(s_n)$  to a real number  $s$  means that, for any  $\varepsilon > 0$ , all but finitely many members  $s_n$  will be inside the  $\varepsilon$ -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

$$|s - s_n| < \varepsilon \quad \text{for all } n \geq n_0$$

(see (11.11) on p.213). According to (11.90) we can write that as

$$\left| \sum_{k=n+1}^{\infty} a_k \right| < \varepsilon \quad \text{for all } n \geq n_0,$$

which is the same as (11.89.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

$$|s_i - s_j| < \delta \quad \text{for all } i, j \geq n_0$$

(see (11.28) on p.226) which ensures convergence to some number  $s$  without specifying what it might actually be. Again we use (11.90) and obtain, assuming without loss of generality that  $i < j$ ,

$$\left| \sum_{k=i+1}^j a_k \right| < \delta \quad \text{for all } j > i \geq n_0 \quad \blacksquare$$

**Corollary 11.1.** *If a series  $\sum a_j$  converges then  $\lim_{n \rightarrow \infty} a_n = 0$ .*

*PROOF:* Let  $\varepsilon > 0$ . It follows from 11.89b that there is some  $n_0 \in \mathbb{N}$  such that  $|a_m - 0| = \left| \sum_{k=m}^m a_k \right| < \varepsilon$  for all  $m \geq n_0$ . But this means that the sequence  $a_n$  converges to zero.  $\blacksquare$

Here is a second corollary.

**Corollary 11.2** (Dominance criterion <sup>125</sup>). 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.

In particular, if  $|b_k| \leq a_k$  for all  $k \in \mathbb{N}$  then  $\left| \sum_{k=1}^{\infty} b_j \right| \leq \sum_{k=1}^{\infty} a_j$

PROOF: Let  $\varepsilon > 0$ . It follows from 11.89b that there is some  $n_0 \in \mathbb{N}$  such that  $\left| \sum_{k=m}^n a_k \right| < \varepsilon$  for all  $m, n \geq n_0$ . Let  $M := \max(n_0, N)$ . We obtain

$$\left| \sum_{k=i+1}^j b_j \right| \leq \sum_{k=i+1}^j |b_j| \leq \sum_{k=i+1}^j a_j < \varepsilon \quad \text{for all } j > i \geq M.$$

We conclude from (11.89b) that  $\sum b_k$  converges.

Now assume that  $|b_k| \leq a_k$  for all  $k \in \mathbb{N}$ . Let

$$s_n := \sum_{k=1}^n |a_k|, \quad s := \lim_{n \rightarrow \infty} s_n, \quad t_n := \sum_{k=1}^n b_k, \quad t := \lim_n t_n.$$

It follows from the triangle inequality for real numbers that  $|t_n| \leq s_n$  for all  $n \in \mathbb{N}$ . We apply cor.9.4 on p.167 to deduce that  $|t| = \lim_n |t_n| \leq \lim_n s_n = s$ . This completes the proof. ■

**Remark 11.24.** It is very important to remember that a series either converges to a finite number or it diverges. If it diverges it may be the case that  $\sum_{k=1}^{\infty} a_k = \infty$  or  $\sum_{k=1}^{\infty} a_k = -\infty$  or there is no limit at all. As an example for a series which has no limit, look at the oscillating sequence

$$(11.91) \quad a_0 = 1; \quad a_1 = -1; \quad a_2 = 1; \quad a_3 = -1; \dots \quad s_n = \sum_{k=0}^n (-1)^k$$

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  $s_0 = s_2 = s_4 = \dots = s_{2j} = 0$ .

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. <sup>126</sup>

Just so you understand the difference between limits and contact points (see (def.11.25) on p.222): 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.

<sup>125</sup> This is a generalization of [1] B/G (Beck/Geoghegan) prop.12.3, p.115.

<sup>126</sup> 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 (11.89a) on p.249) is not satisfied, hence no convergence of the series.

**Definition 11.39** (Finite permutations). Let  $N \in \mathbb{N}$  and let  $[N] := \{1, 2, 3, \dots, N\}$  denote the set of the first  $N$  integers.<sup>127</sup> A **permutation** of  $[N]$  is a mapping

$$\pi(\cdot) : [N] \rightarrow [N]; \quad j \mapsto \pi(j)$$

which is both surjective: each element  $k$  of  $[N]$  is the image  $\pi(j)$  for a suitable  $j \in [N]$  and injective: different arguments  $i \neq j \in [N]$  will always map to different images  $\pi(i) \neq \pi(j) \in [N]$  (see (5.9) on p.93). Remember that

$$\text{surjective} + \text{injective} = \text{bijective}$$

and that under our assumptions the inverse mapping

$$\pi^{-1}(\cdot) : [N] \rightarrow [N]; \quad \pi(j) \mapsto \pi^{-1}\pi(j) = j,$$

which associates with each image  $\pi(j)$  the unique argument  $j$  which maps into  $\pi(j)$ , exists (see def. 5.9 on p.93 for properties of the inverse mapping).

It is customary to write

$$i_1 \text{ instead of } \pi(1), \quad i_2 \text{ instead of } \pi(2), \quad \dots, \quad i_j \text{ instead of } \pi(j), \quad \dots \quad \square$$

**Definition 11.40** (Permutations of  $\mathbb{N}$ ). A **permutation** of  $\mathbb{N}$  is a bijective function

$$\pi(\cdot) : \mathbb{N} \rightarrow \mathbb{N}; \quad j \mapsto \pi(j). \quad \square$$

*Permutations are the means of describing a **A rearrangement** or **reordering** of the members of a finite or infinite sequence or series. Look at any sequence  $(a_j)$ . Given a permutation  $\pi(\cdot)$  of the natural numbers, we can form the sequence  $(b_k) := (a_{\pi(k)})$ , i.e.,*

$$b_1 = a_{\pi(1)}, \quad b_2 = a_{\pi(2)}, \quad \dots, \quad b_k = a_{\pi(k)}, \quad \dots$$

*We can use the inverse permutation,  $\pi^{-1}(\cdot)$ , to regain the  $a_j$  from the  $b_j$  because*

$$b_{\pi^{-1}(k)} = a_{\pi^{-1}(\pi(k))} = a_k.$$

**Proposition 11.31.** *Let  $(a_n)$  be a sequence of non-negative members:  $a_n \geq 0$  for all  $n \in \mathbb{N}$ . Then exactly one of the following is true:*

**a.** *the series  $\sum a_n$  converges (to a finite number). In that case*

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\pi(n)} \quad \text{for **any** permutation } \pi(\cdot) \text{ of } \mathbb{N}.$$

**b.** *the series  $\sum_{n=1}^{\infty} a_n$  has limit  $\infty$ . In that case it is true for **any** permutation  $\pi(\cdot)$  of  $\mathbb{N}$  that the reordered series  $\sum_{n=1}^{\infty} a_{\pi(n)}$  also has limit  $\infty$ .*

---

<sup>127</sup> This notation was copied from chapter 13 (Cardinality) of [1] B/G (Beck/Geoghegan). It has nothing to do with equivalence classes!

PROOF of **a**: Let  $b_j := a_{\pi(j)}$  and, hence,  $a_k = b_{\pi^{-1}(k)}$ . Let  $N \in \mathbb{N}$ . Let

$$(11.92) \quad \alpha := \max\{\pi(j) : j \leq N\} \quad \text{and} \quad \beta := \max\{\pi^{-1}(k) : k \leq N\}.$$

Note that  $\alpha \geq N$  and  $\beta \geq N$ . Because all terms  $a_j, b_k$  are non-negative it follows that

$$\begin{aligned} \sum_{j=1}^N b_j &= \sum_{j=1}^N a_{\pi(j)} \leq \sum_{k=1}^{\alpha} a_k \leq \sum_{k=1}^{\alpha} a_k + \sum_{k=\alpha+1}^{\infty} a_k = \sum_{k=1}^{\infty} a_k, \\ \sum_{k=1}^N a_k &= \sum_{k=1}^N b_{\pi^{-1}(k)} \leq \sum_{j=1}^{\beta} b_j \leq \sum_{j=1}^{\beta} b_j + \sum_{j=\beta+1}^{\infty} b_j = \sum_{j=1}^{\infty} b_j. \end{aligned}$$

We take limits as  $N \rightarrow \infty$  and it follows from cor.9.4 on p.167 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 non-negative, the sequence  $s_n$  of the partial sums is non-decreasing and hence has a limit  $s$ .  $s \notin \mathbb{R}$  because we assumed that  $\sum a_j$  is not convergent and we can rule out  $s = -\infty$  because  $s \geq a_0 \geq 0$ . 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 11.41** (absolutely convergent series). A series  $\sum a_j$  is **absolutely convergent** if the corresponding series  $\sum |a_j|$  of its absolute values converges. □

**Proposition 11.32.** Let  $\sum a_k$  be an absolutely convergent series. Then  $\sum a_k$  converges and

$$(11.93) \quad \left| \sum_{k=1}^{\infty} a_k \right| \leq \sum_{k=1}^{\infty} |a_k|.$$

PROOF: This follows from the dominance criterion (cor.11.2) ■

It follows from prop.11.31 on p.251 that if a series of non-negative 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 <sup>128</sup> that the reverse is also true: Any series whose value is invariant under rearrangements is absolutely convergent.

**Theorem 11.17.** <sup>129</sup> Let  $\sum a_k$  be an absolutely convergent series. Let  $\pi : \mathbb{N} \rightarrow \mathbb{N}$  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  $\sum a_k$ . <sup>130</sup>

<sup>128</sup> see cor.11.4 on p.259

<sup>129</sup> This was proved by the German mathematician Peter Gustav Lejeune Dirichlet (1805-1859).

<sup>130</sup>  $\sum a_k$  converges according to prop.11.32.

PROOF: Let  $\varepsilon > 0$ . Since  $\sum |a_k|$  converges, there exists  $n_0 \in \mathbb{N}$  such that

$$(11.94) \quad \sum_{k=n_0+1}^{n_0+m} |a_k| \leq \sum_{k=n_0+1}^{\infty} |a_k| < \varepsilon \quad \text{for all } m \in \mathbb{N}.$$

For  $n \in \mathbb{N}$  let  $s_n := \sum_{k=1}^n a_k$  and  $t_n := \sum_{k=1}^n b_k$ .

Let  $A := \{\pi(j) : 1 \leq j \leq n_0\}$  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  $t_p$ .

We recall that  $p_0$  was chosen so big that each of  $a_1, \dots, a_{n_0}$  is one of  $b_1, \dots, b_{p_0}$ .

It follows from all this that each of  $a_1, \dots, a_{n_0}$  is a term both of  $s_n$  and  $t_p$ , hence none of those terms appears in the difference  $s_n - t_p$ . We obtain for big enough  $m \in \mathbb{N}$  (the bigger of  $\max(\{\pi(j) : 1 \leq j \leq n\})$  and  $p$ )

$$|s_n - t_p| \leq \sum_{k=n_0+1}^{n_0+m} |a_k| < \varepsilon.$$

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 (11.94) that  $|s - s_n| < \varepsilon$ , hence  $|s - t_p| < 2\varepsilon$ .

We remember how  $p$  was defined and see that it could be any integer  $\geq p_0$ , a number which depends (via  $n_0$ ) only on  $\varepsilon$ .

To summarize: for all  $\varepsilon > 0$  there exists  $p_0$  such that  $p \geq p_0$  implies  $|s - t_p| < 2\varepsilon$ . But then  $\lim_{p \rightarrow \infty} t_p = s$ .

On the other hand,  $\lim_{p \rightarrow \infty} t_p = t = \sum_{k=1}^{\infty} b_k$ .

This concludes the proof that  $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} b_k$ . ■

There are series which are convergent but not absolutely convergent. Such series are given a special name:

**Definition 11.42** (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 11.43** (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 11.33** (Leibniz test for alternating series). Let  $\sum a_k$  be an alternating series such that the sequence  $|a_k|$  is non-increasing:  $|a_1| \geq |a_2| \geq |a_3| \geq \dots$ . Then  $\sum a_k$  converges.

PROOF: Left as an exercise. ■

**Example 11.14** (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:  $\sum_{j=1}^{\infty} \frac{1}{n} = \infty$ . On the other hand, according to the Leibniz test,  $\sum(-1)^n/n$  converges. It follows that the alternating harmonic series is convergent but not absolutely convergent, i.e., it is conditionally convergent. □

We are going to prove Riemann's Reordering Theorem, from which it can be easily deduced that if  $\sum a_j$  is conditionally convergent and  $x \in \mathbb{R}$ , a rearrangement  $\sum a_{\pi_j}$  can be found which converges to  $x$ . In preparation we must prove a lemma.

**Lemma 11.2.** Let  $\sum a_k$  be a series. We split it into two series  $\sum p_k$  and  $\sum q_k$  as follows.

$p_j$  is the  $j$ th strictly positive member of the sequence  $(a_k)_k$  and  $q_j$  is the  $j$ th strictly negative member of that sequence.

The following is true:

- a. If  $\sum a_k$  is absolutely convergent then both  $\sum p_k$  and  $\sum q_k$  are (absolutely) convergent.
- b. If  $\sum a_k$  is convergent but not absolutely convergent then  $\sum p_k$  has limit  $\infty$  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  $j$ th (not  $m$ th!) 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  $\sum_{i=1}^{\infty} p_i < \infty$ . 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 non-decreasing, 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

$$a_{M_n} = p_n \quad (\star) \quad \text{and} \quad M_n \geq n. \quad (\star\star)$$

Let

$$I_n := \{i \leq M_n : a_i > 0\}, \quad J_n := \{j \leq M_n : a_j < 0\}.$$

Then

$$(11.95) \quad \sum_{k=1}^{M_n} a_k = \sum_{i \in I_n} a_k + \sum_{j \in J_n} a_k \stackrel{(\star)}{=} \sum_{i=1}^n p_i + \sum_{j \in J_n} a_k \geq \sum_{i=1}^n p_i + \sum q_k.$$

It follows from  $(\star\star)$  that if  $n \rightarrow \infty$  then  $M_n \rightarrow \infty$ . Hence, from (11.95),

$$\sum a_k = \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k = \lim_{n \rightarrow \infty} \sum_{k=1}^{M_n} a_k \geq \lim_{n \rightarrow \infty} \left( \sum_{i=1}^n p_i + \sum q_k \right) = \infty,$$

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 11.18** (Riemann's Rearrangement Theorem). <sup>131</sup>

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

$$\liminf_{n \rightarrow \infty} \sum_{k=1}^n b_k = \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} \sum_{k=1}^n b_k = \beta.$$

PROOF: We may assume that  $a_j \neq 0$  for all  $j \in \mathbb{N}$  because those terms do not contribute anything to the partial sums, hence leave the limit of the series and any rearrangement unchanged.

We split  $\sum a_j$  into the series  $\sum p_j$  of its positive members and  $\sum q_j$  of its negative members in the same way as was done in lemma 11.2:

$p_j$  is the  $j$ th strictly positive member of the sequence  $(a_k)_k$ ;

$q_j$  is the  $j$ th strictly negative member of  $(a_k)_k$ .

It was proved in lemma 11.2 that  $\sum_{k=1}^{\infty} p_k = \infty$  and  $\sum_{k=1}^{\infty} q_k = -\infty$ .

**case 1:**  $\beta \geq 0$ .

Let  $U_1 := \{k \in \mathbb{N} : p_1 + p_2 + \dots + p_k > \beta\}$ .  $U_1$  is not empty because  $\sum p_j$  has limit  $\infty$ , 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  $\sum b_k$ .

<sup>131</sup> This was proved by the German mathematician Bernhard Riemann (1826-1866).

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  $\beta$ .

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  $\alpha$ .

Let  $U_2 := \left\{k \in \mathbb{N} : k > u_1 \text{ and } \sigma_2 + \sum_{j=u_1+1}^{u_1+k} p_j > \beta\right\}$ .  $U_2$  is not empty because  $\sum_{j=u_1+1}^{\infty} p_j$  has limit  $\infty$ , hence  $u_2 := \min(U_2)$  exists. We call  $p_{u_1+1}, p_{u_1+2}, \dots, p_{u_2}$  the **second upcrossing** of  $\sum b_k$ .

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  $\beta$ .

Let  $L_2 := \left\{k \in \mathbb{N} : k > l_1 \text{ and } \sigma_3 + \sum_{j=l_1+1}^{l_1+k} q_j < \alpha\right\}$ .  $L_2$  is not empty because  $\sum_{j=l_1+1}^{\infty} q_j$  has limit  $-\infty$ , hence  $l_2 := \min(L_2)$  exists. We call  $q_{l_1+1}, q_{l_1+2}, \dots, q_{l_2}$  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  $\alpha$ .

It should be clear how we proceed. Let us assume that we have constructed the  $N$ th upcrossing  $p_{u_{N-1}+1}, p_{u_{N-1}+2}, \dots, p_{u_N}$  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  $N$ th downcrossing



$q_{l_{N-1}+1}, q_{l_{N-1}+2}, \dots, q_{l_N}$  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  $U_{N+1} := \left\{ k \in \mathbb{N} : k > u_N \text{ and } \sigma_{2N} + \sum_{j=u_N+1}^{u_N+k} p_j > \beta \right\}$ .  $U_{N+1}$  is not empty because  $\sum_{j=u_N+1}^{\infty} p_j$  has limit  $\infty$ , hence  $u_{N+1} := \min(U_{N+1})$  exists. We call  $p_{(u_{N+1})}, p_{(u_{N+1}+2)}, \dots, p_{u_{(N+1)}}$  the  $(N+1)$ th upcrossing of  $\sum b_k$ .

We add more terms to  $b_1, b_2, \dots, b_{n_{2N}}$ .

$$\begin{aligned} n_{(2N+1)} &:= n_{(2N)} + u_{(N+1)}; \\ b_{(n_{(2N)}+1)} &:= p_{(u_{N+1})}, b_{(n_{(2N)}+2)} := p_{(u_{N+1}+2)}, \dots, b_{(n_{(2N+1)})} := p_{u_{(N+1)}}, \\ \sigma_{(2N+1)} &:= \sum_{j=1}^{n_{(2N+1)}} b_j. \end{aligned}$$

Let  $L_{N+1} := \left\{ k \in \mathbb{N} : k > l_N \text{ and } \sigma_{2N+1} + \sum_{j=l_N+1}^{l_N+k} q_j < \alpha \right\}$ .  $L_{N+1}$  is not empty because  $\sum_{j=l_N+1}^{\infty} q_j$  has limit  $\infty$ , hence  $l_{N+1} := \min(L_{N+1})$  exists. We call  $q_{(l_{N+1})}, q_{(l_{N+1}+2)}, \dots, q_{l_{(N+1)}}$  the  $(N+1)$ th downcrossing of  $\sum b_k$ .

We add more terms to  $b_1, b_2, \dots, b_{n_{(2N+1)}}$ .

$$\begin{aligned} n_{(2(N+1))} &:= n_{(2N+1)} + l_{(N+1)}; \\ b_{(n_{(2N+1)}+1)} &:= q_{(l_{N+1})}, b_{(n_{(2N+1)}+2)} := q_{(l_{N+1}+2)}, \dots, b_{(n_{(2(N+1))})} := q_{l_{(N+1)}}, \\ \sigma_{2(N+1)} &:= \sum_{j=1}^{n_{2(N+1)}} b_j. \end{aligned}$$

We have defined by recursion  $\sum_{k=1}^{n_N} b_k$  for all  $N \in \mathbb{N}$

We now show that the increasing sequence  $(n_N)_{N \in \mathbb{N}}$  is not bounded above. We observe that  $n_{(2N)}$  is the number of terms that belong to the first  $N$  upcrossings plus the first  $N$  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  $\alpha$  to above  $\beta$  and at least one term  $q_j$  is needed to move a partial sum from above  $\beta$  to below  $\alpha$ . Hence  $n_{2N} \geq 2N$  and this proves that the sequence  $(n_N)_{N \in \mathbb{N}}$  is indeed not bounded above.

It follows that  $\sum b_k$  indeed has infinitely many terms.

We note that all positive terms  $p_j$  and all negative terms  $q_j$  are being used in sequence, starting with the first one. 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.9.2 (Characterization of limsup and liminf) on p.164 that

$$(11.96) \quad \liminf_{n \rightarrow \infty} s_n \leq \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} s_n \geq \beta.$$

We now prove that for any  $\varepsilon > 0$

$$(11.97) \quad \liminf_{n \rightarrow \infty} s_n \geq \alpha - \varepsilon \quad \text{and} \quad \limsup_{n \rightarrow \infty} s_n \leq \beta + \varepsilon.$$

Let  $\varepsilon > 0$ . The terms  $(a_n)_n$  of the original series  $\sum a_k$  converge to zero because  $\sum a_k$  converges (see cor.11.1 on p.249). 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

$$(11.98) \quad |p_j| = p_j < \varepsilon \quad \text{and} \quad |q_j| = -q_j < \varepsilon \quad \text{for all } j \geq n_0.$$

$|p_j| = p_j < \varepsilon$  is true whenever  $j \geq n_0$  because  $p_j$  is the  $j$ th 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  $j$ th negative member of  $(a_n)_n$ , hence  $q_j = a_i$  for some  $i \geq j \geq n_0$ . We have proved (11.98).

We recall that  $n_1, n_3, n_5, \dots$  are precisely the integers  $n$  for which  $s_n > \beta$ , so

$$s_{(n_1-1)} \leq \beta, \quad s_{(n_3-1)} \leq \beta, \quad \dots, \quad s_{(n_{(2j-1)}-1)} \leq \beta, \quad \dots$$

But then  $s_{(n_{(2j-1)})} \leq \beta + \varepsilon$  because less than  $\varepsilon$  was added to the previous term (which is no bigger than  $\beta$ ) for any  $j$  so big that the last item in the  $j$ th upcrossing is less than  $\varepsilon$

It follows from (11.98) 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 (11.97) and this implies, together with (11.96), that

$$(11.99) \quad \liminf_{n \rightarrow \infty} s_n = \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} s_n = \beta.$$

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  $\alpha$  rather than an upcrossing to obtain a rearrangement  $\sum c_k$  for which a partial sum  $\sum_{j=1}^n a_j$  exceeds  $\alpha$  when  $n$  is the last term of an upcrossing and it drops below  $\beta$  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  $\varepsilon$ . For all others the partial sums cannot exceed  $\beta$  or drop below  $\alpha$  by more than  $\varepsilon$  and we conclude as before that

$$(11.100) \quad \liminf_{n \rightarrow \infty} \sum_{k=1}^n c_k = \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} \sum_{k=1}^n c_k = \beta. \quad \blacksquare$$

**Corollary 11.3.** 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

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n b_k = \alpha.$$

PROOF: We apply Riemann's Reordering Theorem to the special case  $\beta = \alpha$ : There is a rearrangement  $\sum b_j$  of  $\sum a_j$  such that

$$\liminf_{n \rightarrow \infty} \sum_{k=1}^n b_k = \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} \sum_{k=1}^n b_k = \alpha.$$

It follows now from thm.9.3 on p.166 that  $\sum b_j$  converges to  $\alpha$ . ■

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 11.4.** Let  $\sum a_k$  be a convergent series with limit  $\alpha \in \mathbb{R}$  such that each rearrangement  $\sum b_k$  also converges to  $\alpha$ .

Then  $\sum a_k$  is absolutely convergent.

PROOF: We assume to the contrary that  $\sum a_k$  is not absolutely convergent. This series is convergent but not absolutely convergent, hence conditionally convergent. We apply Riemann's Reordering Theorem and find that there is a rearrangement of  $\sum a_j$  which converges to a different real number, contrary to our assumption. ■

**Corollary 11.5** (Dichotomy for convergent series). Let series  $\sum a_k$  be a convergent series. Then either **a** or **b** is true:

- a. All rearrangements of  $\sum a_k$  converge to the same limit.
- b. For any  $\alpha \in \mathbb{R}$  there is a rearrangement of  $\sum a_k$  which converges to  $\alpha$ .

PROOF: Either  $\sum a_k$  is absolutely convergent and **a** is true according to Riemann's Reordering Theorem or the series it is conditionally convergent and **b** is true according to cor.11.3. ■

## 11.4 Exercises for Ch.11

### 11.4.1 Exercises for Ch.11.1

#### 11.4.1.1 Exercises for Ch.11.1.1 (Definition and Examples of Metric Spaces)

**Exercise 11.1.** Prove prop.11.1 on p.205: Let  $(X, d)$  be a metric space. Let  $n \in \mathbb{N}$  and  $x_1, x_2, \dots, x_n \in X$ . Then

$$(11.101) \quad d(x_1, x_n) \leq \sum_{j=1}^{n-1} d(x_j, x_{j+1}) = d(x_1, x_2) + d(x_2, x_3) + \dots + d(x_{n-1}, x_n). \quad \square$$

**Exercise 11.2.** Prove thm.11.1 (Norms define metric spaces) on p.206: Let  $(V, \|\cdot\|)$  be a normed vectors space. Then the function

$$(11.102) \quad d_{\|\cdot\|}(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}_{\geq 0}; \quad (x, y) \mapsto d_{\|\cdot\|}(x, y) := \|y - x\|$$

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. □

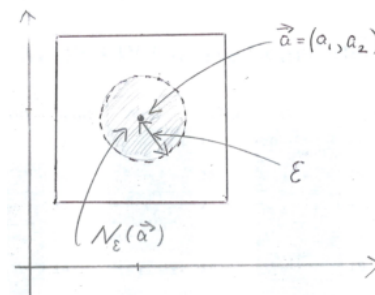
**11.4.1.2 Exercises for Ch.11.1.2** (Measuring the Distance of Real Valued Functions)

**11.4.1.3 Exercises for Ch.11.1.3** (Neighborhoods and Open Sets)

**Exercise 11.3.** Prove prop.11.5 on p.211: 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 11.4.** 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  $\varepsilon$  such that  $N_\varepsilon(\vec{a}) \subseteq A$



$\square$

**Exercise 11.5.** 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  $(\mathbb{R}, d_{|\cdot|})$ .  $\square$

**Exercise 11.6. 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 and the set of all irrational numbers.  $\square$

**11.4.1.4 Exercises for Ch.11.1.4** (Convergence)

**Exercise 11.7.** Given is a metric space  $(X, d)$ .

Prove the following: A sequence  $(x_n)$  of elements of  $X$  converges to  $a \in X$  as  $n \rightarrow \infty$  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 def.9.14 (Tail sets of a sequence) on p.159.)  $\square$

**Exercise 11.8.** Prove prop.11.7 on p.214:

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

$$\lim_{n \rightarrow \infty} x_n = L \Leftrightarrow \lim_{n \rightarrow \infty} y_n = L. \quad \square$$

**Exercise 11.9.** Prove prop.11.6 on p.214:

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  $\lim_{n \rightarrow \infty} (y_n)_n = L$ .  $\square$

**Exercise 11.10.** Let  $f_n, f \in \mathcal{B}([0, 1], \mathbb{R})$   $n \in \mathbb{N}$  be continuous such that  $f = \lim_{n \rightarrow \infty} f_n$  in  $(\mathcal{B}([0, 1], d_{\|\cdot\|_\infty})$  (!) Prove  $\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \int_0^1 f(x) dx$ . You must use the  $\varepsilon, N$  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  $\alpha := \min_{x \in [0, 1]} h(x), \beta := \max_{x \in [0, 1]} h(x)$ . Then  $\exists \lambda \in [\alpha, \beta]$  such that

$\lim_{n \rightarrow \infty} \int_0^1 h(x) dx = \lambda (= \lambda(1 - 0))$ . **c.** Use without proof that  $\left| \int_0^1 h(x) dx \right| \leq \int_0^1 |h(x)| dx$  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  $\alpha_n, \lambda_n, \beta_n$ ).

□

#### 11.4.1.5 Exercises for Ch.11.1.5 (Abstract Topological spaces)

**Exercise 11.11.** It was stated in example.11.5 on p.215 that the discrete topology which is induced by the discrete metric  $d(x, y) = 1$  if  $x \neq y$  and  $0$  if  $x = y$  is the entire powerset  $2^X$  of  $X$ . Prove it. □

#### 11.4.1.6 Exercises for Ch.11.1.6 (Bases and Neighborhood Bases)

**Exercise 11.12.** 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  $(X, \mathfrak{U}_d)$ . □

#### 11.4.1.7 Exercises for Ch.11.1.7 (Metric and Topological Subspaces)

**Exercise 11.13.** Prove prop.11.12 on p.221: 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 the definition def.11.12 on p.214 of an abstract topological space. □

#### 11.4.1.8 Exercises for Ch.11.1.8 (Bounded Sets and Bounded Functions)

#### 11.4.1.9 Exercises for Ch.11.1.9 (Contact Points and Closed Sets)

**Exercise 11.14.** Let  $A \subseteq \mathbb{R}$  be a closed, nonempty set which is bounded above. Prove that the maximum of  $A$  exists and that  $\sup(A) = \max(A)$ . □

**Exercise 11.15.** Prove prop.11.18 on p.225: Let  $(X, \mathfrak{U})$  be a topological space and  $A \subseteq X$ . Then  $\partial A = \bar{A} \cap \overline{A^c}$ . □

**Exercise 11.16.** Prove prop.11.16 on p.225: Let  $(X, \mathfrak{U})$  be a topological space and  $A \subseteq B \subseteq X$ . Then  $\bar{A} \subseteq \bar{B}$ . □

**Exercise 11.17.** Prove rem.11.10 on p.216: Let  $(X, \mathfrak{U})$  be a topological space and  $A \subseteq B \subseteq X$ . Then  $A^\circ \subseteq B^\circ$ . □

**Exercise 11.18.** Prove parts **c** and **d** of prop.11.20 (Closure of a set as a hull operator) on p.226: Let  $A$  be a subset of a topological space  $(X, \mathfrak{U})$ . Then **c.**  $\bar{\bar{A}} = \bar{A}$ , **d.**  $\overline{A \cup B} = \bar{A} \cup \bar{B}$ . □

### 11.4.1.10 Exercises for Ch.11.1.10 (Completeness in Metric Spaces)

**Exercise 11.19.** Let  $(X, d)$  be a metric space and  $A \subseteq X, A \neq \emptyset$ . Let

$$(11.103) \quad \gamma := \gamma(A) := \inf\{d(x, y) : x, y \in A \text{ and } x \neq y\}.$$

a. Prove that if  $\gamma > 0$  then  $A$  is complete.

b. The reverse is not true. Find a counterexample.  $\square$

**Exercise 11.20.** Let  $(X, d)$  be a metric space and let  $A \subseteq X$  be a finite subset. Prove that  $A$  is complete.  $\square$

**Exercise 11.21.** 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 11.22.** 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.  $\square$

### 11.4.2 Exercises for Ch.11.2

**Exercise 11.23.** Prove prop.11.8 (Opposite of continuity) on p.214:

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**  $j$ .  $\square$

**Exercise 11.24.** Prove prop.11.27 (All polynomials are continuous) on p.239.  $\square$

**Exercise 11.25.** Let  $n \in \mathbb{N}$  and let  $p(x) := \sum_{j=0}^n a_j x^j$  ( $a_j \in \mathbb{R}, a_n \neq 0$ ) 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  $x_n \rightarrow \infty$  or  $x_n \rightarrow -\infty$ . Examine  $y_k := \frac{p(x_k)}{a_n x_k^n}$  and use prop.9.11 (rules of arithmetic for limits).  $\square$

**Exercise 11.26.** In example 11.13 on p.246 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  $n \in \mathbb{N}$ .

$\square$

## 12 Compactness

### 12.1 Introduction: Closed and Bounded sets in Euclidean space

One of the results that are true for  $N$ -dimensional space is the “sequence compactness” of closed and bounded subsets: Any sequence that lives in such a set has a convergent subsequence. We will discuss that next.

**Theorem 12.1** (Convergent subsequences in closed and bounded sets of  $\mathbb{R}$ ). *Let  $A$  be a bounded and closed set of real numbers and let  $(z_n)$  be an arbitrary sequence in  $A$ . Then there exists  $z \in A$  and a subset*

$$n_1 < n_2 < \dots < n_j < \dots \text{ of indices such that } z = \lim_{j \rightarrow \infty} z_{n_j}$$

*i.e., the subsequence  $(z_{n_j})$  converges to  $z$ .*

*PROOF:* Let  $m$  be the midpoint between  $a := \inf(A)$  and  $b := \sup(A)$ . Because  $A$  is bounded,  $a$  and  $b$  must exist as finite numbers. Let

$$(12.1) \quad A_{\star 1} := A \cap [a, m]; \quad A^{\star 1} := A \cap [m, b].$$

Then at least one of  $A_{\star 1}$ ,  $A^{\star 1}$  must contain infinitely many of the  $z_n$  because  $A_{\star 1}$  and  $A^{\star 1}$  form a “covering” of  $A$  (the formal definition will be given later in def.12.4 on p.271), i.e.,  $A_{\star 1} \cup A^{\star 1} \supseteq A$ . We pick one with infinitely many elements and call it  $A_1$ . In case both sets contain infinitely many of the  $z_n$ , it does not matter which one we pick. We observe that  $\text{diam}(A_1) \leq \text{diam}(A)/2$

Let  $m_1$  be the midpoint between  $a_1 := \inf(A_1)$  and  $b_1 := \sup(A_1)$ . Let

$$(12.2) \quad A_{\star 2} := A_1 \cap [a_1, m_1]; \quad A^{\star 2} := A_1 \cap [m_1, b_1].$$

Then at least one of  $A_{\star 2}$ ,  $A^{\star 2}$  must contain infinitely many of the  $z_n$ . We pick one with infinitely many elements and call it  $A_2$ . In case both sets contain infinitely many of the  $z_n$ , it does not matter which one we pick. Note that

$$\text{diam}(A_2) \leq \text{diam}(A_1)/2 \leq \text{diam}(A)/2^2$$

We keep picking the midpoints  $m_j$  of the sets  $A_j$  each of which has at most half the diameter of the previous one. (Why?) In other words, we have constructed a sequence

$$(12.3) \quad A \supseteq A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots \quad \text{such that} \\ \text{diam}(A) \geq 2\text{diam}(A_1) \geq 2^2\text{diam}(A_2) \dots \geq 2^n\text{diam}(A_n) \dots$$

which means that  $\text{diam}(A_n) \leq \text{diam}(A)/2^n \rightarrow 0$  as  $n \rightarrow \infty$ .

We pick a subsequence  $(x_j)_j = (z_{n_j})_j$  of the original sequence  $(z_n)_n$  such that  $z_{n_j} \in A_j$  for all  $j \in \mathbb{N}$ . This is possible because the sets  $A_j$  were picked in such a way that each one of them contains infinitely many of the  $z_k$ .

The sequence of sets  $(A_j)$  is “nested”:  $A_j \subseteq A_{j-1}$  for each index  $j$ . It follows that  $A_m$  contains all  $A_k$  for any  $k > m$  and this implies that  $A_m$  contains all members  $x_k = z_{n_k}$ , for all  $k > m$ . Thus

$$|x_m - x_k| \leq \text{diam}(A_m) \leq \frac{\text{diam}(A)}{2^m} \quad \text{for all } m \text{ and } k \text{ such that } k > m.$$

It follows that  $(x_n)$  is a Cauchy sequence (p.226). According to theorem 11.8 about the completeness of  $\mathbb{R}$  (p.227) there is a contact point  $x$  such that  $x_n \rightarrow x$  for  $n \rightarrow \infty$ .

Because  $A$  is a closed set it contains all its contact points. It follows that  $x \in A$  and we have found a subsequence of the original sequence  $(z_n)$  which converges to an element of  $A$ . ■

**Theorem 12.2** (Convergent subsequences in closed and bounded sets of  $\mathbb{R}^N$ ). Let  $A$  be a bounded and closed set of  $\mathbb{R}^N$  and let  $(\vec{z}_n)$  be an arbitrary sequence of  $N$ -dimensional vectors in  $A$ . Then there exists  $\vec{z} \in A$  and a subset

$$n_1 < n_2 < \dots < n_j < \dots \text{ of indices such that } \vec{z} = \lim_{j \rightarrow \infty} \vec{z}_{n_j}$$

i.e., the subsequence  $(\vec{z}_{n_j})$  converges to  $\vec{z}$ .

*Proof (outline):* We review the above proof for  $\mathbb{R}$ :

The base idea was to chop  $A$  in half during each step to obtain a sequence of sets  $A_n$  which become smaller and smaller in diameter but yet contain infinitely many points of the original sequence  $z_n$ .

In higher dimensions we would still find the center point  $\vec{m}_n$  which is determined by the fact that it is the center of a  $\gamma$ -neighborhood ( $N$ -dimensional ball) that contains  $A_n$  and does so with the smallest radius possible. We then take the minimal square (in  $\mathbb{R}^2$ ) or the minimal  $N$ -dimensional cube (in  $\mathbb{R}^N$ ) that is parallel to the coordinate axes and still contains that sphere or ball.

We then divide that  $N$ -dimensional cube (a square in 2 dimensions, a cube in 3 dimensions) into  $2^N$  sectors (4 quadrants in  $\mathbb{R}^2$ , 8 sectors in  $\mathbb{R}^3$ ) and partition  $A_n$  into at most  $2^N$  pieces by intersecting it with those  $2^N$  sectors).

The set  $A_{n+1}$  is then chosen from one of those pieces which contain infinitely many of the  $z_n$ . Again, we get a nested sequence  $A_n$  whose diameters contract towards 0. You'll find more detail about the messy calculations required in the proof of prop.12.2 on p.266.

Each  $A_n$  contains infinitely many of the  $\vec{z}_k$ . Now pick  $\vec{x}_j := \vec{z}_{n_j}$  where  $\vec{z}_{n_j}$  is one of the infinitely many members of the original sequence  $(\vec{z}_n)$  which are contained in  $A_j$ .

Because  $A_j \subseteq A_K$  for  $j \geq K$  and  $\lim_{K \rightarrow \infty} \text{diam}(A_K) = 0$ , we do the following for a given  $\varepsilon > 0$ : choose  $K$  so big that  $\text{diam}(A_K) \leq \varepsilon$ . Note that

$$\text{if } i, j \geq K \text{ then } d(\vec{x}_i, \vec{x}_j) = d(\vec{z}_{n_i}, \vec{z}_{n_j}) \leq \text{diam}(A_K) \leq \varepsilon.$$

This is true because  $n_i \geq i$  (and  $n_j \geq j$ ), hence  $\vec{x}_i, \vec{x}_j \in A_K$ . It follows that the sequence  $(\vec{x}_j)$  is Cauchy. We have seen in thm.11.9 on p.228 that  $\mathbb{R}^N$  is complete. It follows that  $\vec{L} := \lim_{j \rightarrow \infty} \vec{x}_j$  exists in  $\mathbb{R}^N$ .

The proof is complete if it can be shown that  $\vec{L} \in A$ .

But we know that all  $\vec{x}_i = \vec{z}_{n_i}$  belong to  $A$ .  $\vec{L}$  is a contact point of  $A$  according to the sequence criterion for contact points in metric spaces (thm.11.6 on p.222).

any neighborhood  $N_\varepsilon(\vec{L})$  contains an entire tail set of the sequence  $(\vec{x}_i)_i$ . As the closed set  $A$  owns all its contact points, it follows that  $L \in A$  and the theorem is proved. ■

**Theorem 12.3.** Let  $A$  be a bounded and closed set of real numbers and let  $f(\cdot) : A \rightarrow \mathbb{R}$  be a continuous function on  $A$ . Then  $f(\cdot)$  is a bounded function.

*PROOF:* Let us assume that  $f(\cdot)$  is not bounded and show that this leads to a contradiction.

An unbounded function is not bounded from above, from below, or both. We might as well assume that  $f(\cdot)$  is not bounded from above because otherwise it is not bounded from below and we can work with  $-f(\cdot)$  which then is not bounded from above. We conclude that there is a sequence  $(z_n) \in A$  such that

$$(12.4) \quad f(z_n) > n \quad \text{for all } n \in \mathbb{N}.$$



According to thm.12.1 (Convergent subsequences in closed and bounded sets) there exists a subsequence  $(x_j) = (z_{n_j})$  and  $x_0 \in A$  such that  $x_0 = \lim_{n \rightarrow \infty} x_n$ .

In particular,  $f(x_0)$  exists as a finite value and  $f(x_n) \rightarrow f(x_0)$  because  $f(\cdot)$  is continuous in  $x_0$ . But the  $x_n$  were constructed as a subsequence of the  $z_j$  which have the property that  $f(z_j) > j$  for all  $j$ .

The subsequence  $(f(x_n))$  cannot converge to  $f(x_0)$  because  $f(x_j) = f(z_{n_j}) > n_j$ , i.e.,  $\lim_{j \rightarrow \infty} f(z_{n_j}) = \infty$ .

We have reached a contradiction and it follows that  $f(\cdot)$  is bounded. ■

**Corollary 12.1.** Let  $a < b$  be two real numbers and let  $f(\cdot) : [a, b] \rightarrow \mathbb{R}$  be a continuous function on  $[a, b]$ . Then  $f(\cdot)$  is a bounded function.

PROOF: The interval  $[a, b]$  is closed and bounded ( $\text{diam}([a, b]) = b - a$ ). The proof follows from thm.12.3. ■

## 12.2 Four Definitions of Compactness

We now look at ways to extend those results to general metric spaces by looking at the concept of compactness.

Compact sets are a wonderful thing to deal with because they allow you in some sense to go from dealing with “arbitrarily many” to dealing with “countably many” and even “finitely many”. This chapter will show that A, B and C below are equivalent statements for any subspace  $(K, d|_{K \times K})$  of a metric space  $(X, d)$ :

- A. any sequence in  $K$  has a convergent subsequence
- B.  $K$  is complete and contains only finitely many point of a grid of length  $\varepsilon$
- C. any open covering of  $K$  has a finite subcovering
- D.  $K$  is bounded and closed - ONLY works in  $\mathbb{R}^N$ !

Such metric spaces  $K$  will be called “compact” (see def.12.3 (Sequence compactness) on p.268 and def.12.5 (Compact sets) on p.272).

We can now state theorems 12.1 and 12.2 of the introduction as follows: In  $\mathbb{R}^N$  statement D implies A.

When you take a course on real analysis you will probably be given the definition of compactness as that in C: any open covering of  $K$  has a finite subcovering. In this document this definition is pushed into the background as it is the most difficult to understand. But full proofs will be given of the equivalence of sequence compactness (def.A) on the one hand and completeness plus “total boundedness” (def.B) on the other hand.

One of the important results of this chapter on compactness is that, if you look at  $\mathbb{R}^N$  with the Euclidean norm and its associated metric

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (\vec{x} = (x_1, x_2, \dots), \vec{y} = (y_1, y_2, \dots) \in \mathbb{R}^N)$$

(see (10.3) on p.177) then all four statements A, B, C, D coincide.

## 12.3 $\varepsilon$ -Nets and Total Boundedness

We now briefly discuss  $\varepsilon$ -nets and decreasing sequences of closed sets which contract to a single point.

**Definition 12.1** ( $\varepsilon$ -nets). Let  $\varepsilon > 0$ . Let  $(X, d)$  be a metric space and  $A \subseteq X$ . Let  $G \subseteq A$  be a subset of  $A$  with the following property:

$$(12.5) \quad \text{For each } x \in A \text{ there exists } g \in G \text{ such that } x \in N_\varepsilon(g).$$

In other words, the points of  $G$  form a “grid” or “net” fine enough so that no matter what point  $x$  of  $A$  you choose, you can always find a “grid point”  $g$  with distance less than  $\varepsilon$  to  $x$ , because that is precisely the meaning of  $x \in N_\varepsilon(g)$ .

We call  $G$  an  $\varepsilon$ -net or  $\varepsilon$ -grid for  $A$  and we call  $g \in G$  a **grid point** of the net.  $\square$

**Proposition 12.1** ( $\varepsilon$ -nets and coverings). Let  $\varepsilon > 0$ . Let  $(X, d)$  be a metric space and  $A \subseteq X$ . Let  $G \subseteq A$  be an  $\varepsilon$ -grid for  $A$ . Then  $\{N_\varepsilon(g)\}_{g \in G}$  is an open covering of  $A$  in the sense of def.12.4 on p.271: It is a collection of open sets whose union “covers”, i.e., contains,  $A$ .

PROOF: Let  $x \in A$ . We can choose a point  $g = g(x) \in G$  such that  $x \in N_\varepsilon(g(x))$ . It follows from  $\{x\} \subseteq N_\varepsilon(g(x))$  and  $g(x) \in G$  for all  $x \in A$  that

$$A = \bigcup_{x \in A} \{x\} \subseteq \bigcup_{x \in A} N_\varepsilon(g(x)) \subseteq \bigcup_{g \in G} N_\varepsilon(g). \blacksquare$$

**Proposition 12.2** ( $\varepsilon$ -nets in  $\mathbb{R}^N$ ). Let  $(X, d)$  be  $\mathbb{R}^N$  with the Euclidean metric.

A. Let

$$\mathbb{Z}^N = \{\vec{z} = (z_1, z_2, \dots, z_N) : z \in \mathbb{Z}\}$$

In other words,  $\mathbb{Z}^N$  is the set of all points in  $\mathbb{R}^N$  with integer coordinates. <sup>132</sup>  
Then  $\mathbb{Z}^N$  is a  $\sqrt{N}$ -net of  $\mathbb{R}^N$ .

B. Let  $\varepsilon > 0$  and  $G_\varepsilon^{\mathbb{R}^N} := \{\varepsilon \vec{z} : \vec{z} \in \mathbb{Z}^N\}$ . Then  $G_\varepsilon^{\mathbb{R}^N}$  is an  $\varepsilon\sqrt{N}$ -net of  $\mathbb{R}^N$ .

C. Let  $A$  be a bounded set in  $\mathbb{R}^N$  and  $\varepsilon > 0$ . Then there is  $n \in \mathbb{N}$  and  $g_1, \dots, g_n \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$  such that

$$A \subseteq N_\varepsilon(g_1) \cup N_\varepsilon(g_2) \cup \dots \cup N_\varepsilon(g_n),$$

i.e.,  $A$  is covered by finitely many  $\varepsilon$ -neighborhoods of points in the  $(\varepsilon/\sqrt{N})$ -grid  $G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$ .

**(Skip this proof!)** (all three parts A, B, C)

PROOF of A.

Let  $\vec{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ . For each  $x_j$  let  $x_j^*$  be the integer part of  $x_j$ : If  $x_j$  has decimal representation  $x_j = m + \sum_{i=1}^{\infty} d_{ji} 10^{-i}$  where  $m \in \mathbb{Z}$  and each  $d_{ji}$  is a digit, then  $x_j^* = m$ .

Before we continue, here is an example. If  $N = 5$  and  $\vec{x} = (12.85, -12.35, \frac{2}{3}, 9, -\pi)$  then the associated grid point is  $\vec{x}^* = (12, -12, 0, 9, -3)$ . The distance is:

$$d(\vec{x}, \vec{x}^*) = \sqrt{.85^2 + .35^2 + (2/3)^2 + 0 + (\pi - 3)^2} \leq \sqrt{1 + 1 + 1 + 0 + 1} \leq \sqrt{N}.$$

We see that part **A** of the lemma is true for this specific example.

<sup>132</sup>That is as intuitive a grid as you can think of, especially if you look at the 2-dimensional plane or 3-dimensional space.

Now to the real proof. It is not really more complicated if you notice that  $|x_j - x_j^*| < 1$  for all  $1 \leq j \leq N$ . We get

$$d(\vec{x}, \vec{x}^*) = \sqrt{\sum_{j=1}^N (x_j - x_j^*)^2} < \sqrt{N \cdot 1} = \sqrt{N}$$

So, for each point you can find a grid point with integer coordinates at a distance of less than  $\sqrt{N}$ . That proves that  $\mathbb{Z}^N$  is a  $\sqrt{N}$ -net of  $\mathbb{R}^N$ .

PROOF of B.

Let  $\vec{y} \in \mathbb{R}^N$ . Let  $\vec{x} := (\sqrt{N}/\varepsilon)\vec{y}$  and let  $\vec{x}^*$  be the vector where we discard the decimal parts of  $\vec{x}$ . According to part A. we know that  $\|\vec{x} - \vec{x}^*\|_2 = d(\vec{x}^*, \vec{x}) < \sqrt{N}$ . Thus

$$d(\vec{y}, \frac{\varepsilon}{\sqrt{N}} \vec{x}^*) = d(\frac{\varepsilon}{\sqrt{N}} \vec{x}, \frac{\varepsilon}{\sqrt{N}} \vec{x}^*) = \|\frac{\varepsilon}{\sqrt{N}} \vec{x} - \frac{\varepsilon}{\sqrt{N}} \vec{x}^*\|_2 = \frac{\varepsilon}{\sqrt{N}} \|\vec{x} - \vec{x}^*\|_2 < \frac{\varepsilon}{\sqrt{N}} \sqrt{N} = \varepsilon$$

In other words, for any  $\vec{y} \in \mathbb{R}^N$  there is a vector  $\vec{x}^* \in \mathbb{Z}^N$  such that  $d(\vec{y}, (\varepsilon/\sqrt{N})\vec{x}^*) < \varepsilon$ .

Rephrase that: For any  $\vec{y} \in \mathbb{R}^N$  there is a vector  $\vec{g} \in G_{(\varepsilon/\sqrt{N})}^{\mathbb{R}^N} = \{(\varepsilon/\sqrt{N})\vec{z} : \vec{z} \in \mathbb{Z}^N\}$

such that  $d(\vec{y}, \vec{g}) < \varepsilon$  (choose  $\vec{g} = \vec{x}^*$ ).

So, for each point you can find a grid point in  $G_{\varepsilon}^{\mathbb{R}^N}$  at a distance of less than  $\varepsilon\sqrt{N}$ . It follows that  $G_{\varepsilon}^{\mathbb{R}^N}$  is an  $\varepsilon\sqrt{N}$ -net of  $\mathbb{R}^N$ .

PROOF of C.

Intuitively clear but very messy. Here is an outline.

a. You can choose a radius  $R_1$  so big that  $A \subseteq N_{R_1}(\vec{0})$  (see prop.11.13 on p.221).

b. We enlarge the radius by  $\varepsilon$ : Let  $R := R_1 + \varepsilon$ . The enlarged "N-dimensional ball" of radius  $R$   $N_R(\vec{0})$  is contained in the "N-dimensional cube"

$$Q_R := \{\vec{x} = (x_1, x_2, \dots, x_N) : -R \leq x_j \leq R \text{ for all } 1 \leq j \leq N\}.$$

This is true because if  $\vec{x} = (x_1, x_2, \dots, x_N) \in N_R(\vec{0})$  then  $|x_j| = \sqrt{x_j^2} \leq \sqrt{\sum_i x_i^2} = \|\vec{x}\|_2 = R$ .

c. Let  $\vec{z} = (z_1, z_2, \dots, z_N)$  be a grid point of  $G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$ , i.e., for the  $j$ -th coordinate ( $1 \leq j \leq N$ ) there exists  $m_j \in \mathbb{Z}$  such that  $z_j = m_j \frac{\varepsilon}{\sqrt{N}}$ . We observe that there are only finitely many integers  $m$ , say  $K$ , for which  $-R \leq m \frac{\varepsilon}{\sqrt{N}} \leq R$ . We further observe that  $K$  does not depend on  $j$ .

d. Hence there are only  $K$  possible values for the first coordinate  $z_1 = m_1 \frac{\varepsilon}{\sqrt{N}}$ . For each one of those there are only  $K$  possible values for  $z_2$ , so there are at most  $K^2$  possible combinations  $(z_1, z_2)$  for which  $\vec{z} \in A$ . We keep going and find that there are at most  $K^N$  possible grid points  $\vec{z} \in Q_R$ .

e. Any point in  $\mathbb{R}^N$  with distance less than  $\varepsilon$  from some point in  $A$  must belong to  $B_R(\vec{0})$  (now you know why we chose to augment  $R_1$  by  $\varepsilon$ ). In particular, all grid points  $g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}$  whose neighborhoods  $N_{\varepsilon}(g)$  intersect  $A$  belong to  $B_R(\vec{0})$  and hence to  $Q_R$ . We conclude that  $A \cap N_{\varepsilon}(g) = \emptyset$  for all grid points outside  $Q_R$ , hence

$$(12.6) \quad \bigcup [A \cap N_{\varepsilon}(g) : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}] = \bigcup [(A \cap N_{\varepsilon}(g)) \cap Q_R : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}].$$

We know from part **B** which was already proved that  $\mathbb{R}^N = \bigcup [N_\varepsilon(g) : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}]$ . Hence,

$$\begin{aligned} A &= A \cap \mathbb{R}^N = A \cap \bigcup [N_\varepsilon(g) : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}] \\ &= \bigcup [A \cap N_\varepsilon(g) : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}] = \bigcup [(A \cap N_\varepsilon(g)) \cap Q_R : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N}]. \end{aligned}$$

We obtain the third equality from prop.6.2 (Distributivity of unions and intersections) on p.112 and the last one follows from (12.6) above.

It follows that  $A \subseteq \bigcup [N_\varepsilon(g) : g \in G_{\varepsilon/\sqrt{N}}^{\mathbb{R}^N} \cap Q_R]$ . We have proved **C** as there are only finitely many grid points in  $Q_R$ . ■

**Remark 12.1.** The observant reader will have noted that, in part **C** of the previous proposition, it was not stated that the gridpoints belong to the subset  $A$  of  $\mathbb{R}^N$ . Here is a trivial counterexample. Look at the “standard”  $\varepsilon$ -grid  $G_\varepsilon^{\mathbb{R}^N} = \{\varepsilon \vec{z} : \vec{z} \in \mathbb{Z}^N\}$  defined in prop.12.2, part **B**. Take any  $A \subseteq \mathbb{R}^N$  you like and look at  $B := A \setminus G_\varepsilon^{\mathbb{R}^N}$ , i.e., we have removed all grid points. It is clear that  $B$  cannot be covered by  $\sqrt{N}\varepsilon$  balls belonging to grid points in  $B$ . □

**Definition 12.2** (Total boundedness). Let  $(X, d)$  be a metric space and let  $A$  be a subset of  $X$ . We say that  $A$  is **totally bounded** if for each  $\varepsilon > 0$  there is a finite collection  $\mathcal{G}_\varepsilon := \{g_1, \dots, g_n\}$  of points in  $A$  whose  $\varepsilon$ -balls  $N_\varepsilon(g_j)$  cover  $A$ : For any  $a \in A$  there is  $j = j(a)$  such that  $d(a, g_j) < \varepsilon$ . □

We will use this definition in connection with sequence compactness which is defined in the next section.

## 12.4 Sequence Compactness

We saw in the introductory section that, for the space  $\mathbb{R}^N$  with the Euclidean metric, closed and bounded sets have the property that any sequence contains a convergent subsequence. We named this property in section 12.2, p.265 on Four Definitions of Compactness “sequence compactness” and we will examine that property in this chapter.

**Definition 12.3** (Sequence compactness). Let  $(X, d)$  be a metric space and let  $A$  be a subset of  $X$ . We say that  $A$  is **sequence compact** or **sequentially compact** if it has the following property: Given any sequence  $(x_n)$  of elements of  $A$ , there exists  $L \in A$  and a subset

$$n_1 < n_2 < \dots < n_j < \dots \quad \text{of indices such that} \quad L = \lim_{j \rightarrow \infty} x_{n_j},$$

i.e., there exists a subsequence<sup>133</sup>  $(x_{n_j})$  which converges to  $L$ . □

**Proposition 12.3** (Sequence compactness implies total boundedness). Let  $(X, d)$  be a metric space and let  $A$  be a sequentially compact subset of  $X$ . Then  $A$  is totally bounded.

*PROOF:* This is proved by contradiction. We may assume that  $A \neq \emptyset$  as nothing needs to be shown otherwise. **a.** Assume that  $A$  is not totally bounded. Then 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  $\bigcup_{1 \leq j \leq n} N_\varepsilon(z_j)$  does not cover  $A$ : There exists  $z \in A$  outside any one of those  $\varepsilon$ -neighborhoods, i.e.,  $z \in A \setminus \bigcup [N_\varepsilon(z_j) : 1 \leq j \leq n]$ .

<sup>133</sup> See def.5.20 on p.103.

This allows us to create an infinite sequence  $(x_j)_{j \in \mathbb{N}}$  such that  $d(x_j, x_n) \geq \varepsilon$  for all  $j, n \in \mathbb{N}$  such that  $j \neq n$ , say,  $j < n$ , as follows: We pick

$$x_1 \in A; \quad x_2 \in A \setminus N_\varepsilon(x_1); \quad x_3 \in A \setminus (N_\varepsilon(x_1) \cup N_\varepsilon(x_2)); \quad \dots \quad x_n \in A \setminus \bigcup_{j < n} N_\varepsilon(x_j); \quad \dots$$

**b.** The proof is done if we can show that  $(x_j)_{j \in \mathbb{N}}$  does not possess a convergent subsequence. Assume to the contrary that there is  $L \in A$  and  $n_1 < n_2 < \dots$  such that  $\lim_{j \rightarrow \infty} x_{n_j} = L$ .

We pick the number  $\varepsilon > 0$  that was used in part **a** of the proof. There exists  $N = N(\varepsilon)$  such that  $d(x_{n_m}, L) < \varepsilon/2$  for all  $m \geq N$ . Because  $i \leq n_i$  and  $j \leq n_j$ , it follows for all  $i, j \geq N$  that

$$d(x_{n_i}, x_{n_j}) \leq d(x_{n_i}, L) + d(L, x_{n_j}) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

But the  $x_n$  were constructed in such a fashion that  $d(x_m, x_k) \geq \varepsilon$  for **all**  $m \neq k$ , in particular for  $m := n_i$  and  $k := n_j$  if  $i \neq j$ . We have arrived at a contradiction because  $n_i \neq n_j$  whenever  $i \neq j$ . ■

**Proposition 12.4** (Sequence compact implies completeness). *Let  $(X, d)$  be a metric space and let  $A$  be a sequence compact subset of  $X$ . Then  $A$  is complete, i.e., any Cauchy sequence  $(x_{n_j})$  in  $A$  converges to a limit  $L \in A$ .*

*PROOF:* Let  $(x_n)$  be a Cauchy sequence in  $A$ . Because  $A$  is sequence compact, we can extract a subsequence  $z_j := x_{n_j}$  and find  $L \in A$  such that  $z_j \rightarrow L$  as  $j \rightarrow \infty$ .

(12.7) **EMPTY LABEL! MUST BE REMOVED!**

(12.8) **EMPTY LABEL! MUST BE REMOVED!**

It follows from prop.11.23 on p.231 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. The reverse is also true:

**Theorem 12.4** (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.12.3 on p.268 and prop.12.4 on p.269 that if  $A$  is sequentially compact then  $A$  is totally bounded and complete. We now show the other direction.

Let  $A$  be totally bounded and complete and let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $A$ . All we need to show is the existence of a subsequence  $z_j = x_{n_j}$  which is Cauchy: As  $A$  is complete, such a Cauchy sequence must converge to a limit  $L \in A$ , i.e.,  $x_{n_j} \rightarrow L$  as  $n \rightarrow \infty$ . We extract a Cauchy subsequence  $(x_{n_j})_j$  from  $(x_n)_n$  as follows.

**a.** Because  $A$  is totally bounded, there will be a finite net for  $\varepsilon = 1/2$ : there exists

$$\mathcal{G}_1 = \{g_{1,1}, g_{1,2}, \dots, g_{1,k_1}\} \subseteq A \text{ such that } A \subseteq U_1 := \bigcup [N_{1/2}(g_{1,j}) : j \leq k_1].$$

It follows that  $x_k \in U_1$  for each  $k$ . There are infinitely many indices  $k$  for our sequence but only finitely many points in  $\mathcal{G}_1$ . Hence there exists some  $g_1 \in \mathcal{G}_1$  such that  $B_1 := N_{1/2}(g_1) \cap A$  contains  $x_{1,j} := x_{n_j}$  for an entire (infinite) subsequence  $n_j$ .<sup>134</sup>

<sup>134</sup> Note that it is not claimed that there would be infinitely many different points  $x_{n_j}$ , only infinitely many indices  $n_j$ . Indeed, what would you do if the original Cauchy sequence was chosen to be  $x_1 = x_2 = \dots = a$  for some  $a \in A$ ?

**b.** Because  $A$  is totally bounded, there will be a finite net for  $\varepsilon = 1/3$ : there exists

$$\mathcal{G}_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,k_2}\} \subseteq A \text{ such that } A \subseteq U_2 := \bigcup [N_{1/3}(g_{2,j}) : j \leq k_2].$$

It follows that  $x_{1,k} \in U_2$  for each  $k$ . There are infinitely many indices  $k$  for our sequence but only finitely many points in  $\mathcal{G}_2$ . Hence there exists some  $g_2 \in \mathcal{G}_2$ , such that  $N_{1/3}(g_2)$  contains  $x_{2,j} := x_{1,n_j}$  for an entire subsequence  $n_j$ . As the entire sequence  $(x_{1,k})$  belongs to  $B_1$ , it follows that our new subsequence  $(x_{2,j})$  of  $(x_{1,k})$  belongs to  $B_2 := B_1 \cap N_{1/3}(g_2)$ .

**c.** Having constructed a subsequence  $(x_{n-1,j})$  of the original sequence  $(x_k)$  which lives in a set  $N_{n-1}$  contained in  $N_{1/n}(g_{n-1})$  for a suitable  $g_{n-1} \in A$ , total boundedness of  $A$ , guarantees the existence of a finite net for  $\varepsilon = 1/(n+1)$ : there exists

$$\mathcal{G}_n = \{g_{n,1}, g_{n,2}, \dots, g_{n,k_n}\} \subseteq A \text{ such that } A \subseteq U_n := \bigcup [N_{1/(n+1)}(g_{n,j}) : j \leq k_n].$$

It follows that  $x_{n,k} \in U_n$  for each  $k$ . There are infinitely many indices  $k$  for our sequence but only finitely many points in  $\mathcal{G}_n$ . Hence there must be at least one of those which we name  $g_n$ , such that  $N_{1/(n+1)}(g_n)$  contains  $x_{n,j} := x_{n-1,n_j}$  for an entire subsequence  $n_j$ . As the entire sequence  $(x_{n-1,k})$  belongs to  $N_{n-1}$ , it follows that our new subsequence  $(x_{n,j})$  of  $(x_{n-1,k})$  belongs to  $B_n := N_{n-1} \cap N_{1/(n+1)}(g_n)$ .

We note that the maximal distance  $d(x_{n,i}, x_{n,j})$  between any two members of that new subsequence is bounded by  $2/(n+1)$ , the diameter of  $N_{1/(n+1)}$ . It follows that

$$\text{diam}(B_n) \leq 2/(n+1).$$

**d.** Diagonalization procedure: The following trick is employed quite frequently in real analysis. We create the "diagonal sequence"  $z_1 := x_{1,1}, z_2 := x_{2,2}, \dots$  which is a subsequence of the original sequence  $(x_n)$ . If we can show that it is Cauchy then the proof is complete.

By construction, if  $j \geq n$  then

$$z_j \in B_j \subseteq B_n \subseteq N_{n-1} \subseteq \dots \subseteq B_2 \subseteq B_1 \quad \text{and} \quad \text{diam}(B_j) \leq \frac{2}{j+1} \leq \frac{2}{n+1} \quad (\text{see } ??).$$

Let  $\varepsilon > 0$ . We can find  $N \in \mathbb{N}$  such that  $\frac{1}{N+1} < \frac{\varepsilon}{2}$ . We remember from part c of this proof that

$$B_j = N_{j-1} \cap N_{1/(j+1)}(g_j) \subseteq N_{1/(j+1)}(g_j) \text{ for some suitable } g_j \in A,$$

hence all its points have distance from  $g_j$  bounded by  $(j+1)^{-1}$ . We obtain for any  $i, j \geq N$  that

$$d(z_i, z_j) \leq d(z_i, g_N) + d(g_N, z_j) \leq \frac{1}{N+1} + \frac{1}{N+1} < \varepsilon.$$

It follows that  $(z_n)_n$  is indeed Cauchy and the proof is completed. ■

**Corollary 12.2** (Sequence compact sets are complete). Let  $(X, d)$  be a metric space and let  $K$  be a sequence compact subset of  $X$ . Then  $K$  is complete.

*PROOF:* Immediate from the last theorem.

**Theorem 12.5** (Sequence compact sets are closed and bounded). Let  $A$  be sequence compact subset of a metric space  $(X, d)$ . Then  $A$  is a bounded and closed set.

**a. Proof of boundedness:**

We may assume that  $A$  is not empty because otherwise there is nothing to prove. We assume that  $A$  is not bounded, i.e.,  $\text{diam}(A) = \infty$ . It will be proved by induction that there exists a sequence  $x_n \in A$  such that  $d(x_i, x_j) \geq 1$  for any  $i \neq j$ .

*Base case:* Let  $x_0 \in A$ . There exists  $x_1 \in A$  such that  $r_1 := d(x_0, x_1) \geq 1$ .

*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 already been chosen. Let  $\beta := \max\{d(x_0, x_j) : j \leq n\}$  and  $r := \beta + 1$ . As  $A$  is not bounded, we can pick  $x_{n+1} \in A \setminus N_r(x_0)$ . We obtain

$$\beta + 1 \leq d(x_{n+1}, x_0) \leq d(x_{n+1}, x_j) + d(x_j, x_0) \leq d(x_{n+1}, x_j) + \beta, \quad \text{i.e., } 1 \leq d(x_{n+1}, x_j).$$

We have constructed a sequence  $(x_n)$  for which any two items have distance no less than 1. It follows that there is no Cauchy subsequence, hence no convergent subsequence and we have a contradiction to the sequence compactness of  $A$ .

**b. Proof of closedness:**

If  $A$  is not closed then  $A$  has a contact point  $x \in A^c$ .

As  $N_{1/m}(x) \cap A \neq \emptyset$  we can pick  $x_m \in N_{1/m}(x) \cap A$  such that  $d(x_m, x) < 1/m$  for all  $m \in \mathbb{N}$ .

Clearly the sequence  $x_m$  converges to  $x \in A^c$ . Sequence compactness of  $A$  allows us to extract a subsequence  $z_j = x_{n_j}$  which converges to  $z \in A$ .

Both  $z$  and  $x$  are limit of  $z_j$ . According to thm.11.3 on p.213,  $x = z$ . It follows that both  $x \in A^c$  and  $x \in A$ , a contradiction. This proves that sequence compact sets are closed. ■

**Corollary 12.3** (Sequence compact sets are bounded). REMOVED

REMOVED – REMOVED – REMOVED – REMOVED

REMOVED – REMOVED – REMOVED – REMOVED

**Remark 12.2.** It follows from the results of this chapter and the introductory chapter on Closed and bounded sets in Euclidean space (12.1 on p.263) that the following three of the four definitions of compactness given in section 12.2 on (p.265 are equivalent in  $\mathbb{R}^N$ :

A subset of  $\mathbb{R}^N$  is sequentially compact  
 $\Leftrightarrow$  it is totally bounded and complete  
 $\Leftrightarrow$  it is bounded and closed.

We will see later that any metric space is sequentially compact if and only if it is compact, i.e., covering compact (thm.12.8 on p.274).

In other words, in  $\mathbb{R}^N$  all four of the definition given in section 12.2 on p.265 coincide. □

## 12.5 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 (5.18) on p.102).

**Definition 12.4** (Coverings and open coverings). Let  $X$  be an arbitrary nonempty set and  $A \subseteq X$ . Let  $(U_i)_{i \in I}$  be an indexed family of subsets of  $X$  such that  $A \subseteq \bigcup_{i \in I} U_i$ . Then we call  $(U_i)_{i \in I}$  a **covering** of  $A$ .



A **finite subcovering** of a covering  $(U_i)_{i \in I}$  of the set  $A$  is a finite collection

$$(12.9) \quad U_{i_1}, U_{i_2}, U_{i_3}, \dots, U_{i_n} \quad (i_j \in I \text{ for } 1 \leq j \leq n) \quad \text{such that} \quad A \subseteq U_{i_1} \cup U_{i_2} \cup \dots \cup U_{i_n}.$$

Assume in addition that  $X$  is a topological space — this includes any normed vector space or metric space. If all members  $U_i$  of the family  $(U_i)_{i \in I}$  are open then we call this family an **open covering** of  $A$ .  $\square$

**Remark 12.3.**

- a. Partitions <sup>135</sup> are coverings.
- b. 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

$$(12.10) \quad \bigcup_{j \in I} (U_j \cap A) = A. \quad \square$$

**Definition 12.5** (Compact sets). Let  $(X, \mathfrak{U})$  be a topological space and  $K \subseteq X$ . We say that  $K$  is **compact** if it 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

$$i_1, i_2, \dots, i_n \in I \quad \text{such that} \quad A \subseteq \bigcup_{j=1}^n U_{i_j}. \quad \square$$

**Remark 12.4.**

- a. An open covering for the entire space  $X$  is a collection of open sets  $(U_i)_{i \in I}$  such that  $X = \bigcup [U_i : i \in I]$ .
- b. Any subcovering of an open covering necessarily consists exclusively of open sets, i.e., it is again an open covering of  $A$ .
- c. Let  $(X, d)$  be a metric space. Then

$K \subseteq (X, d)$  is compact if and only if the metric subspace  $(K, d|_{K \times K})$  is compact,

i.e., for any collection of subsets  $(U_i)_{i \in I}$  of  $K$  which are open in  $K$  there exist finitely many indices  $i_1, \dots, i_n \in I$  such that  $K = U_{i_1} \cup \dots \cup U_{i_n}$ . This is true because the open subsets of  $(K, d)$  are the traces in  $K$  of sets which are open in  $(X, d)$  (see def.11.22 on p.220).  $\square$

**Example 12.1.** Here are some simple examples.

- a. Any finite topological space is compact.
- b. Any topological space that only contains finitely many open sets is compact. In particular a set with the indiscrete topology (example 11.6 on p.215) is compact
- c. A space with the discrete metric (def.11.3 on p.207) 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  $(\] \frac{1}{n}, 1[)_{n \in \mathbb{N}}$ .  $\square$

---

<sup>135</sup>see def.6.3 on p.111



**Example 12.2.** Here are the corresponding results for metric spaces, restated in terms of sequence compactness.

- a. Any finite metric space is sequence compact.
- b. Any metric space that only contains finitely many open sets is sequence compact. <sup>136</sup>
- 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  $]0, 1[$ ).  $\square$

We will now see that the correspondence in the above two examples is not a coincidence. We will see that (subspaces of) metric spaces are compact if and only if they are sequentially compact. We will prove each direction separately.

**Theorem 12.6** (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  $F := \{x \in X : x = x_j \text{ for some } j \in \mathbb{N}\}$  <sup>137</sup> 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 12.2 on p.277).

It follows from  $\{z\} \subseteq U_z$  that  $(U_z)_{z \in X}$  is an open covering of  $X$ .  $X$  is compact and we can extract a finite subcovering  $U_1, U_2, \dots, U_k$ . But each  $U_j$  contains at most finitely many distinct members of the sequence  $(x_n)_n$ . Hence this entire sequence consists of only finitely many distinct members.

We conclude that at least one of those members, say  $x_{k^*}$ , will appear infinitely often in that sequence: there is  $k_1 < k_2 < \dots$  such that  $x_{k_1} = x_{k_2} = \dots = x_{k^*}$ . This constant subsequence converges (to  $x_{k^*}$ ). We have reached a contradiction.  $\blacksquare$

**Proposition 12.5.** *Let  $(X, d)$  be a sequence compact metric space. Let  $(U_i)_{i \in I}$  be an open covering of  $X$ . Then there exists  $\rho > 0$  as follows: For each  $x \in X$  there exists  $i \in I$  such that  $N_\rho(x) \subseteq U_i$ .* <sup>138</sup>

*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 11.3 on p.211 that  $N_{\varepsilon/2}(x_{n_j}) \subseteq N_{\varepsilon/2 + \varepsilon/2}(x) = N_\varepsilon(x)$ .

<sup>136</sup>We had to remove the example of the indiscrete topology because this topology does not come from a metric.

<sup>137</sup>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!

<sup>138</sup>The number  $\lambda = 2\rho$  is called a **Lebesgue number** of  $(U_i)_{i \in I}$ . In other words, the Lebesgue number is the diameter of the  $\rho$ -neighborhoods. Note that if  $\lambda$  is a Lebesgue number of an open covering then any  $\lambda'$  which satisfies  $0 < \lambda' < \lambda$  also is a Lebesgue number. of that same cover.

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  $U_i$ . ■

We now can prove the converse of thm.12.6.

**Theorem 12.7.** *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.12.5 there exists  $\rho > 0$  as follows: For each  $x \in X$  there exists  $i(x) \in I$  such that  $N_\rho(x) \subseteq U_{i(x)}$ .

Since  $X$  is totally bounded (see thm.12.4 on p.269) there exist finitely many  $x_1, \dots, x_k \in X$  such that  $\{N_\rho(x_j) : j = 1, \dots, k\}$  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 12.8** (Sequence compact is same as compact in metric spaces). *Let  $(X, d)$  be a metric space and let  $A$  be a subset of  $X$ . Then  $A$  is sequence compact if and only if  $A$  is compact, i.e., every open covering of  $A$  has a finite subcovering.*

PROOF: Theorems 12.6 and 12.7. ■

An easy consequence is the Heine–Borel theorem.

**Theorem 12.9** (Heine–Borel). *Let  $(X, d)$  be  $\mathbb{R}^N$  with the Euclidean norm and its associated metric. A subset  $K \subseteq \mathbb{R}^N$  is compact if and only if it is closed and bounded.*

For a general metric space it is still true that any compact subset is closed and bounded.

PROOF: We have seen in thm.12.2 on p.264 that closed and bounded subsets of  $\mathbb{R}^N$  are sequence compact. We have further seen that sequence compact sets are closed and bounded (thm.12.5 on p.270). Our assertion now follows from the equivalence of compactness and sequence compactness in metric spaces (thm.12.8 on p.274). ■

## 12.6 Continuous Functions and Compact Spaces

**Theorem 12.10** (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, i.e., 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 def.12.5 on p.272.*

PROOF: Let  $(U_j)_{j \in J}$  be a family of sets open in  $A$  whose union is  $A$ . According to def.11.23 on p.221 there are open sets  $V_j$  in  $X$  such that  $U_j = V_j \cap A$ . It follows that  $\bigcup_{j \in J} V_j \supseteq A$ , 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  $A$  (def.11.22 on p.220)

$$U_{i_1} = V_{i_1} \cap A, U_{i_2} = V_{i_2} \cap A, \dots, U_{i_n} = V_{i_n} \cap A$$

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 12.4** (Closed subsets of compact metric spaces are compact). *Let  $A$  be a closed subset of a compact metric space  $(X, d)$ . Then  $(A, d|_{A \times A})$  is a compact subspace.*

PROOF: Immediate from thm.12.10. ■

Let  $(X, \mathfrak{U})$  and  $(Y, \mathfrak{V})$  be topological spaces and  $A \subseteq X$ . We recall that continuity for functions  $f : A \rightarrow (Y, \mathfrak{V})$  was defined in def.11.32 on p.236.

**Theorem 12.11** (Continuous images of compact topological spaces are compact). *Let  $(X, \mathfrak{U})$  and  $(Y, \mathfrak{V})$  be two topological spaces. and let  $f : X \rightarrow Y$  be continuous on  $X$ . If  $X$  is compact then the direct image  $f(X)$  is compact, i.e., 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

$$(12.11) \quad \bigcup_{j \in J} U_j = \bigcup_{j \in J} f^{-1}(V_j) = f^{-1}\left(\bigcup_{j \in J} V_j\right) \supseteq f^{-1}(f(X)) = X.$$

The second equation above follows from prop. 7.2 ( $f^{-1}$  is compatible with all basic set ops) on p.120 and the last one follows from the fact that  $f^{-1}(f(\Gamma)) \supseteq \Gamma$  for any subset  $\Gamma$  of the domain of  $f$  (see cor. 7.1 on p. 123). The “ $\supseteq$ ” relation follows from the assumption that  $\bigcup [V_j : j \in J] \supseteq f(X)$

According to prop.11.24 (“ $f^{-1}(\text{open}) = \text{open}$ ” continuity) on p.236, each  $U_j$  is open as the inverse image of the open set  $V_j$  under the continuous function  $f$ .

It follows from (12.11) 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 (7.19) on p.121) that

$$\begin{aligned} f(X) &= f(U_{j_1} \cup \dots \cup U_{j_n}) = f(U_{j_1}) \cup \dots \cup f(U_{j_n}) \\ &= f(f^{-1}(V_{j_1})) \cup \dots \cup f(f^{-1}(V_{j_n})) \subseteq V_{j_1} \cup \dots \cup V_{j_n}. \end{aligned}$$

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.7.6 on p. 123).

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 12.5** (Continuous images of compact metric spaces are compact). *Let  $(X, d_1)$  and  $(Y, d_2)$  be two metric spaces. and let  $f : X \rightarrow Y$  be continuous on  $X$ . If  $X$  is compact then the direct image  $f(X)$  is compact, i.e., the metric subspace  $(f(X), d_2)$  of  $Y$  is compact.*

PROOF: Immediate from thm.12.11 ■

Read the following remark for an easier way to prove the above theorem.

**Remark 12.5.** We could have proved the last two theorems 12.10 12.11 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.12.4 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 11.6 (Sequence criterion for contact points in metric spaces) on p.222 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.12.5 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 12.6** (Continuous real valued functions attain their max and min on a compact domain). Let  $(X, \mathfrak{U})$  be a topological space and let  $A \subseteq X$  be a compact subspace. Let  $f : A \rightarrow \mathbb{R}$  be continuous on  $A$ . Then there exist  $x_*, x^* \in A$  such that

$$f(x_*) = \min_{x \in A} f(x) \quad \text{and} \quad f(x^*) = \max_{x \in A} f(x).$$

*PROOF:* It follows from thm.12.11 on p.275 and thm.12.5 on p.270 that  $f(X)$  is closed and bounded in  $\mathbb{R}$ . It follows from exercise 11.14 on p.261 that  $\min(f(X))$  and  $\max(f(X))$  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. <sup>139</sup>

**Theorem 12.12** (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 real valued function on  $A$  is uniformly continuous on  $A$ .

*PROOF:* Let us assume that  $f(\cdot)$  is continuous but not uniformly continuous and find a contradiction. Because  $f(\cdot)$  is not uniformly continuous, there exists  $\varepsilon > 0$  such that no  $\delta > 0$ , however small, will satisfy (11.67) on p.241 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

$$(12.12) \quad d_1(x_j, x'_j) < \frac{1}{j} \quad \text{but} \quad d_2(f(x_j), f(x'_j)) \geq \varepsilon.$$

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

$$(12.13) \quad d_1(x'_{j_k}, x) \leq d_1(x'_{j_k}, x_{j_k}) + d_1(x_{j_k}, x) \leq \frac{1}{j_k} + d_1(x_{j_k}, x).$$

Both right hand terms converge to zero as  $k \rightarrow \infty$ . 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 (12.13) that  $(x'_{j_k})$  also converges to  $x$ . It follows from the ordinary continuity of  $f(\cdot)$  that

$$f(x) = \lim_{k \rightarrow \infty} f(x'_{j_k}) = \lim_{k \rightarrow \infty} f(x_{j_k}).$$

<sup>139</sup> See def.11.34 on p.241.

It follows from the “ordinary” (non-uniform) convergence of sequences that there exist  $N, N' \in \mathbb{N}$  such that

$$d_2(f(x), f(x_{j_k})) < \frac{\varepsilon}{2} \text{ for } k \geq N; \quad d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} \text{ for } k \geq N'.$$

Both inequalities are true whenever  $k \geq \max(N, N')$ . It follows for all such  $k$  that

$$(12.14) \quad d_2(f(x_{j_k}), f(x'_{j_k})) < d_2(f(x_{j_k}), f(x)) + d_2(f(x), f(x'_{j_k})) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and we have a contradiction to (12.12). ■

**Corollary 12.7** (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 on  $[a, b]$ :

For any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that

$$(12.15) \quad |f(x) - f(y)| < \varepsilon \text{ for all } x, y \in [a, b] \text{ such that } |x - y| < \delta$$

*PROOF:* This follows from the previous theorem (12.12) because closed intervals  $[a, b]$  are closed and bounded sets and, in  $\mathbb{R}$ , any closed and bounded set is sequence compact. ■

## 12.7 Exercises for Ch.12

**Exercise 12.1.** 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:

- Show sequence compactness to prove that  $X$  is compact.
- Show that  $X$  has the “extract finite open subcovering” property to prove that it is compact.
- Show that  $X$  is complete and totally bounded to prove that it is compact. □

### Hints:

- ANY sequence in  $X$  possesses a constant subsequence (WHY?)
- 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?
- Prop.11.23 on p.231 should prove useful.

**Exercise 12.2.** Prove the following which was used in the proof of thm.12.6 (Compact metric spaces are sequence compact) on p.273: 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  $d(x_{n_j}, z) < \frac{1}{j}$ . □

## 13 Applications of Zorn's Lemma

### 13.1 More on Partially Ordered Sets (Study this!)

*Some of this was copied almost literally from [6] Dudley.*

**Definition 13.1** (Linear orderings). Given is a nonempty set  $X$  and a partial ordering  $\preceq$  on  $X$  (see def.5.4 on p.84).  $\preceq$  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.  $\square$

**Example 13.1.**

- The real numbers line  $(\mathbb{R}, \leq)$  with its usual " $\leq$ " ordering is a linearly ordered set. So is  $(\mathbb{R}, \geq)$  (!)
- If  $X$  is a set with at least two elements then set inclusion is **not** a linear order on  $2^X$ .
- If  $m \in X$  is a maximum of  $X$  then this implies that  $m$  must be related to all other elements of  $X$ .  $\square$

**Definition 13.2.** A **chain** is a linearly ordered subset of a partially ordered set (with the same ordering). In a partially ordered set  $(X, \preceq)$ , an element  $m$  of  $X$  is called **maximal** iff there is no  $x \neq m$  with  $m \preceq x$ . A **maximum** of  $X$  is an  $m \in X$  such that  $x \preceq m$  for all  $x \in X$ .  $\square$

**Note 13.1** (Notes on maximal elements and maxima).

- If  $X$  is not linearly ordered, it may have many maximal elements. For example, for the trivial partial ordering  $x \preceq y$  if and only if  $x = y$ , every element is maximal. A maximum is a maximal element, but the converse is often not true.
- If an ordering is not specified, then we always mean set inclusion.
- If  $m \in X$  is a maximum of  $X$  then this implies that  $m$  must be related to all other elements of  $X$ .  $\square$

**Example 13.2** (Maximal elements and maxima). Let  $X$  be the collection of all intervals  $[a, b] \in \mathbb{R}$  of length  $b - a \leq 2$  such that  $a \leq b$ . These intervals are partially ordered by inclusion. Any interval of length equal to 2 is a maximal element. There is no maximum.  $\square$

**Axiom 13.1** (Zorn's Lemma). A hundred years ago the following was seen as extremely controversial by mathematicians who specialize in the foundations of mathematics.

**Zorn's Lemma:** Let  $(X, \preceq)$  be a partially ordered set with the **ZL property**:

Every chain  $C \subseteq X$ , possesses an upper bound  $u \in X$ , i.e.,  $x \preceq u$  for all  $x \in C$ . **(ZL)**

Then  $X$  has a maximal element.  $\square$

**Remark 13.1.** 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 this statement when building their mathematical theories. Two notes on that remark:

a. Today the mathematicians who refuse to accept proofs which make use, directly or indirectly, of Zorn's lemma, are a very small minority.



b. It can be proved that if one accepts (rejects) Zorn's lemma as a mathematical tool then this is equivalent to accepting (rejecting) the **Axiom of Choice**<sup>140</sup> which is so abstract in its precise formulation that it was relegated to a footnote.

c. Moreover the Axiom of Choice, hence Zorn's lemma, is equivalent to prop.5.6b on p.96: If  $A, B$  are not empty and  $\varphi : A \rightarrow B$  is surjective then  $\varphi$  has a right inverse, i.e., a function  $\psi : B \rightarrow A$  such that  $\varphi \circ \psi = id_B$ . For a proof see <http://planetmath.org/surjectionandaxiomofchoice>.  $\square$

We will see now how Zorn's lemma allows a surprisingly simple proof to the effect that **any** vector space has a basis.

## 13.2 Existence of Bases in Vector Spaces (Study this!)

The following is thematically a continuation of the material in chapter 10 (Vectors and vector spaces).

We now focus on proving that every vector space, even if it does not possess a finite subset which spans the entire space, possesses a basis.

For the remainder of this chapter we assume that  $V$  is a vector space and that  $\mathfrak{B}$  denotes the set

$$(13.1) \quad \mathfrak{B} := \{A \subseteq V : A \text{ is linearly independent} \}.$$

$\mathfrak{B}$  is a partially ordered set with respect to set inclusion. The next lemma allows us to apply Zorn's lemma.

**Lemma 13.1.** Every chain<sup>141</sup>  $\mathfrak{C}$  in  $(\mathfrak{B}, \subseteq)$  possesses an upper bound.

**Proof:**

Let  $U := \bigcup \{C : C \in \mathfrak{C}\}$ . 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

$$(13.2) \quad \sum_{j=1}^k \alpha_j x_j = 0.$$

We must show that each  $\alpha_j$  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  $\alpha_1 = \dots = \alpha_k = 0$ .  $\blacksquare$

**Theorem 13.1.** Every vector space  $V$  has a basis.

PROOF: It follows from lemma 13.1 and Zorn's Lemma (axiom 13.1 on p. 278) 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  $\text{span}(B) = V$ .

Let us assume to the contrary that there exists  $y \in \text{span}(B)^c$ . It follows from lemma 10.2 on p.186 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  $(\mathfrak{B}, \subseteq)$ .  $\blacksquare$

<sup>140</sup> **Axiom of Choice:** Let  $\Omega$  be any nonempty set. Then there is a function  $f : 2^\Omega \setminus \{\emptyset\} \rightarrow \Omega$  such that  $f(A) \in A$  for each  $A \in 2^\Omega \setminus \{\emptyset\}$ , i.e., for each nonempty  $A \subseteq \Omega$ .  $f$  is referred to as a **choice function** because it chooses from any nonempty subset  $A$  one of its elements.

<sup>141</sup> see def.13.2 on p.278.

### 13.3 The Cardinal Numbers are a totally ordered set

As another application of Zorn's Lemma we now prove thm.8.8, p.143, of ch.8.1.1 (Cardinality as a Partial Ordering).

**Theorem 13.2.** Let  $X, Y \subseteq \Omega$ . Then  $\text{card}(X) \leq \text{card}(Y)$  or  $\text{card}(Y) \leq \text{card}(X)$

PROOF: <sup>142</sup> The result is immediate if  $X = \emptyset$  or  $Y = \emptyset$ . Assume  $X$  and  $Y$  are not empty. Let

$$\mathcal{F} := \{D_f \xrightarrow{f} C_f : D_f \subseteq X, C_f \subseteq Y, \text{ and } f \text{ is bijective}\}$$

be the set of all bijective functions with domain contained in  $X$  and codomain contained in  $Y$ . We define a partial order  $\preceq$  on  $\mathcal{F}$  as follows: Let  $D_f \xrightarrow{f} C_f$  and  $D_g \xrightarrow{g} C_g$ . Then

$$f \preceq g \quad \text{if and only if } D_f \subseteq D_g, C_f \subseteq C_g, \text{ and } g|_{D_f} = f.$$

Let  $\mathfrak{C}$  be a chain in  $\mathcal{F}$ . 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  $g'(x) = g(x)$ .

We also notice that  $u$  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  $u(x_1) \neq u(x_2)$ .

$u$  also 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  $u$  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  $(\mathcal{F}, \preceq)$  possesses the ZL property (see axiom 13.1 (Zorn's Lemma) on p.278), hence there exists a maximal element  $D_h \xrightarrow{h} C_h$  in  $\mathcal{F}$ .

We claim that  $D_h = X$  or  $C_h = Y$  (or both). Otherwise there would be  $x_0 \in D_h^c$  and  $y_0 \in C_h^c$  and the function

$$\psi : D_h \cup \{x_0\} \longrightarrow C_h \cup \{y_0\}, \quad \begin{cases} x \mapsto h(x) & \text{if } x \in D_h, \\ x \mapsto y_0 & \text{if } x = x_0 \end{cases}$$

is an injective extension of  $h$ , i.e.,  $\psi \in \mathcal{F}$ . This contradicts the maximality of  $h$ .

**Case 1:**  $D_h = X$ . Then  $h$  is an injective function which maps  $X$  into  $Y$ , i.e.,  $\text{card}(X) \leq \text{card}(Y)$ .

**Case 2:**  $C_h = Y$ . Then  $h : D_h \rightarrow Y$  is a bijective function whose inverse  $f^{-1} : Y \rightarrow D_h$  is an injection from  $Y$  into the subset  $D_h$  of  $X$ , i.e.,  $\text{card}(Y) \leq \text{card}(X)$ . ■

<sup>142</sup> See [7] Haaser/Sullivan: Real Analysis.



### 13.4 Extensions of Linear Functions in Arbitrary Vector Spaces

We now turn our attention to extending a linear real valued function  $f$  from a subspace  $F \subseteq V$  to the entire vector space  $V$ . Note that setting  $f(x) = 0$  for all  $x \in F^c$  does not yield a linear extension of  $f$ . See exercise 13.3 on p288.

**Lemma 13.2.** Let  $V$  be a vector space and let  $F$  be a (linear) subspace of  $V$ . Let  $f : F \rightarrow \mathbb{R}$  be linear.

Let  $\mathcal{G} := \{(W, f_W) : W \text{ is a subspace of } V, W \supseteq F, f_W : W \rightarrow \mathbb{R} \text{ is a linear extension of } f \text{ to } W\}$ .

Then the following defines a partial ordering on  $\mathcal{G}$ :  $(U, f_U) \preceq (W, f_W) \Leftrightarrow U \subseteq W$  and  $f_W|_U = f_U$ .

Moreover this ordering satisfies the requirements of Zorn's Lemma: Every chain in  $(\mathcal{G}, \preceq)$  possesses an upper bound (in  $\mathcal{G}$ ).

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  $\mathcal{G}$ .

Now let  $\mathcal{C}$  be a chain in  $\mathcal{G}$ . We must find an upper bound for  $\mathcal{C}$ . Let  $W := \bigcup \{U : (U, f_U) \in \mathcal{C}\}$ .

We show that  $W$  is a subspace of  $V$ : If  $x, y \in W$  and  $\lambda \in \mathbb{R}$  then there are  $(C_1, f_{C_1}), (C_2, f_{C_2}) \in \mathcal{C}$  such that  $x \in C_1$  and  $y \in C_2$ . Because  $\mathcal{C}$  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  $f_W : W \rightarrow \mathbb{R}$  be defined as follows: If  $x \in W$  then there is some  $(C, f_C) \in \mathcal{C}$  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  $\mathcal{C}$ . To see this let  $(C, f_C), (D, f_D) \in \mathcal{C}$  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  $f_D|_C = f_C$  we conclude that  $f_C(x) = f_D(x)$ , i.e., the definition of  $f_W(\cdot)$  is unambiguous. The above specifically holds for  $x \in F$  and we note that  $f_W$  is an extension of  $f$ .

Next we show linearity of  $f_W$ . Let  $x, y \in W$  and  $\alpha \in \mathbb{R}$ . Then there are  $(C, f_C), (D, f_D) \in \mathcal{C}$  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  $(W, f_W) \in \mathcal{G}$  and  $(W, f_W) \succeq (C, f_C)$  for all  $(C, f_C) \in \mathcal{C}$ . It follows that  $(W, f_W)$  is an upper bound of  $\mathcal{C}$ . ■

**Theorem 13.3** (Extension theorem for linear real valued functions). Let  $V$  be a vector space and let  $F$  be a (linear) subspace of  $V$ . Let  $f : F \rightarrow \mathbb{R}$  be a linear mapping.

Then there is an extension of  $f$  to a linear mapping  $\tilde{f} : V \rightarrow \mathbb{R}$ .

PROOF:

Let  $\mathcal{G} := \{(W, f_W) : W \text{ is a subspace of } V, W \supseteq F, f_W : W \rightarrow \mathbb{R} \text{ is a linear extension of } f \text{ to } W\}$

and let  $(U, f_U) \preceq (W, f_W) \Leftrightarrow U \subseteq W$  and  $f_W|_U = f_U$ .

We have seen in lemma 13.2 that  $\preceq$  is a partial ordering on  $\mathcal{G}$  such that any chain in  $(\mathcal{G}, \preceq)$  possesses an upper bound. We apply Zorn's Lemma (axiom 13.1 on p.278) and conclude that  $\mathcal{G}$  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 lemma 10.9 on p.??, applied with  $V' = \mathbb{N}$  and  $y_0 = \alpha$ , extend  $f'$  to a linear function  $\tilde{f}$  on  $F' \uplus \{a\}$ . It follows that  $(F' \uplus \{a\}, \tilde{f}) \in \mathcal{G}$  and  $(F', f') \not\preceq (F' \uplus \{a\}, \tilde{f})$ .

This contradicts the maximality of  $(F', f')$ . and we have reached a contradiction. ■

### 13.5 The Hahn-Banach Extension Theorem (★)

*Note that this chapter is starred, hence optional.*

Let  $V$  be a vector space and let  $F$  be a (linear) subspace of  $V$ . Let  $f : F \rightarrow \mathbb{R}$  be linear mapping. This chapter shows how to extend  $f$  from its domain  $F$  to the entire space  $V$  subject to some condition which guarantees that if  $V$  is a normed space (hence a metric space) and if  $f$  is continuous on  $F$  then this linear extension will be continuous on all of  $V$ . The Hahn-Banach Extension Theorem provides such conditions.

In preparation for this we must study sublinearity, a generalization of linearity and norms.

#### 13.5.1 Sublinear Functionals

**Definition 13.3** (Sublinear functionals). Let  $V$  be a vector space and  $p : V \rightarrow \mathbb{R}$  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  $V$ . □

**Proposition 13.1.** Let  $V$  be a vector space and  $p : V \rightarrow \mathbb{R}$  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 13.3** (Norms are sublinear). Let  $(V, \|\cdot\|)$  be a normed vector space. Then the function  $p(x) := \|x\|$  is sublinear.

Indeed, norms are homogenous: We have  $\|\lambda x\| = |\lambda| \cdot \|x\|$  not only for  $\lambda \geq 0$  but for all  $\lambda \in \mathbb{R}$ .

Further, subadditivity is just the validity of the triangle inequality. □

**Example 13.4** (Linear functions are sublinear). Let  $V$  be a vector space and let  $f := V \rightarrow \mathbb{R}$  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 13.6 on p.286

#### 13.5.2 The Hahn-Banach extension theorem and its Proof

*This chapter follows closely [5] Choquet.*

As mentioned previously, the subject of the Hahn-Banach extension theorem is extension of a continuous, linear, real-valued function from a subspace to the entire vector space. Theorem 13.4 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  $V$ , then  $f$  can be linearly extended to  $V$  in such a way that  $p$  dominates this extension everywhere on  $V$ .

The following remark is about first extending it to “one more dimension”.

**Remark 13.2.** Let  $V$  be a vector space, let  $F$  be a linear subspace of  $V$  and let  $f : F \rightarrow \mathbb{R}$  be a linear function. Let  $a \in V \setminus F$ . Then any linear extension  $\tilde{f}$  of  $f$  to  $\text{span}(F \uplus \{a\})$  is uniquely determined by its value  $k := \tilde{f}(a)$ .

Indeed, any  $x \in \text{span}(F \uplus \{a\})$  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

$$(13.3) \quad \tilde{f}(x + \lambda a) = \tilde{f}(x) + \lambda \tilde{f}(a) = f(x) + \lambda k.$$

□

**Theorem 13.4** (Hahn–Banach extension theorem). Let  $V$  be a vector space and  $p : V \rightarrow \mathbb{R}$  a sublinear function. Suppose  $F$  is a (linear) subspace of  $V$  and  $f : F \rightarrow \mathbb{R}$  is a linear mapping with  $f \leq p$  on  $F$ . Then there is an extension of  $f$  to a linear map  $\tilde{f} : V \rightarrow \mathbb{R}$  such that  $\tilde{f} \leq p$  on  $V$ .

Before proving this theorem, first we prove two lemmata.

**Lemma 13.3.** Suppose  $F$  is a subspace of  $V$ ,  $f : F \rightarrow \mathbb{R}$  is a linear mapping and  $a \in V \setminus F$ . Let  $k \in \mathbb{R}$  and

$$(13.4) \quad \tilde{f}(x + \lambda a) := f(x) + \lambda k, \quad \text{i.e., } k = \tilde{f}(a)$$

be the linear extension of  $f$  to  $\text{span}(F \uplus \{a\})$  (see prop.10.9 on p.187, applied to  $V' = \mathbb{R}$  and  $y_0 = k$ ). Then

$$(13.5) \quad k \leq \inf_{u \in F} \{p(u + a) - f(u)\} \Leftrightarrow \tilde{f}(x + \lambda a) \leq p(x + \lambda a) \text{ for all } \lambda > 0 \text{ and } x \in F,$$

$$(13.6) \quad k \geq \sup_{v \in F} \{f(v) - p(v - a)\} \Leftrightarrow \tilde{f}(x + \lambda a) \leq p(x + \lambda a) \text{ for all } \lambda < 0 \text{ and } x \in F.$$

Further,

$$(13.7) \quad \sup_{v \in F} \{f(v) - p(v - a)\} \leq k \leq \inf_{u \in F} \{p(u + a) - f(u)\} \\ \Leftrightarrow \tilde{f}(x + \lambda a) \leq p(x + \lambda a) \text{ for all } \lambda \in \mathbb{R} \text{ and } x \in F,$$

**Proof of (13.5),  $\Rightarrow$ :** 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 side of (13.5),

$$\begin{aligned} \tilde{f}(x + \lambda a) &= f(x) + \lambda k = \lambda(f(x/\lambda) + k) = \lambda(f(u) + k) \leq \lambda\left(f(u) + (p(u + a) - f(u))\right) \\ &= \lambda p(u + a) = \lambda p(x/\lambda + a) = p(x + \lambda a) \end{aligned}$$

The inequality follows from the left hand side of (13.5) and we used the positive homogeneity of  $p$  for the last equation.

**Proof of (13.6),  $\Rightarrow$ :** 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 side of (13.6) and  $\lambda < 0$  and positive homogeneity of  $p$ ,

$$\begin{aligned} k \geq f(v) - p(v - a) &\Rightarrow \lambda k \leq f(\lambda v) - \lambda p(v - a) \\ &\Rightarrow -f(\lambda v) + \lambda k \leq (-\lambda)p(v - a) = p((-\lambda)(v - a)) = p((-\lambda)v + \lambda a). \end{aligned}$$

Since  $x = \lambda v$  and linearity of  $f$  implies  $-f(x) = f(-x)$ ,

$$f(-x) + \lambda k = -f(x) + \lambda k \leq p(-x + \lambda a), \text{ hence } \tilde{f}(-x + \lambda a) = f(-x) + \lambda k \leq p(-x + \lambda a)$$

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 (13.5),  $\Leftarrow$ ):** we assume  $\tilde{f}(x + \lambda a) \leq p(x + \lambda a)$  for all  $\lambda > 0$  and  $x \in F$ . We now show that  $k = \tilde{f}(a) \leq p(u + a) - f(u)$  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 (13.6),  $\Leftarrow$ ):** 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  $v \in F$ .

$$-p(v - a) + f(v) \leq -\tilde{f}(v - a) + f(v) = \tilde{f}(a - v) + f(v) = \tilde{f}(a) - \tilde{f}(v) + f(v) = \tilde{f}(a) = k.$$

**Proof of (13.7),  $\Rightarrow$ ):** Let us assume that  $\lambda \in \mathbb{R}$  and  $x \in F$ . For  $\lambda \neq 0$ , validity of the right hand side of (13.7) follows from (13.5) and (13.6). 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 (13.7),  $\Leftarrow$ ):** This is immediate from (13.5) and (13.6). ■

**Lemma 13.4.** Let  $V$  be a vector space and  $p : V \rightarrow \mathbb{R}$  a sublinear function. Let  $F \subset V$  be a genuine subspace of  $V$  and  $a \in V \setminus F$ . Let  $f : F \rightarrow \mathbb{R}$  be a linear mapping with  $f \leq p$  on  $F$ . Let  $G := \text{span}(F \uplus \{a\})$  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  $\sup_{v \in E} \{f(v) - p(v - a)\} = \inf_{u \in E} \{p(u + a) - f(u)\}$ .

**Proof of a.** For  $u, v \in F$  we have

$$f(u) + f(v) = f(u + v) \leq p(u + v) = p((u + a) + (v - a)) \leq p(u + a) + p(v - a),$$

and hence  $f(v) - p(v - a) \leq p(u + a) - f(u)$ . Therefore

$$(13.8) \quad \sup_{v \in F} \{f(v) - p(v - a)\} \leq \inf_{u \in F} \{p(u + a) - f(u)\}.$$

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

$$(13.9) \quad \sup_{v \in F} \{f(v) - p(v - a)\} \leq k \leq \inf_{u \in F} \{p(u + a) - f(u)\}$$

which will conclude the proof of **a** since such a  $k$  exists by 13.8.

Our claim holds because  $f(x) + \lambda k = \tilde{f}(x + \lambda a) \leq p(x + \lambda a)$  for all  $\lambda$  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 13.3.

**Proof of b.** From (13.9) we deduce that  $k$  is unique if and only if

$$\sup_{v \in E} \{f(v) - p(v - a)\} = \inf_{u \in E} \{p(u + a) - f(u)\}.$$

This exactly the case where there is a unique  $\tilde{f}$  which is defined as  $\tilde{f}(x + \lambda a) = f(x) + \lambda k$ . ■

PROOF of thm.13.4 (Hahn-Banach):

Let  $\mathcal{G} = \{(V, g) : V \text{ is a subspace of } E, V \supseteq F, g : V \rightarrow \mathbb{R} \text{ is linear, } g|_F = f, \text{ and } g \leq p \text{ on } V\}$ .

We define a partial order " $\preceq$ " on  $\mathcal{G}$  as follows:

$$(13.10) \quad (V_1, g_1) \preceq (V_2, g_2) \Leftrightarrow V_1 \subseteq V_2 \text{ and } g_2 \text{ is an extension of } g_1.$$

Note that  $\mathcal{G}$  is not empty because  $f \in \mathcal{G}$ .

**a.** We first prove that any chain  $\mathcal{C} \subseteq (\mathcal{G}, \preceq)$  has an upper bound: Let  $W := \bigcup [V : V \in \mathcal{C}]$ . Then  $W$  is a subspace of  $E$  because if  $x, y \in W$  and  $\lambda \in \mathbb{R}$  then there are  $(V_1, g_1), (V_2, g_2) \in \mathcal{C}$  such that  $x \in V_1$  and  $y \in V_2$ . Because  $\mathcal{C}$  is a chain we have  $V_1 \subseteq V_2$  or  $V_2 \subseteq V_1$ , say,  $V_1 \subseteq V_2$ . It follows that  $x, y \in V_2$ . But  $V_2$  is a subspace of  $E$  and we conclude that  $x + \lambda y \in V_2$ , hence  $x + \lambda y \in W$ . It follows that  $W$  is a subspace of  $E$ .

Next we construct a linear  $h : W \rightarrow \mathbb{R}$  such that  $h \leq p$  on  $W$  and  $h|_V = g$  for all  $(V, g) \in \mathcal{C}$ , i.e.,  $h$  is a linear extension of  $g$  for all  $(V, g) \in \mathcal{C}$ . If we find such  $h$  then it follows that  $(W, h) \in \mathcal{G}$  and  $(W, h)$  is an upper bound of  $\mathcal{C}$ .

Let  $x \in W$ . Then  $x \in V_1$  for some  $(V_1, g_1) \in \mathcal{C}$ . We define  $h(x) := g_1(x)$ . This assignment is unambiguous because if  $x \in V_2$  for some other  $(V_2, g_2) \in \mathcal{C}$  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  $(V_1, g_1) \in \mathcal{G}$  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 of  $E$  to conclude that both  $x, y$  belong to some subspace  $V$  of  $E$  such that  $(V, g) \in \mathcal{C}$ . 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  $(W, h) \in \mathcal{G}$ . Let  $(V, g) \in \mathcal{C}$ . Clearly,  $V \subseteq W = \bigcup [U : U \in \mathcal{C}]$ . We have seen that  $h$  is linear, dominated by  $p$ , and is constructed in such a manner that  $h(x) = g(x)$  for all  $x \in V$ . It follows that  $(W, h) \succeq (V, g)$  for all  $(V, g) \in \mathcal{C}$  and we have proved that  $\mathcal{C}$  has an upper bound in  $(\mathcal{G}, \preceq)$ .

**b.** It follows that we can apply Zorn's Lemma (axiom 13.1 on p.278) and hence conclude that  $\mathcal{G}$  possesses a maximal element  $(F', f')$ .

We show that  $F' = E$ . Assume to the contrary that there exists  $a \in E \setminus F'$ . According to lemma 13.4 on p.284, we can extend  $f'$  to a linear function  $\tilde{f}$  on  $F' \uplus \{a\}$  in such a fashion that  $\tilde{f} \leq p$ . It follows that  $(F' \uplus \{a\}, \tilde{f}) \in \mathcal{G}$ . Further,  $(F', f') \not\preceq (F' \uplus \{a\}, \tilde{f})$  because  $(F' \subsetneq F' \uplus \{a\})$ . This contradicts the maximality of  $(F', f')$  and we have reached a contradiction. Thus  $F' = E$ .

The proof is finished:  $f'$  is the linear extension of  $f$  to  $E$  that we were looking for. ■

**Corollary 13.1** (Continuous extensions of continuous linear functions). Let  $(V, \|\cdot\|)$  be a normed vector space. Let  $F$  be a (linear) subspace of  $V$  and let  $f : F \rightarrow \mathbb{R}$  be a continuous linear mapping on  $F$ . Then there is an extension of  $f$  to a continuous linear map  $\tilde{f} : V \rightarrow \mathbb{R}$ .

PROOF: Let  $p(x) := \|f\| \cdot \|x\|$  (see def.11.35 on p.242). Because  $p$  is a positive multiple of the norm  $\|\cdot\|$  on all of  $V$ , it also is a norm on  $V$  (see prop.10.14) on p.191), hence sublinear by example 13.3 on p.282.

According to the Hahn-Banach extension theorem there exists a linear extension  $\tilde{f}$  of  $f$  to all of  $V$  such that

$$(13.11) \quad \tilde{f}(x) \leq p(x) \text{ for all } x \in V.$$

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 (13.11) this shows that

$$-p(x) \leq \tilde{f}(x) \leq p(x) \text{ for all } x \in V$$

and thus

$$(13.12) \quad |\tilde{f}(x)| \leq p(x) = \|f\| \cdot \|x\| \text{ for all } x \in V$$

It follows from (13.12) that  $f$  has been extended in such a way that  $\|\tilde{f}\| \leq \|f\|$ . We apply the continuity criterion for linear functions (thm.11.15 on p.243) 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  $\tilde{f}$ . ■

### 13.6 Convexity (★)

*Note that this chapter is starred, hence optional.*

**Definition 13.4** (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  $\alpha$  and  $\beta$ . Let  $f : I \rightarrow \mathbb{R}$ .

- The **epigraph** of  $f$  is the set  $\text{epi}(f) := \{(x_1, x_2) \in I \times \mathbb{R} : x_2 \geq f(x_1)\}$  of all points in the plane that lie above the graph of  $f$ .
- $f$  is **convex** if for any two vectors  $\vec{a}, \vec{b} \in \text{epi}(f)$  the entire line segment  $S := \{\lambda \vec{a} + (1 - \lambda)\vec{b} : 0 \leq \lambda \leq 1\}$  is contained in  $\text{epi}(f)$ .
- 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 non-decreasing. □

**Proposition 13.2** (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

$$(13.13) \quad f(x_0) \leq S(x_0).$$

Note that any  $x_0$  between  $a$  and  $b$  can be written as  $x_0 = \lambda a + (1 - \lambda)b$  for some  $0 \leq \lambda \leq 1$  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  $0 \leq \lambda \leq 1$  it is true that

$$(13.14) \quad f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b).$$



PROOF of “ $\Rightarrow$ ”: Any line segment  $S$  that connects the points  $(a, f(a))$  and  $(b, f(b))$  in such a way that is entirely contained in the epigraph of  $f$  will satisfy  $f(x_0) \leq S(x_0)$  for all  $a \leq x_0 \leq b$  and it follows that convexity of  $f$  implies (13.13).

PROOF of “ $\Leftarrow$ ”: Let (13.13) be valid for all  $c, d \in I$ . Let  $\vec{a} = (a_1, a_2), \vec{b} = (b_1, b_2) \in \text{epi}(f)$ , i.e.,

$$(13.15) \quad a_2 \geq f(a_1) \text{ and } b_2 \geq f(b_1).$$

We must show that the entire line segment  $S := \{\lambda\vec{a} + (1 - \lambda)\vec{b} : 0 \leq \lambda \leq 1\}$  is contained in  $\text{epi}(f)$ .

Let  $\vec{a}' := (a_1, f(a_1))$ . Let  $S' := \{\lambda\vec{a}' + (1 - \lambda)\vec{b} : 0 \leq \lambda \leq 1\}$  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  $S'' := \{\lambda\vec{a}' + (1 - \lambda)\vec{b}'' : 0 \leq \lambda \leq 1\}$  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  $T(\cdot) : [a_1, b_1] \rightarrow \mathbb{R}$  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  $a \leq x_0 \leq b$  we have  $f(x_0) \leq S''(x_0)$ . We conclude from  $S(\cdot) \geq S'(\cdot) \geq S''(\cdot)$  that  $S(x_0) \geq f(x_0)$ , i.e.  $(x_0, S(x_0)) \in \text{epi}(f)$ . As this is true for any  $a \leq x_0 \leq b$  it follows that the line segment  $S$  is entirely contained in the epigraph of  $f$ . ■

**Example 13.5** (Sublinear functions are convex). Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be sublinear. Then  $f$  is convex.

Let  $a, b \in \mathbb{R}$  and  $0 \leq \lambda \leq 1$ . It follows from subadditivity and positive homogeneity of  $f$  that

$$f(\lambda a + (1 - \lambda)b) \leq f(\lambda a) + f((1 - \lambda)b) = \lambda f(a) + (1 - \lambda)f(b).$$

According to prop.13.2 this implies convexity of  $f$ . □

**Proposition 13.3** (Convex vs concave-up). Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  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

$$(13.16) \quad m = \frac{S(b) - S(x_0)}{b - x_0} > \frac{S(b) - f(x_0)}{b - x_0} = \frac{f(b) - f(x_0)}{b - x_0} = f'(\xi)$$

for some  $x_0 < \xi < b$  (according to the mean value theorem for derivatives). Further

$$(13.17) \quad m = \frac{S(x_0) - S(a)}{x_0 - a} < \frac{f(x_0) - S(a)}{x_0 - a} = \frac{f(x_0) - f(a)}{x_0 - a} = f'(\eta)$$

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 (13.16) and (13.17) we obtain

$$m < f'(\eta) \leq f'(x_0) \leq f'(\xi) < m$$

and we have reached a contradiction. ■

**Proposition 13.4** (Sublinear functions are convex). *Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be sublinear. Then  $f$  is concave-up.*

PROOF: Let  $0 \leq \lambda \leq 1$  and  $x, y \in \mathbb{R}$ . Then

$$(13.18) \quad p(\lambda x + (1 - \lambda)y) \leq p(\lambda x) + p((1 - \lambda)y) = \lambda p(x) + (1 - \lambda)p(y)$$

and it follows that  $f$  is concave-up. ■

### 13.7 Exercises for Ch.13

**Exercise 13.1.** It was stated in example 13.1 on p.278 that  $(\mathbb{R}, \geq)$  is a linearly ordered set. Prove it. (Prove first that this is a POset.) □

**Exercise 13.2.** Let  $A \subseteq \mathbb{R}^2$  and  $\vec{x}_1 = (x_1, y_1), \vec{x}_2 = (x_2, y_2) \in A$ . Let

$$\vec{x}_1 \preceq \vec{x}_2 \Leftrightarrow x_1 \leq x_2 \text{ and } y_1 \leq y_2.$$

- Prove that " $\preceq$ " defines a partial order on  $A$ .
- Prove that no maximal elements exist if  $A = \mathbb{R}^2$ .
- What are the maximal elements of  $A = \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\|_2 = 1\}$ ? □

**Exercise 13.3.** The following was stated at the beginning of ch.13.4 (Extensions of Linear Functions in Arbitrary Vector Spaces) on p.281.

Let  $F \subseteq V$  be a subspace of a vector space  $V$  and let  $f : F \rightarrow \mathbb{R}$  be linear. Let

$$g : F \rightarrow \mathbb{R}; \quad x \mapsto \begin{cases} f(x) & \text{if } x \in F, \\ 0 & \text{if } x \in F^c. \end{cases}$$

Then  $g$  is linear only if  $f(x) = 0$  for all  $x \in F$ . □



## 14 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 [1] B/G forms a group.

Note that this chapter is starred and hence optional.

### 14.1 More on Groups (★)

This chapter adds to the material of ch.3.1 (Semigroups and Groups) on p.27.

**Example 14.1.** 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  $(\mathbb{Z}, \cdot)$  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  $\mathbb{Q}^*$  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  $\mathbb{R}^*$  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  $\mathbb{C}^*$  of all non-zero complex numbers is an abelian group.  $\square$

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 14.2.** Let  $(G, \diamond)$  and  $(H, \bullet)$  be defined as follows:

$$(14.1) \quad G := \{g \in \mathbb{R} : g = e^x \text{ for some } x \in \mathbb{R}\}, \quad e^x \diamond e^y := e^x \cdot e^y = e^{x+y},$$

$$(14.2) \quad H := \{h \in \mathbb{R} : h = \ln(x) \text{ for some } x \in ]0, \infty\}, \quad \ln u \bullet \ln v := \ln u + \ln v = \ln(xy).$$

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  $\varphi$  and  $\psi$  be defined as follows:

$$(14.3) \quad \varphi : (G, \diamond) \rightarrow (H, \bullet), \quad \varphi(g) := \ln g,$$

$$(14.4) \quad \psi : (H, \bullet) \rightarrow (G, \diamond), \quad \psi(h) := e^h.$$

Then  $\varphi$  and  $\psi$  satisfy the following:

$$(14.5) \quad \varphi(g_1 \diamond g_2) = \varphi(g_1) \bullet \varphi(g_2), \quad \varphi(1) = 0, \quad \varphi(g^{-1}) = \varphi(g)^{-1},$$

$$(14.6) \quad \psi(h_1 \bullet h_2) = \psi(h_1) \diamond \psi(h_2), \quad \psi(0) = 1, \quad \psi(h^{-1}) = \psi(h)^{-1}.$$

Further, the functions  $\varphi$  and  $\psi$  are inverse to each other, i.e.,

$$(14.7) \quad \psi(\varphi(g)) = g \quad \text{and} \quad \varphi(\psi(h)) = h$$

for all  $g \in G$  and  $h \in H$ .  $\square$

If you talk about  $\varphi$  and  $\psi$  as “the functions” and  $\diamond$  and  $\bullet$  as “the operations” you might state the results (14.5) and (14.5) as follows:

The functions are **structure compatible** with the operations on their domains and codomains:

*It does not matter whether you first apply the operation to two items in the domain and then apply the function to the result or whether you first map those two items into the codomain and then apply the operation to the two function values. Further, the inverse of the function value is the function value of the inverse and the function maps the neutral element to the neutral element.*

Such functions between two groups have a special name:

**Definition 14.1** (Homomorphisms and isomorphisms). Let  $(G, \diamond)$  and  $(H, \bullet)$  be two groups with neutral elements  $e_G$  and  $e_H$  and let us write  $g^{-1}$  and  $h^{-1}$  for the inverses (in the sense of def. 3.3).

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

$$(14.8) \quad \varphi(g_1 \diamond g_2) = \varphi(g_1) \bullet \varphi(g_2), \quad \varphi(e_G) = e_H, \quad \varphi(g^{-1}) = \varphi(g)^{-1}.$$

Then we call  $\varphi$  a **homomorphism**, more specifically, a **group homomorphism**, from the group  $(G, \diamond)$  to the group  $(H, \bullet)$ .

Let  $\psi : (H, \bullet) \rightarrow (G, \diamond)$  be a group homomorphism from  $(H, \bullet)$  to  $(G, \diamond)$  such that  $\varphi$  and  $\psi$  are inverse to each other.<sup>143</sup>

We call such a bijective homomorphism an **isomorphism**.  $\square$

## 14.2 More on Commutative Rings and Integral Domains

This chapter adds to the material of ch.3.2 (Commutative Rings and Integral Domains) on p.32.

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., [1] Beck/Geoghegan ch.6 for details.

**Theorem 14.1** (Division Algorithm for Integers). See [1] 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

$$(14.9) \quad m = qn + r$$

$$(14.10) \quad 0 \leq r < n$$

$q$  and  $r$  are uniquely determined by  $m$  and  $n$ .

**PROOF:** Will not be given.  $\blacksquare$

<sup>143</sup> They then of course also are bijective functions (see def. 5.9).

**Example 14.3.** 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  $0 \leq 0 < 7$ .  $\square$

## 15 Construction of the Number Systems

### 15.1 The Peano Axioms (★)

*Note that this chapter is starred, hence optional.*

**Definition 15.1** (Set of non-negative integers). We define the set  $\mathbb{N}_0$  (the non-negative integers) axiomatically as follows:

- Ax.1** There is an element “0” contained in  $\mathbb{N}_0$ .
- Ax.2** There is a function  $\sigma : \mathbb{N}_0 \rightarrow \mathbb{N}_0$  such that
  - Ax.2.1**  $\sigma$  is injective,
  - Ax.2.2**  $0 \notin \sigma(\mathbb{N}_0)$  (range of  $\sigma$ ),
  - 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  $\mathbb{N} := \mathbb{N}_0 \setminus \{0\}$ .  $\square$

**Definition 15.2** (Iterative function composition). Let  $X \neq \emptyset$  and  $f : X \rightarrow X$ . We now use the induction axiom above to define  $f^n$  for an arbitrary function  $f : X \rightarrow X$ .

- a.**  $f^0 := \text{id}_X : x \mapsto x$ , **b.**  $f^1 := f$ , **c.**  $f^2 := f \circ f$  (function composition), **c.**  $f^{\sigma(n)} := f \circ f^n$ .  $\square$

**Proposition 15.1.**  $f^n$  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  $f^0 = \text{id}_A$  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  $U = \mathbb{N}_0$ .  $\blacksquare$

**Remark 15.1** ( $\sigma(\cdot)$  as successor function). Of course the meaning of  $\sigma(n)$  will be that of  $n + 1$ :

$$0 \xrightarrow{\sigma} 1 \xrightarrow{\sigma} 2 \xrightarrow{\sigma} 3 \xrightarrow{\sigma} \dots \quad \square$$

**Definition 15.3** (Addition and multiplication on  $\mathbb{N}_0$ ). Let  $m, n \in \mathbb{N}_0$ . Let

$$(15.1) \quad m + n := \sigma^n(m),$$

$$(15.2) \quad m \cdot n := (\sigma^m)^n(0).$$

Note that we know the meaning of  $(\sigma^m)^n$ :  $f := \sigma^m$  is a function  $A \rightarrow A$  and we have established in prop.15.1 the meaning of  $f^n$ , i.e.,  $(\sigma^m)^n$ .  $\square$

**Proposition 15.2.** Addition and multiplication satisfy all rules of arithmetic we learned in high school such as

$$(15.3) \quad m + n = n + m \quad \text{commutativity of addition}$$

$$(15.4) \quad k + (m + n) = (k + m) + n \quad \text{associativity of addition}$$

$$(15.5) \quad m \cdot n = n \cdot m \quad \text{commutativity of multiplication}$$

$$(15.6) \quad k \cdot (m \cdot n) = (k \cdot m) \cdot n \quad \text{associativity of multiplication}$$

$$(15.7) \quad k \cdot (m + n) = k \cdot m + k \cdot n \quad \text{distributivity of addition}$$

$$(15.8) \quad n \cdot 1 = 1 \cdot n = n \quad \text{neutral element for multiplication}$$

$$(15.9)$$

Here 1 is defined as  $1 = \sigma(0)$ .

PROOF: Drudge work. ■

**Definition 15.4** (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 15.3.** “ $<$ ” 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. ■

## 15.2 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 15.5** (Integers as equivalence classes). We define the following equivalence relation  $(m_1, n_1) \sim (m_2, n_2)$  on the cartesian product  $\mathbb{N}_0 \times \mathbb{N}_0$ :

$$(15.10) \quad (m_1, n_1) \sim (m_2, n_2) \Leftrightarrow m_1 + n_2 = n_1 + m_2$$

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 (15.10).

We “embed”  $\mathbb{N}_0$  into  $\mathbb{Z}$  with the following injective function  $e : \mathbb{N}_0 \rightarrow \mathbb{Z}$ :  $e(m) := [(m, 0)]$ .

From this point forward we do not distinguish between  $\mathbb{N}_0$  and its image  $e(\mathbb{N}_0) \subseteq \mathbb{Z}$  and we do not distinguish between  $\mathbb{N}$  and its image  $e(\mathbb{N}) \subseteq \mathbb{Z}$ . In particular we do not distinguish between the two zeroes 0 and  $[(0, 0)]$  and between the two ones 1 and  $[(1, 0)]$ .

Finally we write  $-n$  for the integer  $[(0, n)]$ . □

With those abbreviation we then obtain

**Proposition 15.4** (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 15.2. 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 (15.10) 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 15.6** (Addition, multiplication and subtraction on  $\mathbb{Z}$ ). Let  $[(m_1, n_1)]$  and  $[(m_2, n_2)] \in \mathbb{Z}$ . We define

$$(15.11) \quad -[(m_1, n_1)] := [n_1, m_1],$$

$$(15.12) \quad [(m_1, n_1)] + [(m_2, n_2)] := [(m_1 + m_2, n_1 + n_2)]$$

$$(15.13) \quad [(m_1, n_1)] \cdot [(m_2, n_2)] := [(m_1 m_2 + n_1 n_2, m_1 n_2 + n_1 m_2)]$$

We write  $[(m_1, n_1)] - [(m_2, n_2)]$  (“ $[(m_1, n_1)]$  minus  $[(m_2, n_2)]$ ”) as an abbreviation for  $[(m_1, n_1)] + (-[(m_2, n_2)])$ .

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  $\mathbb{Z}_{\geq 0} = \mathbb{N}_0$ .  $\square$

*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 15.5.** *Let  $m, n \in \mathbb{N}_0$ . Then*

$$(15.14) \quad [(m, n)] + [(0, 0)] = [(0, 0)] + [(m, n)] = [(m, n)],$$

$$(15.15) \quad (-[(m, n)]) + [(m, n)] = [(m, n)] + (-[(m, n)]) = [(0, 0)]$$

$$(15.16) \quad [(m, n)] \cdot [(1, 0)] = [(1, 0)] \cdot [(m, n)] = [(m, n)],$$

*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 15.3.** Again, if it had been possible to define subtraction  $m - n$  for all  $m, n \in \mathbb{N}_0$  then it would be easier to see why addition and multiplication have been defined as you see it in def.15.6:

Addition is defined such that  $(m_1 - n_1) + (m_2 - n_2) = (m_1 + m_2) - (n_1 + n_2)$   
and multiplication:  $(m_1 - n_1) \cdot (m_2 - n_2) = (m_1 m_2 + (-n_1)(-n_2)) - (m_1 n_2 + n_1 m_2)$ .  $\square$

### 15.3 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 15.7** (Fractions as equivalence classes). We define the following equivalence relation  $(p, q) \sim (r, s)$  on the cartesian product  $\mathbb{Z} \times \mathbb{Z}_{\neq 0}$ :

$$(15.17) \quad (p, q) \sim (r, s) \Leftrightarrow p \cdot s = q \cdot r$$

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 (15.17).

We “embed”  $\mathbb{Z}$  into  $\mathbb{Q}$  with the injective function  $e : \mathbb{Z} \rightarrow \mathbb{Q}$  defined as  $e(z) := [(z, 1)]$ .  $\square$

**Remark 15.4. 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 (15.17) 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.  $\square$

**Definition 15.8** (Addition, multiplication, subtraction and division in  $\mathbb{Q}$ ). Let  $[(p_1, q_1)]$  and  $[(p_2, q_2)] \in \mathbb{Q}$ . We define

$$(15.18) \quad -[(p_1, q_1)] := [(-p_1, q_1)],$$

$$(15.19) \quad [(p_1, q_1)] + [(p_2, q_2)] := [(p_1q_2 + q_1p_2, q_1q_2)]$$

$$(15.20) \quad [(p_1, q_1)] - [(p_2, q_2)] := [(p_1, q_1)] + (-[(p_2, q_2)])$$

$$(15.21) \quad [(p_1, q_1)] \cdot [(p_2, q_2)] := [(p_1p_2, q_1q_2)]$$

$$(15.22) \quad [(p_1, q_1)]^{-1} := [(1, 1)]/[(p_1, q_1)] := [(q_1, p_1)] \text{ (if } p_1 \neq 0),$$

$$(15.23) \quad [(p_1, q_1)]/[(p_2, q_2)] := [(p_1q_2, q_1p_2)] = [(p_1, q_1)] \cdot [(p_2, q_2)]^{-1} \text{ (if } p_2 \neq 0) \quad \square$$

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. <sup>144</sup>

Further we have

**Proposition 15.6** (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.  $\blacksquare$

## 15.4 Constructing the Real Numbers via Dedekind Cuts

The material presented here, including the notation, follows [10] Rudin, Walter: Principles of Mathematical Analysis.

Note that in this section small greek letters denote **sets** of rational numbers!

The idea behind real numbers as intervals of rational numbers with no lower bounds, called Dedekind cuts, is as follows:

<sup>144</sup> This was shown for multiplication  $[(p_1, q_1)] \cdot [(p_2, q_2)] = [(p_1p_2, q_1q_2)]$  in exercise 5.10 on p.107.

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

$$(15.24) \quad \Phi : x \mapsto \Phi(x) := \{q \in \mathbb{Q} : q < x\}$$

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  $\Phi$  for free if we take as codomain the set of all cuts. Because  $\Phi$  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 15.9** (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  $\alpha^c \neq \emptyset$
- b. Let  $p, q \in \mathbb{Q}$  such that  $p \in \alpha$  and  $q < p$ . Then  $q \in \alpha$ .
- c.  $\alpha$  does not have a max:  $\forall p \in \alpha \exists q \in \alpha$  such that  $p < q$ .

Given a cut  $\alpha$ , let  $p \in \alpha$  and  $q \in \alpha^c$ . We call  $p$  a **lower number** of the cut  $\alpha$  and we call  $q$  an **upper number** of  $\alpha$ .  $\square$

**Theorem 15.1.** (Rudin thm.1.5)

Let  $\alpha \subseteq \mathbb{Q}$  be a cut. Let  $p \in \alpha, q \in \alpha^c$ . 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  $\alpha$  or both belong to its complement, a contradiction to our assumption. Or we have  $q < p$ . It then follows from  $p \in \alpha$  and def.15.9.b that  $q \in \alpha$ , contrary to our assumption.  $\blacksquare$

**Theorem 15.2.** (Rudin thm.1.6)

Let  $r \in \mathbb{Q}$ . Let  $r^* := \{p \in \mathbb{Q} : p < r\}$ . Then  $r^*$  is a cut and  $r = \min((r^*)^c)$ .

PROOF: In the following let  $p, q, r \in \mathbb{Q}$ .

PROOF of def.15.9.a:  $r - 1 < r \Rightarrow r - 1 \in r^* \Rightarrow r^* \neq \emptyset$ . Further,  $r \in (r^*)^c \Rightarrow (r^*)^c \neq \emptyset$ .

PROOF of def.15.9.b: Let  $q < p$  and  $p \in r^*$ . Then also  $q \in r^* = \{p' \in \mathbb{Q} : p' < r\}$ .

PROOF of def.15.9.c: Let  $p \in r^*$ . Then  $p < (p+r)/2 < r$ , hence  $(p+r)/2 \in r^*$  and  $r$  cannot be the max of  $r^*$ .  $\blacksquare$

**Definition 15.10** (Rational cuts). Let  $r \in \mathbb{Q}$ . The cut  $r^* = \{p \in \mathbb{Q} : p < r\}$  from the previous theorem is called the **rational cut** associated with  $r$ .  $\square$

**Remark 15.5.** If we define intervals in  $\mathbb{Q}$  in the usual way for  $p, q \in \mathbb{Q}$ :

$$]p, q[ := \{r \in \mathbb{Q} : p < r < q\}, \quad [p, q] := \{r \in \mathbb{Q} : p \leq r \leq q\}, \quad \text{etc.}$$

then rational cuts  $r^* (r \in \mathbb{Q})$  are those for which  $r^* = ]-\infty, r[$  and  $(r^*)^c = [r, \infty[$  whereas for non-rational cuts  $\alpha$  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  $\alpha$  as a subset of those real numbers.  $\square$



**Definition 15.11** (Ordering Dedekind cuts). (Rudin def.1.9) Let  $\alpha, \beta$  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.,  $\alpha \subseteq \beta$ .  $\square$

**Proposition 15.7** (Trichotomy of the cuts). (Rudin thm.1.10)

Let  $\alpha, \beta$  be two cuts. Then either  $\alpha < \beta$  or  $\alpha > \beta$  or  $\alpha = \beta$ .

*PROOF:* We only need to show that if  $\alpha \not\subseteq \beta$  then  $\beta \subsetneq \alpha$ .

So let  $\alpha \not\subseteq \beta$ . 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 def.15.9.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.,  $\beta \subsetneq \alpha$ .  $\blacksquare$

**Theorem 15.3** (Addition of two cuts). (Rudin thm.1.12) Let  $\alpha, \beta$  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  $0^*$ , the rational cut corresponding to  $0 \in \mathbb{Q}$ ) and a suitably defined cut  $-\alpha$  for a given cut  $\alpha$  which satisfies  $\alpha + (-\alpha) = (-\alpha) + \alpha = 0^*$

Having defined negatives  $-\alpha$  for all cuts we then also can define their absolute values

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

*PROOF:* Not given here.  $\blacksquare$

**Theorem 15.4** (Multiplication of two cuts). Let  $\alpha \geq 0^*, \beta \geq 0^*$  be two non-negative cuts. Let

$$\alpha \cdot \beta := \begin{cases} \{q \in \mathbb{Q} : q < 0\} \cup \{ab : a \in \alpha, b \in \beta\} & \text{if } \alpha \geq 0^*, \beta \geq 0^*, \\ -|\alpha| \cdot |\beta| & \text{if } \alpha < 0^*, \beta \geq 0^* \text{ or } \alpha \geq 0^*, \beta < 0^*, \\ |\alpha| \cdot |\beta| & \text{if } \alpha < 0^*, \beta < 0^*. \end{cases}$$

Then the set  $\alpha \cdot \beta$  is a cut, called the product of  $\alpha$  and  $\beta$ .

It can be proved that for each cut  $\alpha \neq 0^*$  there is a cut  $\alpha^{-1}$  uniquely defined by the equation  $\alpha \cdot \alpha^{-1} = 1^*$ .

**Theorem 15.5** (The set of all cuts forms a field). 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  $0^*$  is the neutral element for addition and the cut  $1^*$  is the neutral element for multiplication.

$-\alpha$  is the additive inverse of any cut  $\alpha$  and  $\alpha^{-1}$  is the multiplicative inverse of  $\alpha \neq 0^*$ .

Further the set  $\mathbb{R}_{>0} := \{\alpha \in \mathbb{R} : \alpha > 0^*\}$  satisfies B/G axiom 8.26.

*PROOF:* It follows from prop.15.7 on p.297 that  $\mathbb{R}_{>0}$  satisfies B/G axiom 8.26. Proofs of the other properties of  $\mathbb{R}$  are not given here.  $\blacksquare$

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 15.6.** (Rudin thm.1.29) Let  $\alpha, \beta \in \mathbb{R}$  and let  $\alpha < \beta$ . Then there exists  $q \in \mathbb{Q}$  such that  $\alpha < q^* < \beta$

PROOF: Any  $q \in \beta \setminus \alpha$  will do. ■

**Theorem 15.7.** (Rudin thm.1.30) Let  $\alpha \in \mathbb{R}, p \in \mathbb{Q}$ . Then  $p \in \alpha \Leftrightarrow p^* < \alpha$ , i.e.,  $p^* \subsetneq \alpha$

PROOF of  $\Leftarrow$ ): Let  $p \in \alpha$ . it follows for any  $q \in p^*$  that  $q < p \in \alpha$ , hence  $q \in \alpha$ , hence  $p^* \subseteq \alpha$ . As  $p \notin p^* = \{p' \in \mathbb{Q} : p' < p\}$  but  $p \in \alpha$  we have strict inclusion  $p^* \subsetneq \alpha$ .

PROOF of  $\Rightarrow$ ): As  $p^* \subsetneq \alpha$  there exists  $q \in \alpha \setminus p^*$ . As  $q \geq p$  and  $q \in \alpha$  we obtain  $p \in \alpha$  from def.15.9.b. ■

**Theorem 15.8** (Dedekind's Theorem). (Rudin thm.1.32) Let  $\mathbb{R} = A \uplus B$  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  $\gamma$ .

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.15.6 on p.298 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  $\gamma' \in A = B^c$ . We have reached a contradiction and conclude that  $\gamma$  must be unique.

PROOF of existence of  $\gamma$ : Let  $\gamma := \bigcup [\alpha : \alpha \in A]$ .

Step 1: We now show that  $\gamma$  is a cut.

We first show that def.15.9.a is satisfied. As  $B \neq \emptyset$  there is some  $\beta \in B$ . As  $\beta^c \neq \emptyset$  there is some  $q \in \beta^c$ . It follows from  $\alpha \subseteq \beta$  for all  $\alpha \subseteq \gamma = \bigcup [\alpha : \alpha \in A]$  that  $\gamma \subseteq \beta$ , hence  $\gamma^c \supseteq \beta^c$ . It follows from  $q \in \beta^c$  that  $q \in \gamma^c$ , hence  $\gamma^c \neq \emptyset$ . Further, it follows from  $A \neq \emptyset$  that  $\gamma \neq \emptyset$ . We conclude that def.15.9.a is satisfied.

Next we show the validity of def.15.9.b. Let  $p \in \gamma$ , i.e.,  $p \in \alpha_0$  for some  $\alpha_0 \in A$ . Let  $q < p$ . Then  $q \in \alpha_0 \subseteq \bigcup [\alpha : \alpha \in A]$ , i.e.,  $q \in \gamma$ . We conclude that def.15.9.b is satisfied.

Now we show the validity of def.15.9.c. Let  $p \in \gamma$ , i.e.,  $p \in \alpha_0$  for some  $\alpha_0 \in A$ . As the cut  $\alpha_0$  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  $\gamma$  does not have a max and this shows that def.15.9.c is satisfied. We conclude that  $\gamma$  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  $\gamma := \bigcup [\alpha : \alpha \in A]$ .

To show that  $\gamma \leq \beta$  for all  $\beta \in B$  we prove that the opposite statement that

$$(15.25) \quad \gamma > \beta, \text{ i.e., } \gamma \setminus \beta \neq \emptyset \text{ for some cut } \beta \in B$$

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  $\alpha_0 \not\leq \beta$  even though  $\alpha_0 \in A$  and  $\beta \in B$ , contrary to the assumptions about the partitioning  $A \uplus B$  of  $\mathbb{R}$ . ■

**Corollary 15.1.** 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.,  $\alpha \not\subseteq \beta$ ).

Then either  $\max(A)(= l.u.b.(A))$  exists or  $\min(B)(= g.l.b.(B))$  exists.

PROOF: According to thm.15.8 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  $\gamma$  is an upper bound of  $\alpha$  and a lower bound of  $\beta$ . 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 15.9** (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  $A := B^c = \{\alpha \in \mathbb{R} : \alpha < x \text{ i.e., } \alpha \not\subseteq x \text{ for some } x \in E\}$ . In other words,  $\alpha \in A$  if and only if  $\alpha$  is not an upper bound of  $E$ .

$A$  is not empty either: As  $E \neq \emptyset$  there is some  $x \in E$ . Let  $\alpha := x - 1$ . Clearly  $x \leq \alpha$  is not true for all  $x \subseteq E$ . It follows that  $\alpha$  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  $\beta$ , 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.15.8). 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 that the assumption  $\gamma \in A$  leads to a contradiction. As  $\gamma$  is not an upper bound of  $A$  there exists  $x \in E$  such that  $\gamma < x$ . According to thm.15.6 on p. 298 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.15.1 that  $\gamma = \min(B)$ , i.e.,  $\gamma = \sup(E)$ . ■

## 15.5 Constructing the Real Numbers via Cauchy Sequences

This chapter was created after discussions with Nguyen-Phan Tam about teaching the Math 330 course: she plans to construct the real numbers from the rationals by means of equivalence classes of Cauchy sequences in  $\mathbb{Q}$ .

In the following we always assume that  $i, j, k, m, n \in \mathbb{N}$ ,  $\varepsilon, p, q, r, s, p_n, p_{i,j}, \dots \in \mathbb{Q}$ ,  $x, y, z, x_n, x_{i,j}, \dots \in \mathbb{R}$ .

- a. def. convergence in  $\mathbb{Q}$ :  $\lim_{n \rightarrow \infty} q_n = q \Leftrightarrow \forall \text{ pos. } \varepsilon \in \mathbb{Q} \exists N \in \mathbb{Q} \text{ such that if } n \geq N \text{ then } |q_n - q| < \varepsilon$ .
- b. def. Cauchy seqs. in  $\mathbb{Q}$ :  $(q_n)_n$  is Cauchy  $\Leftrightarrow \forall \text{ pos. } \varepsilon \in \mathbb{Q} \exists N \in \mathbb{Q} \text{ such that if } i, j \geq N \text{ then } |q_i - q_j| < \varepsilon$ .
- c. Let  $\mathcal{C} := \{ \text{all Cauchy sequences in } \mathbb{Q} \}$ . For  $(q_n)_n, (r_n)_n$  we define  $(q_n)_n \sim (r_n)_n$  iff  $\lim_{n \rightarrow \infty} (r_n - q_n) = 0$ .

- d.** Let  $q \in \mathbb{Q}$  and  $q_n := q \forall n$ . Write  $q$  for  $[(q_n)_n]$ .
- e.** Let  $\mathbb{R} := \mathcal{C}/\sim$ . Show that for  $[(p_n)], [(q_n)] \in \mathcal{C}$  the operations  $([(p_n)_n], [(q_n)_n]) \mapsto [(p_n + q_n)_n]$  and  $([(p_n)_n], [(q_n)_n]) \mapsto [(p_n \cdot q_n)_n]$  are well defined (do not depend on the particular members chosen from the equivalence classes).
- f.** Let  $[(p_n)_n] \neq 0$  (i.e.,  $\lim_n p_n \neq 0$ ), i.e., we may assume  $p_n \neq 0$  for all  $n$ . Show  $-[(q_n)_n] := [(-q_n)_n]$  and  $[(p_n)_n]^{-1} := [(1/p_n)_n]$  are additive and multiplicative inverses
- g1.** Define  $[(p_n)_n] < [(q_n)_n]$  iff  $\exists \varepsilon > 0$  and  $N \in \mathbb{N}$  such that  $q_n - p_n \geq \varepsilon \forall n \geq N$ .
- 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  $\bar{q}_n \rightarrow [(q_j)_j]$
- 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.**  $\lim_n x_n = [(q_n)_n]$ .
- 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  $Q_n := \{i/j : i, j \in \mathbb{Z} \text{ and } j \leq n\}$ . Let  $U_n := U \cap Q_n$ . Let  $u_n := \min(U_n)$  (exists because  $n \cdot U_n \subset \mathbb{Z}$  is bounded below and has a min. Easy to see that  $u_n$  is Cauchy (in  $\mathbb{Q}$  and, because  $\text{distance}(u_n, A) \leq 1/n$ ,  $[(u_n)_n]$  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 [7] Haaser/Sullivan: Real Analysis.

## 16 Measure Theory (★)

*Note that this entire section is starred, hence optional.*

**Introduction:**

The following are the best known examples of measures ( $a_j, b_j \in \mathbb{R}$ ):

$$\text{Length : } \lambda^1([a_1, b_1]) := b_1 - a_1,$$

$$\text{Area : } \lambda^2([a_1, b_1] \times [a_2, b_2]) := (b_1 - a_1)(b_2 - a_2),$$

$$\text{Volume : } \lambda^3([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) := (b_1 - a_1)(b_2 - a_2)(b_3 - a_3).$$

Then there also are probability measures:  $P\{\text{a die shows a 1 or a 6}\} = 1/3$ .

We will explore in this chapter some of the basic properties of measures.

### 16.1 Basic Definitions

**Definition 16.1** (Extended real valued functions).

$$\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\} = \{x \in \mathbb{R} : x \geq 0\} \cup \{+\infty\}$$

is the set of all non-negative real numbers augmented by the element  $\infty$ .

A mapping  $F$  whose codomain is a subset of

$$\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$$

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:

$$(16.1) \quad 0 \cdot \pm\infty = \pm\infty \cdot 0 = 0. \quad \square$$

**Definition 16.2** (Abstract measures). Let  $\Omega$  be a nonempty set and let  $\mathfrak{F}$  be a set that contains some, but not necessarily all, subsets of  $\Omega$ .  $\mathfrak{F}$  is called a  **$\sigma$ -algebra** or  **$\sigma$ -field** for  $\Omega$  if it satisfies the following:

$$(16.2a) \quad \emptyset \in \mathfrak{F} \quad \text{and} \quad \Omega \in \mathfrak{F}$$

$$(16.2b) \quad A \in \mathfrak{F} \quad \implies \quad \complement A \in \mathfrak{F}$$

$$(16.2c) \quad (A_n)_{n \in \mathbb{N}} \in \mathfrak{F} \quad \implies \quad \bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{F} \quad \text{and} \quad \bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{F}$$

The pair  $(\Omega, \mathfrak{F})$  is called a **measurable space**. Note that  $\mathfrak{F}$  is a set whose elements themselves are sets! The elements of  $\mathfrak{F}$  are called **measurable sets**.

A **measure** on  $\mathfrak{F}$  is an extended real valued function

$$\mu(\cdot) : \mathfrak{F} \rightarrow \bar{\mathbb{R}}_+ \quad A \mapsto \mu(A)$$

with the following properties:

$$(16.3a) \quad \mu(\emptyset) = 0$$

$$(16.3b) \quad A, B \in \mathfrak{F} \text{ and } A \subseteq B \implies \mu(A) \leq \mu(B) \quad (\text{monotony})$$

$$(16.3c) \quad (A_n)_{n \in \mathbb{N}} \in \mathfrak{F} \text{ mutually disjoint} \implies \mu\left(\biguplus_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n) \quad (\sigma\text{-additivity})$$

where mutual disjointness means that  $A_i \cap A_j = \emptyset$  for any  $i, j \in \mathbb{N}$  such that  $i \neq j$  (see def.2.4 on p.13). The triplet  $(\Omega, \mathfrak{F}, \mu)$  is called a **measure space**

We call  $\mu$  a **finite measure** on  $\mathfrak{F}$  if  $\mu(\Omega) < \infty$ .

A measure space can support many different measures.

If  $\mu(\Omega) = 1$  then  $\mu(\cdot)$  is called a **probability measure**. Traditionally, mathematicians write  $P(A)$  rather than  $\mu(A)$  for probability measures and the elements of  $\mathfrak{F}$  (the measurable subsets) are thought of as **events** for which  $P(A)$  is interpreted as the probability with which the event  $A$  might happen.  $\square$

**Example 16.1** (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

$$(16.4) \quad \begin{aligned} \lambda^1([a, b]) &:= b - a, \\ \lambda^2([a_1, b_1] \times [a_2, b_2]) &:= (b_1 - a_1)(b_2 - a_2), \\ \lambda^3([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) &:= (b_1 - a_1)(b_2 - a_2)(b_3 - a_3), \\ \lambda^n([a_1, b_1] \times \cdots \times [a_n, b_n]) &:= (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n) \end{aligned}$$

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  $\sigma$ -algebra  $\mathcal{B}^n$  generated by those parallelepipeds <sup>145</sup>  $\lambda^n$  is called  **$n$ -dimensional Lebesgue measure**

Note that Lebesgue measure is not finite.  $\square$

**Example 16.2.** You can easily verify that the following set function defines a measure on an arbitrary nonempty set  $\Omega$  with an arbitrary  $\sigma$ -field  $\mathfrak{F}$ .

$$\mu(\emptyset) := 0; \quad \mu(A) := \infty \text{ if } A \neq \emptyset$$

Keep this example in mind if you contemplate infinity of measures.  $\square$

**Remark 16.1** (Finite disjoint unions). The  $\sigma$ -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 (16.2c) for  $\sigma$ -algebras is required for exactly that reason: you cannot take advantage of the  $\sigma$ -additivity of a measure  $\mu$  if its domain does not contain countable unions and intersections of all its constituents.

<sup>145</sup> This is not entirely correct: we must demand that the measure is  $\sigma$ -finite, i.e., there are measurable sets with finite measure whose union is the entire space. Such is the case for Lebesgue measure: Let  $A_k := [-k, k]^n$ . The union of those sets is  $\mathbb{R}^k$  and  $\lambda^n(A_k) = (2k)^n < \infty$ .

Note that if we have only finitely many sets then “ $\sigma$ -additivity” which stands for “additivity of countably many” becomes simple additivity. We obtain the following by setting  $A_{N+1} = A_{N+2} = \dots = 0$ :

$$(16.5) \quad \begin{aligned} &A_1, A_2, \dots, A_N \in \mathfrak{F} \text{ mutually disjoint} \\ &\Rightarrow \mu(A_1 \uplus A_2 \uplus \dots \uplus A_N) = \mu(A_1) + \mu(A_2) + \dots + \mu(A_N) \quad (\text{additivity}). \end{aligned}$$

In the case of only two disjoint measurable sets  $A$  and  $B$  the above simply becomes

$$\mu(A \uplus B) = \mu(A) + \mu(B). \quad \square$$

*In many circumstances you have a set function on a  $\sigma$ -algebra which behaves like a measure but you can only prove that it is additive instead of  $\sigma$ -additive. You should not be surprised that there is a special name for those “generalized measures”:*

**Definition 16.3** (Contents as additive measures). Let  $\Omega$  be a nonempty set and let  $\mathfrak{F}$  be a  $\sigma$ -algebra for  $\Omega$ .

A **content** on  $\mathfrak{F}$  is a real valued function  $m(\cdot) : \mathfrak{F} \rightarrow \mathbb{R}$ ,  $A \mapsto m(A)$  which satisfies

$$(16.6a) \quad m(\emptyset) = 0 \quad (\text{positivity})$$

$$(16.6b) \quad A, B \in \mathfrak{F} \text{ and } A \subseteq B \Rightarrow m(A) \leq m(B) \quad (\text{monotony})$$

$$(16.6c) \quad A_1, A_2, \dots, A_N \in \mathfrak{F} \text{ mutually disjoint} \Rightarrow m\left(\biguplus_{n=1}^N A_n\right) = \sum_{n=1}^N m(A_n) \quad (\text{additivity}). \quad \square$$

*Note that  $\mu(\Omega) < \infty$  for a content  $\mu$ . After this digression on contents let us go back to measures.*

**Proposition 16.1** (Simple properties of measures). **A** : Let  $A, B, \in \mathfrak{F}$  and let  $\mu$  be a measure on  $\mathfrak{F}$ . Then

$$(16.7a) \quad \mu(A) \geq 0 \quad \text{for all } A \in \mathfrak{F}$$

$$(16.7b) \quad A \subseteq B \Rightarrow \mu(B) = \mu(A) + \mu(B \setminus A)$$

$$(16.7c) \quad \mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$$

*If  $\mu$  is finite then we can write*

$$(16.8a) \quad A \subseteq B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$$

$$(16.8b) \quad \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

**PROOF of A** : The first property follows from the fact that  $\mu(\emptyset) = 0$ ,  $\emptyset \subseteq A$  for all  $A \in \mathfrak{F}$  and (16.3b). To prove the second property, observe that  $B = A \uplus (B \setminus A)$ .

*Proving the third property is more complicated because neither  $A$  nor  $B$  may be a subset of the other. We first note that because  $A \setminus B \subseteq A$ ,  $B \setminus A \subseteq A$  and  $A \cap B \subseteq A$ ,  $\mu(A \cup B) = \infty$  can only be true if  $\mu(A) = \infty$  or  $\mu(B) = \infty$ . In this case (16.7c) is obviously true. Hence we may assume that  $\mu(A \cup B) < \infty$ . We have*

$$(16.9a) \quad A \cup B = (A \cap B) \uplus (B \setminus A) \uplus (A \setminus B)$$

$$(16.9b) \quad A \cup B = A \uplus (B \setminus A) = B \uplus (A \setminus B)$$

*It follows from (16.9a) that*

$$(16.10) \quad \mu(A \cup B) = \mu(A \cap B) + \mu(B \setminus A) + \mu(A \setminus B)$$

It follows from (16.9b) that

$$(16.11) \quad 2 \cdot \mu(A \cup B) = \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B)$$

We subtract the left and right sides of (16.10) from those of (16.11) and obtain

$$\begin{aligned} \mu(A \cup B) &= \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) - \mu(A \cap B) - \mu(B \setminus A) - \mu(A \setminus B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B) \end{aligned}$$

and the third property is proved. ■

**Proposition 16.2** (Minimal sigma-algebras). *Let  $\Omega$  be a nonempty set.*

**A:** *The intersection of arbitrarily many  $\sigma$ -algebras is a  $\sigma$ -algebra.*

**B:** *Let  $\mathfrak{K}$  be a set which contains subsets of  $\Omega$ . It is not assumed that  $\mathfrak{K}$  is a  $\sigma$ -algebra. Then there exists a  $\sigma$ -algebra which contains  $\mathfrak{K}$  and is minimal in the sense that it is contained in any other  $\sigma$ -algebra that also contains  $\mathfrak{K}$ . We name this  $\sigma$ -algebra  $\sigma(\mathfrak{K})$  because it clearly depends on  $\mathfrak{K}$ . It is constructed as follows:*

$$(16.12) \quad \sigma(\mathfrak{K}) = \bigcap \{ \mathfrak{A} : \mathfrak{A} \supseteq \mathfrak{K} \text{ and } \mathfrak{A} \text{ is a } \sigma\text{-algebra for } \Omega \}.$$

**PROOF of A:**

We must prove (16.2a), (16.2b) and (16.2c). Let  $(\mathfrak{A}_\alpha)_\alpha$  be an arbitrary family of  $\sigma$ -algebras for  $\Omega$ . Let

$$\mathfrak{A} := \bigcap_{\alpha} \mathfrak{A}_\alpha.$$

$\emptyset$  and  $\Omega$  belong to each  $\sigma$ -algebra according to (16.2a). It follows that they both belong to the intersection  $\bigcap_{\alpha} \mathfrak{A}_\alpha$ , i.e.,  $\mathfrak{A}$  satisfies (16.2a). Let  $A \in \mathfrak{A}$ . Then  $A \in \mathfrak{A}_\alpha$  for each  $\alpha$ .  $\complement A$  belongs to each  $\sigma$ -algebra according to (16.2b). It follows that  $\complement A \in \bigcap_{\alpha} \mathfrak{A}_\alpha$ , i.e.,  $\mathfrak{A}$  satisfies (16.2b). Finally, let  $A_n \in \mathfrak{A}$  for all  $n \in \mathbb{N}$ . Then  $A_n \in \mathfrak{A}_\alpha$  for all  $n \in \mathbb{N}$  and for each  $\alpha$ ,  $\bigcup_{n \in \mathbb{N}} A_n$  and  $\bigcap_{n \in \mathbb{N}} A_n$  both belong to each  $\sigma$ -algebra according to (16.2c). It follows that they both belong to the intersection  $\bigcap_{\alpha} \mathfrak{A}_\alpha$ , i.e.,  $\mathfrak{A}$  satisfies (16.2c). It follows that  $\mathfrak{A}$  is a  $\sigma$ -algebra.

**PROOF of B:**

First of all, we know that  $\sigma(\mathfrak{K})$  is an intersection of  $\sigma$ -algebras and, according to part A of this proposition, really is a  $\sigma$ -algebra. We now prove that  $\sigma(\mathfrak{K})$  contains  $\mathfrak{K}$  and is the minimal  $\sigma$ -algebra with that property. First let us prove that  $\sigma(\mathfrak{K}) \supseteq \mathfrak{K}$ . But that is obvious because it is the intersection of sets all of which contain  $\mathfrak{K}$ . On the other hand,  $\sigma(\mathfrak{K})$  is the intersection of **all**  $\sigma$ -algebras that contain  $\mathfrak{K}$ , so it is impossible for any other  $\sigma$ -algebra to both be a strict subset of  $\sigma(\mathfrak{K})$  and also contain  $\mathfrak{K}$ . ■

## 16.2 Sequences of Sets – limsup and liminf

**Assumption 16.1** (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 16.4** (Monotone set sequences). A sequence  $A_k$  of arbitrary subsets of  $X$  is called

$$(16.13a) \quad \text{non-decreasing} \quad \text{if } A_1 \subseteq A_2 \subseteq \dots$$

$$(16.13b) \quad \text{non-increasing} \quad \text{if } A_1 \supseteq A_2 \supseteq \dots$$

$$(16.13c) \quad \text{strictly increasing} \quad \text{if } A_1 \subsetneq A_2 \subsetneq \dots$$

$$(16.13d) \quad \text{strictly decreasing} \quad \text{if } A_1 \supsetneq A_2 \supsetneq \dots$$

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 16.5** (Limits of monotone set sequences). Given are sets  $A_n, B_n \subseteq X$  ( $n \in \mathbb{N}$ ). Assume that

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \quad \text{and let } A := \bigcup_{k \in \mathbb{N}} A_k$$

$$B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots \quad \text{and let } B := \bigcap_{k \in \mathbb{N}} B_k$$

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

$$(16.14a) \quad A = \lim_{n \rightarrow \infty} A_n \quad \text{or} \quad A_n \nearrow A \text{ for } n \rightarrow \infty,$$

$$(16.14b) \quad B = \lim_{n \rightarrow \infty} B_n \quad \text{or} \quad B_n \searrow B \text{ for } n \rightarrow \infty. \quad \square$$

The above are not terribly useful definitions. What does it matter whether we write  $A = \lim_{n \rightarrow \infty} A_n$  or  $A = \bigcup_{k \in \mathbb{N}} A_k$ ? 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 16.6** (lim inf and lim sup of set sequences). Given are sets  $A_n, B_n \subseteq X$  ( $n \in \mathbb{N}$ ). Let

$$(16.15) \quad \liminf_{n \rightarrow \infty} A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k \quad \text{(limit inferior)}$$

$$(16.16) \quad \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \quad \text{(limit superior)}$$

In general those two will not coincide. But if they do then we define

$$(16.17) \quad \lim_{n \rightarrow \infty} A_n := \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$$

We call  $\lim_{n \rightarrow \infty} A_n$  the **limit** of the sequence  $(A_n)$  and we write

$$A_n \rightarrow A \quad \text{for } n \rightarrow \infty. \quad \square$$

The following comments should make matters easier to understand if you abbreviate

**Lemma 16.1** (lim inf and lim sup as monotone limits). Given are sets  $A_n, B_n \subseteq X$  ( $n \in \mathbb{N}$ ). Let

$$(16.18) \quad A_{\star n} := \bigcap_{k \geq n} A_k \quad \text{Then } A_{\star n} \nearrow \liminf_{n \rightarrow \infty} A_n$$

$$(16.19) \quad A^{\star n} := \bigcup_{k \geq n} A_k \quad \text{Then } A^{\star n} \searrow \limsup_{n \rightarrow \infty} A_n$$

PROOF: Let  $m, n \in \mathbb{N}$  such that  $m < n$ . Then

$$\begin{aligned}
 A_{*m} &= \bigcap_{k=m}^{n-1} A_k \cap \bigcap_{k \geq n} A_k = \bigcap_{k=m}^{n-1} A_k \cap A_{*n} \subseteq A_{*n} \\
 A^*_m &= \bigcup_{k=m}^{n-1} A_k \cup \bigcup_{k \geq n} A_k = \bigcup_{k=m}^{n-1} A_k \cup A^*_n \supseteq A^*_n
 \end{aligned}$$

This proves that  $A_{*n}$  is non-decreasing and  $A^*_n$  is non-increasing. By the very definition of the limit of a monotone sequence of sets it is true that

$$\begin{aligned}
 \lim_{n \rightarrow \infty} A_{*n} &= \bigcup_{n \in \mathbb{N}} A_{*n} = \liminf_{n \rightarrow \infty} A_n \\
 \lim_{n \rightarrow \infty} A^*_n &= \bigcap_{n \in \mathbb{N}} A^*_n = \limsup_{n \rightarrow \infty} A_n
 \end{aligned}$$

■

### 16.3 Conditional Expectations as Generalized Averages

EMPTY!!

EMPTY!!

EMPTY!!

## 17 Appendix: Addenda to Beck/Geoghegan's "The Art of Proof"

This chapter contains extensions of material found in [1] 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 Alert:

Even though the subject matter of this chapter is primarily the book [1] 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 "[1] B/G prop.9.7" or something similar.

### 17.1 AoP Ch.1: Integers

#### 17.1.1 Ch.1.1 – Axioms

**Remark 17.1.** 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 def.3.7 on p.35 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.3.7 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 principal domain  $\mathbb{Z}$  is an ordered principal domain with positive cone  $\mathbb{N}$ . See def.3.8 on p.36. Finally, B/G ax.2.15, the induction axiom, is added. Only at this point is  $\mathbb{Z}$  completely specified.  $\square$

#### Remark 17.2.

It follows from the last remark that all of the material in B/G ch.1, i.e., all of the propositions and corollaries and definitions inbetween B/G prop.1.6 and prop.1.27, extend to any integral domain  $(R, \oplus, \odot)$ ;

Here are some examples to illustrate rem.17.2.

**Proposition 17.1** (B/G ch.1 prop.1.6). Let  $(R, \oplus, \odot)$  be an integral domain and  $a, b, c \in R$ . Then

$$(17.1) \quad (a \oplus b) \oplus c = a \oplus (b \oplus c).$$

*PROOF:* See [1] Beck/Geoghegan: *The Art of Proof*, ch.1 prop.1.6.  $\blacksquare$

**Proposition 17.2** (B/G ch.1 prop.1.12). Let  $(R, \oplus, \odot)$  be an integral domain and  $x \in R$ . If  $a \oplus x = a$  for all  $a \in R$  then  $x = 0$ .

*PROOF:* See [1] Beck/Geoghegan: *The Art of Proof*, ch.1 prop.1.12.  $\blacksquare$

**Proposition 17.3** (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  $a \odot x = a \odot x'$ . It follows from cor.3.1 on p.35 that  $x = x'$  ■

**Remark 17.3.** Note that the equation  $y = a \odot x \oplus b$  need not have a solution. For example, there is no integer  $x$  which satisfies the equation  $2x + 0 = 1$ . □

We now switch to material from ch.1 of the B/G book.

**Proposition 17.4.** Let  $m, n \in \mathbb{Z}$  such that  $m|n$  and  $m \neq 0$ , 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

$$n = j' \cdot m = j \cdot m, \quad \text{hence } (j' - j)m = 0.$$

B/G Proposition 1.26 states that if  $a, b \in \mathbb{Z}$  and if  $a \cdot b = 0$ , then  $a = 0$  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 17.1** (Quotients). Let  $m, n \in \mathbb{Z}$  such that  $m|n$  and  $m, n \neq 0$ . Let  $j \in \mathbb{Z}$  be the unique integer for which  $n = j \cdot m$ . We write either of

$$(17.2) \quad \frac{n}{m}, \quad n/m, \quad n \div m$$

instead of  $j$  and we call  $n$  the **dividend** or **numerator**,  $m$  the **divisor** or **denominator**, and  $j$  the **quotient** of the expression  $n/m$ . □

**Note 17.1.** Note that the assignment  $(m, n) \mapsto n/m$  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  $n/m$  is only defined for those  $m, n \in \mathbb{Z}$  which satisfy the condition  $m|n$ . □

## 17.2 AoP Ch.2: Natural Numbers and Induction

The following remark is also part of rem.17.1 on p.307.

**Remark 17.4.** 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 principal domain with positive cone  $\mathbb{N}$ . See def.3.8 on p.36. □

**Remark 17.5.**

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  $(R, \oplus, \odot, P)$ . □

Here is an example for rem.17.5b.

**Proposition 17.5** (B/G ch.1 prop.2.2). Let  $(R, \oplus, \odot, P)$  be an ordered integral domain with the order  $\leq$  induced by the positive cone  $P$ . Then property **d** of def.3.8 (ordered integral domains) on p.36 can be “sharpened” as follows: One and only one of the following is true for any  $a \in R$

$$a \in P; \quad -a \in P; \quad a = 0.$$

PROOF: See [1] Beck/Geoghegan: *The Art of Proof*, ch.1 prop.2.2. ■

### 17.2.1 AoP Ch.2.2 (Ordering the Integers)

The material given here complements ch.2.2 (Ordering the Integers) of [1] Beck/Geoghegan.

**Remark 17.6.** 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.*

The next two propositions belong after B/G prop.2.3 ( $1 \in \mathbb{N}$ ).

**Proposition 17.6.** If  $k \in \mathbb{Z}$  then  $k + 1 > k$ .

*Proof:* Left as an exercise. ■

**Corollary 17.1.**  $1 > 0$ .

PROOF: This follows from  $1 = 1 + 0$  and prop.17.6, applied to  $k := 0$ . ■

The next proposition belongs after B/G prop.2.4 (“ $\leq$ ” is transitive).

**Proposition 17.7.** Let  $m, n, p \in \mathbb{Z}$  such that

$$(17.3) \quad m \leq n, \quad n \leq p.$$

Then  $m \leq p$ .

PROOF: There are four cases.

1.  $m < n, n < p$ : It follows from B/G prop.2.4 (transitivity of “ $<$ ”) that  $m < p$ , in particular,  $m \leq p$ .
2.  $m < n, n = p$ : It follows that  $m < p$ . This implies  $m \leq p$ .
3.  $m = n, n < p$ : It follows again that  $m < p$ . This implies  $m \leq p$ .
4.  $m = n, n = p$ : It follows that  $m = p$ . This implies  $m \leq p$ . ■

### 17.2.2 AoP Ch.2.3 (Induction)

The material given here complements ch.2.3 (Induction) of [1] Beck/Geoghegan.

**Proposition 17.8** (B/G Prop.2.20). If  $k \in \mathbb{N}$  then

$$(17.4) \quad k \geq 1.$$

*PROOF:* The proof is done by induction on  $k$ .

**Base case:** For  $k = 1$  (17.4) reads  $1 \geq 1$ . This is true because “=” implies “ $\geq$ ”.

**Induction step:** Let  $k \in \mathbb{N}$  and assume that  $k \geq 1$ . We are done if we can prove under this assumption that

$$(17.5) \quad k + 1 \geq 1.$$

According to prop.17.6,  $k + 1 > k$ , hence  $k + 1 \geq k$ . Further  $k \geq 1$  (our induction assumption.) It follows from the transitivity of “ $\leq$ ” (prop.17.7) that  $k + 1 \geq 1$ , i.e., (17.5) is valid. ■

**Proposition 17.9** (B/G Prop.2.21). There exists no  $x \in \mathbb{Z}$  such that  $0 < x < 1$ .

*PROOF:* Assume to the contrary that there is some  $x \in \mathbb{Z}$  such that  $0 < x$ , i.e.,  $x > 0$ , and  $x < 1$ .

It follows from  $x > 0$  that  $x \in \{k \in \mathbb{Z} : k > 0\}$ . But according to B/G prop.2.13 this set equals  $\mathbb{N}$ . So  $x \in \mathbb{N}$ . It now follows from B/G prop.2.20 that  $x \geq 1$ , contrary to the assumption that  $x < 1$ . ■

**Corollary 17.2** (B/G Cor.2.22). Let  $n \in \mathbb{Z}$ . There exists no  $x \in \mathbb{Z}$  such that  $n < x < n + 1$ .

*PROOF:* Assume to the contrary that such  $x$  exists. According to B/G prop.2.7(i) an inequality between integers is preserved if one adds the same item to both of its sides. It follows that

$$n - n < x - n < (n + 1) - n, \quad \text{i.e., } 0 < x - n < 1.$$

The integer  $x' := x - n$  satisfies  $0 < x' < 1$ , a contradiction to prop.17.9. ■

**Proposition 17.10** (sharpening of B/G Prop.2.13).  $\mathbb{N} = \{k \in \mathbb{Z} : k \geq 1\}$ .

*PROOF:* This follows from  $1 > 0$  (cor.17.1 on p.309) and prop.17.9 above. ■

### 17.2.3 Bounded Sets in $\mathbb{Z}$

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

**Definition 17.2** (Upper and lower bounds, maxima and minima in  $\mathbb{Z}$ ). Let  $A \subseteq \mathbb{Z}$ .<sup>146</sup> Let  $l, u \in \mathbb{Z}$ . We call  $l$  a **lower bound** of  $A$  if  $l \leq a$  for all  $a \in A$ .

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

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

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

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

<sup>146</sup> The definitions given here also apply to subsets of  $\mathbb{Q}$  and  $\mathbb{R}$ . See def.9.1 on p.145.

**Proposition 17.11.** Let  $A \subseteq \mathbb{Z}$ . If  $A$  has a maximum then it is unique. If  $A$  has a minimum then it is unique.

PROOF: Left as an exercise.

■

**Definition 17.3** (Translation and dilation of sets of integers). Let  $A \subseteq \mathbb{Z}$ . and  $\alpha, b \in \mathbb{Z}$ . We define

$$(17.6) \quad \lambda A + b := \{\lambda a + b : a \in A\}.$$

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

$$(17.7) \quad A + b = \{a + b : a \in A\},$$

$$(17.8) \quad -A = \{-a : a \in A\}. \quad \square$$

**Remark 17.7.** Note that the above makes sense for any algebraic structure<sup>147</sup> with binary operations “+” and/or “·”, e.g., groups or vector spaces (in the latter case “ $(\lambda, a) \mapsto \lambda a$ ” would be the scalar product between a real number (scalar)  $\lambda$  and a vector  $a$ .  $\square$

**Theorem 17.1** (Generalization of the Well-Ordering Principle).

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

*Proof (outline):*

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

b. We assume that  $B$  is bounded above. Let  $b^*$  be an upper bound of  $B$ . Let  $B' := -B$ . Then  $B'$  has  $-b^*$  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 ■

**Example 17.1** (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  $\pm 2$ ) but has neither min (would have to be  $-\sqrt{2} \notin A$ ) nor max (would have to be  $+\sqrt{2} \notin A$ ).  
**but:**  $-\sqrt{2}$  is a special lower bound: it is the **greatest lower bound** or **infimum**  $\inf(A)$  of  $A$  and  $\sqrt{2}$  is a special upper bound: it is the **least** (smallest) **upper bound** or **supremum**  $\sup(A)$  of  $A$ .

<sup>147</sup> See the optional chapter 14 (Algebraic structures) on p.289

b.  $\mathbb{Q}$ : The set  $B := \{x \in \mathbb{Q} : x^2 < 2\} = A \cap \mathbb{Q}$  is bounded in  $\mathbb{Q}$  (by  $\pm 2$ ) and also has neither min nor max for the same reasons as  $A$ .

**Further:**  $-\sqrt{2}$  is **not** a lower bound of  $B$  and  $\sqrt{2}$  is **not** an upper bound of  $B$  because those numbers are not in our “universe”  $\mathbb{Q}$ . The set  $B$  has neither min, max, inf, sup!  
□

**Theorem 17.2** (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:* Assume to the contrary that there exists an upper bound of  $\mathbb{N}$ . According to thm17.1 (Generalization of the Well-Ordering Principle) on p.311  $\mathbb{N}$  has a maximum. Let  $u^* := \max(\mathbb{N})$ . Then  $u^* + 1$  belongs to  $\mathbb{N}$  as the sum of two natural numbers. It follows from  $u^* + 1 > u^*$  that  $u^*$  is not the largest element of  $\mathbb{N}$  and we have reached a contradiction. ■

#### 17.2.4 Exercises for Ch.17.2

**Exercise 17.1.** Use everything up to AND including B/G prop.2.3, to prove prop.17.6 on p.309 of this document: If  $k \in \mathbb{Z}$  then  $k + 1 > k$ . □

### 17.3 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.39 of this document.

a. Universal quantifier  $\forall$ , existential quantifier  $\exists$ , unique existential quantifier  $\exists!$ : ch.4.5.1 on p.64.

b. “ $(\forall x)(\forall y)$  has the same meaning as  $(\forall x \text{ and } y)$ ”: part a of def.4.17 (Doubly quantified expressions) in ch.4.5.2 on p.65.

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 “ $P \rightarrow Q$ ” 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”,<sup>148</sup> for  $P$  to be true and  $Q$  to be false. “ $P \Rightarrow Q$ ” has the combination (  $P$  is true,  $Q$  is false ) marked as irrelevant (logically impossible); the remaining three combinations give the same outcome as for  $P \rightarrow Q$ , i.e., they all evaluate to **true**.

Example: If  $P$  : “ $x = 100$ ” and  $Q$  : “ $x > 10$ ” is correct because 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.50. There you also find the definition of the **converse** and the **contrapositive** of an implication.

---

<sup>148</sup>See ch.4.2.2, def.4.7 on p.45.



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.56). Do you see the connection to De Morgan’s law for sets? (see thm.6.1 on p.111).
- 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.
- e. In the above observe the “binding power” or preference of 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)*.”
- e. xxx

## 17.4 AoP Ch.4: Recursion

Some of the recursive definitions in chapter 4 of AoP, such as  $\sum x_j$  and  $\prod x_j$  were defined to start at  $j = 1$ . One cannot deduce from those definitions what it means to start at some index other than 1. We fix this shortcoming in this chapter.

### 17.4.1 Generalization of Some Recursive Definitions

**Definition 17.4** (Extension of the definition  $\sum_{j=1}^n x_j$ ). We modify the definition of “ $\Sigma$ ” (B/G p.34, 35) as follows: Let  $k \in \mathbb{Z}$  and let  $(x_j)_{j=k}^{\infty}$  be a sequence of integers.

For each  $n \in \mathbb{Z}$  such that  $k \leq n$ , we define an integer called  $\sum_{j=k}^n x_j$  as follows. <sup>149</sup>

$$(17.9) \quad \text{(i)} \quad \sum_{j=k}^k x_j = x_k, \quad \text{(ii)} \quad \sum_{j=k}^{n+1} x_j = \sum_{j=k}^n x_j + x_{n+1}.$$

We call  $\sum_{j=k}^n x_j$  the **sum** of the integers  $x_k, x_{k+1}, \dots, x_{n-1}, x_n$ .  $\square$

**Definition 17.5** (Extension of the definition  $\prod_{j=1}^n x_j$ ). We modify the definition of “ $\Pi$ ” (B/G p.34, 35) as follows: Let  $k \in \mathbb{Z}$  and let  $(x_j)_{j=k}^{\infty}$  be a sequence of integers.

---

<sup>149</sup>  $\sum_{j=k}^n x_j$  can also be written  $x_k + x_{k+1} + \dots + x_n$ .

For each  $n \in \mathbb{Z}$  such that  $k \leq n$ , we define an integer called  $\prod_{j=k}^n x_j$  as follows. 150

$$(17.10) \quad (i) \quad \prod_{j=k}^k x_j = x_k, \quad (ii) \quad \prod_{j=k}^{n+1} x_j = \prod_{j=k}^n x_j \cdot x_{n+1}.$$

We call  $\prod_{j=k}^n x_j$  the **prod** of the integers  $x_k, x_{k+1}, \dots, x_{n-1}, x_n$ .  $\square$

**Proposition 17.12** (Generalized Law of Associativity). *Let  $x_1, x_2, \dots, x_n \in \mathbb{Z}$ . Then the formula for associativity  $x_1 + (x_2 + x_3) = (x_1 + x_2) + x_3$  extends 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  $\sum_{j=1}^n x_n$ .*

PROOF:

The proof is done by induction on  $n$ .

Basis: The proof is obvious for  $k = 1, 2, 3$ .

Induction assumption: Any grouping with parentheses of  $x_1 + \dots + x_n$  equals  $\sum_{j=1}^n x_n$ .

Let us now assume that we have a sum of  $x_1, \dots, x_n$ , grouped by parentheses. Let  $A$  denote that sum. We may assume 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 **(c)** are excluded, we have either of

$$\text{case 1: } x_1 + (\dots) \quad \text{or} \quad \text{case 2: } (\dots) + x_{n+1}.$$

Case 1: it follows from the induction assumption that  $A = x_1 + \sum_{j=2}^{n+1} x_j$ .

Case 2: it follows from the induction assumption that  $A = \sum_{j=1}^n x_j + x_{n+1}$ .

In either case we conclude that  $A = \sum_{j=1}^n x_{n+1}$  with the help of B/G Prop. 4.16(i): If  $a, b, c \in \mathbb{Z}$  such that  $1 \leq a \leq b < c \leq n + 1$  then

$$\sum_{j=a}^c x_j = \sum_{j=a}^b x_j + \sum_{j=b+1}^c x_j. \quad \blacksquare$$

## 17.5 AoP Ch.5: Underlying Notions in Set Theory

There are no addenda at this point in time.

---

<sup>150</sup>  $\prod_{j=k}^n x_j$  can also be written  $x_k \cdot x_{k+1} \cdots x_n$  or  $x_k x_{k+1} \cdots x_n$ .

## 17.6 AoP Ch.6: Equivalence Relations and Modular Arithmetic

### 17.6.1 Equivalence Relations

**Remark 17.8.** Note that B/G defines in this chapter the absolute value for integers as usual: (see def.2.15 on p.18)

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

### 17.6.2 The Division Algorithm

The following simple proposition about the absolute value of the difference of two non-negative numbers is very useful in chapters 6 and 7 of the B/G text.

**Proposition 17.13.** Let  $m, n \in \mathbb{Z}_{\geq 0}$ . Then

$$(17.11) \quad |n - m| \leq \max(m, n), \text{ i.e.,}$$

$$(17.12) \quad -\max(m, n) \leq n - m \leq \max(m, n).$$

*Proof:* Left as an exercise. ■

**Corollary 17.3.** Let  $x, y, n \in \mathbb{Z}$  such that  $0 \leq x, y < n$ . Then

$$(17.13) \quad -n < y - x < n.$$

*PROOF:* Obvious. ■

As an easy corollary of the uniqueness of prime factorizations up to ordering we obtain the following proposition

**Proposition 17.14.** 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, and  $m$  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. ■

**Proposition 17.15** (Division Algorithm for Polynomials (B/G prop. 6.18)). Let  $\alpha, \beta \in \mathbb{Z}_{\geq 0}$  and let

$$(17.14) \quad n(x) := \sum_{j=0}^{\alpha} a_j x^j, \quad m(x) := \sum_{j=0}^{\beta} b_j x^j,$$

be two polynomials with real coefficients  $a_j, b_j$  such that  $p(x)$  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  $\alpha$  or  $r(x) = 0$  (and hence has no degree), such that

$$(17.15) \quad m(x) := q(x)n(x) + r(x).$$

**PROOF:**

**Case 1:**  $\alpha = 0$ , i.e.,  $n(x) = n = \text{const}$ .

Note that  $n \neq 0$  because we assume that  $n(x)$  is not the null polynomial. Let  $q(x) := \frac{m(x)}{n}$  and  $r(x) := 0$ . Then  $m(x) = n(x)q(x) + r(x)$  yields the desired decomposition 17.15 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 17.15.

**Case 3:**  $\alpha > 0$  and  $\beta \geq \alpha$ .

We handle this case using strong induction on  $\beta$ . First some notation. Let

$$(17.16) \quad A := \sum_{j=0}^{\alpha-1} \frac{b_{\beta} a_j}{a_{\alpha}} x^{j+\beta-\alpha}, \quad B := \sum_{j=0}^{\beta-1} b_j x^j,$$

$$(17.17) \quad p(x) := m(x) - \frac{b_{\beta}}{a_{\alpha}} x^{\beta-\alpha} n(x) = b_{\beta} x^{\beta} + B - \frac{b_{\beta}}{a_{\alpha}} x^{\beta-\alpha} a_{\alpha} x^{\alpha} - A = B - A.$$

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  $r(x)$ .

Base case:  $\beta = \alpha$ .

(17.17) yields  $p(x) = m(x) - \frac{b_{\beta}}{a_{\alpha}} n(x)$ . We have  $m(x) = p(x) + \frac{b_{\beta}}{a_{\alpha}} n(x)$ . Let  $q(x) := \frac{b_{\beta}}{a_{\alpha}}$  and  $r(x) := p(x)$ . Then  $m(x) = n(x)q(x) + r(x)$  yields the desired decomposition 17.15 of  $m(x)$ .

**Induction assumption:** Any polynomial  $m'(x)$  with degree less than  $\beta$  can be written as  $m'(x) = q'(x)n(x) + r'(x)$  such that  $r'(x)$  has degree less than  $\alpha$  or  $r'(x) = 0$ .

We use again the notation of (17.16) and (17.17). We have seen that the degree of  $p(x)$  is less than  $\beta$  (unless it has no degree because  $p(x) = 0$ ). It follows from (17.17) 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 17.15. ■

### 17.6.3 The Integers Modulo $n$

In this chapter we assume that  $n \in \mathbb{N}$  is fixed and that for two integers  $a$  and  $b$ ,

$$a \sim b \text{ means } a \equiv b \pmod{n}, \text{ if } n|(a - b).$$

We also define  $[a]^0 := [1]$ . The following cannot be found in the B/G text.

**Proposition 17.16** (Arithmetic mod  $n$ ). Let  $m_1, m_2, \dots, m_k \in \mathbb{Z}$ . Then

$$(17.18) \quad [m_1 + m_2 + \dots + m_k] = [m_1] \oplus [m_2] \oplus \dots \oplus [m_k],$$

$$(17.19) \quad [m_1 \cdot m_2 \cdot \dots \cdot m_k] = [m_1] \odot [m_2] \odot \dots \odot [m_k],$$

$$(17.20) \quad \left[ \sum_{j=1}^k a_j x^j \right] = \sum_{j=1}^k [a_j] \odot [x]^j.$$

*PROOF:* We only give the proof of (17.19). It is done by induction on the number of factors  $k$ . The proof of (17.18) is similar and (17.20) is a simple consequence of the two first equations.

*Basis:* The proof is obvious for  $k = 1$ . We note that (17.19) is true for two factors (B/G prop.6.25 and the definition of " $a \odot b$ " that follows it).

*Induction assumption:* We assume that (17.19) holds for some  $k \in \mathbb{N}$ . We then obtain for  $k + 1$  that

$$\begin{aligned} [m_1 \cdot m_2 \cdot \dots \cdot m_{k+1}] &= [(m_1 \cdot m_2 \cdot \dots \cdot m_k) \cdot m_{k+1}] \\ &= ([m_1 \cdot m_2 \cdot \dots \cdot m_k]) \odot [m_{k+1}] \quad (\text{B/G def. of } "a \odot b") \\ &= ([m_1] \odot [m_2] \odot \dots \odot [m_k]) \odot [m_{k+1}]. \quad (\text{Induction assumption (17.19)}) \quad \blacksquare \end{aligned}$$

### 17.6.4 Prime Numbers

**Definition 17.6** (Prime numbers and prime factorizations). Let  $p \in \mathbb{N}, p \geq 2$ .  $p$  is a **prime number** or  $p$  is **prime** if  $q \in \mathbb{Z}$  and  $q|p$  implies that  $q = \pm 1$  or  $q = \pm p$ .  $p$  is called a **composite number** or just a **composite** if  $p$  is not prime.

We note that 1 is **not** prime.

Let  $m \in \mathbb{Z}$ . If there are primes  $p_1, \dots, p_k$  such that  $m = \pm p_1 \cdot p_2 \cdot \dots \cdot p_k$  then  $p_1, \dots, p_k$  are called **factors** or **prime factors** of  $m$  and  $\pm p_1 \cdot p_2 \cdot \dots \cdot p_k$  is called a **prime factorization** or just a **factorization** of  $m$ .  $\square$

**Proposition 17.17** (B/G prop.6.28). Let  $n \in \mathbb{N}$  such that  $n > 1$ . Then  $n$  has a prime factorization.

The proof is left as exercise 17.3.

**Lemma 17.1.** Let  $p$  be prime and let  $n \in \mathbb{N}$ . We have the following:

- a.  $\gcd(p, n) = 1$  or  $\gcd(p, n) = p$ .
- b. If  $p \nmid n$  ( $p$  does not divide  $n$ ) then  $\gcd(p, n) = 1$ .

*PROOF:* Let  $g := \gcd(p, n)$ . Then  $g \in \mathbb{N}, g|n$  and  $g|p$  (B/G prop.6.29(i)). Only 1 and  $p$  divide  $p$ , so either  $g = p$  or  $g = 1$ . This proves a.

We have  $p \nmid n$  and  $g|n$ , hence  $g \neq p$ . It follows from a which was already proven that the only remaining choice for  $g$  is  $g = 1$ , and b follows.  $\blacksquare$

We next look at Euclid's Lemma and the uniqueness of prime number factorizations. B/G thm 6.32 states the following: Every integer  $\geq 2$  can be factored uniquely (i.e. up to permutation) into primes. The proof of this theorem (thm.17.3 of this chapter) requires the generalization of Euclid's lemma to more than two factors:

**Proposition 17.18** (Euclid's Lemma for more than two factors). Let  $p$  be prime and  $m_1, m_2, \dots, m_k \in \mathbb{N}$ . If  $p|(m_1 m_2 \cdots m_k)$  then  $p|m_j$  for some  $1 \leq j \leq k$ .

PROOF: Done by strong induction on the number of factors  $k$ .

Basis: There is nothing to prove for  $k = 1$  and B/G prop. 6.31 (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  $p|n_i$  for some  $1 \leq i \leq j$ .

To prove that  $p|m_i$  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 B/G prop. 6.31 that  $p|m_k$  or  $p|(m_1 m_2 \cdots m_{k-1})$ . 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 17.3** (Uniqueness of prime factorizations (B/G thm 6.32)). Every integer  $\geq 2$  can be factored uniquely (i.e. up to permutation) into primes.

PROOF: by 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  $n$ .  $p_1|q_1 q_2 \cdots q_l$ , hence  $p_1|q_{j_0}$  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  $n_2 := n/p_1 = p_2 p_3 \cdots p_k = q_2 q_3 \cdots q_l$  is an integer less than  $n$  and it follows from the induction assumption that  $q_2 \cdots q_l$  are just a reordering of  $p_2 \cdots p_k$ . ■

**Proposition 17.19** (Corollary to Fermat's Little Theorem). Let  $p$  be prime and let  $m \in \mathbb{N}$  such that  $p \nmid m$ . Then

$$m^{p-1} \equiv 1 \pmod{p}.$$

PROOF: It follows from Fermat's Little Theorem (B/G thm.6.35) that  $p|(m^p - m)$ , i.e.,  $p|m(m^{p-1} - 1)$ . It follows from Euclid's lemma and  $p \nmid m$  that  $p|(m^{p-1} - 1)$ , i.e.,  $m^{p-1} \equiv 1 \pmod{p}$ . ■

### 17.6.5 Exercises for Ch.17.6

#### Exercise 17.2. <sup>151</sup>

Let  $K \in \mathbb{N}$  such that  $K > 1$  and  $n \in \mathbb{Z}_{\geq 0}$ . Prove that  $K^n > n$ . □

**Exercise 17.3.** Prove prop.17.17 on p.317: Any integer  $\geq 2$  has a prime factorization. □

<sup>151</sup> Note that this exercise generalizes B/G prop.7.1: If  $n \in \mathbb{N}$  then  $n < 10^n$ .

## 17.7 AoP Ch.7: Arithmetic in Base Ten

### 17.7.1 Base-Ten Representation of Integers

We repeat here the definition of  $\nu(n)$ , the number of digits of a non-negative integer  $n$ , and some of its core properties.

**Definition 17.7** (Number of digits w.r.t. base 10). For  $n \in \mathbb{Z}_{\geq 0}$  let

$$(17.21) \quad E := E(n) := \{e \in \mathbb{N} : n < 10^e\},$$

$$(17.22) \quad \nu(n) := \min(E(n)) \quad (n \in \mathbb{N}).$$

We call  $\nu(n)$  the **number of digits of  $n$  with respect to base 10**.  $\square$

**Remark 17.9.**

- "E" stands for "Exponent":  $E(n)$  is the set of all base 10 exponents  $e$  which are so big that  $10^e$  dominates  $n$ .
- It follows from  $0 < 10 = 10^1$  that  $\nu(0) = \min(\mathbb{N}) = 1$ .
- If  $n \in \mathbb{N}$  then  $n < 10^n$  (see B/G prop.7.1), hence  $E(n) \neq \emptyset$ . It follows from thm.17.1 (the Well-Ordering Principle) on p.311, that  $E(n)$  indeed possesses a minimum.
- No gaps property of  $E$**   $E = \{k \in \mathbb{N} : k \geq \nu(N)\}$ . This follows from the fact that the function  $k \mapsto 10^k$  is strictly increasing: If  $k, k' \in \mathbb{N}$  such that  $k < k'$ , then

$$10^{k'} - 10^k = 10^k(10^{k'-k} - 1)$$

. Both factors are positive (see B/G prop.7.1), hence  $10^k < 10^{k'}$ .

**Proposition 17.20** (B/G Prop.7.5). Let  $n, q, r \in \mathbb{N}$  such that  $n = 10q + r$  and  $0 \leq r < 10$  (the unique division algorithm decomposition of  $n$  for division by 10.) Then  $\nu(n) = \nu(q) + 1$ .

*PROOF:* We abbreviate  $k := \nu(q)$ . The proof will be done as follows.

- We prove by contradiction that  $n < 10^{k+1}$ .
- We prove by contradiction that  $10^k \leq n$ .
- The proof is done because we obtain from **a** and **b** that  $10^k \leq n < 10^{k+1}$  and it follows from B/G prop.7.3 that  $\nu(n) = k + 1 = \nu(q) + 1$ .

*PROOF of a.*

If we had instead that  $n \geq 10^{k+1}$  then

$$10q = n - r \geq 10^{k+1} - r > 10^{k+1} - 10 = 10(10^k - 1).$$

It follows that  $q > 10^k - 1$ , hence  $q \geq 10^k$ . This contradicts  $k = \nu(q)$ .

*PROOF of b.*

Assume to the contrary that  $n < 10^k$ . Then

$$10q = n - r < 10^k - r \leq 10^k; \text{ hence } q < 10^{k-1},$$

which contradicts our assumption that  $k = \nu(q)$ .  $\blacksquare$

**Proposition 17.21** (B/G prop.7.9: Uniqueness of base-ten representation for positive integers). *Let  $n \in \mathbb{N}$  and assume that*

$$(17.23) \quad n = \sum_{j=0}^p x_j 10^j = \sum_{j=0}^q y_j 10^j$$

where  $p, q \in \mathbb{Z}_{\geq 0}$ , each  $x_i$  and each  $y_i$  is a digit,  $x_p \neq 0$  and  $y_q \neq 0$ . Then  $p = q$  and  $x_i = y_i$  for all  $i$ .

*PROOF:* By allowing leading digits to be zero, we may assume  $p = q$ . Let  $j^*$  be the largest index  $j$  such that  $x_j \neq y_j$ . We obtain

$$(17.24) \quad 0 = \sum_{j=0}^p x_j 10^j - \sum_{j=0}^q y_j 10^j = \sum_{j=0}^{j^*} (x_j - y_j) 10^j.$$

It follows from cor.17.3 on p.315 and B/G prop.7.8 that

$$(17.25) \quad \left| \sum_{j=0}^{j^*-1} (x_j - y_j) 10^j \right| \leq \sum_{j=0}^{j^*-1} 9 \cdot 10^j = 10^{j^*} - 1.$$

One of  $x_{j^*}, y_{j^*}$  is larger than the other. We may assume  $x_{j^*} > y_{j^*}$ , and hence  $x_{j^*} - y_{j^*} \geq 1$ . We obtain

$$\begin{aligned} 0 &= \sum_{j=0}^{j^*} (x_j - y_j) 10^j = (x_{j^*} - y_{j^*}) 10^{j^*} + \sum_{j=0}^{j^*-1} (x_j - y_j) 10^j \\ &\stackrel{(17.25)}{\geq} (x_{j^*} - y_{j^*}) 10^{j^*} - (10^{j^*} - 1) \geq 1 \cdot 10^{j^*} - 10^{j^*} + 1 = 1. \end{aligned}$$

We have reached a contradiction. ■

**Proposition 17.22** (B/G Prop.7.11). *Let  $n := \sum_{j=0}^{\nu(n)-1} x_j 10^j$ , where each  $x_j$  is a digit; then*

$$(17.26) \quad n = x_0 + x_1 + \cdots + x_{\nu(n)-1} \pmod{3}.$$

*PROOF:* We define  $a \sim b$  iff  $3|b - a$ . It follows from  $3|9$  that  $[10] = [9] \oplus [1] = [0] \oplus [1] = [1]$ , hence  $[10]^j = [1]^j = [1^j] = [1]$ . We apply repeatedly prop.17.16 (Arithmetic mod  $n$ ) on p.317 and obtain

$$\begin{aligned} [n] &= \left[ \sum_{j=0}^{\nu(n)-1} x_j 10^j \right] \\ &= \sum_{j=0}^{\nu(n)-1} [x_j] \odot [10]^j \\ &= \sum_{j=0}^{\nu(n)-1} [x_j] \odot [1] \\ &= \sum_{j=0}^{\nu(n)-1} [x_j] \\ &= [x_0 + x_1 + \cdots + x_{\nu(n)-1}]. \end{aligned}$$

It follows that  $n = x_0 + x_1 + \cdots + x_{\nu(n)-1} \pmod{3}$  ■



### 17.7.2 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 17.10** (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  $0, 1, 2, \dots, 9$ . 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,  $\text{Add}(n, x_n, y_n, z_n)$ , whose task it is to compute the  $n$ -th digits  $z_n$  for the sum  $z := \sum_{n=0}^{K+1} z_n \cdot 10^n := x + y$ .

**Subroutine**  $\text{Add}(n, x_n, y_n, i_{n-1}, z_n, i_n)$  :

/\*

/\* Inputs:  $n, x_n, y_n, i_{n-1}$

/\* Output:  $z_n, i_n$

/\*

If  $n = 0$  then {

$i_{-1} := 0;$

}

$$i_n := \begin{cases} 0 & \text{if } x_n + y_n + i_{n-1} < 10 \\ 1 & \text{if } x_n + y_n + i_{n-1} \geq 10 \end{cases}$$

$$z_n := (x_n + y_n + i_{n-1}) - i_n \cdot 10$$

**end-of-Subroutine**

Note that the “sum digit”  $z_n$  and the “carry”  $i_n$  are associated with the “Euclidian 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:

$$i_n = q \quad \text{and} \quad z_n = r. \quad \square$$

## 17.8 AoP Ch.8: Real Numbers

B/G ch.8 axiomatically defines the set  $\mathbb{R}$  of all real numbers. It was stated in ch.2.2 (Numbers) on p.15 that real numbers are the same as decimal numbers. A proof of this will be given in B/G ch.12 (Decimal Expansions).

## 17.8.1 Axioms

### Remark 17.11.

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  $\mathbb{Z}$ .  $\square$

The following proposition belongs between B/G prop.8.7:

Let  $x, y, z \in \mathbb{R}$  and  $x \neq 0$ . If  $xy = xz$  then  $y = z$ ;

and B/G prop.8.8:

Let  $x, y, z \in \mathbb{R}$ . Then  $(x + y)z = xz + yz$ .

**Proposition 17.23** (Uniqueness of multiplicative inverse). Let  $a \in \mathbb{R}$  and  $a \neq 0$ . Then the equation  $ax = 1$  has  $a^{-1}$  as a unique solution.

*PROOF:* Let  $x := a^{-1}$ . Then  $ax = aa^{-1} = 1$ , i.e.,  $a^{-1}$  is a solution.

Uniqueness: Let  $x' \in \mathbb{R}$  such that  $ax' = 1$ . It follows from  $ax = 1 = ax'$  and B/G prop.8.7 that  $x = x'$ .  $\blacksquare$

The following proposition belongs at the end of B/G ch.8.1.

**Proposition 17.24** (Uniqueness of  $ax = b$ ). Let  $a, b \in \mathbb{R}$  and  $a \neq 0$ . Then the equation  $ax = b$  has a solution  $x \in \mathbb{R}$  and this solution is unique.

*PROOF:* Existence of a solution: Let  $x := a^{-1}b$ . Then  $ax = a(a^{-1}b) = (aa^{-1})b = b$ , i.e.,  $x$  is a solution.

Uniqueness: Let  $x' \in \mathbb{R}$  such that  $ax' = b$ . It follows from  $0 = b - b = (ax) - (ax') = a(x - x')$ . Hence, from B/G prop.8.7, we obtain  $x + (-x') = 0$ . It follows from the uniqueness of the additive inverse that  $x = x'$ .  $\blacksquare$

## 17.9 AoP Ch.9: Embedding $\mathbb{Z}$ in $\mathbb{R}$

**Remark 17.12.** Note that B/G prop.9.10 and prop.9.12 are covered by thm.5.2 on p.97.  $\square$

## 17.10 AoP Ch.10: Limits and Other Consequences of Completeness

**Proposition 17.25** (Generalization of B/G prop.10.5). Let  $x, y \in \mathbb{R}_{\geq 0}$ . 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  $x^2 < y^2$ .

*PROOF of a,  $\Rightarrow$ ):* Let  $0 \leq x \leq y$ . It follows from B/G prop.8.37(iv) that this implies both  $x \cdot x \leq x \cdot y$  and  $x \cdot y \leq y \cdot y$ . Transitivity of " $\leq$ "<sup>152</sup> then yields  $x^2 \leq y^2$ .

<sup>152</sup> Proved for integers in prop.17.7 on p.309.

PROOF of **a**,  $\Leftarrow$ ): Let  $x^2 \leq y^2$ . Then  $(y - x)(y + x) = y^2 - x^2 \geq 0$ .

**Case 1:**  $x^2 = y^2$ , i.e.,  $(y - x)(y + x) = 0$ . It follows from B/G prop.8.23 that  $y - x = 0$  or  $y + x = 0$ .  $y - x = 0$  implies  $x = y$  and  $y + x = 0$  implies  $x = y = 0$  because neither  $x$  nor  $y$  is allowed to be negative. Either way we have  $x = y$ , hence  $x \leq y$  and we are done.

**Case 2:**  $x^2 < y^2$ , i.e.,  $(y - x)(y + x) > 0$ . We have seen in case 1 that this necessitates  $x \neq y$ , and we have seen in the proof of **a**,  $\Rightarrow$ ) that  $y \leq x$  implies  $y^2 \leq x^2$ , so  $x^2 < y^2$  implies  $0 \leq x < y$ .

But then  $x + y \geq y > 0$ . We conclude from B/G prop.8.36 that  $y - x > 0$ , hence  $x < y$ .

PROOF of **b**: It is obvious that  $x = y$  implies  $x^2 = y^2$ . We have seen in the proof of **a**,  $\Leftarrow$ ), case 1, that  $y^2 = x^2$  implies  $x = y$ .

PROOF of **c**: This is a consequence of **a** and **b**. ■

**Proposition 17.26** (Alternate proof of 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  $A_{\text{uppb}}$  for the set of upper bounds of  $A$  (see def.9.2 on p.145).

First, we prove the existence of  $r$ .

It follows from  $\frac{14}{10} > 0$  and  $(\frac{14}{10})^2 < 2$  that  $A \neq \emptyset$ . It follows from  $\frac{15}{10} > 0$  and  $(\frac{15}{10})^2 > 2$  and  $a^2 < 2$  for all  $a \in A$  that  $a^2 < (\frac{15}{10})^2$ .

Hence, from prop.17.25 (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  $x_n \notin A_{\text{uppb}}$ , 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.9.11) that

$$\lim_{n \rightarrow \infty} y_n^2 - \lim_{n \rightarrow \infty} x_n^2 = \lim_{n \rightarrow \infty} \frac{4r}{n} = 4r \cdot \lim_{n \rightarrow \infty} \frac{1}{n},$$

and the latter expression is zero because  $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$  (see example 9.4.a on p.150). 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.9.13 (Domination Theorem for Limits) on p.156 that

$$\lim_{n \rightarrow \infty} x_n^2 \leq \lim_{n \rightarrow \infty} 2 = 2 \leq \lim_{n \rightarrow \infty} y_n^2 = \lim_{n \rightarrow \infty} x_n^2,$$

hence  $\lim_{n \rightarrow \infty} x_n^2 = 2$ . We use again the rules of arithmetic for limits and  $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$  to conclude

$$\lim_{n \rightarrow \infty} x_n = r + \lim_{n \rightarrow \infty} \frac{1}{n} = r, \text{ hence, } 2 = \lim_{n \rightarrow \infty} x_n^2 = \left( \lim_{n \rightarrow \infty} x_n \right)^2 = r^2. \quad \blacksquare$$

## 17.11 AoP Ch.11: Rational and Irrational Numbers

No material so far for addenda to B/G ch.11.

## 17.12 AoP Ch.12: Decimal Expansions

**Notation Alert:**

For all of this chapter let  $\mathcal{D} := \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$  denote the set of digits.

Much of the following proposition is part of B/G prop.12.4 and prop.12.5 on p.116.

**Proposition 17.27** (Geometric series for decimals). Let  $n \in \mathbb{N}$  and  $d_j \in \mathcal{D}$  for  $j \geq n$ . Then

- a.  $0 \leq 9 \sum_{j=n}^{\infty} 10^{-j} = \frac{1}{10^{n-1}},$   
 b.  $\sum_{j=n}^{\infty} d_j 10^{-j} \leq \frac{1}{10^{n-1}},$   
 c.  $\sum_{j=n}^{\infty} d_j 10^{-j} = \frac{1}{10^{n-1}} \Leftrightarrow d_j = 9 \forall j \geq n.$

*PROOF of a and b:*

$$\sum_{j=n}^{\infty} d_j 10^{-j} \leq 9 \cdot \sum_{j=n}^{\infty} 10^{-j} = 9 \cdot 10^{-n} \sum_{k=0}^{\infty} 10^{-k} = \frac{9}{10^n} \cdot \frac{1}{9/10} = 10^{n-1}$$

*PROOF of c:* For  $k \geq n$  let  $s_k = \sum_{j=n}^k d_j 10^{-j}$ ,  $s = \sum_{j=n}^{\infty} d_j 10^{-j}$ ,  $s'_k = 9 \sum_{j=n}^k 10^{-j}$ ,  $s' = 9 \sum_{j=n}^{\infty} 10^{-j}$ .

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.9.13 on p.156) and the fact that the difference of the limits is the limit of the differences, and we obtain

$$s' - s = \lim_{k \rightarrow \infty} (s'_k - s_k) \geq 10^{-j_0}.$$

This proves " $\Rightarrow$ ". The " $\Leftarrow$ " direction is immediate from part a. ■

**Lemma 17.2.** Let  $x = m + \sum_{j=1}^{\infty} d_j 10^{-j}$  ( $m \in \mathbb{Z}_{\geq 0}$ ,  $d_j \in \mathcal{D}$  for all  $j \in \mathbb{N}$ ) be a decimal expansion of a non-negative real number  $x$ . Then

- a.  $m \leq x \leq m + 1,$   
 b.  $x = m$  iff  $d_j = 0$  for all  $j \in \mathbb{N},$   
 c.  $x = m + 1$  iff  $d_j = 9$  for all  $j \in \mathbb{N}.$

*PROOF:*

Let  $s := \sum_{j=1}^{\infty} d_j 10^{-j}.$

Part a is immediate from prop.17.27.a.

Part b is clear because if  $d_n > 0$  for some  $n$  then  $s \geq s_n > d_n 10^{-n} > 0$ .

Part c follows from prop.17.27.c with  $n = 1$ . ■

The above lemma is a generalization of the following proposition.

**Proposition 17.28** (Uniqueness of Decimal Expansions (B/G prop.12.7)). Let  $m \in \mathbb{Z}$  and  $d_j \in \mathcal{D}$  for all  $j \in \mathbb{N}$  such that  $m + \sum_{j=1}^{\infty} \frac{d_j}{10^j} = 1$ . 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 17.2.c with  $m = 0$ . ■

**Theorem 17.4** (B/G thm.12.8). Let  $x \in \mathbb{R}_{\geq 0}$  have two different decimal representations

$$(17.27) \quad x = d_0 + \sum_{j=1}^{\infty} \frac{d_j}{10^j} = e_0 + \sum_{j=1}^{\infty} \frac{e_j}{10^j}$$

where  $d_0, e_0 \in \mathbb{Z}_{\geq 0}$  and  $d_j, e_j \in \mathcal{D}$  for all  $j \in \mathbb{N}$ .

Let  $K$  be the smallest subscript such that  $d_K \neq e_K$ . If  $d_K < e_K$  then  $e_K = d_K + 1$ ,  $e_j = 0$ , and  $d_j = 1$  for all  $j > K$ .

PROOF: It follows from (17.27) and  $d_j = e_j$  for  $0 \leq j < K$  that

$$(17.28) \quad 10^K \sum_{j=K}^{\infty} \frac{d_j}{10^j} = 10^K \sum_{j=K}^{\infty} \frac{e_j}{10^j}, \quad \text{i.e.,} \quad d_K + \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} = e_K + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j},$$

$$(17.29) \quad \text{hence} \quad \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} = (e_K - d_K) + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j}.$$

$$(17.30) \quad \text{Thus} \quad \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j} \leq \sum_{j=1}^{\infty} \frac{9}{10^j} = 1 \leq (e_K - d_K) + \sum_{j=1}^{\infty} \frac{e_{K+j}}{10^j} = \sum_{j=1}^{\infty} \frac{d_{K+j}}{10^j}.$$

Here the first equality of (17.30) follows from prop.17.28 (with  $m = 0$ ). We obtain the second inequality of (17.30) from  $(e_K - d_K) \geq 1$  and  $e_{K+j} \geq 0$ . The last equality of (17.30) follows from (17.29).

(17.30) 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.17.28 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 17.13.** Note that if  $x \in \mathbb{R}$  has a finite decimal expansion  $x = m + \sum_{j=1}^K \frac{d_j}{10^j}$  ( $K \in \mathbb{N}$ ), 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 17.4.** If  $x \in \mathbb{R}$  has different decimal expansions then  $x \in \mathbb{Q}$ .

*PROOF:* It follows from thm.17.4 above that  $x$  has exactly two expansions, one of which has finitely many terms. It then follows from rem.17.13 that  $x$  is rational. ■

The following is copied from B/G for convenience:

**Definition 17.8** (Repeating Decimals). A non-negative decimal

$$x = m.d_1d_2\dots = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad (d_j \in \{0, 1, 2, \dots, 9\})$$

is **repeating** if there are natural numbers  $N$  and  $p$  such that

$$d_{N+n+kp} = d_{N+n} \quad \forall 0 \leq n < p, k \in \mathbb{N}. \quad \square$$

Note that the above definition INCLUDES  $p = 1$  and  $d_N = 0$  (finite expansion!).

**Proposition 17.29** (B/G Prop.12.11, p.119). Every repeating decimal represents a rational number.

*PROOF:* Let  $x = (m, d_1, d_2, \dots) = m + \sum_{j=1}^{\infty} d_j 10^{-j}$  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

$$(17.31) \quad \sum_{j=N}^{N+Kp-1} d_j 10^{-j} = \sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+kp+n} 10^{-(N+kp+n)} = \sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)}$$

The last equation results from the periodicity of length  $p$ . We take limits  $K \rightarrow \infty$  on both sides above <sup>153</sup> and we obtain

$$\begin{aligned} \sum_{j=N}^{\infty} d_j 10^{-j} &= \lim_{K \rightarrow \infty} \sum_{j=N}^{N+Kp-1} d_j 10^{-j} \\ &= \lim_{K \rightarrow \infty} \sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{p-1} d_{N+n} 10^{-(N+kp+n)}. \end{aligned}$$

<sup>153</sup> 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.9.6 (Subsequences of convergent sequences) on p.151.

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 17.2.** We illustrate (17.31) 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  $N = 17$ , Summation of the first four period  $K = 4$  should start at  $j = N = 5$  and stop at  $j = 17 - 1 = 16$ . Formula (17.31) 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

$$\sum_{k=0}^{K-1} \sum_{n=0}^{p-1} d_{N+kp+n} 10^{-(N+kp+n)} = \sum_{k=0}^3 \sum_{n=0}^{3-1} d_{5+3k+n} 10^{-(5+3k+n)} = \sum_{k=0}^3 \sum_{n=0}^2 d_{5+3k+n} 10^{-(5+3k+n)}$$

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,  $k = 4$  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 17.14.** 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

$$2 - 6 - 4 - 5 - 1 - 3 - 2 \dots$$

and once we have the same remainder, we are in an endless loop.

**Note 17.2** (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 zeroes 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  $x \in \mathbb{Q}$ . □

### 17.13 AoP Ch.13: Cardinality

**Proposition 17.30** (B/G prop.13.2). Let  $X, Y \neq \emptyset$  and let  $f : X \xrightarrow{\sim} Y$  be a bijection. Let  $a \in X$  and let

$$(17.32) \quad \tilde{f} : X \setminus \{a\} \rightarrow Y \setminus \{f(a)\}; \quad x \mapsto f(x)$$

be the restriction  $f|_{X \setminus \{a\}}$  of  $f$  to  $X \setminus \{a\}$  (with  $f(a)$  having been removed from the original codomain). Then  $\tilde{f}$  is bijective.

*PROOF:* This is cor.7.4 on p.124. ■

**Proposition 17.31** (B/G prop.13.6.). Every subset of a finite set is finite.

*PROOF:* This is cor.7.3 on p.124. ■

### 17.14 Exercises for Ch.17

All exercises appear at the end of the individual subchapters.



## 18 Other Appendices

### 18.1 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:  $A = \text{Alpha}$   $B = \text{Beta}$ . 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!

$\alpha$ alpha	$\theta$ theta	$\xi$ xi	$\phi$ phi
$\beta$ beta	$\vartheta$ theta	$\pi$ pi	$\varphi$ phi
$\gamma$ gamma	$\iota$ iota	$\rho$ rho	$\chi$ chi
$\delta$ delta	$\kappa$ kappa	$\varrho$ rho	$\psi$ psi
$\epsilon$ epsilon	$\varkappa$ kappa	$\sigma$ sigma	$\omega$ omega
$\varepsilon$ epsilon	$\lambda$ lambda	$\varsigma$ sigma	
$\zeta$ zeta	$\mu$ mu	$\tau$ tau	
$\eta$ eta	$\nu$ nu	$\upsilon$ upsilon	

$\Gamma$ Gamma	$\Lambda$ Lambda	$\Sigma$ Sigma	$\Psi$ Psi
$\Delta$ Delta	$\Xi$ Xi	$\Upsilon$ Upsilon	$\Omega$ Omega
$\Theta$ Theta	$\Pi$ Pi	$\Phi$ Phi	

### 18.2 Notation

This appendix on notation has been provided because future additions to this document may use notation which has not been covered in class. It only covers a small portion but provides brief explanations for what is covered.

For a complete list check the list of symbols and the index at the end of this document.

**Notations 18.1. a)** If two subsets  $A$  and  $B$  of a space  $\Omega$  are disjoint, i.e.,  $A \cap B = \emptyset$ , then we often write  $A \uplus B$  rather than  $A \cup B$  or  $A + B$ . Both  $A^c$  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  $\mathbb{N} = \{1, 2, 3, \dots\}$  of all natural numbers excludes the number zero. We write  $\mathbb{N}_0$  or  $\mathbb{Z}_+$  or  $\mathbb{Z}_{\geq 0}$  for  $\mathbb{N} \uplus \{0\}$ .  $\mathbb{Z}_{\geq 0}$  is the B/G notation. It is very unusual but also very intuitive.  $\square$

**Definition 18.1.** Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence of real numbers. We call that sequence **non-decreasing** or **increasing** if  $x_n \leq x_{n+1}$  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 **non-increasing** 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  $n \in \mathbb{N}$ .  $\square$

## References

- [1] *Matthias Beck and Ross Geoghegan. The Art of Proof. Springer, 1st edition, 2010.*
- [2] *Regina Brewster and Ross Geoghegan. Business Calculus, a text for Math 220, Spring 2014. Binghamton University, 11th edition, 2014.*
- [3] *Matthew Brin and Gerald Marchesi. Linear Algebra, a text for Math 304, Spring 2016. Binghamton University, 13th edition, 2016.*
- [4] *John Bryant and Penelope Kirby. Course Notes for MAD 2104 Discrete Mathematics I. Florida State University.*
- [5] *Gustave Choquet. Lectures on Analysis, Vol. 2: Representation Theory. Benjamin, New York, 1st edition, 1969.*
- [6] *Richard M. Dudley. Real Analysis and Probability. Cambridge University Press, Cambridge, New York, 2nd edition, 2002.*
- [7] *Norman B. Haaser and Joseph A. Sullivan. Real Analysis. publisher = Van Nostrand, year = 1971, 1st edition.*
- [8] *A.N. Kolmogorov and S.V. Fomin. Introductory Real Analysis. Dover, Mineola, 1st edition, 1975.*
- [9] *James R. Munkres. Topology. Prentice-Hal, 1st edition, 2000.*
- [10] *Walter Rudin. Principles of Mathematical Analysis. McGraw-Hill, New York, San Francisco, Toronto, London, 2nd edition, 1964.*
- [11] *James Stewart. Single Variable Calculus. Thomson Brooks Cole, 7th edition, 2012.*
- [12] *Boto Von Querenburg. Mengentheoretische Topologie. Springer, Berlin, Heidelberg, New York, 1st edition, 1973.*
- [13] *Richard E. Williamson, Richard H. Crowell, and Hale F. Trotter. Calculus of Vector Functions. Prentice Hall, Englewood Cliffs, 3rd edition, 1972.*

## List of Symbols

- $(V, \|\cdot\|)$  (normed vector space), 191  
 $-A$ , 148, 311  
 $1_A$  (indicator function of  $A$ ), 125  
 $2^\Omega, \mathfrak{P}(\Omega)$  (power set, 15  
 $A + b$ , 148, 311  
 $A \uplus B$  (disjoint union), 329  
 $A^c$  (complement), 329  
 $A_{lowb}$  (lower bounds of  $A$ ), 145  
 $A_{uppb}$  (upper bounds of  $A$ ), 145  
 $F_0$  (contradiction statement), 47  
 $N_\varepsilon^A(a)$  (Trace of  $N_\varepsilon^A(a)$  in  $A$ ), 220  
 $T_0$  (tautology statement), 47  
 $[a, b[, ]a, b]$  (half-open intervals), 18  
 $[a, b]$  (closed interval), 18  
 $[x]_f$  (fiber of  $f$  over  $f(x)$ ), 122  
 $\Leftrightarrow$  (logical equivalence), 48  
 $f_n(\cdot) \rightarrow f(\cdot)$  (pointwise convergence), 245  
 $f_n(\cdot) \xrightarrow{uc} f(\cdot)$  (uniform convergence), 245  
 $\Rightarrow$  (implication), 50  
 $\|\vec{x}\|_p$  ( $p$ -norm on  $\mathbb{R}^n$ ), 192  
 $\|f\|$  (norm of linear  $f$ ), 242  
 $\|f\|_{L^2}$  ( $L^2$ -norm), 195  
 $\|f\|_{L^p}$  ( $L^p$ -norm on  $\mathcal{C}([a, b], \mathbb{R})$ ), 196  
 $\|x\|_\bullet$  (Norm for  $x \bullet y$ ), 193  
 $\mathfrak{P}(\Omega), 2^\Omega$  (power set, 15  
 $\mathcal{U}$  (universe of discourse), 40  
 $\vec{x}$  vector, 113  
 $\bar{A}$  (closure of  $A$ ), 223, 224  
 $\bigcap [A_i : i \in I]$ , 109  
 $\bigcap_{i \in I} A_i$ , 109  
 $\bigcup [A_i : i \in I]$ , 109  
 $\bigcup_{i \in I} A_i$ , 109  
 $\chi_A$  (indicator function of  $A$ ), 125  
 $\complement A$  (complement), 329  
 $\exists$  (exists), 64  
 $\exists!$  (exists unique), 64  
 $\forall$  (for all), 64  
 $\frac{n}{m}$  divisor, 308  
 $\inf(x_i), \inf(x_i)_{i \in I}, \inf_{i \in I} x_i$  (families), 149  
 $\inf(x_n), \inf(x_n)_{n \in \mathbb{N}}, \inf_{n \in \mathbb{N}} x_n$  (sequences), 149  
 $\inf(A)$  (infimum of  $A$ ), 145  
 $\inf(A)$  infimum, 311  
 $\inf_{x \in A} f(x)$  (infimum of  $f(\cdot)$ ), 149  
 $\inf_A f$  (infimum of  $f(\cdot)$ ), 149  
 $\lambda^1, \lambda^2, \dots, \lambda^n$ , (Lebesgue measure), 302  
 $\leftrightarrow$  (double arrow logic op.), 48  
 $\lim_{n \rightarrow \infty} x_n$ , 109, 150, 213  
 $\liminf_{n \rightarrow \infty} A_n$ , 170  
 $\liminf_{n \rightarrow \infty} f_n$ , 168  
 $\liminf_{n \rightarrow \infty} x_j$  (limit inferior), 160  
 $\limsup_{n \rightarrow \infty} A_n$ , 170  
 $\limsup_{n \rightarrow \infty} f_n$ , 168  
 $\limsup_{n \rightarrow \infty} x_j$  (limit superior), 160  
 $\mathbb{1}_A$  (indicator function of  $A$ ), 125  
 $\mathbb{N}, \mathbb{N}_0$ , 329  
 $\mathbb{R}^+, \mathbb{R}_{>0}$ , 329  
 $\mathbb{R}_+, \mathbb{R}_{\geq 0}$ , 329  
 $\mathbb{R}_{>0}, \mathbb{R}^+$ , 329  
 $\mathbb{R}_{\geq 0}, \mathbb{R}_+$ , 329  
 $\mathbb{Z}_+, \mathbb{Z}_{\geq 0}$ , 329  
 $\text{epi}(f)$  (epigraph), 286  
 $\chi = x_{n^*}, x_{n^*+1}, \dots, x_{n^*}$  list, 113  
 $\max(A), \max A$  (maximum of  $A$ ), 145, 310  
 $\min(A), \min A$  (minimum of  $A$ ), 145, 310  
 $\neg$  (negation), 43  
 $\partial A$  (boundary of  $A$ ), 212, 224  
 $\sup(x_i), \sup(x_i)_{i \in I}, \sup_{i \in I} x_i$  (families), 149  
 $\sup(x_n), \sup(x_n)_{n \in \mathbb{N}}, \sup_{n \in \mathbb{N}} x_n$  (sequences), 149  
 $\sup(A)$  (supremum of  $A$ ), 145  
 $\sup(A)$  supremum, 311  
 $\sup_{x \in A} f(x)$  (supremum of  $f(\cdot)$ ), 149  
 $\sup_A f$  (supremum of  $f(\cdot)$ ), 149  
 $\rightarrow$  (arrow operator), 50  
 $\vee$  (disjunction), 49  
 $|X|$  (size of a set), 82, 130  
 $\wedge$  (conjunction), 44  
 $]a, b[$  (open interval), 18  
 $a < b$  ordered integral domain, 36  
 $a \ominus b$  difference, 33  
 $f(A)$  (direct image), 118  
 $f^{-1}(B)$  (indirect image, preimage), 118  
 $g \circ f$  (function composition), 90  
 $g \circ f(x)$  (function composition), 90  
 $m + n \pmod 2$  (addition mod 2), 127  
 $n/m$  divisor, 308  
 $n \div m$  divisor, 308

$r^*$  (rational cut), 296  
 $x \bullet y$  (inner product), 188  
 $x \bullet y$  (inner product), 188  
 $x \diamond y$  binary operation, 98  
 $x \in X$  (element of a set), 11, 18  
 $x \notin X$  (not an element of a set), 11  
 $x^\bullet$  unary operation, 98  
 $x_n \rightarrow -\infty$ , 151  
 $x_n \rightarrow \infty$ , 151  
 $x_n \rightarrow a$ , 109, 150, 213  
 $\prod_{j=k}^n x_j$  prod, 314  
 $\sum_{j=k}^n x_j$  sum, 313  
 $(X, d(\cdot, \cdot))$  (metric space), 205  
 $(A, d_{A \times A})$  (metric subspace), 219  
 $(\Omega, \mathfrak{F})$  (measurable space), 301  
 $(\Omega, \mathfrak{F}, \mu)$  (measure space), 302  
 $(x, y)^T$  (transpose), 92  
 $(x_1, x_2, \dots, x_N)$  ( $N$ -tuple), 115  
 $(x_1, x_2, x_3, \dots, x_n)$  ( $n$ -dimensional vector), 174  
 $-f(\cdot), -f$  (negative function), 100  
 $-x$  (negative of  $x$ ), 180  
 $0(\cdot)$  (zero function), 100  
 $A \times B$  (cartesian product of 2 sets), 81  
 $A^c$  (complement of  $A$ ), 14  
 $D_f$  (natural domain of  $f$ ), 85  
 $N_\varepsilon(x_0)$  ( $\varepsilon$ -neighborhood), 210  
 $X^I = \prod_{i \in I} X$  (cartesian product), 115  
 $X_1 \times \dots \times X_N$  (cartesian product), 114  
 $[x]_\sim, [x]$  (equivalence class), 82  
 $\Gamma_f, \Gamma(f)$  (graph of  $f$ ), 88  
 $\|f\|_\infty$  (sup-norm), 190  
 $\|x\|$  (norm on a vector space), 191  
 $\mathcal{C}_{\mathcal{B}}(X, \mathbb{R})$ , 239  
 $\mathcal{B}$  (base of a topology), 217  
 $\mathfrak{F}$  ( $\sigma$ -algebra), 301  
 $\mathfrak{N}(x)$  (neighborhood system), 218  
 $\mathfrak{U}_{\|\cdot\|}$  (norm topology), 215  
 $\mathfrak{U}_{d(\cdot, \cdot)}$  (metric topology), 215  
 $\vec{x} + \vec{y}$  (vector sum), 175  
 $\alpha \vec{x}$  (scalar product), 175  
 $\alpha f$  (scalar product of functions), 100  
 $\alpha x, \alpha \cdot x$  (scalar product), 180  
 $\mathbb{C}A$  (complement of  $A$ ), 14  
 $\emptyset$  (empty set), 12  
 $\lambda A + b$  (translation/dilation in  $\mathbb{R}$ ), 148  
 $\lambda A + b$ , 311  
 $\mapsto$  (maps to), 88  
 $\mathbb{N}_0$  (non-negative integers), 18  
 $\mathbb{R}^*$  (non-zero real numbers), 18  
 $\mathbb{R}^+$  (positive real numbers), 18  
 $\mathbb{R}_{>0}$  (positive real numbers), 18  
 $\mathbb{R}_{\geq 0}$  (non-negative real numbers), 18  
 $\mathbb{R}_{\neq 0}$  (non-zero real numbers), 18  
 $\mathbb{R}_+$  (non-negative real numbers), 18  
 $\mathbb{Z}_{\geq 0}$  (non-negative integers), 18  
 $\mathbb{Z}_+$  (non-negative integers), 18  
 $\mathbb{N}$  (natural numbers), 16  
 $\mathbb{Q}$  (rational numbers), 16  
 $\mathbb{R}$  (real numbers), 16  
 $\mathbb{Z}$  (integers), 15, 16  
 $\text{span}(A)$  (linear span), 183  
 $\mu(\cdot)$  (finite measure), 302  
 $\mu(\cdot)$  (measure), 301  
 $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$  (extended real numbers), 168  
 $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$  (extended real numbers), 301  
 $\overline{\mathbb{R}}_+$  (non-negative extended), 301  
 $\pi_j(\cdot)$  ( $j$ th coordinate function), 185  
 $\pi_{i_1, i_2, \dots, i_m}(\cdot)$  ( $m$ -dim projection), 185  
 $\mathcal{B}(X, \mathbb{R})$  (bounded real valued functions on  $X$ ), 181  
 $\mathcal{C}(A, \mathbb{R})$  (cont. real valued functions on  $A \subseteq \mathbb{R}$ ), 181  
 $\mathcal{C}(X, \mathbb{R})$  (continuous real valued functions on  $X$ ), 238  
 $\mathcal{F}(X, \mathbb{R})$  (real valued functions on  $X$ ), 181  
 $\preceq_A$  (partial order on subset), 84  
 $\prod_{i \in I} X_i$  (cartesian product), 115  
 $\therefore$  (therefore), 73  
 $\varepsilon_{x_0}$  (point mass), 185  
 $d_{A \times A}$  (induced/inherited metric), 219  
 $d_{\|\cdot\|}$  (metric induced by norm), 206  
 $f : X \xrightarrow{\sim} Y$  (bijective function), 94  
 $f|_A$  (restriction of  $f$ ), 99  
 $f + g$  (sum of functions), 100  
 $f - g$  (difference of functions), 100  
 $f/g$  (quotient of functions), 100  
 $f^{-1}(\cdot)$  (inverse function), 94  
 $fg, f \cdot g$  (product of functions), 100  
 $xRy$  (equivalent items), 81

$x \preceq y$  (precedes), 84  
 $x \sim y$  (equivalent items), 82  
 $x \succ y$  (succeeds), 84  
 $x + y$  (vector sum), 179  
 $\|\vec{v}\|_2$  (Euclidean norm), 177  
**false**, 39  
**true**, 39  
**xor** (exclusive or), 49  
 $\{\}$  (empty set), 12  
 $(x_i)_{i \in J}$  (family), 102  
 $(A, \mathfrak{U}_A)$  - topol. subspace, 221  
 $[n] = \{1, 2, \dots, n\}$ , 130  
 $\mathfrak{U}_A$  - induced/inherited topology, 221  
 $\mathfrak{U}_A$  - subspace topology, 221  
 $A \cap B$  (A intersection B), 13  
 $A \setminus B$  (A minus B), 14  
 $A \subset B$  (A is strict subset of B), 12  
 $A \subseteq B$  (A is subset of B), 12  
 $A \subsetneq B$  (A is strict subset of B), 12  
 $A \Delta B$  (symmetric difference of A and B), 14  
 $A \uplus B$  (A disjoint union B), 13  
 $B \supset A$  (B is strict superset of A), 12  
 $B \supseteq A$  (B is strict superset of A), 12  
 $g^{-1}$  group: inverse element, 30  
 $(X, \mathfrak{U})$  (topological space), 215  
 $(x_j)$  (sequence), 104  
 $(x_j)$  (sequence), 248  
 $(x_{n_j})$  (subsequence), 104  
 $\liminf_{n \rightarrow \infty} A_n$  (limit inferior for sets), 305  
 $\limsup_{n \rightarrow \infty} A_n$  (limit superior for sets), 305  
 $\sum_{k=1}^{\infty} a_k$  (series), 158  
 $\|\vec{v}\|_2$  (length or Euclidean norm of  $\vec{v}$ ), 176  
 $\mathfrak{U}$  (topology), 215  
 $A \cup B$  (A union B), 13  
 $A \supseteq B$  (A is superset of B), 12  
 $A^\circ$  (interior of A), 216  
 $A_n \nearrow \bigcup_n A_n$ , 170  
 $A_n \searrow \bigcap_n A_n$ , 170  
 $\text{int}(A)$  (interior of A), 216  
 $x_n \nearrow \xi$  ( $n \rightarrow \infty$ ), 152  
 $x_n \searrow \xi$  ( $n \rightarrow \infty$ ), 152  
 $\text{card}(X) < \text{card}(Y)$ , 130  
 $\text{card}(X) = \text{card}(Y)$ , 130  
 $\text{card}(X) > \text{card}(Y)$ , 130  
 $\text{card}(X) \geq \text{card}(Y)$ , 130  
 $\text{card}(X) \leq \text{card}(Y)$ , 130  
 $\dim(V)$  (dimension of vector space V), 187  
**F** (false), 39  
g.l.b.(A) (greatest lower bound of A), 145  
l.u.b.(A) (least upper bound of A), 145  
L/I (logically impossible), 45  
**T** (true), 39

## Index

- $L^2$ -norm, 195
- $N$ -tuple, 115
- $\|\cdot\|_\infty$  distance, 207
- $\sigma$ -algebra, 301
- $\sigma$ -field, 301
- $\varepsilon$ - $\delta$  continuous function, 233
- $\varepsilon$ -closeness, 205
- $\varepsilon$ -grid, 266
- $\varepsilon$ -net, 266
  
- abelian group, 29
- absolute convergence, 252
- absolute value, 18
- abstract integral, 185
- addition, 32
- addition mod 2, 127
- after, 84
- algebra of sets, 116
- algebraic number, 140
- alternating harmonic series, 254
- alternating series, 253
- antecedent, 52
- antiderivative, 25
- antisymmetric relation, 82
- area, 193
  - net area, 193
- argument, 88
- arrow operator, 50
- assertion, 69
  - valid, 69
- assignment operator, 20, 88
- associativity, 27, 180
- axiom, 69
- Axiom of Choice, 97, 279
  
- base (of a topology), 217
- basis, 186
- before, 84
- Bernstein polynomial, 240
- bilinear, 188
- binary operation, 98
- binary operator, 44
- bound variable, 40
- boundary, 212, 216
- boundary (in a topological space), 224
- boundary point, 212
- boundary point (in a topological space), 224
- boundary point (topological space), 216
- bounded, 145, 310
- bounded above, 145, 310
- bounded below, 145, 310
  
- cancellation rule, 34
- cardinality
  - comparison of, 130
  - equality, 130
- cardinality (equivalence class), 141
- cartesian product, 21, 81, 114
- cartesian product of  $N$  sets, 114
- cartesian product of a family, 115
- Cauchy criterion, 226
- Cauchy sequence, 226
- chain, 278
- characteristic function, 125
- choice function, 279
- closed interval, 18
- closed set (in a metric space), 223
- closed set (in a topological space), 224
- closed with respect to an operation, 181
- closure (in a metric space), 223
- closure (in a topological space), 224
- closure operator, 226
- cluster point, 223
- codomain, 20, 88
- commutative group, 29
- commutative ring with unit, 33
- commutativity, 29, 179
- compact, 272
  - covering compact, 272
  - sequentially, 268
- complement, 14
- complete set, 227
- completeness axiom, 145, 296
- composite, 317
- composite number, 317
- composition, 90
- compound statement, 43
- compound statement function, 43
- concave-up, 286
- conclusion, 52

*conditionally convergent series*, 253  
*conjecture*, 69  
*conjugate indices*, 197  
*conjunction operator*, 44  
*connective*, 42  
     *negation*, 43  
*consequent*, 52  
*contact point (in a metric space)*, 222  
*contact point (in a topological space)*, 224  
*content*, 303  
*continuity at  $x_0$* , 235  
*continuous real-valued function*, 156  
*contradiction*, 47  
*contradictory*, 45  
*contrapositive*, 52  
*convergence*, 213  
*convergence in  $\mathbb{R}$* , 109, 150  
*convergence, uniform*, 245  
*converse*, 52  
*convex*, 286  
*coordinate function*, 185  
*corollary*, 70  
*countable set*, 131  
*countably infinite set*, 131  
*covering*, 271  
     *extract finite open subcovering property*, 272  
*cut*, 296  
  
*De Morgan's Law*, 111  
*decimal*, 16  
     *repeating*, 326  
*decimal digit*, 15  
*decimal numeral*, 16  
*decimal point*, 16  
*decreasing sequence*, 329  
*Dedekind cut*, 296  
     *lower number*, 296  
     *upper number*, 296  
*denominator*, 308  
*dense set*, 17  
*difference*, 33  
*digit*, 15  
*dimension*, 174, 187  
*direct image*, 118  
*discrete metric*, 207  
*discrete topology*, 215  
*disjunction operator*, 49  
  
*distributive laws*, 180  
*distributivity*, 33  
*dividend*, 308  
*divisor*, 308  
*domain*, 20, 88  
*dot product*, 188  
*double arrow operator*, 48  
*dummy variable*  
     *functions*, 89  
*dummy variable (setbuilder)*, 11  
  
*element of a set*, 11  
*embed*, 182  
*empty set*, 12  
*epigraph*, 286  
*equality*  
     *arbitrary cartesian products*, 115  
     *cartesian products*, 81  
     *finite cartesian products*, 114  
*equivalence class*, 82  
*equivalence operator*, 48  
*equivalence relation*, 82  
*equivalent*, 82  
*Euclidean norm*, 177  
*event*, 302  
*exclusive or operator*, 49  
*existential quantification*, 64  
*existential quantifier*, 64  
*extended real numbers line*, 168  
*extended real valued function*, 168, 301  
*extension of a function*, 99  
*exterior point*, 216  
*exterior point (topological space)*, 216  
*extract finite open subcovering property*, 272  
  
*factor (prime)*, 317  
*factorization (prime)*, 317  
*family*, 102  
     *mutually disjoint*, 111  
     *supremum*, 149  
*fiber over  $f(x)$* , 122  
*finite list*, 113  
*finite measure*, 302  
*finite set*, 130  
*finite subcovering*, 272  
*first axiom of countability*, 218  
*first countable*, 218

*function*, 20, 88  
 $\| \cdot \|_{\infty}$  distance, 207  
 $\varepsilon$ - $\delta$  continuous, 233  
 argument, 88  
 assignment operator, 20, 88  
 bijective, 94  
 bilinear, 188  
 bounded from above, 148  
 bounded from below, 148  
 bounded function, 148  
 codomain, 20, 88  
 composition, 90  
 constant function, 100  
 continuous in topological spaces, 236  
 convergence, 245  
 difference, 100  
 direct image, 118  
 domain, 20, 88  
 domain, natural, 85  
 extension, 99  
 fiber over  $f(x)$ , 122  
 function value, 88  
 identity, 90  
 image, 88  
 independent variable, 88  
 infimum, 149  
 injective, 94  
 inverse, 94  
 left inverse, 97  
 linear function on  $\mathbb{R}$ , 32  
 maps to operator, 20, 88  
 maximal displacement distance, 207  
 mean distance, 209  
 mean square distance, 209  
 natural domain, 85  
 negative function, 100  
 onto, 94  
 pointwise convergence, 245  
 preimage, 88, 118  
 product, 100  
 quotient, 100  
 range, 88  
 real function, 99  
 real valued function, 89, 99  
 restriction, 99  
 right inverse, 97  
 scalar product, 100  
 sequence continuous, 233  
 sum, 100  
 sup-norm distance, 207  
 supremum, 149  
 surjective, 94  
 target, 88  
 uniform continuity, 241  
 uniform convergence, 245  
 zero function, 100  
 function value, 88  
 graph, 20, 88  
 greatest lower bound, 145, 311  
 greek letters, 329  
 grid point, 266  
 group, 29  
     homomorphism, 290  
     isomorphism, 290  
     structure compatible functions, 290  
     subgroup, 37  
 half-open interval, 18  
 harmonic series, 254  
 Hoelder's inequality, 198  
 Hoelder's inequality in  $\mathbb{R}^n$ , 200  
 homomorphism, 290  
 hypothesis, 52  
 identifying, 16  
 identity, 28, 90  
 identity function, 28  
 iff, 13  
 image, 88  
 implication, 50  
 implication operator, 50  
 impossible, logically, 45  
 inadmissible, 40  
 increasing sequence, 329  
 independent variable, 88  
 index, 101  
 index set, 102  
 indexed family, 102  
 indexed item, 101  
 indicator function, 125  
 indirect image, 118  
 indiscrete topology, 215  
 induced metric, 219



- induced order*, 36
- induced topology by metric*, 221
- infimum*, 145, 311
- infimum of a family*, 149
- infimum of a sequence*, 149
- infinite set*, 130
- inherited metric*, 219
- inherited topology from metric*, 221
- injective function*, 94
- inner point (topological space)*, 216
- inner product*, 188
  - norm, 193
- integer*, 17
- integral*
  - definite, 25
  - indefinite, 25
- integral domain*, 35
  - ordered, 36
  - positive cone, 36
- interior*, 216
- interior point*, 211
- interior point (topological space)*, 216
- intersection*
  - family of sets, 109
- interval*
  - closed, 18
  - half-open, 18
  - open, 18
- inverse*, 30
- inverse function*, 94
- inverse relation*, 85
- irrational number*, 17
- isolated point*, 223
- isomorphism*, 290
  
- L/I (logically impossible)*, 45
- least upper bound*, 145, 311
- Lebesgue measure, n-dimensional*, 302
- Lebesgue number*, 273
- left inverse*, 97
- lemma*, 70
- lim inf*, 160
- lim sup*, 160
- limit*, 109, 150, 213
- limit (set sequence)*, 305
- limit inferior*, 160
- limit inferior (set sequence)*, 305
  
- limit point*, 223
- limit superior*, 160
- limit superior (set sequence)*, 305
- linear combination*, 183
- linear function*, 184
  - norm, 242
- linear function on  $\mathbb{R}$* , 32
- linear mapping*, 184
- linear ordering relation*, 278
- linear space*, 179
- linear span*, 183
- linearly dependent*, 186
- linearly independent*, 186
- linearly ordered set*, 103, 278
- logic*
  - antecedent, 52
  - assertion, 69
  - axiom, 69
  - bound variable, 40
  - compound statement, 43
  - compound statement functions, 43
  - conclusion, 52
  - conjecture, 69
  - consequent, 52
  - contrapositive, 52
  - converse, 52
  - corollary, 70
  - existential quantification, 64
  - existential quantifier, 64
  - hypothesis, 52
  - implication, 50
  - inadmissible, 40
  - L/I, 45
  - lemma, 70
  - predicate, 40
  - premise, 52
  - proof, 69
  - proposition, 39
  - proposition function, 40
  - rule of inference, 69
  - statement, 39
  - statement function, 40
  - theorem, 70
  - truth value, 39
  - unique existential quantification, 64
  - unique existential quantifier, 64

- universal quantification, 64
- universal quantifier, 64
- universe of discourse, 40
- UoD (universe of discourse), 40
- valid assertion, 69
- logic operators
  - arrow, 50
  - conjunction, 44
  - disjunction, 49
  - double arrow, 48
  - equivalence, 48
  - exclusive or, 49
  - implication, 50
- logical equivalence, 44
- logical operator, 42
- logically equivalent, 48
- logically impossible, 45
- lower bound, 145, 310
- lowest terms, 17
- mapping
  - inverse, 94
- mapping (see function), 88
- maps to operator, 20, 88
- maximal displacement distance, 207
- maximal element, 278
- maximum, 145, 147, 278, 310
- mean distance, 209
- mean square distance, 209
- measurable space, 301
- measure, 301
- measure space, 302
- member of a set, 11
- member of the family, 102
- metric, 204
  - inherited topology, 221
  - induced, 219
  - induced topology, 221
  - inherited, 219
- metric associated with a norm, 206
- metric derived from a norm, 206
- metric induced by a norm, 206
- metric of uniform convergence, 248
- metric space, 205
  - $\epsilon$ -closeness, 205
  - continuity at  $x_0$ , 235
- metric subspace, 219
- metric topology, 215
- minimum, 145, 310
- Minkowski's inequality, 199
- Minkowski's inequality for  $(\mathbb{R}^n, \|\cdot\|_p)$ , 201
- monoid, 27
- monomial, 239
- monotone set sequence, 304
- multiplication, 32
- natural domain, 85
- natural number, 17
- negation operator, 43
- negative, 180
- neighborhood, 210, 211, 216
- neighborhood (metric space), 212
- neighborhood base, 218
- neighborhood system, 218
- net area, 193
- neutral element, 27
- no gaps property, 319
- non-decreasing set sequence, 304
- non-increasing set sequence, 304
- non-decreasing sequence, 329
- non-increasing sequence, 329
- norm
  - $L^p$ -norm on  $\mathcal{C}([a, b], \mathbb{R})$ , 196
  - $p$ -norm on  $\mathbb{R}^n$ , 192
  - Euclidean norm, 177
  - sup-norm, 190
  - supremum norm, 190
- norm associated with an inner product, 193
- norm of uniform convergence, 248
- norm on a vector space, 191
- norm topology, 215
- normalized form (polynomials), 240
- normed vector space, 191
- not countable set, 131
- null vector, 180
- nullspace, 180
- number
  - composite, 317
- number of digits w.r.t base 10, 319
- numbers
  - algebraic number, 140
  - integer, 15
  - irrational number, 17
  - natural numbers, 16

- rational numbers, 16
- real numbers, 16
- transcendental number, 140
- numerator, 308
- open covering, 272
  - Lebesgue number, 273
- open exterior, 216
- open exterior (topological space), 216
- open interval, 18
- open neighborhood (metric space), 212
- open set, 211
  - trace, 220
- open set (topological space), 215
- open sets in a subspace, 221
- order induced by positive cone, 36
- ordered integral domain, 36
- ordered pair, 81
- ordering
  - partial, 84
- origin, 115
- parallelepiped, 115
- parallelepiped,  $n$ -dimensional, 302
- partial order
  - reflexive, 84
  - strict, 84
- partial order relation, 84
- partial ordering, 84
  - after, 84
  - before, 84
- partially ordered set, 84
- partition, 25, 111
- partitioning, 25, 111
- period, 16
- period length, 16
- permutation, 251
  - infinite, 251
- point of accumulation, 223
- pointwise convergence, 245
- polynomial, 239
- POset, 84
  - maximal element, 278
  - maximum, 278
- power set, 15
- predicate, 40
- preimage, 118
- premise, 52
- prime, 317
- prime factor, 317
- prime factorization, 317
- prime number, 317
- prod, 314
- projection, 185
- projection on coordinates  $i_1, i_2, \dots, i_m$ , 185
- proof, 69
  - proof by counterexample, 29
- proof by counterexample, 29
- proposition, 39
- proposition function, 40
- punctured neighborhood, 223
- quotient, 290, 308
- range, 88
- rational cut, 296
- rational number, 17
  - lowest terms, 17
- real number, 17
- real valued function, 89
- real-valued function
  - continuous, 156
- rearrangement, 251
- reflexive, 82
- reflexive partial order, 84
- related items  $x$  and  $y$ , 81
- relation, 81
  - antisymmetric, 82
  - empty, 82
  - equivalence relation, 82
  - equivalent items, 82
  - inverse, 85
  - linear ordering, 278
  - partial order, 84
  - reflexive, 82
  - symmetric, 82
  - total ordering, 278
  - transitive, 82
- remainder, 290
- reordering, 251
- repeating decimal, 16, 326
- replacement principle for statements, 56
- restriction of a function, 99
- right inverse, 97

- ring
  - cancellation rule, 34
  - commutative, with unit, 33
  - integral domain, 35
  - zero divisor, 34
- ring of sets, 116
- rule of inference, 69
- scalar, 183
- scalar product, 180
- second axiom of countability, 219
- second countable, 219
- semigroup, 27
- sequence, 103, 248
  - decreasing, 329
  - increasing, 329
  - infimum, 149
  - non-decreasing, 329
  - non-increasing, 329
  - numeric, 158
  - partial sums, 158
  - start index, 103
  - strictly decreasing, 329
  - strictly increasing, 329
  - subsequence, 104
  - supremum, 149
  - tail set, 160
- sequence compact, 268
- sequence continuous function, 233
- sequentially compact, 268
- series, 104, 158
  - absolute convergence, 252
  - alternating, 253
  - alternating harmonic, 254
  - conditionally convergent, 253
  - convergence, 158
  - harmonic, 254
  - limit, 158
  - rearrangement, 251
  - reordering, 251
- set, 11
  - bounded, 221
  - compact, 272
  - complete, 227
  - countable, 131
  - countably infinite, 131
  - covering, 271
  - dense, 17
  - diameter, 221
  - difference, 14
  - difference set, 14
  - disjoint, 13
  - finite, 130
  - finite subcovering, 272
  - infinite, 130
  - intersection, 13
  - limit, 305
  - limit inferior, 305
  - limit superior, 305
  - linearly ordered, 103, 278
  - monotone sequence, 304
  - non-decreasing sequence, 304
  - non-increasing sequence, 304
  - not countable, 131
  - open covering, 272
  - partially ordered, 84
  - POset, 84
  - proper subset, 12
  - proper superset, 12
  - setbuilder notation, 11
  - size, 130
  - strict subset, 12
  - strict superset, 12
  - strictly decreasing sequence, 304
  - strictly increasing sequence, 304
  - subset, 12
  - superset, 12
  - symmetric difference, 14
  - totally ordered, 103, 278
  - uncountable, 131
  - union, 13
- sets
  - limit, 170
  - limit inferior, 170
  - limit superior, 170
  - ring, 116
- simple statement, 43
- size, 130
- source, 88
- span, 183
- start index, 103
- statement, 39
  - logical equivalence, 44, 48

- replacement principle, 56
- statement function, 40
- strict partial order, 84
- strictly decreasing sequence, 329
- strictly decreasing set sequence, 304
- strictly increasing sequence, 329
- strictly increasing set sequence, 304
- structure compatible functions, 290
- subgroup, 37
- sublinear functional, 282
- subsequence, 104
- subspace
  - metric, 219
  - open sets, 221
  - topological, 221
- subspace (of a vector space), 180
- subspace, generated, 184
- subscript, 101
- sum, 179, 313
- sum mod 2, 127
- sup–norm, 190
- sup–norm displacement distance, 207
- supremum, 145, 311
- supremum norm, 190
- supremum of a family, 149
- supremum of a sequence, 149
- surjective function, 94
- symmetric, 82
  
- tail set, 160
- target, 88
- tautology, 47
- theorem, 70
- topological space, 215
  - boundary, 224
  - boundary point, 224
  - closure, 224
  - contact point, 224
  - continuous function, 236
  - first axiom of countability, 218
  - first countable, 218
  - open set, 215
  - second axiom of countability, 219
  - second countable, 219
- topological spaces
  - exterior point, 216
  - inner point, 216
  - interior point, 216
  - open exterior, 216
- topological subspace, 221
- topology, 215
  - discrete topology, 215
  - generated by metric, 215
  - generated by norm, 215
  - indiscrete topology, 215
  - induced by metric, 215, 221
  - induced by norm, 215
  - inherited from metric, 221
  - metric topology, 215
  - norm topology, 215
- total ordering relation, 278
- totally bounded, 268
- totally ordered set, 103, 278
- trace, 220
- transcendental number, 140
- transitive, 82
- transpose, 92
- triangle inequality, 18, 22
- truth table, 44
- truth value, 39
  
- unary operation, 98
- unary operator, 44
- uncountable set, 131
- uniform continuity, 241
- uniform convergence, 245
  - metric, 248
  - norm, 248
- uniformly continuous, 237
- union
  - family of sets, 109
- unique existential quantification, 64
- unique existential quantifier, 64
- universal quantification, 64
- universal quantifier, 64
- universal set, 14
- universe of discourse, 40
- UoD (universe of discourse), 40
- upper bound, 145, 310
  
- valid assertion, 69
- vector, 113
  - Euclidean norm, 176
  - length, 176

*norm, Euclidean, 177*  
*scalar product, 175*  
*sum, 175, 179*  
*transpose, 92*  
*vector (element of a vector space), 180*  
*vector space, 174, 179*  
    *basis, 186*  
    *normed, 191*  
*vector, n-dimensional, 174*

*xor, 49*

*Young's inequality, 197*

*zero divisor, 34*  
*zero element, 180*  
*zero function, 100*  
*zero vector, 180*  
*ZL property (Zorn's Lemma), 278*  
*Zorn's Lemma, 278*