Math 330 - Additional Material

Michael Fochler
Department of Mathematics
Binghamton University

Last update: December 7, 2015

Version: **2015-12-07**

Contents

1 Before you start **4**

2 Notation and preliminaries (Read this!) **5**

3 Some Basics (Understand this!) **6**
3.1 Numbers . . . 6
3.2 First things about sets, Functions (Mappings) and Families . . . 9
3.2.1 Definition of sets . . . 9
3.2.2 Definition of functions, injectivity, surjectivity and bijectivity . . . 10
3.2.3 Sequences, families and functions as families . . . 12
3.3 Basic set operations and Cartesian products . . . 14
3.4 Countable sets . . . 18

4 Sets and Functions, direct and indirect images (Study this!) **22**
4.1 Basic Properties of Sets . . . 22
4.2 Direct images and indirect images (preimages) of a function . . . 22
4.3 Appendix: Cardinality - Alternate approach to Beck/Geoghegan . . . 28

5 Real functions (Understand this!) **33**
5.1 Operations on real functions . . . 33
5.2 Maxima, suprema, limsup ... **(Study this!)** . . . 34

6 Vectors and vector spaces (Understand this!) **43**
6.1 $\mathbb{R}^N$: Euclidian space . . . 43
6.1.1 N–dimensional Vectors . . . 43
6.1.2 Addition and scalar multiplication for N–dimensional vectors . . . 44
6.1.3 Length of N–dimensional vectors, the Euclidean Norm . . . 44
6.2 General vector spaces . . . 48
6.2.1 Vector spaces: Definition and examples . . . 48
6.2.2 Normed vector spaces (Skip this!) . . . 54

7 Convergence and Continuity **59**
7.1 Metric spaces **(Study this!)** . . . 59
7.1.1 Measuring the distance of real functions . . . 62
7.1.2 Bounded sets and bounded functions . . . 64
7.1.3 Neighborhoods and open sets . . . 65
7.1.4 Digression: Abstract topological spaces (Skip starting at def. 7.12: Basis and neighborhood basi)!) . . . 68
7.1.5 Convergence, contact points and closed sets . . . 70
7.1.6 Completeness in metric spaces . . . 73
7.1.7 Appendix: Addenda to chapter 7.1: Metric Spaces . . . 76
7.2 Continuity **(Study this!)** . . . 77
7.2.1 Definition and characterization of continuous functions . . . 77
7.2.2 Continuity of constants and sums and products . . . 79
7.2.3 Function spaces (Understand this!) . . . 81
7.2.4 Continuity of Polynomials (Understand this!) . . . 82

7.3 Function sequences and infinite series . . . 85
7.3.1 Convergence of function sequences **(Study this!)** . . . 85
7.3.2 Infinite Series (Understand this!) . . . 87
7.4 Appendix: Addenda to chapter 7 . . . 91
7.4.1 Convergence (**Study this!**) . . . 91
7.4.2 Completeness (**Study this!**) . . . 91
7.4.3 Uniform convergence and metric subspaces (**Study unif conv, understand subspaces!**) . . . 92
7.4.4 The Hahn-Banach separation theorem (**Skip this!**) . . . 94

8 Compactness (Study this!) 97
8.1 Introduction: Closed and bounded sets in Euclidian space (Understand this!) . . . 97
8.2 Four definitions of compactness . . . 99
8.3 ε-nets and total boundedness . . . 99
8.4 Sequence compactness . . . 102
8.5 Uniform continuity . . . 104
8.6 Open coverings and the Heine–Borel theorem . . . 106
8.6.1 Appendix: Proofs for (open covering) compactness **(Skip this!)** . . . 107
8.7 Appendix: Addenda to chapter 8 . . . 114
8.7.1 Sequence compactness . . . 114
8.7.2 Continuous functions and compact spaces . . . 115

References 117

List of Symbols 118

Index 120

History of Updates: *2015-12-02*

Date	*Topic*
2015-12-07	*Additions to ch.8.: new subchapter 8.7.2*
2015-12-02	*Additions to ch.7.4: Addenda to ch.7.*
2015-11-12	*Appendix to ch.7.1: The opposite to convergence at L.*
	Subchapter on cardinality added to the end of ch. 4.
2015-10-04	*Re-worked ch.3; additional material on families and more*
	Section “Convergence and Continuity” now complements appendix A of B/G.
2015-04-20	*Material starting with “Maxima, suprema, limsup ...” has been reorganized*
	Section “Convergence and Continuity” now complements appendix A of B/G.
2015-04-13	*Major rework of “Maxima, suprema, limsup ...”*
2015-03-25	*Added background material: new chapters “Some Basics”, “Real Functions”,*
	“Vectors and vector spaces”, “Convergence and continuity”
2015-03-17	*Added new section “Maxima, suprema, limsup ...”*
2015-03-08	*Added new section “Basic properties of sets”*

1 Before you start

This write-up provides some additional background on material that cannot found in sufficient detail in the [1] B/G (Beck/Geoghegan) text book or the additional documents I published on the home page of the Math 330 course.

How you know what to focus on:

Scrutinize the table of contents, including the headings for the subchapters:

When you read "Study this", you should understand the material in depth, comparable to the Beck Geoghegan book.

When you read "Understahd this", you should know the definitions, propositions and theorems without worrying about proofs. Chances are that the material will be referred to from essential sections of this write-up and needed for their understanding.

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

All directives apply to the entire subtree and a lower level directive overrides the "parent directives". Example: the "Understand this!" directive of subsection 7.2.4: Continuity of Polynomials overrides the "Study this!" directive of subsection 7.2 on Continuity.

Accordingly, when you do not see any comment, back up in the table of contents until you find one.

The material consists of two very distinct portions.

A. Material directly related to Math 330:

	Topic
1.	*All of ch.4, p.22: "Sets and Functions, direct and indirect images"*
2.	*Ch.5.2, p.34: "Maxima, suprema, limsup ..."*
3.	*Almost all of Ch.7, p.59) on "Convergence and Continuity". Major exception: most of subsection 7.1.4 ("Digression: Abstract topological spaces") on p.68 can be skipped.*
4.	*Ch.8, p.97: "Compactness". Much of this chapter will be relevant starting Monday, April 27, possibly earlier.*

B. Material to help you understand topics taught in the course.
This includes everyting not listed in A above. This material is optional and was provided to you under the theory that, particularly in Math, more words take a lot less time to understand than a skeletal write-up like the one given in the course text.

Accordingly, almost all of the material provided in this document comes with quite detailed proofs. Those proofs are there for you to study. Some of those proofs, notably those in prop. 4.2, make use of " $\iff$ " to show that two sets are equal.

As I said many times in class, you should abstain from using " $\iff$ " between statements in your proofs as you very likely lack the experience to do so without error.

Almost all of the material in A (directly related to the course) was written from scratch with the exception of chapter 7. The remainder was pulled in from a document that was written more than five years ago. I have *made some alterations in the 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.*

Some of those alterations ***that may not have been done with*** 100% ***consistency*** *are:*

a. countable and countably infinite v.s. denumbrable and countable:
We use the B/G definitions: A set A is countable if it is either finite or infinite, but sequentiable (the elements of A can be indexed $a_1, a_2, a_3, \dots$) and "countably infinite" means countable but not finite. Originally I used the term "countable" for what we now call "countably infinite" whereas the term "denumbrable" was used to indicate that A is either finite or countably infinite.

b. Inclusion of sets $B \subseteq A$:
The great majority of all books that I have read use $B \subset A$ to indicate that each element of B also belongs to A whereas the notation $B \subsetneq A$ is used to indicate that, in addition, there is at least one $a \in A$ that does not belong to B. I have converted this to match the B/g notation we also use in the course: $B \subseteq A$ rather than $B \subset A$ means that each element of B also belongs to A. $B \subset A$ means that, in addition, there is at least one $a \in A$ that does not belong to B. I also write $B \subsetneq A$ if i want to emphasize that we deal with strict inclusion that excludes equality of A and B.

c. Neighborhoods $B_\varepsilon(x)$ of "radius" ε around x These sets were originally denoted $N_\varepsilon(x)$ and if you see either this expression or $N_\delta(x)$ then you have found one that I have overlooked.

There is also a difference in style: the original document is written in a much more colloquial style as it was addressed to high school students who had expressed a special interest in studying math.

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

2 Notation and preliminaries (Read this!)

This introductory chapter on the notation used has been provided because future additions to this document may use notation which has not been covered in class.

Notation 2.1. a) If two subsets A and B of a space Ω are disjoint, i.e., $A \cap B = \emptyset$, then we often write $A \biguplus B$ rather than $A \cup B$ or $A + B$. Both $\complement A$ and $A^{\complement}$ denote the complement $\Omega \setminus A$ of A.

b) $\mathbb{R}_{>0}$ or $\mathbb{R}^+$ denotes the interval $]0, +\infty[$, $\mathbb{R}_{\geqq 0}$ or $\mathbb{R}_+$ denotes the interval $[0, +\infty[$,

c) The set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of all natural numbers excludes the number zero. We write $\mathbb{N}_0$ or $\mathbb{Z}_+$ or $\mathbb{Z}_{\geqq 0}$ for $\mathbb{N} \biguplus \{0\}$. $\mathbb{Z}_{\geqq 0}$ is the B/G notation. It is very unusual but also very intuitive.

Definition 2.1. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. We call that sequence **non-decreasing** or **increasing** if $x_n \leqq x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **strictly increasing** if $x_n < x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **non-increasing** or **decreasing** if $x_n \geqq x_{n+1}$ for all n.

We call it **strictly decreasing** if $x_n > x_{n+1}$ for all $n \in \mathbb{N}$.

3 Some Basics (Understand this!)

3.1 Numbers

Remark 3.1 (Classification of numbers)**.** [1]

We call numbers without decimal points such as $3, -29, 0, 3000000, 3 \cdot 10^6, -1, \dots$ **integers** and we write $\mathbb{Z}$ for the set [2] of all integers.

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

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

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

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

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

$$0.\bar{3} \;=\; .\bar{3} \;=\; .\bar{3} \;=\; 0.33333333333\dots \;=\; \frac{1}{3} \;=\; \frac{2}{6} \tag{3.1}$$

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

$$-4 \;=\; -4.000 \;=\; -3.\bar{9} \;=\; -\frac{12}{3} \;=\; -\frac{400}{100} \tag{3.2}$$

We call the barred portion of the decimal digits the **period** of the number and we also talk about **periodic decimals.**

You may have heard that there are numbers which cannot be expressed as integers or fractions or numbers with a finite amount of decimals to the right of the decimal point. Examples for that are $\sqrt{2}$ *and* $\boldsymbol{\pi}$. Those "**irrational numbers**" (really, that what we call them) fill the gaps between the rational numbers. In fact, there is a simple way (but not easy to prove) of characterizing irrational numbers: Rational numbers are those that can be expressed with at most finitely many digits to the right of the decimal point, including periodic decimals such as $1.6\bar{6}$. 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.

Now we can finally give an informal definition of the most important kind of numbers: We call any kind of number, either rational or irrational, a **real number** and we write $\mathbb{R}$ for the set of all

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

[2] You will learn more about sets in the section "3.2" on p.9. All you need to know here is that a set is a collection of stuff called members or elements. 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. Example: $A = \{1, 2.6, \text{the moon}, \text{London}\}$ is the set whose elements are the numbers 1 and 2.6, the moon and the city of London. $B = \{1, 2.6, \text{the moon}, 2.6, \text{London}\}$ is equal to the set A: The second occurrence of 2.6 is simply ignored.

real numbers. It can be shown that there are a lot more irrational numbers than rational numbers, even though $\mathbb{Q}$ is a **dense subset** in $\mathbb{R}$ in the following sense: No matter how small an interval $(a,b) = \{x \in \mathbb{R} : a < x < b\}$ of real numbers you choose, it will contain infinitely many rational numbers.

Definition 3.1 (Types of numbers)**.** We summarize what was said sofar about the classification of numbers:

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

$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \dots\}$ denotes the set of all **integers**.

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

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

$\mathbb{R} \setminus \mathbb{Q}(see^3)$ = {all real numbers which cannot be written as fractions of integers} denotes the set of all **irrational numbers**. There is no special symbol for irrational numbers. Example: $\sqrt{2}$ and π are irrational.

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

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

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

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

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

Assumption 3.1 (Square roots are always assumed non–negative)**.** Remember that for any number a it is true that

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

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

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

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

Proposition 3.1 (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:*

$$(3.3) \qquad Triangle\ Inequality: \boxed{|\boldsymbol{a}+\boldsymbol{b}| \leqq |\boldsymbol{a}| + |\boldsymbol{b}|}$$

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

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

[3] The set difference $X \setminus Y$ (see 3.12 on p.14.) is the set of all elements which belong to X but not to Y.

Proposition 3.2 (The Triangle Inequality for n real numbers)**.** *The above inequality also holds true for more than two real numbers: Let $n \in \mathbb{N}$ such that $n \geqq 2$. Let $a_1, a_2, \dots, a_n \in \mathbb{N}$. Then*

$$|a_1 + a_2 + \ldots + a_n| \leqq |a_1| + |a_2| + \ldots + |a_n| \tag{3.4}$$

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

Definition 3.2 (Principle of proof by complete induction)**.** Actually, "definition" is a misnomer. This principle is a mathematical statement that follows from the structure of the natural numbers which have a starting point to the "left" (a smallest element 1) and then progress in the well understood sequence [4]

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

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

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

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

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

Either you have been explained this principle before and say "Oh, that – what's the big deal?" or you will be mighty confused. So let me explain how it works by walking you through the proof of the triangle inequality for n real numbers (3.4).

Proof of the triangle inequality for n real numbers:

A. For $k_0 = 2$, inequality 3.4 was already shown (see 3.3), so we found a k_0 for which the property is true.

B. Let us assume that 3.4 is true for some $k \geqq 2$. We now must prove the inequality for $k+1$ numbers $a_1, a_2, \dots, a_k, a_{k+1} \in \mathbb{N}$: We abbreviate

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

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

$$|A + a_{k+1}| \leqq |A| + |a_{k+1}| \leqq B + |a_{k+1}| \tag{3.5}$$

In other words,

$$\left|\big(a_1 + a_2 + \ldots + a_k\big) + a_{k+1}\right|\ \leqq\ B + |a_{k+1}|\ =\ \big(|a_1| + |a_2| + \ldots + |a_k|\big) + |a_{k+1}| \tag{3.6}$$

and this is (3.4) for $k+1$ rather than k numbers: We have shown the validity of the triangle inequality for $k+1$ items under the assumption that it is valid for k items. It follows from the induction principle that the inequality is valid for any $k \geqq k_0 = 2$. ■

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

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

Step 1: We know that it's true for $k_0 = 2$ because that was actually proved in A.

Step 2: But according to B, if it's true for k_0, it's also true for the successor $k_0 + 1 = 3$.

Step 3: But according to B, if it's true for $k_0 + 1$, it's also true for the successor $(k_0 + 1) + 1 = 4$.

Step 4: But according to B, if it's true for $k_0 + 2$, it's also true for the successor $(k_0 + 2) + 1 = 5$.

...

Step $53,920$: But according to B, if it's true for k_0+$53,918$, it's also true for the successor $k_0 + 1 = 53,919$.

...

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

3.2 First things about sets, Functions (Mappings) and Families

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, amongst them sets and functions. Once you know about those, you can tackle sets of functions, set functions, sets of set functions, ...

3.2.1 Definition of sets

Definition 3.3 (Sets)**.** You probably know what a set is: 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**.

Example 3.1 (Oscillating sequence)**.** So, 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?

There will be a lot more to be said about sets but it is helpful to have an understanding of functions, also called mappings, before we proceed.

3.2.2 Definition of functions, injectivity, surjectivity and bijectivity

Look at the set $\mathbb{R}$ of all real numbers and the function $y = f(x) = x^2 + 1$ which associates with every real number x (the "argument" or "independent variable") another real number $y = x^2 + 1$ (the "function value" or "dependent variable"):

$$f(0) = 1,\ f(2) = 5,\ f(-2) = 5,\ f(-10) = 101,\ f(1/2) = 1.25,\ f(-2/3) = 4/9 + 1 = 13/9,\ \dots$$

You can think of this function as a rule or law which specifies what real number y will be the output or result of providing the real number x as input. [5]

I am quite sure that you did not have any difficulty following the above because you have already been taught about functions. But let us look a little bit closer at the function $y = f(x) = x^2 + 1$ and its properties:

(a): *There is a function value $f(x)$ for <u>every</u> $x \in \mathbb{R}$.*

(b): *Not every $x \in \mathbb{R}$ is suitable as a function value: A square cannot be negative, hence $x^2 + 1$ will never be less than 1.*

(c): *There is <u>exactly one</u> function value $f(x)$ for every $x \in \mathbb{R}$. Not zero, not two, not 21 y-values belong to a given x but exactly one: $f(2) = 5$ and $f(2)$ is nothing else but 5.*

(d): *On the other hand, given $y \in \mathbb{R}$, there may be zero x-values (e.g., $y = 1/2$), exactly one x-value (if $y = 1$) or two x-values (e.g. $y = 5$ which is obtained as both $f(2)$ and $f(-2)$.*

Here is a complicated way of looking at the example above: Let $X = \mathbb{R}$ and $Y = \mathbb{R}$. Then $y = f(x) = x^2 + 1$ is a rule which "maps" each element $x \in X$ to a <u>uniquely determined</u> number $y \in Y$ which depends on y (in a very simple way: it's 1 plus the square of x).

Mathematicians are very lazy as far as writing is concerned and they figured out long ago that writing "depends on xyz" all the time not only takes too long, but also is aesthetically very unpleasing and makes statements and their proofs hard to understand. So they decided to write "(xyz)" instead of "depends on xyz" and the modern notion of a function or mapping $y = f(x)$ was born.

Here is another example: if you say $f(x) = x^2 - \sqrt{2}$, it's just a short for "I have a rule which maps a number x to a value $f(x)$ which depends on x in the following way: compute $x^2 - \sqrt{2}$." It is crucial to understand from which set X you are allowed to pick the "arguments" x and it is often helpful to state what kinds of objects $f(x)$ the x–arguments are associated with, i.e., what set Y they will belong to.

Put all this together and you see the motivation for the following definition.

Definition 3.4 (Mappings (functions))**.** Given are the two arbitrary sets X and Y each of which has at least one element. We assign to each $a \in X$ exactly one element $y = f(a) \in Y$. Such an association $f(\cdot)$ is called a **function** or **mapping** from X into Y. The set X is called the **domain** or **preimage** and Y is called the **target** or **image set** or **codomain** of the mapping $f(\cdot)$. Domain elements $x \in X$ are called or **independent variables** or **argument** and $f(x) \in Y$ is called the **function value** of x. The subset

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

[5] If you do not know about the different kinds of numbers, review the section "Numbers" on p.6. To get by, it is enough that you know that we call positive integers $\{1, 2, 3, \dots\}$ "natural numbers" and we call any kind of number, including fractions and decimals, "real numbers". We write $\mathbb{N}$ for the set of all natural numbers and $\mathbb{R}$ for the set of all real numbers.

of Y is called the **range** or **image** of the function $f(\cdot)$. [6]

Usually mathematicians simply write f for the function $f(\cdot)$ We shall sometimes follow that convention but ofte include the "$(\cdot)$" part if it helps you to see more easily in a formula that a function rather than a simple element is involved. If the names of the sets involved need to be stressed, mathematicians draw diagrams such as

$$f: \ X \longrightarrow Y \qquad x \longmapsto f(x)$$

.

They say "f maps X into Y" and "f maps the domain value x to the function value $f(x)$".

Remark 3.3 (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 will call a mapping a function only if its target is a subset of the real numbers [7] but the majority does what I'll try to adhere to in this document: I use "mapping" and "function" interchangeably and I'll talk about **real functions** rather than just functions if the codomain is part of $\mathbb{R}$ (see (5.1) on p.33).

Definition 3.5 (identity mapping)**.** Given any non–empty set X, we shall use the symbol id for the **identity** mapping

$$id(\cdot): \ X \longrightarrow X \qquad x \longmapsto x$$

which assigns each element of the domain to itself. If it is necessary to show the name of the set X to avoid confusion, the notation id_X is used.

Definition 3.6 (Surjective, injective, bijective)**. a. Surjectivity:** In general it is not true that $f(X) = Y$. But if it is, we call $f(\cdot)$ **surjective** and we say that f maps X **onto** Y.

b. Injectivity: For each argument $a \in X$ there must be exactly one function value $f(a) \in f(X)$. But it is OK if more than one argument is mapped into one and the same $y \in f(X)$. $f(\cdot)$ is called **injective** if different arguments $x_1 \neq x_2 \in X$ will always be mapped into different values $f(x_1) \neq f(x_2)$.

c. Bijectivity: Assume now that the mapping $f(\cdot)$ from X into Y is both injective and surjective. In that case it is called **bijective**. In other words, a bijective mapping has the following property: For each $y \in Y$ there exists at least one $x \in X$ such that $y = f(x)$ (because f is surjective) but no more than one such x (because f is injective). In other words, not only does each x in the domain uniquely determine its corresponding function value $y = f(x)$, but the reverse also is true: Each y in the codomain uniquely determines an x in the domain that is mapped by f to y.

We write $g(y) = x$ for the mapping that assigns to any $y \in Y$ this unique element $x \in X$ whose image $f(x)$ is y. This assignment $y \mapsto g(y)$ defines indeed a mapping from Y into X.

It is not hard to see that $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$. We call $g(\cdot)$ the **inverse mapping** or **inverse function** of $f(\cdot)$ and write $f^{-1}(\cdot)$.

Many more properties of mappings will be discussed later. Now we shall look at families, sequences and more some additional properties of sets.

[6] We distinguish the image set (codomain) Y of $f(\cdot)$ from its image (range) $f(X)$.

[7] (or if it is a subset of the complex numbers, but we won't discuss complex numbers in this document)

3.2.3 Sequences, families and functions as families

We can turn any set into a "family" by tagging each of its members with an "index". As an example, look at this tagged version of S_2 from example 3.1 on p. 9:

$$F = (a_1, e_1, e_2, i_1, i_2, i_3, o_1, o_2, o_3, o_4, u_A, u_B, u_C, u_D, u_E)$$

I chose on purpose not to tag the five "u-vowels" with numbers $1, 2, 3, 4, 5$ but rather with letters "A, B, C, D, E" just to drive home the point that the nature of the index does not matter. Only the ability to distinguish any two members of the collection by their index does.

Definition 3.7 (Indexed families and sequences)**.** An indexed collection is called an **indexed family** or simply a **family**. In all cases of interest to us such a collection is indexed through the elements of a set which we call the **index set** of the family. If the name of the index set is J, then we can use the notation

$$(x_i)_{i \in J}.$$

A **sequence** (x_j) is nothing but a family of things x_j which are indexed by integers. Usually those integers are the natural numbers $\mathbb{N} = \{1, 2, 3, 4, \dots\}$ or the non-negative integers $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$

Sequences are easier understood than families and you probably have been taught about them already. Here are two examples of sequences:

Example 3.2 (Oscillating sequence)**.** $x_j := (-1)^j$ $(j \in \mathbb{N}_0)$
Try to understand why this is the sequence

$$x_0 = 1, \quad x_2 = -1, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 1, \quad x_5 = -1, \ \dots$$

Example 3.3 (Series (summation sequence))**.** $s_k := 1 + 2 + \ldots + k$ $(k = 1, 2, 3, \dots)$

$$\begin{aligned} s_1 =& 1, \quad s_2 = 1 + 1/2 = 2 - 1/2, \quad s_3 = 1 + 1/2 + 1/4 = 2 - 1/4, \quad \dots, \\ s_k =& 1 + 1/2 + \ldots + 2^{k-1} = 2 - 2^{k-1}; \quad s = 1 + 1/2 + 1/4 + 1/8 +, \dots \quad \text{"infinite sum".} \end{aligned}$$

You obtain s_{k+1} from $s_k = 2 - 2^{k-1}$ by cutting the difference 2^{k-1} to the number 2 in half (that would be 2^k) and adding that to s_k. It is intuitively obvious that the infinite sum s adds up to 2. Such an infinite sum is called a **series**. The precise definition of a series will be given later.

Note 3.1. This is something you should remember: the name of the index variable 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 $(x_\beta)_{\beta \in J}$.

Note 3.2. There is a subtle difference between sequences and families.

a. Sequences:

Let Y be a set that contains all indexed items y_j of a sequence $(y_j)_{j \in \mathbb{N}_0}$. We can always create such a Y by defining $Y := \{y_j : j \geq 0\}$.

We can transport the natural left-to-right ordering

$0, 1, 2, 3 \ldots$

to the indexed family

$y_0, y_1, y_2, y_3 \ldots$

which allows us to reconstruct the assignment

$0 \mapsto y_0,\ 1 \mapsto y_1,\ 2 \mapsto y_2,\ 3 \mapsto y_3\ \ldots.$

In other words, the sequence $y_0, y_1, y_2, y_3 \ldots$ contains just as much information as the more complicated sequence of elements of $\mathbb{N}_0 \times Y$, $(0, y_0), (1, y_1), (2, y_2), (3, y_3) \ldots.$ You should be able to see that this last collection describes a function $f : \mathbb{N} \longrightarrow Y$ which maps its domain elements j as follows: $f(j) := y_j$.

b. Families:

Contrast the above with a family $(y_x)_{x \in X}$. In other words, we have a bunch of "y-items" which are indexed by an index set X which we assume, as usual, to be not empty. There may not be a natural order on X which would allow to rank any two items $x, \tilde{x} \in X$ as x first, $\tilde{x}$ second, or vice versa. Such would be, for example, the case for two-dimensional space $X = \mathbb{R}^2$ (which is bigger: $(3, 5) or (5, 3)$)? We can no longer infer which x_0 was the index for a given y_x, say, $y_x = 129$. To do so we need to pair up the index values with the y items they are indexing:

If we replace the original family $(y_x)_{x \in X}$ with the new one. $\big((x, y_x)\big)_{x \in X}$ then this new family completely and uniquely describes the function $f : X \longrightarrow Y$ which maps its domain elements x as follows: $f(x) := y_x$.

We express this yet another way: any function $f : X \longrightarrow Y$ can be written equivalently as the family $\big((x, f(x))\big)_{x \in X}$. This expression in turn is equivalent to the set $\Gamma_f := \Gamma(f) := \{(x, f(x) : x \in X\}$

We note that there is no issue with the fact that families may contain duplicates whereas sets may not: Even if two items $x, \tilde{x} \in X$ may to the same $y \in Y$, the two pairs $(x, f(x))$ and $(\tilde{x}, f(\tilde{x})$ are considered different because their left sides do not match.

We have laid the groundwork for the following definition.

Definition 3.8 (Mappings as graphs)**.** Given are the two arbitrary sets X and Y each of which has at least one element and a function $f : X \longrightarrow Y$. Then $\Gamma_f := \Gamma(f) := \{(x, f(x)) : x \in X\}$ is called the **graph** of the function f.

Proposition 3.3. *The following three definitions of a function $f : X \longrightarrow Y$ are equivalent:*

a. *assigning to each $a \in X$ exactly one element $y = f(a) \in Y$ (see def. 3.4), p. 10*
b. *f is defined by the family $\big((x, f(x))\big)_{x \in X}$*
c. *f is defined by its graph $\Gamma_f := \Gamma(f) := \{(x, f(x)) : x \in X\}$*

Proof: contained in note 3.1 above. ■

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.

3.3 Basic set operations and Cartesian products

Definition 3.9 (empty set)**.** $\emptyset$ or {} denotes the **empty set**. It is the one set that does not contain any elements.

Definition 3.10 (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 $A \neq B$, i.e., there is at least one $x \in B$ such that $x \notin A$, we can emphasize that by saying that A is a **strict subset** of B. We write "$A \subsetneq B$" or "$A \subset B$". Alternatively we say that B is a **strict superset** of A and we write "$B \supsetneq A$") or "$B \supset A$".

Definition 3.11 (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 $A \cap B = \emptyset$. In this case we can also write $A \uplus B$ (pronounced "A disjoint union B") for the union $A \cup B$ of disjoint sets. We call a family of sets $(A_i)_i$ **mutually disjoint** if any two different sets A_i, A_j have intersection $A_i \cap A_j = \emptyset$. In this case we often write $A \uplus B$ rather than $A \cup B$ for the union of A and B.

Definition 3.12 (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:

$$A \setminus B := \{x \in A : x \notin B\} \tag{3.7}$$

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

$$A \triangle B := (A \cup B) \ \setminus \ (A \cap B) \tag{3.8}$$

Draw some Venn diagrams in which the sets are represented as circles to understand why the following is true for any sets A, B, X where we assume that $A \subseteq X$.

(3.9a) $$A \triangle B = (A \setminus B) \uplus (B \setminus A)$$
(3.9b) $$A \setminus A = \emptyset$$
(3.9c) $$A \triangle A = \emptyset$$
(3.9d) $$X \triangle A = X \setminus A$$
(3.9e) $$A \cup B = (A \triangle B) \cup (A \cap B)$$

After this digression about $A \setminus B$ and $A \triangle B$ we now continue with the set-theoretic notations which are relevant for this article.

Definition 3.13 (Universal set)**.** Usually there always is a big set Ω that contains everything we are interested in and we then deal with all kinds of subsets $A \subseteq \Omega$. Such a set is called a **"universal" set**.

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

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

Definition 3.14 (Complement of a set)**.** The **complement** of a set A consists of all elements of Ω which do not belong to A. We write $\complement A$ or $A^{\complement}$. In other words:

(3.10) $$A^{\complement} := \complement A := \Omega \setminus A = \{\omega \in \Omega : x \notin A\}$$

Remark 3.4 (Complement of empty, all)**.** Note that for any kind of universal set Ω it is true that

(3.11) $$\Omega^{\complement} = \emptyset, \qquad \emptyset^{\complement} = \Omega$$

Example 3.4 (Complement of a set relative to the unit interval)**.** Assume we are exclusively dealing with the unit interval, i.e., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. Let $a \in [0,1]$ and $\delta > 0$ and

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

the δ–neighborhood [8] of a (with respect to $[0,1]$ because numbers outside the unit interval are not considered part of our universe). Then the complement of $B_\delta(a)$ is

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

Theorem 3.1 (De Morgan's Law)**.** *Let there be a universal set Ω (see (3.13) on p.15). Then the following "duality principle" holds for any indexed family $(A_\alpha)_{\alpha \in I}$ of sets:*

(3.13)
$$\begin{aligned} &\textbf{\textit{a)}} \quad \complement\Big(\bigcup_\alpha A_\alpha\Big) = \bigcap_\alpha (\complement A_\alpha) \\ &\textbf{\textit{b)}} \quad \complement\Big(\bigcap_\alpha A_\alpha\Big) = \bigcup_\alpha (\complement A_\alpha) \end{aligned}$$

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.

[8] Neighborhoods of a point will be discussed in the chapter on the topology of $\mathbb{R}^n$ (see (7.6) on p.66) In short, the δ–neighborhood of a is the set of all points with distance less than δ from a.

Generally speaking this leads to the duality principle that 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 by their complements, all unions by intersections and all intersections by unions.

Proof of De Morgan's law, formula a:

First we prove that $\complement(\bigcup_\alpha A_\alpha) \subseteq \bigcap_\alpha(\complement A_\alpha)$*: Assume that* $x \in \complement(\bigcup_\alpha A_\alpha)$*. Then* $x \notin (\bigcup_\alpha A_\alpha)$ *which is the same as saying that* x *does not belong to any of the* A_α*. That means that* x *belongs to each* $\complement A_\alpha$ *and hence also to the intersection* $\bigcap_\alpha(\complement A_\alpha)$*.*

Now we prove that the right hand side set of formula a contains the left hand side set. So let $x \in \bigcap(\complement A_\alpha)$*. Then* x *belongs to each of the* $\complement A_\alpha$ *and hence to none of the* A_α*. Then it also does not belong to the union of all the* A_α *and must therefore belong to the complement* $\complement(\bigcup_\alpha A_\alpha)$*. This completes the proof of formula a. The proof of formula b is not given here because the mechanics are the same.* ■

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.*

Definition 3.15 (Cartesian Product of two sets)**.** The **cartesian product** of two sets A and B is

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

i.e., it consists of all pairs (a,b) with $a \in A$ and $b \in B$.

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

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

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

Example 3.5 (Coordinates in the plane)**.** Here is the most important example of a cartesian product of two sets. Let $A = B = \mathbb{R}$. Then $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \{(x,y) : x,y \in \mathbb{R}\}$ is the set of pairs of real numbers. I am sure you are familiar with what those are: They are just points in the plane, expressed by their x– and y–coordinates.
Examples 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) $(1.234, -\sqrt{2}) \in \mathbb{R}^2$
Now you should understand why we do not allow two pairs to be equal if we flip the coordinates: Of course $(1,0)$ and $(0,1)$ are different points in the xy–plane!

Remark 3.5 (Empty cartesian products)**.** Note that $A \times B = \emptyset$ if and only if $A = \emptyset$ or $B = \emptyset$ or both are empty.

Remark 3.6 (Associativity of cartesian products)**.** Assume we have three sets A, B and C. We can then look at

$$\begin{aligned}(A \times B) \times C \ &= \ \{((a,b),c) : a \in A, b \in B, c \in C\}\\ A \times (B \times C) \ &= \ \{(a,(b,c)) : a \in A, b \in B, c \in C\}\end{aligned}$$

In either case, we are dealing with a triplet of items a, b, c in exactly that order. This means that it does not matter whether we look at $((a,b),c) \in (A \times B) \times C$ or $(a,(b,c)) \in A \times (B \times C)$. and we can simply write

$$A \times B \times C \;:=\; (A \times B) \times C \;=\; A \times (B \times C) \qquad \textbf{associativity} \tag{3.14}$$

Now we know that the next definition makes sense:

Definition 3.16 (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 $X_1, X_2, X_3, \dots, X_N$, we define the **cartesian product** as [9]

$$X_1 \times X_2 \times X_3 \times \ldots \times X_N \;:=\; \{(x_1, x_2, \dots, x_N) : x_j \in X_j \text{ for all } 1 \leqq j \leqq N\}$$

Two elements $(x_1, x_2, \dots, x_N)$ and $(y_1, y_2, \dots, y_N)$ of $X_1 \times X_2 \times X_3 \times \ldots \times X_N$ are called **equal** if and only if $x_j = y_j$ for all j such that $1 \leqq j \leqq N$. In this case we write $(x_1, x_2, \dots, x_N) = (y_1, y_2, \dots, y_N)$.

As a shorthand, we abbreviate $X^N := \underbrace{X \times X \times + \cdots \times X}_{N\text{times}}$.

Example 3.6 (N–dimensional coordinates)**.** Here is the most important example of a cartesian product of N sets. Let $X_1 = X_2 = \ldots = X_N = \mathbb{R}$. Then $\mathbb{R}^N \;=\; \{(x_1, x_2, \dots, x_N) : x_j \in \mathbb{R}\}$ for $1 \leqq j \leqq N$ is the set of points in N–dimensional space. You may not be familiar with what those are unless $N = 2$ (see example 3.5 above) or $N = 3$.
In the 3–dimensional case it is customary to write (x, y, z) rather than (x_1, x_2, x_3) . Each such triplet of real numbers represents a point in (ordinary 3–dimensional) space and we speak of its x–coordinate, y–coordinate and z–coordinate.
For the sake of completeness: If $N = 1$ the item $(x) \in \mathbb{R}^1$ (where $x \in \mathbb{R}$; observe the parentheses around x) is considered the same as the real number x. In other words, we "identify" $\mathbb{R}^1$ with $\mathbb{R}$. Such a "one–dimensional point" is simply a point on the x–axis.

A short word on vectors and coordinates: For $N \leqq 3$ you can visualize the following: Given a point x on the x–axis or in the plane or in 3–dimensional space, there is a unique arrow that starts at the point whose coordinates are all zero (the "origin") and ends at the location marked by the point x. Such an arrow is customarily called a vector.

Because it makes sense in dimensions $1, 2, 3$, an N**–tuple** $(x_1, x_2, \dots, x_N)$ is also called a vector of dimension N. You will read more about this in the chapter 6, p.43, on vectors and vector spaces.

This is worth while repeating: We can uniquely identify each $x \in \mathbb{R}^N$ with the corresponding vector: an arrow that starts in $(\underbrace{0, 0, \dots, 0}_{N\text{times}})$ and ends in x.

[9] 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 $N = 3$:

$$X_1 \times (X_2 \times X_3 \times X_4), \quad (X_1 \times X_2) \times (X_3 \times X_4), \quad X_1 \times (X_2 \times X_3 \times X_4),$$

An exact proof that we can group the sets with parentheses any way we like is very tedious and will not be given here.

Now that we have discussed the cartesian product of finitely many sets, we'll deal with cartesian products of an entire family of sets $(X_i)_{i \in I}$.

Definition 3.17 (Cartesian Product of a family of sets)**.** Let I be an arbitrary, non–empty set (the index set) and let $(X_i)_{i \in I}$ be a family of non–empty sets X_i. The **cartesian product** of the family $(X_i)_{i \in I}$ is the set

$$\prod_{i \in I} X_i \ := \ (\prod X_i)_{i \in I} \ := \ \{(x_i)_{i \in I} : x_i \in X_i \ \forall i \in I\}$$

of all familes $(x_i)_{i \in I}$ each of whose members x_j belongs to the corresponding set X_j. The "$\prod$" is the greek "upper case" letter "Pi" (whose lower case incarnation "π" you are probably more familiar with). As far as I know, it was chosen because it has the same starting "p" sound as the word "product" (as in cartesian product).

Two elements $(x_i)_{i \in I}$ and $(y_k)_{k \in I}$ of $\prod\limits_{i \in I} X_i$ are called **equal** if and only if $x_i = y_i$ for all $i \in I$. In this case we write $(x_i)_{i \in I}$ = $(y_k)_{k \in I}$.

As a shorthand, if all sets X_i are equal to one and the same set X, we abbreviate $X^I := \prod\limits_{i \in I} X$.

It turns out that the very last remark in the preceding definitions fits in very nicely with the next chapter on mappings because the elements $(y_x)_{x \in X}$ *of the cartesian product* Y^X *are nothing but mappings* [10]

$y(\cdot) : X \to Y$. *But before we get there, we take a quick look at countably infinite sets.*

3.4 Countable sets

This brief chapter is not very precise in that we do not talk about an axiomatic approach to finite sets and countably infinite sets. You can find that in ch.13 of [1] (Beck/Geoghegan).

Here are the definitions but they won't be needed in this document.

Everyone understands what a finite set is: It's a set with a finite number of elements:

Definition 3.18 (Finite sets)**.** Let $n \in \mathbb{N}$. we say that a set X has **cardinality** n and we write $\text{card}(X) := |X| := n$ if there is a bijective mapping between X and the set $[n] := \{1, 2, \dots, n\}$ We call such sets **finite** .

In other words, a set X of cardinality n is one whose elements can be enumerated as $x_1, x_2, \dots, x_n$: The cardinality of a finite set is simply the number of elements it contains.

We define the empty set $\emptyset$ to be finite and set $card(\emptyset) := 0$.

You may be surprised to hear this but there are ways to classify the degree of infinity when looking at infinite sets.

[10] Mappings or functions were briefly discussed already in paragraph **??** on p.**??**. Families being functions in disguise explains why, contrary to sets, an item can be listed more often than once (in fact, infinitely often): you keep track of the index i of an item x_i.

The "smallest degree of infinity" is found in sets that can be compared, in a sense, to the set $\mathbb{N}$ of all natural numbers. Look back to definition (3.2) on the principle of complete induction. It is based on the property of $\mathbb{N}$ that there is a starting point a_1 = 1 and from there you can progress in a sequence

$$a_2 = 2;\ a_3 = 3;\ a_4 = 4;\ \dots\ a_k = k;\ a_{k+1} = k+1;\ \dots$$

in which no two elements a_j, a_k are the same for different j and k. We have a special name for inifinite sets whose elements can be arranged into a sequence of that nature.

Definition 3.19 (Countable and countably infinite sets)**.** Let X an arbitrary set such that there is a bijection $f : \mathbb{N} \longrightarrow X$. This means that all of the elements of X can be arranged in a sequence

$$X = \{x_1 = f(1), x_2 = f(2), x_3 = f(3), \dots\}.$$

which is infinite, i.e., we rule out the case of sets with finitely many members. X is called a **countably infinite** set. We call a set that is either finite or countably infinite a **countable set**. and we also say that X is countable.

A set that is neither finite nor countably infinite is called **uncountable** or **not countable**

The proofs given in the remainder of this brief chapter on cardinality are not precise as we do not try to establish, for example in the first proof below, that for any subset B of a countable set there either exists an $n \in \mathbb{N}$ and a bijection from B to $[n]$ or there exists a bijection between B and $\mathbb{N}$. You may be surprised to hear that even the fact that there is no bijection between $[m] = \{1, 2, \dots m\}$ and $[n] = \{1, 2, \dots n\}$ for $m \neq n$ needs a proof that is not entirely trivial.

Theorem 3.2 (Subsets of countable sets are countable)**.** *Any subset of a countable set is countable.*

Proof: It is obvious that any subset of a finite set is finite. So we only need to deal with the case where we take a subset B of a countably infinite set A. Because A is countably infinite, we can arrange its elements into a sequence

$$A = \{a_1,\ a_2, a_3,\ \dots\}$$

where $j_1 = min\{j \geqq 1 : a_{j_1} \in B$ We walk along that sequence and set

$$\begin{aligned} b_1 &:= a_{j_1} \quad \textit{where } j_1 = \min\{j \geqq 1 : a_{j_1} \in B\}, \\ b_2 &:= a_{j_2} \quad \textit{where } j_2 = \min\{j > j_1 : a_{j_2} \in B\}, \\ b_3 &:= a_{j_3} \quad \textit{where } j_3 = \min\{j > j_2 : a_{j_3} \in B\}, \dots \\ &\dots \end{aligned}$$

i.e., b_j is element number j of the subset B. The sequence (b_j) contains exactly all elements of B which means that this set is either finite (in case there is an $n_0 \in \mathbb{N}$ such that b_{n_0} is the last element of that sequence) or it is countably infinite in case that there are infinitely many b_j. ■

The following proposition is proved again more exactly in a later chapter (see thm.4.1 on p.31)

Theorem 3.3 (Countable unions of countable set)**.** *The union of countably many countable sets is countable.*

Proof: In the finite case let the sets be

$$A_1,\ A_2,\ A_3,\ \dots,\ A_N.$$

In the countable case let the sets be

$$A_1,\ A_2,\ A_3,\ \dots,\ A_n,\ A_{n+1},\ \dots$$

In either case we can assume that the sets are mutually disjoint, i.e., any two different sets A_i, A_j *have intersection* $A_i \cap A_j = \emptyset$ *(see definition (3.11) on p.14). This is just another way of saying that no two sets have any elements in common. The reason we may assume mutual disjointness is that if we substitute*

$$B_1 := A_1; \quad B_2 := A_2 \setminus B_1; \quad B_3 := A_3 \setminus B_2; \quad \dots$$

then

$$\bigcup_{j \in \mathbb{N}} A_j \;=\; \bigcup_{j \in \mathbb{N}} B_j$$

(why?) and the B_j *are mutually disjoint. So let us assume the* A_j *are mutually disjoint. We write the elements of each set* A_j *as* $a_{j1}, a_{j2}, a_{j3}, \dots$*.*

A. *Let us first assume that none of those sets is finite. We start the elements of each* A_j *in a separate row and obtain*

$$\begin{tikzcd}
a_{11} \arrow[r] & a_{12} \arrow[dl] & a_{13} \arrow[r] & a_{14} \arrow[dl] & \cdots \\
a_{21} \arrow[d] & a_{22} \arrow[ur] & a_{23} \arrow[dl] & a_{24} & \cdots \\
a_{31} \arrow[ur] & a_{32} \arrow[dl] & a_{33} & a_{34} & \cdots \\
a_{41} & a_{42} & a_{43} & a_{44} & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{tikzcd}$$

Now we create a new sequence b_n *by following the arrows from the start at* a_{11}*. We obtain*

$$b_1 = a_{11};\ b_2 = a_{12};\ b_3 = a_{21};\ b_4 = a_{31};\ \dots$$

You can see that this sequence manages to collect all elements a_{ij} *in that infinite two–dimensional grid and it follows that the union of the sets* A_j *is countable.*

B. *How do we modify this proof if some or all of the* A_i *are finite? We proceed as follows: If the predecessor* A_{i-1} *is finite with* N_{i-1} *elements, we stick the elements* a_{ij} *to the right of the last element* $a_{i-1,N_{i-1}}$*. Otherwise they start their own row. If* A_i *itself is finite with* N_i *elements, we stick the elements* $a_{i+1,j}$ *to the right of the last element* a_{i,N_i}*. Otherwise they start their own row …*

B.1. *If an infinite number of sets has an infinite number of elements, then we have again a grid that is infinite in both horizontal and vertical directions and you create the "diagonal sequence"* b_j *just as before: Start off with the top-left element.Go one step to the right. Down–left until you hit the first column. Then down one step. Then up–right until you hit the first row. Then one step to the right. Down–left until you hit the first* *column. Then down one step. Then up–right until you hit the first row. Then one step to the right. Down–left until you hit the first column. Then down one step. Then up–right until ... I'm sure you get the picture.*

B.2. *Otherwise, if only a finite number of sets has an infinite number of elements, then we have a grid that is infinite in only the horizontal direction. You create the "diagonal sequence" b_j almost as before. The exception: if you hit the bottom row, then must go one to the right rather than one down. Afterward you march again up–right until you hit the first column ...*
■

Corollary 3.1 (The rational numbers are countable)**.**

Proof: Assume we can show that the set $\mathbb{Q} \cap [0, 1[= \{q \in \mathbb{Q} : 0 \leqq q < 1$ *is countable. Then the set* $\mathbb{Q} \cap [z, z + 1[= \{q \in \mathbb{Q} :$ *is countable for any integer* $z \in \mathbb{Z}$*. The reason: once we find a sequence* b_j *that runs through all elements of* $\mathbb{Q} \cap [0, 1[$*, then the sequence* $e_j := b_j + z$ *runs through all elements of* $\mathbb{Q} \cap [z, z + 1[$*. But* $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup \{-k : k \in \mathbb{N}$ *is countable as a union of only three countable sets. Abbreviate* $Q_z := \mathbb{Q} \cap [z, z + 1[$*. Can you see that* $\mathbb{Q} = \bigcup_{z \in \mathbb{Z}} Q_z$*? Good for you, because now that you know that* $\mathbb{Z}$ *is countable, you understand that* $\mathbb{Q}$ *can be written as a countable union of sets* Q_z *each of which is countable. So we are done with the proof ... except we still must prove that the set* Q_0 *of all rational numbers between zero and one is countable.*

We do that now. Let $A_1 := 0$*. Let*

$$
\begin{aligned}
A_2 :=& \{z \in Q_1 : z \textit{ has denominator } 2\} &=& \{\frac{0}{2}, \frac{1}{2}\} \\
A_3 :=& \{z \in Q_1 : z \textit{ has denominator } 3\} &=& \{\frac{0}{3}, \frac{1}{3}, \frac{2}{3}\} \\
A_4 :=& \{z \in Q_1 : z \textit{ has denominator } 4\} &=& \{\frac{0}{4}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}\} \\
& \dots \\
A_n :=& \{z \in Q_1 : z \textit{ has denominator } n\} &=& \{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\} \\
& \dots
\end{aligned}
$$

Then each set is finite and $Q_1 = \bigcup_{k \in \mathbb{N}} A_k$ *is a countable union of countably many finite sets and hence, according to the previous theorem (3.3), countable. We are finished with the proof.* ■

Theorem 3.4 (The real numbers are uncountable)**.** *The real numbers are uncountable: There is no sequence* $(r_n)_{n \in \mathbb{N}}$ *such that* $\{r_n : n \in \mathbb{N}\} = \mathbb{R}$*.*

Proof:

LATER ■

4 Sets and Functions, direct and indirect images (Study this!)

4.1 Basic Properties of Sets

The following trivial lemma (a lemma is a "proof subroutine" which is not remarkable on its own but very useful as a reference for other proofs) is useful if you need to prove statements of the form $A \subseteq B$ or $A = B$ for sets A and B. It is a means to simplity the proofs of [1] B/G (Beck/Geoghegan), project 5.12. You must reference this lemma as the "inclusion lemma" when you use it in your homework or exams. Be sure to understand what it means if you choose $J = \{1, 2\}$ (draw one or two Venn diagrams).

Lemma 4.1 (Inclusion lemma)**.** *Let J be an arbitrary, non-empty index set and let X_j, Y, Z_j, W $(j \in J)$ be sets such that $X_j \subseteq Y \subseteq Z_j \subseteq W$ for all $j \in J$. Then*

$$\bigcap_{j \in J} X_j \subseteq Y \subseteq \bigcup_{j \in J} Z_j \subseteq W. \tag{4.1}$$

Proof:
Let $x \in \bigcap_{j \in J} X_j$. Then $x \in X_j$ for all $j \in J$. But then $x \in Y$ for all $j \in J$ because $X_j \subseteq Y$ for all $j \in J$. But $x \in Y$ for all $j \in J$ implies that $x \in Y$ and the left side inclusion of the lemma is shown.

Now assume $x \in Y$. We note that $Y \subseteq Z_j$ for all $j \in J$ implies $x \in Z_j$ for all $j \in J$. But then certainly $x \in Z_j$ for at least one $j \in J$ (did you notice that we needed to assume $J \neq \emptyset$?) It follows that $x \in \bigcup_{j \in J} Z_j$ and the middle inclusion of the lemma is shown.

Finally, assume $x \in \bigcup_{j \in J} Z_j$ It follows from the definitions of unions that there exists at least one $j_0 \in J$ such that $x \in Z_{j_0}$. But then $x \in W$ as W contains Z_{j_0}. x is an arbitrary element of $\bigcup_{j \in J} Z_j$ and if follows that $\bigcup_{j \in J} Z_j \subseteq W$. This finishes the proof of the rightmost inclusion. ■

4.2 Direct images and indirect images (preimages) of a function

Here are the references for the material below. I took them from a Math 330 course which was held some time ago by Prof. Mazur. You should recognize them from your home page and syllabus:

[6] Author unknown: Introduction to Functions Ch.2. (mazur-330-func-1.pdf)
[7] Author unknown: Properties of Functions Ch.2. (mazur-330-func-2.pdf)
[8] Author unknown: Ch.1: Introduction to Sets and Functions (mazur-330-sets-1.pdf)
[9] Author unknown: Ch.4: Applications of Methods of Proof (mazur-330-sets-2.pdf)
[3] Pete L. Clark: Lecture notes on relations and functions (mazur-330-relat-func.pdf)

Definition 4.1. Let X, Y be two non-empty sets and $f : X \to Y$ be an arbitrary function with domain X and codomain Y. Let $A \subseteq X$ and $B \subseteq Y$. Let

$$1) \quad f(A) = \{f(x) : x \in A\} \tag{4.2}$$

$$2) \quad f^{-1}(B) = \{x \in X : f(x) \in B\} \tag{4.3}$$

We call $f(A)$ the **direct image** of A under f and we call We call $f^{-1}(B)$ the **indirect image** or **preimage** of B under f

Notational conveniences:

If we have a set that is written as $\{\dots\}$ then we may write $f\{\dots\}$ instead of $f(\{\dots\})$ and $f^{-1}\{\dots\}$ instead of $f^{-1}(\{\dots\})$. Specifically for $x \in X$ and $y \in Y$ we get $f^{-1}\{x\}$ and $f^{-1}\{y\}$. Many mathematicians will write $f^{-1}(y)$ instead of $f^{-1}\{y\}$ but this writer sees no advantages doing so whatsover. There seemingly are no savings with respect to time or space for writing that alternate form but we are confounding two entirely separate items: a subset $f^{-1}\{y\}$ of X v.s. the function value $f^{-1}(y)$ of $y \in Y$ which is an elementof X. We can talk about the latter only in case that the inverse function f^{-1} of f exists.

In measure theory and probability theory the following notation is also very common: $\{f \in B\}$ rather than $f^{-1}(B)$ and $\{f = y\}$ rather than $f^{-1}\{y\}$

Let $a < b \in \mathbb{R}$. We write $\{a \leqq f \leqq b\}$ rather than $f^{-1}([a,b])$, $\{a < f < b\}$ rather than $f^{-1}(]a,b[)$, $\{a \leqq f < b\}$ rather than $f^{-1}([a,b[)$ and $\{a < f \leqq b\}$ rather than $f^{-1}(]a,b])$, $\{f \leqq b\}$ rather than $f^{-1}(]-\infty,b])$, etc.

Proposition 4.1. *Some simple properties:*

$$f(\emptyset) = f^{-1}(\emptyset) = 0 \tag{4.4}$$

$$A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2) \tag{4.5}$$

$$B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2) \tag{4.6}$$

$$x \in X \Rightarrow f(\{x\}) = \{f(x)\} \tag{4.7}$$

$$f(X) = Y \iff f \text{ is surjective} \tag{4.8}$$

$$f^{-1}(Y) = X \quad \textit{always!} \tag{4.9}$$

Proof of all properties is immediate. ■

Proposition 4.2 (f^{-1} is compatible with all basic set ops)**.** *In the following we assume that J is an arbitrary index set, and that $B \subseteq Y$, $B_j \subseteq Y$ for all j.*
The following all are true:

$$f^{-1}\Big(\bigcap_{j \in J} B_j\Big) = \bigcap_{j \in J} f^{-1}(B_j) \tag{4.10}$$

$$f^{-1}\Big(\bigcup_{j \in J} B_j\Big) = \bigcup_{j \in J} f^{-1}(B_j) \tag{4.11}$$

$$f^{-1}(B^{\complement}) = f^{-1}(B)^{\complement} \tag{4.12}$$

$$f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2) \tag{4.13}$$

Proof of (4.10): Let $x \in X$. Then

$$
\begin{aligned}
x \in f^{-1}(\bigcap_{j\in J} B_j) &\iff f(x) \in \bigcap_{j\in J} B_j \quad (\textit{def preimage}) \\
&\iff \forall j\ f(x) \in B_j \quad (\textit{def} \cap) \\
&\iff \forall j\ x \in f^{-1}(B_j) \quad (\textit{def preimage}) \\
&\iff x \in \bigcap_{j\in J} f^{-1}(B_j) \quad (\textit{def} \cap)
\end{aligned}
\tag{4.14}
$$

Proof of (4.11): Let $x \in X$. Then

$$
\begin{aligned}
x \in f^{-1}(\bigcup_{j\in J} B_j) &\iff f(x) \in \bigcup_{j\in J} B_j \quad (\textit{def preimage}) \\
&\iff \exists j_0 :\ f(x) \in B_{j_0} \quad (\textit{def} \cup) \\
&\iff \exists j_0 : x \in f^{-1}(B_{j_0}) \quad (\textit{def preimage}) \\
&\iff x \in \bigcup_{j\in J} f^{-1}(B_j) \quad (\textit{def} \cup)
\end{aligned}
\tag{4.15}
$$

Proof of (4.12): Let $x \in X$. Then

$$
\begin{aligned}
x \in f^{-1}(B^{\complement}) &\iff f(x) \in B^{\complement} \quad (\textit{def preimage}) \\
&\iff f(x) \notin B \quad (\textit{def } (\cdot)\complement) \\
&\iff x \notin f^{-1}(B) \quad (\textit{def preimage}) \\
&\iff x \in f^{-1}(B)^{\complement} \quad ((\cdot)\complement)
\end{aligned}
\tag{4.16}
$$

Proof of (4.13): Let $x \in X$. Then

$$
\begin{aligned}
x \in f^{-1}(B_1 \setminus B_2) &\iff x \in f^{-1}(B_1 \cap B_2^{\complement}) \quad (\textit{def} \setminus) \\
&\iff x \in f^{-1}(B_1) \cap f^{-1}(B_2^{\complement}) \quad (\textit{see } (4.10)) \\
&\iff x \in f^{-1}(B_1) \cap f^{-1}(B_2)^{\complement} \quad (\textit{see } (4.12)) \\
&\iff x \in f^{-1}(B_1) \setminus f^{-1}(B_2) \quad (\textit{def} \setminus)
\end{aligned}
\tag{4.17}
$$

■

Proposition 4.3 (Properties of the direct image)**.** *In the following we assume that J is an arbitrary index set, and that $A \subseteq X$, $A_j \subseteq X$ for all j.*
The following all are true:

$$f(\bigcap_{j\in J} A_j) \subseteq \bigcap_{j\in J} f(A_j) \tag{4.18}$$

$$f(\bigcup_{j\in J} A_j) = \bigcup_{j\in J} f(A_j) \tag{4.19}$$

Proof of (4.18): This follows from the monotonicity of the direct image (see 4.5):

$$\bigcap_{j\in J} A_j \subseteq A_i\, \forall i \in J \Rightarrow f(\bigcap_{j\in J} A_j) \subseteq f(A_i)\, \forall i \in J \quad (\textit{see } 4.5)$$
$$\Rightarrow f(\bigcap_{j\in J} A_j) \subseteq \bigcap_{i\in J} f(A_i) \quad (\textit{def} \cap)$$

First proof of (4.19)) - "Expert proof":

$$y \in f(\bigcup_{j\in J} A_j) \iff \exists\, x \in X:\; f(x) = y \textit{ and } x \in \bigcup_{j\in J} A_j \quad (\textit{def } f(A)) \tag{4.20}$$
$$\iff \exists\, x \in X \textit{ and } j_0 \in J:\; f(x) = y \textit{ and } x \in A_{j_0} \quad (\textit{def} \cup) \tag{4.21}$$
$$\iff \exists\, x \in X \textit{ and } j_0 \in J:\; f(x) = y \textit{ and } f(x) \in f(A_{j_0}) \quad (def f 4.2) \tag{4.22}$$
$$\iff \exists\, j_0 \in J:\; y \in f(A_{j_0}) \quad (\textit{def } f(A)) \tag{4.23}$$
$$\iff y \in \bigcup_{j\in J} f(A_j) \quad (\textit{def} \cup) \tag{4.24}$$

Alternate proof of (4.19)) - 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 "$\subseteq$":

$$y \in f(\bigcup_{j\in J} A_j) \Rightarrow \exists\, x \in X:\; f(x) = y \textit{ and } x \in \bigcup_{j\in J} A_j \quad (\textit{def } f(A)) \tag{4.25}$$
$$\Rightarrow \exists\, j_0 \in J:\; f(x) = y \textit{ and } x \in A_{j_0} \quad (\textit{def} \cup) \tag{4.26}$$
$$\Rightarrow y = f(x) \in f(A_{j_0})(\textit{def } f(A)) \tag{4.27}$$
$$\Rightarrow y \in \bigcup_{j\in J} f(A_j) \quad (\textit{def} \cup) \tag{4.28}$$

B. Proof of "$\supseteq$":

This is a trivial consequence from the monotonicity of $A \mapsto f(A)$:

$$A_i \subseteq \bigcup_{j\in J} A_j\; \forall\, i \in J \Rightarrow f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\; \forall\, i \in J \tag{4.29}$$
$$\Rightarrow \bigcup_{i\in J} f(A_i) \subseteq f(\bigcup_{j\in J} A_j)\; \forall\, i \in J \quad (\textit{def} \cup) \tag{4.30}$$

■

You see that the "elementary" proof is barely longer than the first one, but it is so much easier to understand!

Remark 4.1. In general you will not have equality in (4.18). 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 $f(A_1) \cap f(A_2) = [0, \infty]$ which is clearly bigger than $\{0\}$. ■

Proposition 4.4 (Preimage of function compositions)**.** *Let X, Y, Z be a arbitrary, non–empty sets. Let $f : X \to Y$ and $g : Y \to Z$ and let $W \subseteq Z$. Then*

$$(g \circ f)^{-1} \ = \ f^{-1} \circ g^{-1}, \ i.e., \quad (g \circ f)^{-1}(W) \ = \ f^{-1}\big(g^{-1}(W)\big) \ \text{for all } W \subseteq Z.$$

Proof:
a. "$\subseteq$": Let $W \subseteq Z$ and $x \in (g \circ f)^{-1}(W)$. Then $(g \circ f)(x) = g\big(f(x)\big) \in W$, hence $f(x) \in g^{-1}(W)$. But then $x \in f^{-1}\big(g^{-1}(W)\big)$. This proves "$\subseteq$)".

b. "$\supseteq$": Let $W \subseteq Z$ and $x \in f^{-1}\big(g^{-1}(W)\big)$. Then $f(x) \in g^{-1}(W)$, hence $h(x) = g(f(x)) \in W$, hence $x \in h^{-1}(W) = (g \circ f)^{-1}(W)$. This proves "$\supseteq$)". ■

Proposition 4.5 (Indirect image and fibers of f)**.** *We define on X the equivalence relation*

(4.31) $$x_1 \ \sim \ x_2 \iff f(x_1) \ = \ f(x_2), \ i.e.,$$
(4.32) $$[x]_f \ = \{\bar{x} \in X : f(\bar{x}) = f(x)\}, \ \text{are the equivalence classes.}$$

Then the following is true:

(4.33) $$x \in X \ \Rightarrow \ \Big[\ [x]_f \ = \ \{\hat{x} \in X : f(\hat{x} = f(x))\} \ = \ f^{-1}\{f(x)\}\ \Big]$$
(4.34) $$A \subseteq X \ \Rightarrow \ f^{-1}(f(A)) \ = \ \bigcup_{a \in A} [a]_f.$$

Proof of (4.33): The equation on the left is nothing but the definition of the equivalence classes generated by an equivalence relation, the equation on the right follows from the definition of preimages.

Proof of (4.34):

As $f(A) \ = \ f(\bigcup_{x \in A}\{x\}) \ = \ \bigcup_{x \in A}\{f(x)\}$ (see 4.19), it follows that

(4.35) $$f^{-1}(f(A)) \ = f^{-1}(\bigcup_{x \in A}\{f(x)\})$$
(4.36) $$= \bigcup_{x \in A} f^{-1}\{f(x)\} \quad (\text{see } 4.11)$$
(4.37) $$= \bigcup_{x \in A} [x]_f \quad (\text{see } 4.33)$$

■

Corollary 4.1.

(4.38) $$A \in X \ \Rightarrow \ f^{-1}(f(A)) \ \supseteq \ A.$$

Proof: It follows from $x \sim x$ for all $x \in X$ that $x \in [x]_f$, i.e., $\{x\} \in [x]_f$ for all $x \in X$. But then

(4.39) $$A \ = \bigcup_{x \in A}\{x\} \ \subseteq \ \bigcup_{x \in A}[x]_f \ = \ f^{-1}(f(A))$$

where the last equation holds because of (4.34). ■

Proposition 4.6.

$$B \subset Y \;\; \Rightarrow \;\; f(f^{-1}(B)) \;=\; B \cap f(X). \tag{4.40}$$

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) $x_0 \in f^{-1}(B) \;\Rightarrow y = f(x_0) \in B$ (def. of preimage)
b) Of course $x_0 \in X$. Hence $y = f(x_0) \in f(X)$. a and b together imply $y \in B \cap f(X)$.

Proof of "$\supseteq$":

Let $y \in f(X)$ and $y \in B$. We must prove that $y \in f(f^{-1}(B))$. Because $y \in f(X)$ there exists $x_0 \in X$ such that $y = f(x_0)$. Because $y = f(x_0) \in B$ we conclude that $x_0 \in f^{-1}(B)$ (def preimage). Let us abbreviate $A := f^{-1}(B)$. Now it easy to see that

$$x_0 \in f^{-1}(B) = A \;\; \Rightarrow \;\; y \;=\; f(x_0) \in f(f^{-1}(B)). \tag{4.41}$$

We have shown that if $y \in f(X)$ and $y \in B$ then $y \in f(f^{-1}(B))$. The proof is completed. ■

Remark 4.2. Be sure to understand how the assumption $y \in f(X)$ was used.

Corollary 4.2.

$$B \in Y \;\; \Rightarrow \;\; f(f^{-1}(B)) \;\subseteq\; B. \tag{4.42}$$

Trivial as $f(f^{-1}(B)) \;=\; B \cap f(X) \;\subseteq\; B$. ■

4.3 Appendix: Cardinality - Alternate approach to Beck/Geoghegan

At the beginning of this chapter we look at two lemmata that let you replace bijective and surjective functions with more suitable ones that inherit bijectivity or surjectivity. This will come in handy when we prove propositions concerning cardinality.

The first lemma shows how to preserve bijectivity if two function values need to be switched around.

Lemma 4.2. *Let $X, Y \neq \emptyset$, let $f : X \longrightarrow Y$ be bijective and let $x_1, x_2 \in X$. Let*

$$g(x) := \begin{cases} f(x_2) & \text{if } x = x_1, \\ f(x_1) & \text{if } x = x_2, \\ f(x) & \text{if } x \neq x_1, x_2. \end{cases} \tag{4.43}$$

(In other words, we swap two function arguments). Then $g : X \longrightarrow Y$ also is bijective.

Proof: *Let $y_1 := f(x_1, y_2) := f(x_2)$. Let $f^{-1} : Y \longrightarrow X$ be the inverse function of f and define $G : Y \longrightarrow X$ as follows*

$$G(y) := \begin{cases} f^{-1}(y_2) & \text{if } y = y_1, \\ f^{-1}(y_1) & \text{if } y = y_2, \\ f^{-1}(y) & \text{if } y \neq y_1, y_2. \end{cases} \tag{4.44}$$

We shall show that G satisfies $G \circ g = id_X$ and $g \circ G = id_Y$, i.e., g has G as its inverse. This suffices to prove bijectivity of g.

$$\begin{aligned} y \neq y_1, y_2 \quad \Rightarrow \quad g \circ G(y) &= g(f^{-1}(y)) = f(f^{-1}(y)) = y \text{ as } f^{-1}(y) \neq x_1, x_2, \\ g \circ G(y_1) &= g(f^{-1}(y_2)) = g(x_2) = f(x_1) = y_1 \text{ as } f^{-1}(y_2) = x_2, \\ g \circ G(y_2) &= g(f^{-1}(y_1)) = g(x_1) = f(x_2) = y_2 \text{ as } f^{-1}(y_1) = x_1. \end{aligned}$$

Further,

$$\begin{aligned} x \neq x_1, x_2 \quad \Rightarrow \quad G \circ g(x) &= G(f(x)) = f^{-1}f(x)) = y \text{ as } f(x) \neq y_1, y_2, \\ G \circ g(x_1) &= G(f(x_2)) = G(y_2) = f^{-1}(y_1)) = x_1 \text{ as } f(x_1) = y_1, \\ G \circ g(x_2) &= G(f(x_1)) = G(y_1) = f^{-1}(y_2)) = x_2 \text{ as } f(x_2) = y_2. \end{aligned}$$

We have proved that g has an inverse, the function G. ■

Note that the validity of $G \circ g = id_X$ and $g \circ G = id_Y$ is obvious without the use of any formalism: g differs from f only in that it switches around the function values $f(x_1)$ and $f(x_2)$. and G differs from f^{-1} only in that this switch is reverted.

A more general version of the above shows how to preserve surjectivity if two function values need to be switched around.

Lemma 4.3. *Let $X, Y \neq \emptyset$ and assume that Y contains at least two elements. Let $f : X \longrightarrow Y$ be surjective and let $y_1, y_2 \in Y$. Let $A_1 := f^{-1}\{y_1\}$, $A_2 := f^{-1}\{y_2\}$, and $B := X \setminus (A_1 \cup A_2)$. Let*

$$g(x) := \begin{cases} y_2 & \text{if } x \in A_1, \\ y_1 & \text{if } x \in A_2, \\ f(x) & \text{if } x \in B. \end{cases} \tag{4.45}$$

(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 $g : X \longrightarrow Y$ also is surjective.

Proof:

We notice that A_1, A_2, B partition X into three mutually exclusive parts: $X = B \uplus A_1 \uplus A_2$ and that the sets $f(A_1) = \{y_1\}$, $f(A_2) = \{y_2\}$, $f(B) = Y \setminus \{y_1, y_2\}$ partition Y into $Y = f(B) \uplus f(A_1) \uplus f(A_2)$. (Do you see why $f(B) = Y \setminus \{y_1, y_2\}$?) B and hence $f(B)$ might be empty but none of the other four sets are. It follows that there is indeed a function value $g(x)$ for each $x \in X$ and there is exactly one such value, i.e., g in fact defines a mapping from X to Y. The surjectivity of g follows from that of f and the fact that

$$Y = f(B) \cup f(A_1) \cup f(A_2) = g(B) \cup g(A_2) \cup g(A_1) \tag{4.46}$$

(see (4.19) on p. 24 in prop. 4.3 (Properties of the direct image)). ■

The definitions of finite, countable, countably infinite and uncountable sets were given at the beginning of ch. 3.4: "Countable sets" on p. 18.

Definition 4.2 (cardinality comparisons)**.** Given two arbitrary sets X and Y we say **cardinality of** $X \leqq$ **cardinality of** Y and we write $\mathbf{card}(X) \leqq \mathbf{card}(Y)$ if there is an injective mapping $f : X \longrightarrow Y$.

We say X, Y **have same cardinality** and we write $\mathbf{card}(X) = \mathbf{card}(Y)$ if both $\text{card}(X) \leqq \text{card}(Y)$ and $\text{card}(Y) \leqq \text{card}(X)$, i.e., there is a bijective mapping $f : X \xrightarrow{\sim} Y$.

Finally we say **cardinality of** X $<$ **cardinality of** Y and we write $\mathbf{card}(X) < \mathbf{card}(Y)$ if both $\text{card}(X) \leqq \text{card}(Y)$ and $\text{card}(Y) = \text{card}(X)$, i.e., there is an injective mapping but not a bijection $f : X \longrightarrow Y$.

Proposition 4.7. *Let $m, n \in \mathbb{N}$. Let $\emptyset \neq A \subseteq [m]$. If $m < n$ then there is no surjection from A to $[n]$.*

Proof by induction on n:

Base case: Let $n = 2$. This implies $m = 1$ and $A = [1]$ (no other non-empty subset of $[1]$). If there was surjective $f : A \longrightarrow [2]$ then either $f(1) = 1$ in which case $2 \notin f(A)$ or $f(1) = 2$ in which case $1 \notin f(A)$. This proves the base case.

Induction assumption: Fix $n \in \mathbb{N}$ and assume that for any $\tilde{m} < n$ and non-empty $\tilde{A} \subseteq [\tilde{m}]$ there is no surjective $\tilde{f} : \tilde{A} \longrightarrow [n]$.

We must now prove the following: Let $m \in \mathbb{N}$ and $\emptyset \neq A \subseteq [m]$. If $m < n + 1$ then there is no surjection from A to $[n + 1]$. Let us assume contrary to assumption that a surjective $f : A \longrightarrow [n + 1]$ exists.

***case 1:** $n \notin A$**:** As $m < n + 1$ this implies both $n, n + 1 \notin A$, hence $A \subseteq [n - 1]$. Let $\tilde{A} := A \setminus f^{-1}\{n + 1\}$. Then $\tilde{A} \subseteq A \subseteq [n-1]$ and, as the surjective f "hits" every integer between 1 and $n+1$ and we only removed those $a \in A$ which map to $n + 1$, the restriction $\tilde{f}$ of f to $\tilde{A}$ still maps to any integer between 1 and n. In other words, $\tilde{f} : \tilde{A} \longrightarrow [n]$ is surjective, contradictory to our induction assumption.*

***case 2:** $n \in A$ **and** $f(n) = n + 1$**:** As in case 1, let $\tilde{A} := A \setminus f^{-1}\{n + 1\}$. Then $\tilde{A} \subseteq A \subseteq [n - 1]$ because n was discarded from A as an element of $f^{-1}\{n + 1\}$. Again, the surjective f "hits" every integer between 1 and $n + 1$ and again, we only removed those $a \in A$ which map to $n + 1$. It follows as in case 1 that $\tilde{f} : \tilde{A} \longrightarrow [n]$ is surjective, contradictory to our induction assumption.*

case 3: $n \in A$ ***and*** $f(n) \neq n+1$***:** According to lemma 4.7 on p. 29 we can replace f by a surjective function g which maps n to $n+1$. This function satisfies the conditions of case 2 above, for which it was already proved that no surjective mapping from A to $[n+1]$ exists. We have reached a contradiction.* ■

Corollary 4.3 (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. 4.7 with $A := [m]$. ■

Corollary 4.4 (Pigeonhole Principle)**.** *B/G Prop.13.5: Let $m, n \in \mathbb{N}$. If $m < n$ then there is no injective $f : [n] \longrightarrow [m]$.*

Proof: Otherwise g would have a (surjective) left inverse $g : [m] \longrightarrow [n]$ in contradiction to the preceding proposition. ■

Proposition 4.8 (B/G Prop.13.6, p.122: Subsets of finite sets are finite)**.** *Let $\emptyset \neq B \subseteq A$ and let A be finite. Then B is finite.*

Proof: Done by induction on the cardinality n of sets:

Base case: $n = 1$ or $n = 2$: Proof obvious.

Induction assumption: Assume that all subsets of sets of cardinality less than n are finite.

Now let A be a set of $card(A) = n$. there is a bijection $a(\cdot) : [n] \longrightarrow A$. Let $B \subseteq A$.

Case 1: *$a(n) \in B$***:** Let $B_n := B \setminus \{a(n)\}$ and $A_n := A \setminus \{a(n)\}$. Then the restriction of $a(\cdot)$ to $[n-1]$ is a bijection $[n-1] \longrightarrow A_n$ according to B/G prop.13.2. As $card(A_n) = n-1$ and $B_n \subseteq A_n$ it follows from the induction assumption that B_n is finite: there exists $m \in \mathbb{N}$ and a bijection $b(\cdot) : [m] \longrightarrow B_n$. We now extend $b(\cdot)$ to $[m+1]$ by defining $b(m+1) := a(n)$. It follows that this extension remains injective and it is also surjective if we choose as codomain $B_n \cup \{a(n)\} \ = B$. It follows that B is finite.*

Case 2: *$a(n) \notin B$***:** We pick an arbitrary $b \in B$. Let $j := a^{-1}(b)$. Clearly $j \in [n]$. Now we modify the mapping $a(\cdot)$ by switching the function values for j and n. We obtain another bijection $f : [n] \longrightarrow A$ (see lemma 4.2 on p. 28) for which $b(n) = a(j) = b \in B$. We now can apply what was proved in case 1 and obtain that B is finite.* ■

Proposition 4.9 (B/G Cor.13.16, p.122)**.** *$\mathbb{N}^2$ is countable.*

Proof: Done by directly specifiying a bijection $F : \mathbb{N}^2 \xrightarrow{\sim} \mathbb{N}$.

We think of $\mathbb{N}^2$ as a matrix with "infinitely many rows and columns"

$$(1,1)\ (1,2)\ (1,3)\ \ldots \tag{4.47}$$
$$(2,1)\ (2,2)\ (2,3)\ \ldots \tag{4.48}$$
$$(3,1)\ (3,2)\ (3,3)\ \ldots \tag{4.49}$$

Let For an integer $n \geqq 2$ let $D_n := \ \{(i,j) \in \mathbb{N}^2 : i+j\}$ be the "n-th diagonal" of $\mathbb{N}^2$. It is clear that

$\mathbb{N}^2 = \bigcup_{n \geqq 2} D_n$. *We reorganize the elements of* $\mathbb{N}^2$ *into an ordinary sequence* $(a_j)_{j \in \mathbb{N}}$ *as follows:*

$$a_1 := (1,1), \tag{4.50}$$
$$a_2 := (1,2),\ a_3 := (2,1), \tag{4.51}$$
$$a_4 := (1,3),\ a_5 := (2,2),\ a_6 := (3,1), \tag{4.52}$$
$$a_7 := (1,4),\ a_8 := (2,3),\ a_9 := (3,2),\ a_{10} := (4,1), \tag{4.53}$$
$$\dots \tag{4.54}$$

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* $(n-1, 1)$). *In formal terms: For* $k \in \mathbb{N}$ *let* $s_k := \sum_{j=1}^{k} j = k(k+1)/2$. *As* $card(D_n) = n - 1$ *and the first element of* D_{n+1} *becomes part of the sequence* (a_j) *after the last element of* D_n *was assigned, we obtain*

$$a_{s_{n-1}+j} = (j, (n+1) - j) \quad (1 \leqq j < n+1) \tag{4.55}$$

If we interpret the sequence (a_j) *as a function* $a(\cdot) : \mathbb{N} \longrightarrow \mathbb{N}^2$ *then we show that is bijective by giving the inverse* $g : \mathbb{N}^2 \longrightarrow \mathbb{N}$*:*

$$F(1,1) = 2, \tag{4.56}$$

$$F(i, (n+1) - i) = s_{n-1} + i \quad \left(n \geqq 2,\ 1 \leqq i < n+1,\ s_k = \sum_{j=1}^{k} j = \frac{k(k-1)}{2}\right) \tag{4.57}$$

■

Lemma 4.4. *Let* $A_1,\ A_2,\ A_3,\ \dots$ *be a sequence of sets. Let*

$$B_1 := A_1; \quad B_2 := A_2 \setminus A_1; \quad B_3 := A_3 \setminus (A_1 \cup A_2); \quad B_{n+1} := A_{n+1} \setminus (A_1 \cup A_2 \cup \dots \cup A_n).$$

Then the sets B_j *are mutually disjoint and*

$$\bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{n} B_j \quad (n \in \mathbb{N}) \tag{4.58}$$
$$\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} B_j \tag{4.59}$$

Proof: Left to the reader. ■

The following proposition was proved informally before (see thm.3.3 on p.19)

Theorem 4.1 (B/G prop.13.6: Countable unions of countable sets)**.** *The union of countably many countable sets is countable.*

Proof: Let the sets be

$$A_1,\ A_2,\ A_3,\ \dots \quad \text{and let } A := \bigcup_{n \in \mathbb{N}} A_i.$$

We assume that at least one of those sets is not empty: otherwise their union is empty, hence finite, hence countable and we are done. We may assume, on account of lemma 4.4 that the sets are mutually disjoint, i.e., any two different sets A_i, A_j have intersection $A_i \cap A_j = \emptyset$ (see definition (3.11) on p.14).

A. *As each of the non-empty A_i is countable, either this set 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 $f : A \longrightarrow \mathbb{N}^2$ as follows: Let $a \in A$. As the A_j are disjoint there is a unique index i such that $a \in A_i$ and, as sets do not contain duplicates of their elements, there is a unique index j such that $a = a_i(j)$. In other words, for any $a \in A$ there exists a unique pair $(i_a, j_a) \in \mathbb{N}^2$ such that $a = a_{i_a}(j_a)$ and the assignment $a \mapsto (i_a, j_a)$ defines an injective function $f : A \longrightarrow \mathbb{N}^2$. But then this same assignment gives us a bijective function $F : A \xrightarrow{\sim} f(A)$. $f(A)$ is countable as a subset of the countable set $\mathbb{N}^2$ and this proves the theorem as any subset of a countable set is countable (see B/G prop.13.10).* ■

Corollary 4.5. *Let the set X not be countable and let $A \subseteq X$ be countable. Then its complement $A^{\complement}$ is not countable.*

Instructor's Note:
Proof: Left to the reader.

Definition 4.3 (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 4.10 (B/G Prop.13.21, p.125: All algebraic numbers are countable)**.** *All algebraic numbers are countable*

Proof: Let P be the set of all integer polynomials and Z the set of zeroes for all such polynomials. Let

$$P_n := \{\text{polynomials } p(x) = \sum_{j=1}^{k} a_j x^j : a_j \in \mathbb{Z} \text{ and } -n \leqq a_j \leqq n\}. \tag{4.60}$$

Then P_n is finite and

$$Z_n := \{x \in \mathbb{R} : p(x) = 0 \text{ for some } p \in P_n\} \tag{4.61}$$

also is finite as a polynomial of degree n has at most n zeroes. ■

Here are some trivial consequences of the fact that $\mathbb{R}$ is not countable (see thm. 3.4, p.3.4 and B/G Thm.13.22).

Proposition 4.11 (The transcendental numbers are not countable)**.** *All transcendental numbers are not countable.*

Proof: the uncountable real numbers are the disjoint union of the countable algebraic numbers with the transcendentals. ■

5 Real functions (Understand this!)

5.1 Operations on real functions

Definition 5.1 (real functions)**.** If the codomain Y of a mapping

$$f(\cdot): \quad X \longrightarrow Y \qquad x \longmapsto f(x)$$

is a subset of $\mathbb{R}$, then we call $f(\cdot)$ a **real function** or **real valued function**.

Remember that this definition does not exclude the case $Y = \mathbb{R}$ because $Y \subseteq \mathbb{R}$ is in particular true if both sets are equal.

Real functions are a pleasure to work with because, given any fixed argument x_0, the object $f(x_0)$ is just an ordinary number. In particular you can add, subtract, multiply and divide real functions. Of course, division by zero is not allowed:

Definition 5.2 (Operations on real functions)**.** Let X an arbitrary non-empty set.
Given are two real functions $f(\cdot), g(\cdot) : X \to (R)$ and a real number α. The **um** $f + g$, **difference** $f - g$, **product** fg or $f \cdot g$, **quotient** f/g or $\frac{f}{g}$, and **scalar product** αf are defined by doing the operation in question with the numbers $f(x)$ and $g(x)$ for each $x \in X$.

$$\begin{aligned}
(f+g)(x) &:= f(x) + g(x) \\
(f-g)(x) &:= f(x) - g(x) \\
(fg)(x) &:= f(x)g(x) \\
(f/g)(x) &:= f(x)/g(x) \qquad \text{for all } x \in X \text{ where } g(x) \neq 0 \\
(\alpha f)(x) &:= \alpha \cdot g(x)
\end{aligned} \tag{5.1}$$

Before we list some basic properties of addition and scalar multiplication of functions (the operations that interest us the most), let us have a quick look at constant functions.

Definition 5.3 (Constant functions)**.** Let a be an ordinary real number. You can think of a as a function from any non-empty set X to $\mathbb{R}$ as follows:

$$a(\cdot): \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto a$$

In other words, the function $a(\cdot)$ assigns to each $x \in X$ one and the same value a. We call such a function a **constant function**.

The most important constant function is the **zero function** $0(\cdot)$ which maps any $x \in X$ to the number zero. We usually just write 0 for this function unless doing so would confuse the reader. Note that scalar multiplication $(\alpha f)(x) = \alpha \cdot g(x)$ is a special case of multiplying two functions $(gf)(x) = g(x)f(x)$: Let $g(x) = \alpha$ (constant function α).

We do not need to assume that $f(\cdot)$ is a real function. We call any mapping f from X to Y constant if its image $f(X) \subseteq Y$ is a singleton, i.e, it consists of exactly one element.

One last definition before we finally get so see some examples:

Definition 5.4 (Negative function)**.** Let X be an arbitrary, non-empty set and let

$$f(\cdot): \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto f(x)$$

be a real function on X. The function

$$-f(\cdot): \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto -f(x)$$

which assigns to each $x \in X$ the value $-f(x)$ is called **negative** f or **minus** f. Sometimes we write $-f$ rather than $-f(\cdot)$.

All those last definitions about sums, products, scalar products, ... of real functions are very easy to understand if you remember that, for any fixed $x \in X$, you just deal with ordinary numbers!

Example 5.1 (Arithmetic operations on real functions)**.** For simplicity, we set $X := \mathbb{R}_+ = \{x \in \mathbb{R} : x \geqq 0\}$. Let

$$\begin{aligned}
f(\cdot): &\quad \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto (x-1)(x+1) \\
g(\cdot): &\quad \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto x-1 \\
h(\cdot): &\quad \mathbb{R}_+ \longrightarrow \mathbb{R} \qquad x \longmapsto x+1
\end{aligned}$$

Then

$$\begin{aligned}
(f+h)(x) &= (x-1)(x+1)+x+1 &&= x^2-1+x+1 = x(x+1) && \forall x \in \mathbb{R}_+ \\
(f-g)(x) &= (x-1)(x+1)-(x-1) &&= x^2-1-x+1 = x(x-1) && \forall x \in \mathbb{R}_+ \\
(gh)(x) &= (x-1)(x+1) &&= f(x) && \forall x \in \mathbb{R}_+ \\
(f/h)(x) &= (x-1)(x+1)/(x+1) &&= x-1 = g(x) && \forall x \in \mathbb{R}_+ \\
(f/g)(x) &= (x-1)(x+1)/(x-1) &&= x+1 = h(x) && \forall x \in \mathbb{R}_+ \setminus \{1\}
\end{aligned}$$

It is really, really important for you to understand that $f/g(\cdot)$ and $h(\cdot)$ are **not the same functions** on $\mathbb{R}_+$. Matter of fact, $f/g(\cdot)$ is not defined for all $x \in \mathbb{R}_+$ because for $x = 1$ you obtain $\frac{(1-1)(1+1)}{1-1} = 0/0$. The domain of f/g is different from that of h and both functions thus are different.

5.2 Maxima, suprema, limsup ... (Study this!)

Definition 5.5 (Upper and lower bounds, maxima and minima)**.** Let $A \subseteq \mathbb{R}$. Let $l, u \in \mathbb{R}$. We call l a **lower bound** of A if $l \leqq a$ for all $a \in A$. We call u an **upper bound** of A if $u \geqq a$ for all $a \in A$.

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

The next proposition will show that minimum and maximum 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 5.1. *Let $A \subseteq \mathbb{R}$. If A has a maximum then it is unique. If A has a minimum then it is unique.*

Proof for maxima: Let u_1 and u_2 be two maxima of A: both are upper bounds of A and both belong to A. As u_1 is an upper bound, it follows that $a \leqq u_1$ for all $a \in A$. Hence $u_2 \leqq u_1$. As u_2 is an upper bound, it follows that $u_1 \leqq u_2$ and we have equality $u_1 = u_2$. The proof for minima is similar. ■

Definition 5.6. Given $A \subseteq \mathbb{R}$ we define

$$(5.2)\qquad \begin{aligned} A_{\mathit{lowb}} &:= \{l \in \mathbb{R} : l \text{ is lower bound of } A\} \\ A_{\mathit{uppb}} &:= \{u \in \mathbb{R} : u \text{ is upper bound of } A\}. \end{aligned}$$

We say that A is **bounded above** if $A_{\mathit{uppb}} \neq \emptyset$ and we say that A is **bounded below** if $A_{\mathit{lowb}} \neq \emptyset$.

Axiom 5.1 ($\mathbb{R}$ is complete)**.** (see [1] B/G axiom 8.52, p.83).
Let $A \subseteq \mathbb{R}$.
If A_{uppb} is not empty then A_{uppb} has a minimum.

Remark 5.1. A_{lowb} and/or A_{uppb} may be empty: $A = \mathbb{R}$, $A = \mathbb{R}_{>0}$, $A = \mathbb{R}_{<0}$,

Definition 5.7. Let $A \subseteq \mathbb{R}$. If A_{uppb} is not empty then $\min(A_{\mathit{uppb}})$ exists by axiom 5.1 and it is unique by prop. 5.5. We write $\sup(A)$ or l.u.b.(A) for $\min(A_{\mathit{uppb}})$ and call it the **supremum** or **least upper bound** of A.

We shall see in cor.5.1 that, if A_{lowb} is not empty, then $\max(A_{\mathit{lowb}})$ exists and it is unique by prop. 5.5. We write $\inf(A)$ or g.l.b.(A) for $\max(A_{\mathit{lowb}})$ and call it the **infimum** or **greatest lower bound** of A.

Proposition 5.2 (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 $-A = \{-y : y \in A\}$:*

$$(5.3)\qquad \begin{aligned} &-x \textit{ is a lower bound of } A \iff x \textit{ is an upper bound of -}A \quad \textit{and vice versa},\\ &-x \in A_{\mathit{uppb}} \iff x \in (-A)_{\mathit{lowb}} \quad \textit{and vice versa},\\ &-x = \sup(A) \iff x = \inf(-A) \quad \textit{and vice versa},\\ &-x = \max(A) \iff x = \min(-A) \quad \textit{and vice versa}. \end{aligned}$$

Proof: A simple consequence of

$$-x \leq y \iff x \geqq -y \quad \textit{and} \quad -x \geqq y \iff x \leqq -y. \qquad \blacksquare$$

Corollary 5.1. *Let $A \subseteq \mathbb{R}$. If A has lower bounds then $\inf(A)$ exists.*

Proof: According to the duality proposition prop.5.2, 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.5.2 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 5.2 (Example a: Maximum exists)**.** Let $X_1 := \{t \in \mathbb{R} : 0 \leqq t \leqq 1\}$.
For each $x \in X_1$ we have $|f(x) - g(x)| = g(x) - f(x) = 2x$ and the biggest possible such difference is $g(1) - f(1) = 2$, so $d(f,g) = 2$.

Example 5.3 (Example b: Supremum is finite)**.** Let $X_2 := \{t \in \mathbb{R} : 0 \leqq t < 1\}$, i.e., we now exclude the right end point 1 at which the maximum difference was attained. For each $x \in X$ we have

$$|f(x) - g(x)| = g(x) - f(x) = 2x$$

and the biggest possible such difference is certainly bigger than

$$g(0.9999999999) - f(0.9999999999) = 1.9999999998.$$

If you keep adding $5,000$ 9s to the right of the argument x, then you get the same amount of 9s inserted into the result $2x$, so $2x$ comes closer than anything you can imagine to the number 2, without actually being allowed to reach it. The supremum is still considered in a case like this to be 2. This precisely is the difference in behavior between the supremum $s := \sup(A)$ and the **maximum** $m := max(A)$ of a set $A \subseteq \mathbb{R}$ of real numbers: For the maximum there must actually be at least one element $a \in A$ so that $a = max(A)$. For the supremum it is sufficient that there is a sequence $a_1 \leqq a_2 \leqq \dots$ which approximates s from below in the sense that the difference $s - a_n$ "drops down to zero" as n approaches infinity. I will not be more exact than this because doing so would require us to delve into the concept of convergence and contact points.

Example 5.4 (Example c: Supremum is infinite)**.** Let $X_3 := \{t \in \mathbb{R} : 0 \leqq t\}$. For each $x \in X_1$ 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 amongst all $x \in X_3$. By the way, you should understand that even though $\sup(A)$ as best approximation of the largest value of $A \subseteq \mathbb{R}$ is allowed to take the "value" $+\infty$ or $-\infty$ this cannot be allowed for $\max(A)$. How so? The infinity values are not real numbers, but, by definition of the maximum, if $\alpha := max(A)$ exists, then $\alpha \in A$. In particular, the max must be a real number.

That last example motivates the following definition.

Definition 5.8 (Supremum and Infimum of unbounded and empty sets)**.** If A is not bounded from above then we define

$$\sup A = \infty \tag{5.4}$$

If A is not bounded from below then we define

$$\inf A = -\infty \tag{5.5}$$

Finally we define

$$\sup \emptyset = -\infty, \qquad \inf \emptyset = +\infty \tag{5.6}$$

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 5.9 (supremum and infimum of functions)**.** Let X be an arbitrary set, $A \subseteq X$ a subset of X, $f : X \to \mathbb{R}$ a real function on X. Look at the set $f(A) = \{f(x) : x \in A\}$, i.e., the image of A under $f(\cdot)$.

The **supremum of** $f(\cdot)$ **on** A is then defined as

$$\sup_A f := \sup_{x \in A} f(x) = \sup\left(f(A)\right) \tag{5.7}$$

The **infimum of** $f(\cdot)$ **on** A is then defined as

(5.8)
$$\inf_A f \ := \ \inf_{x \in A} f(x) \ = \ \inf\left(f(A)\right)$$

Definition 5.10 (supremum and infimum of families)**.** Let $(x_i)_{i \in I}$ be an indexed family of real numbers x_i. Remember that if $I \subseteq \mathbb{Z}$ we call (x_i) a sequence!

The **supremum of** $(x_i)_{i \in I}$ is then defined as

(5.9)
$$\sup\left(x_i\right) \ := \ \sup\left(x_i\right)_{i \in I} \ := \ \sup_{i \in I} x_i \ = \ \sup\left\{x_i : i \in I\right\}$$

The **infimum of** $(x_i)_{i \in I}$ is then defined as

(5.10)
$$\inf\left(x_i\right) \ := \ \inf\left(x_i\right)_{i \in I} \ := \ \inf_{i \in I} x_i \ = \ \inf\left\{x_i : i \in I\right\}$$

The definition above for families is consistent with the one given earlier for sequences (the special case of countable families). We repeat it here for your convenience.

Definition 5.11 (supremum and infimum of sequences)**.** Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of real numbers x_n. The **supremum of** $(x_n)_{n \in \mathbb{N}}$ is then defined as

(5.11)
$$\sup\left(x_n\right) \ := \ \sup\left(x_n\right)_{n \in \mathbb{N}} \ := \ \sup_{n \in \mathbb{N}} x_n \ = \ \sup\left\{x_n : n \in \mathbb{N}\right\}$$

The **infimum of** $(x_n)_{n \in \mathbb{N}}$ is then defined as

(5.12)
$$\inf\left(x_n\right) \ := \ \inf\left(x_n\right)_{n \in \mathbb{N}} \ := \ \inf_{n \in \mathbb{N}} x_n \ = \ \inf\left\{x_n : n \in \mathbb{N}\right\}$$

We note that the "duality principle" for min and max, sup and inf is true in all cases above: You flip the sign of the items you examine and the sup/max of one becomes the inf/min of the other and vice versa.

Definition 5.12 (Tail sets of a sequence)**.** Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}$. Let

(5.13)
$$T_n \ := \ \{x_j : j \geqq n\} \ = \ \{x_n, x_{n+1}, x_{n+2}, x_{n+3}, \dots\}$$

be what remains in the sequence after we discard the first $n-1$ elements. We call $(T_n)_{n \in \mathbb{N}}$ the sequence of **tail sets** of the given sequence $(x_k)_{k \in \mathbb{N}}$.

Remark 5.2. Some simple properties of tail sets:

a. We deal with sets and not with sequences T_n: If, e.g., $x_n = (-1)^n$ then each $T_n = \{-1, 1\}$ only contains two items and not infinitely many.

b. The tail set sequence T_n is "decreasing": If $m < n$ then $T_m \supseteq T_n$.

We recall the following: Let x_n be a sequence of real numbers that is non-decreasing, i.e., $x_n \leqq x_{n+1}$ for all n (see def. 2.1, p.5) and bounded above. Then $\lim\limits_{n\to\infty} x_n$ exists and coincides with $\sup\{x_n : n \in \mathbb{N}\}$ (see the proof of [1] B/G thm 10.19, p.101). And, for a sequence y_n of real numbers that is non-increasing, i.e., $y_n \geqq y_{n+1}$ for all n and bounded below, the analogous result is that $\lim\limits_{n\to\infty} y_n$ exists and coincides with $\inf\{y_n : n \in \mathbb{N}\}$. It follows that

$$
\begin{aligned}
\inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big) &= \lim_{n\to\infty}\big(\sup(T_n)\big) = \lim_{n\to\infty}\big(\sup\{x_j : j \in \mathbb{N}, j \geqq n\}\big),\\
\sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big) &:= \lim_{n\to\infty}\big(\inf(T_n)\big) = \lim_{n\to\infty}\big(\inf\{x_j : j \in \mathbb{N}, j \geqq n\}\big).
\end{aligned} \tag{5.14}
$$

An expression like $\sup\{x_j : j \in \mathbb{N}, j \geqq n\}$ can be written more compactly as $\sup\limits_{j\in\mathbb{N}, j\geqq n}\{x_j\}$. Moreover, when dealing with sequences (x_n), it is understood in most cases that $n \in \mathbb{N}$ or $n \in \mathbb{Z}_{\geqq 0}$ and the last expression simplifies to $\sup\limits_{j\geqq n}\{x_j\}$. This can also be written as $\sup\limits_{j\geqq n}(x_j)$ or $\sup\limits_{j\geqq n} x_j$.

In other words, (5.14) becomes

$$
\begin{aligned}
\inf_{n\in\mathbb{N}}\big(\sup_{j\geqq n} x_j\big) &= \inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big) = \lim_{n\to\infty}\big(\sup(T_n)\big) = \lim_{n\to\infty}\big(\sup_{j\geqq n} x_j\big),\\
\sup_{n\in\mathbb{N}}\big(\inf_{j\geqq n} x_j\big) &= \sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big) = \lim_{n\to\infty}\big(\inf(T_n)\big) = \lim_{n\to\infty}\big(\inf_{j\geqq n} x_j\big).
\end{aligned} \tag{5.15}
$$

The above justifies the following definition:

Definition 5.13. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and let $T_n = \{x_j : j \in \mathbb{R}, j \geqq n\}$ be the tail set for x_n. Assume that T_n is bounded above for some $n_0 \in \mathbb{N}$ (and hence for all $n \geqq n_0$). We call

$$\limsup_{n\to\infty} x_j := \lim_{n\to\infty}\big(\sup_{j\geqq n} x_j\big) = \inf_{n\in\mathbb{N}}\big(\sup_{j\geqq n} x_j\big)$$

the **lim sup** or **limit superior** of the sequence (x_n). If, for each n, T_n is not bounded above then we say $\limsup\limits_{n\to\infty} x_j = \infty$. Assume that T_n is bounded below for some n_0 (and hence for all $n \geqq n_0$). We call

$$\liminf_{n\to\infty} x_j := \lim_{n\to\infty}\big(\inf_{j\geqq n} x_j\big) = \sup_{n\in\mathbb{N}}\big(\inf_{j\geqq n} x_j\big)$$

the **lim inf** or **limit inferior** of the sequence (x_n). If, for each n, T_n is not bounded below then we say $\liminf\limits_{n\to\infty} x_j = -\infty$.

Proposition 5.3. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded above with tail sets T_n.*

A. Let

$$
\begin{aligned}
\mathscr{U} &:= \{y \in \mathbb{R} : T_n \cap [y,\infty] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\\
\mathscr{U}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \geqq y\},\\
\mathscr{U}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \cdots \in \mathbb{N} \text{ such that } x_{n_j} \geqq y \text{ for all } j \in \mathbb{N}\},\\
\mathscr{U}_3 &:= \{y \in \mathbb{R} : x_n \geqq y \text{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{5.16}
$$

Then $\mathscr{U} = \mathscr{U}_1 = \mathscr{U}_2 = \mathscr{U}_3$.

B. There exists $z = z(\mathscr{U}) \in \mathbb{R}$ such that $\mathscr{U}$ is either an interval $]-\infty, z]$ or an interval $]-\infty, z[$.

C. Let $u := \sup(\mathscr{U})$. Then $u = z = z(\mathscr{U})$ as defined in part B. Further, u is the only real number such that
C1.

$$u - \varepsilon \in \mathscr{U} \quad \text{and} \quad u + \varepsilon \notin \mathscr{U} \quad \text{for all } \varepsilon > 0. \tag{5.17}$$

C2. *There exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $u = \lim_{j\to\infty} x_{n_j}$ and u is the largest real number for which such a subsequence exists.*

Proof of A:

A.1 - $\mathscr{U} = \mathscr{U}_1$: This equality is valid by definition of tailsets of a sequence:

$$x \in T_n \iff x = x_j \text{ for some } j \geqq n \iff x = x_{n+k} \text{ for some } k \in \mathbb{Z}_{\geqq 0}$$

from which it follows that $x \in T_n \cap [y, \infty] \iff x = x_{n+k}$ for some $k \geqq 0$ and $x_{n+k} \geqq y$.

A.2 - $\mathscr{U}_1 \subseteq \mathscr{U}_2$: Let $y \in \mathscr{U}_1$ and $n \in \mathbb{N}$. We prove the existence of $(n_j)_j$ by induction on j. Base case $j = 1$: As $T_2 \cap [y, \infty] \neq \emptyset$ there is some $x \in T_2$ such that $y \leqq x < \infty$, i.e., $x \geqq y$. Because $x \in T_2 = \{x_2, x_3, \dots\}$ we have $x = x_{n_1}$ for some integer $n_1 > 1$ and we have proved the existence of n_1. Induction assumption: Assume that $n_1 < n_2 < \cdots < n_{j_0}$ have already been picked. Induction step: Let $n = n_{j_0}$. As $y \in \mathscr{U}_1$ there is $k \in \mathbb{N}$ such that $x_{n_{j_0}+k} \geqq y$. We set $n_{j_0+1} := n_{j_0} + k$. As this index is strictly larger than n_{j_0}, the induction step has been proved.

A.3 - $\mathscr{U}_2 \subseteq \mathscr{U}_3$: This is trivial: Let $y \in \mathscr{U}_2$. The strictly increasing subsequence $n_1 < n_2 < n_3 < \cdots \in \mathbb{N}$ constitutes the infinite set of indices that is required to grant y membership in $\mathscr{U}_3$.

A.2 - $\mathscr{U}_3 \subseteq \mathscr{U}$: Let $y \in \mathscr{U}_3$. Fix some $n \in \mathbb{N}$. Let $J = J(y) \subseteq \mathbb{N}$ be the infinite set of indices j for which $x_j \geqq y$. At most finitely many of those j can be less than that given n and there must be (infinitely many) $j \in J$ such that $j \geqq n$ Pick any one of those, say j'. Then $x_{j'} \in T_n$ and $x_{j'} \geqq y$. It follows that $y \in \mathscr{U}$

We have shown the following sequence of inclusions:

$$\mathscr{U} = \mathscr{U}_1 \subseteq \mathscr{U}_2 \subseteq \mathscr{U}_3 \subseteq \mathscr{U}$$

It follows that all four sets are equal and part A of the proposition has been proved.

Proof of B: Let $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $y_2 \in \mathscr{U}$. It follows from $[y_2, \infty] \subseteq [y_1, \infty]$ that, because $T_n \cap [y_2, \infty] \neq \emptyset$ for all $n \in \mathbb{N}$, we must have $T_n \cap [y_1, \infty] \neq \emptyset$ for all $n \in \mathbb{N}$, i.e., $y_1 \in \mathscr{U}$. But that means that $\mathscr{U}$ must be an interval of the form $]-\infty, z]$ or $]-\infty, z[$ for some $z \in \mathbb{R}$.

Proof of C: Let $z = z(\mathscr{U})$ as defined in part B and $u := \sup(\mathscr{U})$.

Proof of C.1 - (5.17) part 1, $u - \varepsilon \in \mathscr{U}$: As $u - \varepsilon$ is smaller than the least upper bound u of $\mathscr{U}$, $u - \varepsilon$ is not an upper bound of $\mathscr{U}$. Hence there is $y > u - \varepsilon$ such that $y \in \mathscr{U}$. It follows from part B that $u - \varepsilon \in \mathscr{U}$.✓

Proof of C.1 - (5.17) part 2, $u + \varepsilon \notin \mathscr{U}$: This is trivial as $u + \varepsilon > u = \sup(\mathscr{U})$ implies that $y \leqq u < u + \varepsilon$ for all $y \in \mathscr{U}$. But then $y \neq u$ for all $y \in \mathscr{U}$, i.e., $u \notin \mathscr{U}$. This proves $u + \varepsilon \notin \mathscr{U}$.

Proof of C.2: We construct by induction a sequence $n_1 < n_2 < \dots$ of natural numbers such that

$$u - 1/j \leq x_{n_j} \leqq u + 1/j. \tag{5.18}$$

Base case: We have proved as part of C.1 that $x_n \geqq u + 1$ for at most finitely many indices n. Let K be the largest of those. As $u - 1 \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geqq u - 1$. Infinitely many of them must exceed K. We pick one of them and that will be n_1. Clearly, n_1 satisfies (5.18) and this proves the base case.

Let us now assume that $n_1 < n_2 < \dots < n_k$ satisfying (5.18) have been constructed. $x_n \geqq u + 1/(k+1)$ is possible for at most finitely many indices n. Let K be the largest of those. As $u - 1/(k+1) \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geqq u - 1/(k+1)$. Infinitely many of them must exceed $\max(K, n_k)$. We pick one of them and that will be n_{k+1}. Clearly, n_{k+1} satisfies (5.18) and this finishes the proof by induction.

We now show that $\lim\limits_{j\to\infty} x_{n_j} = u$. Given $\varepsilon > 0$ there is $N = N(\varepsilon)$ such that $1/N < \varepsilon$. It follows from (5.18) that $|x_{n_j} - u| \leqq 1/j < 1/N < \varepsilon$ for all $j \geqq n$ and this proves that $x_{n_j} \to u$ as $j \to \infty$.

We will be finished with the proof of C.2 if we can show that if $w > u$ then there is no sequence $n_1 < n_2 < \dots$ such that $x_{n_j} \to w$ as $j \to \infty$. Let $\varepsilon := (w-u)/2$. According to (5.17), $u + \varepsilon \notin \mathscr{U}$. But then, by definition of $\mathscr{U}$, there is $n \in \mathbb{N}$ such that $T_n \cap [u+\varepsilon, \infty[= \emptyset$. But $u + \varepsilon = w - \varepsilon$ and we have $T_n \cap [w - \varepsilon, \infty[= \emptyset$. This implies that $|w - x_j| \geqq \varepsilon$ for all $j \geqq n$ and that rules out the possibility of finding n_j such that $\lim\limits_{j\to\infty} x_{n_j} = w$. ■

Corollary 5.2. *As in prop.5.3, let $u := \sup(\mathscr{U})$. Then $\mathscr{U} =]-\infty, u]$ or $\mathscr{U} =]-\infty, u[$.*

Further, u is determined by the following property: For any $\varepsilon > 0$, $x_n > u - \varepsilon$ for infinitely many n and $x_n > u + \varepsilon$ for at most finitely many n.

Proof: This follows from $\mathscr{U} = \mathscr{U}_3$ and parts B and C of prop.5.3. ■

When we form the sequence $y_n = -x_n$ then the roles of upper bounds and lower bounds, max and min, inf and sup will be reversed. Example: x is an upper bound for $\{x_j : j \geqq n$ if and only if $-x$ is a lower bound for $\{y_j : j \geqq n$.

The following "dual" version of prop. 5.3 is a direct consequence of the duality of upper/lower bounds, min/max, inf/sup proposition prop.5.2, p.35.

Proposition 5.4. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ which is bounded below with tail sets T_n.*

A. Let

$$\begin{aligned}
\mathscr{L} &:= \{y \in \mathbb{R} : T_n \cap [-\infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\\
\mathscr{L}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \leqq y\},\\
\mathscr{L}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \text{ such that } x_{n_j} \leqq y \text{ for all } j \in \mathbb{N}\},\\
\mathscr{L}_3 &:= \{y \in \mathbb{R} : x_n \leqq y \text{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{5.19}$$

Then $\mathscr{L} = \mathscr{L}_1 = \mathscr{L}_2 = \mathscr{L}_3$.

B. There exists $z = z(\mathscr{L}) \in \mathbb{R}$ such that $\mathscr{L}$ is either an interval $[z, \infty[$ or an interval $]z, \infty[$.

C. Let $l := \inf(\mathscr{L})$. Then $l = z = z(\mathscr{L})$ as defined in part B. Further, l is the only real number such that

C1.

$$l + \varepsilon \in \mathscr{L} \quad \text{and} \quad l - \varepsilon \notin \mathscr{L} \tag{5.20}$$

C2. *There exists a subsequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $l = \lim\limits_{j\to\infty} x_{n_j}$ and l is the smallest real number for which such a subsequence exists.*

Proof: Let $y_n = -x_n$ and apply prop.5.3. ■

Proposition 5.5. *Let (x_n) be a bounded sequence of real numbers. As in prop. 5.3 and prop 5.4, let*

$$
\begin{aligned}
u &= \sup(\mathscr{U}) = \sup\{y\in\mathbb{R} : T_n\cap[y,\infty[\neq\emptyset \text{ for all } n\in\mathbb{N}\},\\
l &= \inf(\mathscr{L}) = \inf\{y\in\mathbb{R} : T_n\cap]-\infty,y]\neq\emptyset \text{ for all } n\in\mathbb{N}\},
\end{aligned}
\tag{5.21}
$$

Then

$$u = \limsup_{n\to\infty} x_j \quad \text{and} \quad l = \liminf_{n\to\infty} x_j.$$

Proof that $u = \limsup\limits_{n\to\infty} x_j$: Let

$$\beta_n := \sup_{j\geqq n} x_j, \quad \beta := \inf_n \beta_n = \limsup_{n\to\infty} x_n. \tag{5.22}$$

We shall prove that β has the properties listed in prop.5.3.C that uniquely characterize u: For any $\varepsilon > 0$, we have

$$\beta-\varepsilon\in\mathscr{U} \quad \text{and} \quad \beta+\varepsilon\notin\mathscr{U}$$

An other way of saying this is that

$$b\in\mathscr{U} \text{ for } b<\beta \quad \text{and} \quad a\notin\mathscr{U} \text{ for } a>\beta. \tag{5.23}$$

We now shall prove the latter characterization. Let $a\in\mathbb{R}, a>\beta=\inf\{\beta_n : n\in\mathbb{N}\}$. Then a is not a lower bound of the β_n: $\beta_{n_0} < a$ for some $n_0\in\mathbb{N}$. As the β_n are not increasing in n, this implies strict inequality $\beta_j < a$ for all $j\geqq n_0$. By definition, β_j is the least upper bound (hence an upper bound) of the tail set T_j. We conclude that $x_j < a$ for all $j\geqq n_0$. From that we conclude that $T_n\cap[a,\infty]=\emptyset$ for all $j\geqq n_0$. It follows that $a\notin\mathscr{U}$.

Now let $b\in\mathbb{R}, b<\beta=g.l.b\{\beta_n : n\in\mathbb{N}\}$. As $\beta\leqq\beta_n$ we obtain $b<\beta_n$ for all n. In other words, $b<\sup(T_n)$ for all n: It is possible to pick some $x_k\in T_n$ such that $b<x_k$. But then $T_n\cap[b,\infty]\neq\emptyset$ for all n and we conclude that $b\in\mathscr{U}$.

We put everything together and see that β has the properties listed in (5.23). This finishes the proof that $u=\limsup\limits_{n\to\infty} x_j$. The proof that $l=\liminf\limits_{n\to\infty} x_j$ follows again by applying what has already been proved to the sequence $(-x_n)$. ■

We have collected everything to prove

Theorem 5.1 (Characterization of limsup and liminf)**.** *Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}$. Then*

A1. $\limsup\limits_{n\to\infty} x_n$ is the largest of all real numbers x for which a sequence $n_1 < n_2 < \cdots \in\mathbb{N}$ can be found such that $x = \lim\limits_{j\to\infty} x_{n_j}$.

A2. $\limsup\limits_{n\to\infty} x_n$ *is the only real number* u *such that, for all* $\varepsilon > 0$*, the following is true:*
$x_n > u + \varepsilon$ *for at most finitely many* n *and* $x_n > u - \varepsilon$ *for infinitely many* n*.*

B1. $\liminf\limits_{n\to\infty} x_n$ *is the smallest of all real numbers* x *for which a sequence* $n_1 < n_2 < \dots \in \mathbb{N}$ *can be found such that* $x = \lim\limits_{j\to\infty} x_{n_j}$*.*

B2. $\liminf\limits_{n\to\infty} x_n$ *is the only real number* l *such that, for all* $\varepsilon > 0$*, the following is true:*
$x_n < l - \varepsilon$ *for at most finitely many* n *and* $x_n < l + \varepsilon$ *for infinitely many* n*.*

Proof: We know from prop.5.5 on p.41 that $\limsup\limits_{n\to\infty} x_n$ *is the unique number* u *described in part C of prop.5.3, p.38:* $u - \varepsilon \in \mathscr{U}$ *and* $u + \varepsilon \notin \mathscr{U}$ *for all* $\varepsilon > 0$ *and* u *is the largest real number for which there exists a subsequence* $(n_j)_{j\in\mathbb{N}}$ *of integers such that* $u = \lim\limits_{j\to\infty} x_{n_j}$*.*

$u - \varepsilon \in \mathscr{U} = \mathscr{U}_3$ *(see part A of prop.5.5) means that there are infinitely many* n *such that* $x_n \geqq u - \varepsilon$ *and* $u + \varepsilon \notin \mathscr{U} = \mathscr{U}_3$ *means that there are at most finitely many* n *such that* $x_n \geqq u + \varepsilon$*. This proves A1 and A2.*

We also know from prop.5.5 that $\liminf\limits_{n\to\infty} x_n$ *is the unique number* l *described in part C of prop.5.4, p.40:* $l + \varepsilon \in \mathscr{L}$ *and* $l - \varepsilon \notin \mathscr{L}$ *for all* $\varepsilon > 0$ *and* l *is the smallest real number for which there exists a subsequence* $(n_j)_{j\in\mathbb{N}}$ *of integers such that* $u = \lim\limits_{j\to\infty} x_{n_j}$*.*

$l + \varepsilon \in \mathscr{L} = \mathscr{L}_3$ *(see part A of prop.5.5) means that there are infinitely many* n *such that* $x_n \leqq l + \varepsilon$ *and* $l - \varepsilon \notin \mathscr{L} = \mathscr{L}_3$ *means that there are at most finitely many* n *such that* $x_n \leqq l - \varepsilon$*. This proves B1 and B2.*
■

Theorem 5.2 (Characterization of limits via limsup and liminf)**.** *Let* $(x_n)_{n\in\mathbb{N}}$ *be a bounded sequence in* $\mathbb{R}$*. Then* (x_n) *converges to a real number if and only if liminf and limsup for that sequence coincide and we have*

$$\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n. \tag{5.24}$$

Proof of "⇒": Let $L := \lim\limits_{n\to\infty} x_n$*. Let* $\varepsilon > 0$*. There is* $N = N(\varepsilon) \in \mathbb{N}$ *such that* $T_k \subseteq \,]\, L - \varepsilon, L + \varepsilon \,[$ *for all* $k \geqq N$*. But then*

$$L - \varepsilon \leqq \alpha_k := \inf(T_k) \leq \beta_k := \sup(T_k) \leqq L + \varepsilon \text{ for all } k \geqq N.$$

It follows from $T_j \subseteq T_k$ *for all* $j \geqq k$ *that*

$$\begin{aligned} L - \varepsilon &\leq \alpha_k \leqq \alpha_j \leqq \beta_j \leqq \beta_k \leqq L + \varepsilon, \quad \text{hence} \\ L - \varepsilon &\leq \lim_{k\to\infty} \alpha_k = \liminf_{k\to\infty} x_k \leqq \limsup_{k\to\infty} x_k = \lim_{k\to\infty} \beta_k \leqq L + \varepsilon. \end{aligned}$$

The equalities above result from prop.5.5. We have shown that, for any $\varepsilon > 0$*,* $\liminf\limits_{k\to\infty} x_k$ *and* $\limsup\limits_{k\to\infty} x_k$ *differ by at most* 2ε*, hence they are equal.*

Proof of "⇐": Let $L := \liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$*. Let* $\varepsilon > 0$*. We know from (5.17), p.39 and (5.20), p.40 that* $L + \varepsilon/2 \notin \mathscr{U}$ *and* $L - \varepsilon/2 \notin \mathscr{L}$ *But then there are at most finitely many* n *for which* x_n *has a distance from* L *which exceeds* $\varepsilon/2$*. Let* N *be the maximum of those* n*. It follows that* $|x_n - L| < \varepsilon$ *for all* $n > N$*, hence* $L = \lim\limits_{n\to\infty} x_n$*.* ■

6 Vectors and vector spaces (Understand this!)

6.1 $\mathbb{R}^N$: Euclidian space

6.1.1 N–dimensional Vectors

*This following definition of a vector is much more specialized than what is usually understood amongst mathematicians. For them, a vector is an element of a **"vector space"** . You can find later in the document the definition of a vector space ((6.4) on p.49) What you see here is a definition of vectors of "finite dimension".*

Definition 6.1 (N–dimensional vectors)**.** A **vector** is a finite, ordered collection $\vec{v} = (x_1, x_2, x_3, \dots, x_N)$ of real numbers $x_1, x_2, x_3, \dots, x_N$. "Ordered" means that it matters which number comes first, second third, ... If the vector has N elements then we say that it is N**–dimensional** . The set of all N–dimensional vectors is written as $\mathbb{R}^N$.

You are encouraged to go back to the section on cartesian products (3.16 on p.17) to review what was said there about $\mathbb{R}^N = \underbrace{\mathbb{R} \times \mathbb{R} \times + \cdots \times \mathbb{R}}_{N times}$. *Here are some examples of vectors:*

Example 6.1 (Two–dimensional vectors)**.** The two–dimensional vector with coordinates $x = -1.5$ and $y = \sqrt{2}$ is written $(-1.5, \sqrt{2})$ and we have $(-1.5, \sqrt{2}) \in \mathbb{R}^2$. Order matters, so this vector is different from $(\sqrt{2}, -1.5) \in \mathbb{R}^2$.

Example 6.2 (Three–dimensional vectors)**.** The three–dimensional vector $\vec{v_t} = (3 - t, 15, \sqrt{5t^2 + \frac{22}{7}}) \in \mathbb{R}^3$ with coordinates $x = 3 - t$, $y = 15$ and $z = \sqrt{5t^3 + \frac{22}{7}}$ is an example of a parametrized vector (parametrized by t). To be picky, Each specific value of t defines an element of $\in \mathbb{R}^3$, e.g., $\vec{v}_{-2} = (5, 15, \sqrt{20 + \frac{22}{7}})$.
Can you see that

$$F(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R}^3 \qquad t \longmapsto F(t) = \vec{v_t}$$

defines a mapping from $\mathbb{R}$ into $\mathbb{R}^3$ in the sense of definition (3.4) on p.10? Each argument s has assigned to it one and only one argument $\vec{v_s} = (3 - s, 15, \sqrt{5s^2 + \frac{22}{7}}) \in \mathbb{R}^3$.
Or, is it rather that we have three functions

$$x(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R} \qquad t \longmapsto x(t) = 3 - t$$
$$y(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R} \qquad t \longmapsto y(t) = 15$$
$$z(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R} \qquad t \longmapsto x(t) = \sqrt{5t^2 + \frac{22}{7}}$$

and $t \to \vec{v_t} = (x(t), y(t), z(t))$ is a vector of three real valued functions $x(\cdot), y(\cdot), z(\cdot)$?

Both points of view are correct and it depends on the specific circumstances how you want to interpret $\vec{v_t}$

Example 6.3 (One–dimensional vectors)**.** Let us not forget about the one–dimensional case: A one-dimensional vector has a single coordinate. $\vec{w_1} = (-3) \in \mathbb{R}^1$ with coordinate $x = -3 \in \mathbb{R}$ and

$\vec{w_2} = (5.7a) \in \mathbb{R}^1$ with coordinate $x = 5.7a \in \mathbb{R}$ are one–dimensional vectors. $\vec{w_2}$ is not a fixed number but parametrized by a.

Mathematicians do not distinguish between the one–dimensional vector (x) and its coordinate value, the real number x. For brevity, they will simply write $\vec{w_1} = -3$ and $\vec{w_2} = 5.7a$.

Example 6.4 (Vectors as functions)**.** An N–dimensional vector $\vec{x} = (x_1, x_2, x_3, \cdots, x_N)$ can be interpreted as a real function (remember: a real function is one which maps it arguments into $\mathbb{R}$)

$$\begin{aligned} f_{\vec{x}}(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R} \qquad m \mapsto x_m \\ f_{\vec{x}}(1) = x_1,\ f_{\vec{x}}(2) = x_2,\ \cdots,\ f_{\vec{x}}(N) = x_N, \end{aligned} \tag{6.1}$$

i.e., as a real function whose domain is the natural numbers $1, 2, 3, \cdots, N$. This goes also the other way around: given a real function $f(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R}$ we can associate with it the vector

$$\begin{aligned} \vec{v}_{f(\cdot)} := (f(1), f(2), f(3), \cdots, f(N)) \\ \vec{v}_{f_1} = f(1),\ \vec{v}_{f_2} = f(2),\ ,\cdots, \vec{v}_{f_N} = f(N) \end{aligned} \tag{6.2}$$

6.1.2 Addition and scalar multiplication for N–dimensional vectors

Definition 6.2 (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 α. We define the **sum** $\vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ as the vector $\vec{z}$ with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_N = x_N + y_N; \tag{6.3}$$

We define the **scalar product** $\alpha\vec{x}$ of α and $\vec{x}$ as the vector $\vec{w}$ with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_N = \alpha x_N; \tag{6.4}$$

The following picture describes vector addition:

Adding two vectors $\vec{v}$ and $\vec{w}$ means that you take one of them, say $\vec{v}$, and shift it in parallel (without rotating it in any way or flipping its direction), so that its starting point moves from the origin to the endpoint of the other vector $\vec{w}$. Look at the picture and you see that the vectors $\vec{v}$, $\vec{w}$ and $\vec{v}$ shifted form three pages of a parallelogram. $\vec{v} + \vec{w}$ is then the diagonal of this parallelogram which starts at the origin and ends at the endpoint of $\vec{v}$ shifted.

6.1.3 Length of N–dimensional vectors, the Euclidean Norm

*It is customary to write $\|\vec{v}\|$ for the length, sometimes also called the **norm** of the vector $\vec{v}$.*

Length of one–dimensional vectors: *For a vector $\vec{v} = x \in \mathbb{R}$ its length is its absolute value $\|\vec{v}\| = |x|$. This means that $\|-3.57\| = |-3.57| = 3.57$ and $\|\sqrt{2}\| = |\sqrt{2}| \approx 1.414$.*

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*

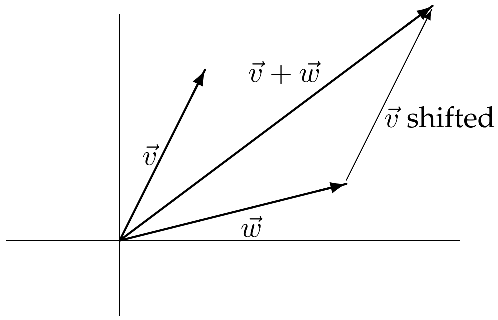

Figure 1: Adding two vectors.

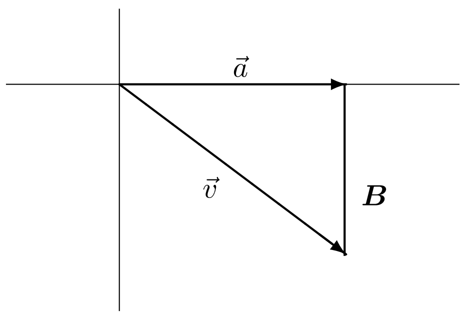

Figure 2: Length of a 2–dimensional vectors.

arrow which starts at the origin and ends at the point with coordinates $x = 4$ *and* $y = -3$*. How long is that arrow?*

Think of it as the hypothenuse of a right angle triangle whose two other sides are the horizontal arrow from $(0, 0)$ *to* $(4, 0)$ *(the vector* $\vec{a} = (4, 0)$ *) and the vertical line* $\boldsymbol{B}$ *between* $(4, 0)$ *and* $(4, -3)$*. Note that* $\boldsymbol{B}$ *is not a vector because it does not start at the origin! Obviously (I hope it's obvious) we have* $\|\vec{a}\| = 4$ *and length–of(*$\boldsymbol{B}$*)* $= 3$*. Pythagoras tells us that*

$$\|\vec{v}\|^2 = \|\vec{a}\|^2 + \textit{length–of}(\boldsymbol{B})^2$$

and we obtain for (4, -3): $\|\vec{v}\| = \sqrt{16 + 9} = 5$.

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* $\|(x, y)\|^2 = x^2 + y^2$ *which* *becomes, after taking square roots on both sides,*

$$\|(x,y)\| = \sqrt{x^2+y^2} \tag{6.5}$$

Length of three–dimensional vectors: *This is not so different from the two–dimensional case above. We build on the previous example. Let $\vec{v} = (4,-3,12)$. Think of an xyz-coordinate system with origin (the spot where x-axis, y-axis and z-axis intersect) $(0,0,0)$. Then $\vec{v}$ is represented by an arrow which starts at the origin and ends at the point with coordinates $x = 4$, $y = -3$ and $z = 12$. How long is that arrow?*

Remember what the standard 3–dimensional coordinate system looks like: The x-axis goes from west to east, the y-axis goes from south to north and the z-axis goes vertically from down below to the sky. Now drop a vertical line $\boldsymbol{B}$ from the point with coordinates $(4,-3,12)$ to the xy–plane which is "spanned" by the x-axis and y-axis. This line will intersect the xy–plane at the point with coordinates $x = 4$ and $y = -3$ (and $z = 0$. Why?) Note that $\boldsymbol{B}$ is not a vector because it does not start at the origin! It should be clear that length–of$(\boldsymbol{B}) = |z| = 12$. Now we connect the origin $(0,0,0)$ with the point $(4,-3,0)$ in the xy–plane which is the endpoint of $\boldsymbol{B}$.

We can forget about the z–dimension because this arrow is entirely contained in the xy–plane. Matter of fact, it is a genuine two–dimensional vector $\vec{a} = (4,-3)$ because it starts in the origin. Observe that $\vec{a}$ has the same values 4 and -3 for its x– and y–coordinates as the original vector $\vec{v}$. [11] *We know from the previous example about two–dimensional vectors that*

$$\|\vec{a}\|^2 = \|(x,y)\|^2 = x^2+y^2 = 16+9 = 25.$$

At this point we have constructed a right angle triangle with a) hypothenuse $\vec{v} = (x,y,z)$ where we have $x = 4$, $y = -3$ and $z = 12$, b) a vertical leg with length $|z| = 12$ and c) a horizontal leg with length $\sqrt{x^2+y^2} = 5$. Pythagoras tells us that

$$\|\vec{v}\|^2 = z^2 + \|(x,y)\|^2 = 144+25 = 169 \quad \textit{or} \quad \|\vec{v}\| = 13.$$

None of what we just did depended on the specific values 4, -3 and 12. Any vector $(x,y,z) \in \mathbb{R}^3$ is the hypothenuse of a right triangle where the square lengths of the legs are z^2 and x^2+y^2. This means we have proved the general formula $\|(x,y,z)\|^2 = x^2+y^2+z^2$ or

$$\|(x,y,z)\| = \sqrt{x^2+y^2+z^2} \tag{6.6}$$

The previous examples provide the motivation for the following definition:

Definition 6.3 (Euclidean norm)**.** Let $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ be an n–dimension vector. The **Euclidean norm** $\|\vec{v}\|$ of $\vec{v}$ is defined as follows:

$$\|\vec{v}\| \quad = \quad \sqrt{{x_1}^2+{x_2}^2+\ldots+{x_n}^2} \quad = \quad \sqrt{\sum_{j=1}^{n} {x_j}^2} \tag{6.7}$$

[11] You will learn in the chapter on vector spaces that the vector $\vec{a} = (4,-3)$ is the projection on the xy–coordinates $\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2 \quad (x,y,z) \mapsto (x,y)$ of the vector $\vec{v} = (4,-3,12)$. (see Example C(6.16) on p.54)

This definition is important enough to write the special cases for $n = 1, 2, 3$ where $\|\vec{v}\|$ coincides with the length of $\vec{v}$:

$$
\begin{aligned}
(6.8) \qquad & 1-dim: \quad \|(x)\| = \sqrt{x^2} = |x| \\
& 2-dim: \quad \|(x,y)\| = \sqrt{x^2+y^2} \\
& 3-dim: \quad \|(x,y,z)\| = \sqrt{x^2+y^2+z^2}
\end{aligned}
$$

Proposition 6.1 (Properties of the Euclidian norm)**.** *Let $n \in \mathbb{N}$. Then the Euclidean norm, viewed as a function*

$$\|\cdot\|: \ \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \vec{v} = (x_1, x_2, \dots, x_n) \longmapsto \|\vec{v}\| = \sqrt{\sum_{j=1}^{n} {x_j}^2}$$

has the following three properties:

(6.9a) $\quad \|\vec{v}\| \geqq 0 \quad \forall \vec{v} \in \mathbb{R}^n$ *and* $\|\vec{v}\| = 0 \iff \vec{v} = 0$ ***positive definite***

(6.9b) $\quad \|\alpha\vec{v}\| = |\alpha| \cdot \|\vec{v}\| \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}$ ***homogeneity***

(6.9c) $\quad \|\vec{v} + \vec{w}\| \leqq \|\vec{v}\| + \|\vec{w}\| \quad \forall\, \vec{v}, \vec{w} \in \mathbb{R}^n$ ***triangle inequality***

Proof:
a. It is certainly true that $\|\vec{v}\| \geqq 0$ for any n–dimensional vector $\vec{v}$ because it is defined as $+\sqrt{K}$ where the quantity K is, as a sum of squares, non–negative. If 0 is the zero vector with coordinates $x_1 = x_2 = \ldots = x_n = 0$ then obviously $\|0\| = \sqrt{0 + \ldots + 0} = 0$. Conversely, let $\vec{v} = (x_1, x_2, \dots, x_n)$ be a vector in $\mathbb{R}^n$ such that $\|\vec{v}\| = 0$. This means that $\sqrt{\sum_{j=1}^{n} {x_j}^2} = 0$ which is only possible if everyone of the non–negative x_j is zero. In other words, $\vec{v}$ must be the zero vector 0.

b. Let $\vec{v} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then

$$
\begin{aligned}
\|\alpha\vec{v}\| &= \sqrt{\sum_{j=1}^{n} (\alpha x_j)^2} = \sqrt{\sum_{j=1}^{n} \alpha^2 \alpha {x_j}^2} = \sqrt{\alpha^2 \sum_{j=1}^{n} \alpha {x_j}^2} = \sqrt{\alpha^2}\sqrt{\sum_{j=1}^{n} \alpha {x_j}^2} \\
&= \sqrt{\alpha^2}\|\vec{v}\| = |\alpha| \cdot \|\vec{v}\|
\end{aligned}
$$

because it is true that $\sqrt{\alpha^2} = |\alpha|$ for any real number α (see assumption 3.1 on p.7).

c. The proof will only be given for $n = 1, 2, 3$.
***n* = 1** *: Property (6.9.c) simply reduces to the triangle inequality for real numbers (see 3.1 on 7) and we are done.*
***n* = 2, 3** *: Look back at the picture about addition of vectors in the plane or in space (see p.45). Remember that for any two vectors $\vec{v}$ and $\vec{w}$ you can always build a triangle whose sides have length $\|\vec{v}\|$, $\|\vec{w}\|$ and $\|v \vec{+} w\|$. 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 $\|v \vec{+} w\| \leqq \|\vec{v}\| + \|\vec{w}\|$ and we are done with the following limitation:*

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 (6.9.c). An exact proof will be given as a consequence of the so–

called Cauchy–Schwartz inequality which you will find further down (theorem (6.1) on p.57) in the section which discusses inner products on vector spaces. ■

6.2 General vector spaces

6.2.1 Vector spaces: Definition and examples

Mathematicians are very fond of looking at very different objects and figuring out what they have in common. They then create an abstract concept whose items have those properties and examine what they can conclude. For those of you who have had some exposure to object oriented programming: It's like defining a base class, e.g., "mammal", that possesses the core properties of several concrete items such as "horse", "pig", "whale" (sorry – can't require that all mammals have legs). We have looked at the following items that seem to be quite different:

real numbers
N–dimensional vectors
real functions

Well, that was sort of disingenuous. I took great pains to explain that real numbers and one–dimensional vectors are sort of the same (see 6.3 on p.43). Besides I also explained that N–dimensional vectors can be thought of as real functions on a special domain X, namely $1, 2, 3, \cdots, N$. (see 6.4 on p.44). 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 (6.1.2) on p.44). Given are two N–dimensional vectors $\vec{x} = (x_1, x_2, \dots, x_N)$ and $\vec{y} = (y_1, y_2, \dots, y_N)$ and a real number α. Then the sum $\vec{z} = \vec{x} + \vec{y}$ of $\vec{x}$ and $\vec{y}$ is the vector with the components

$$z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_N = x_N + y_N;$$

and the scalar product $\vec{w} = \alpha\vec{x}$ of α and $\vec{x}$ is the vector with the components

$$w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_N = \alpha x_N;$$

Example 6.5 (Vector addition and scalar multiplication)**.** We use $N = 2$ in this example:
Let $a = (-3, 1/5)$, $b = (5, \sqrt{2})$ We add those vectors by adding each of the coordinates separately:

$$a + b = (2, 1/5 + \sqrt{2})$$

and we multiply a with a scalar $\lambda \in \mathbb{R}$, e.g. $\lambda = 100$, by multiplying each coordinate with λ:

$$100a = (-300, 20).$$

In the last example I have avoided using the notation "$\vec{x}$" with the cute little arrows on top for vectors. I did that on purpose because this notation is not all that popular in Math even for N–dimensional vectors and definitely not for the more abstract vectors as elements of a vector space. Here now is the definition of a vector space, taken almost word for word from the book "Introductory Real Analysis" (Kolmogorov/Fomin [4]). This definition is rather lengthy because a set needs to satisfy many rules to be a vector space.

Definition 6.4 (Vector spaces (linear spaces))**.** A non–empty set L of elements $x, y, z, \dots$ is called a **vector space** or **linear space** if it satisfies the following:

A. Any two elements $x, y \in L$ uniquely determine a third element $x + y \in L$, 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 $0 \in L$, called the **zero element**, or **zero vector**, or **null vector**, with the property that $x + 0 = x$ for each $x \in L$;
4. For every $x \in L$, there exists an element $-x$, called the **negative** of x, with the property that $x + (-x) = 0$ for each $x \in L$. When adding negatives, then there is a convenient short form. We write $x - y$ as an abbreviation for $x + (-y)$;

B. Any real number α and element $x \in L$ together uniquely determine an element $\alpha x \in L$ (sometimes also written $\alpha \cdot x$ for clarity), called the **scalar product** of α and x. It has the following properties:

1. $\alpha(\beta x) \ = \ (\alpha\beta)x$;
2. $1x \ = \ x$;

C. The operations of addition and scalar multiplication obey the two **distributive laws**

1. $(\alpha + \beta)x \ = \ \alpha x + \beta x$;
2. $\alpha(x + y) \ = \ \alpha x + \alpha y$;

The elements of a vector space are called **vectors**

Definition 6.5 (Subspaces of vector spaces)**.** Let L be a vector space and let $A \subseteq L$ be a non–empty subset of L with the following property: For any $x, y \in A$ and $\alpha \in \mathbb{R}$ the sum $x + y$ and the scalar product αx also belong to A. Note that if $\alpha = 0$ then $\alpha x = 0$ and it follows that the null–vector belongs to A.

A is called a **subspace** of L.

We ruled out the case $A = \emptyset$ but did not ask that A be a strict subset of L ((3.10) on p.14). In other words, L is a subspace of itself.

The set $\{0\}$ which contains the null–vector 0 of L as its single element also is a subspace, the so called **nullspace**

Proposition 6.2 (Subspaces are vector spaces)**.** *A subspace of a vector space is a vector space, i.e., it satisfies all requirements of definition (6.4).*

Proof: None of the equalities that are part of the definition of a vector space magically ceases to be valid just because we look at a subset. The only thing that could go wrong is that some of the expressions might not belong to A anymore. I'll leave it to you to figure out why this won't be the case, but I'll show you the proof for the second distributive law of part C.

We must prove that for any $x, y \in A$ *and* $\lambda \in \mathbb{R}$

$$\lambda(x+y) \;=\; \lambda x + \lambda y :$$

First, $x + y \in A$ *because a subspace contains the sum of any two of its elements. It follows that* $\lambda(x+y)$ *as product of a real number with an element of* A *again belongs to* A *because it is a subspace. Hence the left hand side of the equation belongs to* A.

Second, both λx *and* λy *belong to* A *because each is the scalar product of* λ *with an element of* A *and this set is a subspace. Hence the right hand side of the equation belongs to* A.

Equality of $\lambda(x+y)$ *and* $\lambda x + \lambda y$ *is true because it is true if we look at* x *and* y *as elements of* L. ■

Remark 6.1 (Closure properties)**.** If a subset B of a larger set X has the property that certain operations on members of B will always yield elements of B, then we say that B is **closed** with respect to those operations.

We can now express the definition of a linear subspace as follows:

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 6.6 (A: vector space $\mathbb{R}$)**.** The real numbers $\mathbb{R}$ are a vector space if you take the ordinary addition of numbers as "+" and the ordinary multiplication of numbers as scalar multiplication.

Example 6.7 (B: vector space $\mathbb{R}^n$)**.** More general, the sets $\mathbb{R}^n$ of n–dimensional vectors are vector spaces when you define addition and scalar multiplication as in (6.2) on p.44. To see why, just look at each component (coordinate) separately and you just deal with ordinary real numbers.

Example 6.8 (C: vector space of real functions)**.** The set

$$\mathscr{F}(X, \mathbb{R}) = \{f(\cdot) : f(\cdot) \text{ is a real function on } X\}$$

of all real functions on an arbitrary non–empty set X is a vector space if you define addition and scalar multiplication as in (5.2) on p.33. The reason is that you can verify the properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ and again, you just deal with ordinary real numbers. The "sup–norm"

$$\|f(\cdot)\| \;=\; \sup\{|f(x)| : x \in X\}$$

is **not a real function** on all of $\mathscr{F}(X, \mathbb{R})$ because $\|f(\cdot)\| = +\infty$ for any unbounded $f(\cdot) \in \mathscr{F}(X, \mathbb{R})$.

The subset

$$\mathscr{B}(X, \mathbb{R}) = \{h(\cdot) : h(\cdot) \text{ is a bounded real function on } X\}$$

(see (7.1) on p. 63) is a subspace of the vector space of all real functions on X. On this subspace the sup–norm truly is a real function in the sense that $\|f(\cdot)\| < \infty$.

And here are some more examples:

Example 6.9 (D: subspace $\{(x,y) : x = y\}$). The set $L := \{(x,x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ of all vectors in the plane with equal x and y coordinates has the following property: For any two vectors $\vec{x} = (a,a)$ and $\vec{y} = (b,b) \in L$ ($a, b \in \mathbb{R}$) and real number α the sum $\vec{x} + \vec{y} = (a+b, a+b)$ and the scalar product $\alpha\vec{x} = (\alpha a, \alpha a)$ have equal x–and y–coordinates, i.e., they again belong to L. Moreover the zero–vector 0 with coordinates $(0,0)$ belongs to L. It follows that the subset L of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (see (6.5) on p.49).

I won't show the following even though it is not hard:

Example 6.10 (E: subspace $\{(x,y) : y = \alpha x\}$). Any subset of the form

$$L_\alpha := \{(x,y) \in \mathbb{R}^2 : y = \alpha x\}$$

is a subspace of $\mathbb{R}^2$ ($\alpha \in \mathbb{R}$). Draw a picture: L_α is the straight line through the origin in the xy–plane with slope α.

Example 6.11 (F: Embedding of linear subspaces). The last example was about the subspace of a bigger space. Now we switch to the opposite concept, the **embedding** of a smaller space into a bigger space. We can think of the real numbers $\mathbb{R}$ as a part of the xy–plane $\mathbb{R}^2$ or even 3–dimensional space $\mathbb{R}^3$ by identifying a number a with the two-dimensional vector $(a,0)$ or the three-dimensional vector $(a,0,0)$. Let $M < N$. It is not a big step from here that the most natural way to uniquely associate an N–dimensional vector with an M–dimensional vector $\vec{x} := (x_1, x_2, \dots, x_M)$ by adding zero-coordinates to the right:

$$\vec{x} := (x_1, x_2, \dots, x_M, \underbrace{0, 0, \dots, 0}_{N-M\text{times}})$$

Example 6.12 (G: All finite–dimensional vectors). Let

$$\mathfrak{S} \;:=\; \bigcup_{n \in \mathbb{N}} \mathbb{R}^n \;=\; \mathbb{R}^1 \cup \mathbb{R}^2 \cup \ldots \cap \mathbb{R}^n \cup \ldots$$

be the set of all vectors of finite (but unspecified) dimension.

We can define addition for any two elements $\vec{x}, \vec{y} \in \mathfrak{S}$ as follows: If $\vec{x}$ and $\vec{y}$ both happen to have the same dimension N then we add them as usual: the sum will be $x_1 + y_1, x_2 + y_2, \dots, x_N + y_N,$. If not, then one of them, say $\vec{x}$ will have dimension M smaller than the dimension N of $\vec{y}$. We now define the sum $\vec{x} + \vec{y}$ as the vector

$$\vec{z} := (x_1 + y_1, x_2 + y_2, \dots, x_M + y_M, y_{M+1}, y_{M+2}, \dots, y_N)$$

which is hopefully what you expected me to do.

Example 6.13 (H: All sequences of real numbers). Let $\mathbb{R}^{\mathbb{N}} = \prod_{j \in \mathbb{N}} \mathbb{R}$ (see (3.17) on p.18). Is this the same set as $\mathfrak{S}$ from the previous example? The answer is No. Can you see why? I would be surprised if you do, so let me give you the answer: Each element $x \in \mathfrak{S}$ is of some finite dimension, say N, meaning that that it has no more than N coordinates. Each element $y \in \mathbb{R}^{\mathbb{N}}$ is a collection of numbers $y_1, y_2, \dots$ none of which need to be zero. In fact, $\mathbb{R}^{\mathbb{N}}$ is the vector space of all sequences of real numbers. Addition is of course done coordinate by coordinate and scalar multiplication with $\alpha \in \mathbb{R}$ is done by multiplying each coordinate with α.

There is again a natural way to embed $\mathfrak{S}$ into $\mathbb{R}^{\mathbb{N}}$ as follows: We transform an N–dimensional vector $(a_1, a_2, \dots, a_N)$ into an element of $\mathbb{R}^{\mathbb{N}}$ (a sequence $(a_j)_{j \in \mathbb{N}}$) by setting $a_j = 0$ for $j > N$.

Definition 6.6 (linear combinations)**.** Let L be a vector space and let $x_1, x_2, x_3, \dots, x_n \in L$ be a finite number of vectors in L. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. We call the finite sum

$$\sum_{j=0}^{n} \alpha_j x_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_n x_n \tag{6.10}$$

a **linear combination** of the vectors x_j . The multipliers $\alpha_1, \alpha_2, \dots$ are called **scalars** in this context.

In other words, linear combinations are sums of scalar products. You should understand that the expression in (6.10) always is an element of L, no matter how big $n \in \mathbb{N}$ was chosen:

Proposition 6.3 (Vector spaces are closed w.r.t. linear combinations)**.** *Let L be a vector space and let $x_1, x_2, x_3, \dots, x_n \in L$ be a finite number of vectors in L. Let $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in \mathbb{R}$. Then the linear combination $\sum\limits_{j=0}^{n} \alpha_j x_j$ also belongs to L. Note that this is also true for subspaces because those are vector spaces, too.*

Proof: This is another example of a proof by complete induction (see def. 3.2, 8). Each scalar product $\alpha_j x_j$ is an element of L because part B of the definition of a vector space demands it. The sum of two such expressions belongs to L because part A demands it. Then (6.10) must be true for $n = 3$ because, if we set $z := \alpha_1 x_1 + \alpha_2 x_2$, then $\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = z + \alpha_3 x_3$ can be written as the sum of two elements of L and therefore belongs to L. But then $\sum\limits_{j=0}^{4} \alpha_j x_j = \sum\limits_{j=0}^{3} \alpha_j x_j + \alpha_4 x_4$ can be written as the sum of two elements of L (we just saw that $\sum\limits_{j=0}^{3} \alpha_j x_j$ as the sum of three elements of L belongs to L) and therefore belongs to L.

We keep going with $n = 5, 6, 7, \dots$ (an exact proof needs induction) and conclude that $\sum\limits_{j=0}^{n} \alpha_j x_j = \sum\limits_{j=0}^{n-1} \alpha_j x_j +$ $\alpha_n x_n$ can be written as the sum of two elements of L (we just saw that $\sum\limits_{j=0}^{n-1} \alpha_j x_j$ as the sum of $n-1$ elements of L belongs to L) and therefore belongs to $L \dots$. ■

Definition 6.7 (linear mappings)**.** Let L_1, L_2 be two vector spaces.

Let $f(\cdot) : L_1 \to L_2$ be a mapping with the following properties:

$$f(x+y) = f(x) + f(y) \quad \forall x, y \in L_1 \qquad \textbf{additivity} \tag{6.11a}$$

$$f(\alpha x) = \alpha f(x) \quad \forall x \in L_1,\ \forall \alpha \in \mathbb{R} \qquad \textbf{homogeneity} \tag{6.11b}$$

Then we call $f(\cdot)$ a **linear mapping**.

Note 6.1 (Note on homogeneity)**.** We encountered homogeneity when looking at the properties of the Euclidian norm ((6.9) on p.47), but homogeneity is defined differently there in that you had to take the absolute value $|\alpha|$ instead of α.

Remark 6.2 (Linear mappings are compatible with linear combinations)**.** We saw in the last proposition that vector spaces are closed with respect to linear combinations. Linear mappings and linear combinations go together very well in the following sense:

Remember that for any kind of mapping $x \mapsto f(x)$, $f(x)$ was called the image of x. Now we can express what linear mappings are about like this:

> A: The image of the sum is the sum of the image
>
> B: The image of the scalar product is the scalar product of the image
>
> C: The image of the linear combination is the linear combination of the image

Mathematicians express this by saying that linear mappings **preserve** or are **compatible with** linear combinations.

Proposition 6.4 (Linear mappings preserve linear combinations)**.** *Let L_1, L_2 be two vector spaces.*

Let $f(\cdot) : L_1 \to L_2$ be a linear map and let $x_1, x_2, x_3, \dots, x_n \in L_1$ be a finite number of vectors in the domain L_1 of $f(\cdot)$. Let $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n \in \mathbb{R}$. Then $f(\cdot)$ preserves any such linear combination:

$$f(\sum_{j=0}^{n} \lambda_j x_j) \;=\; \sum_{j=0}^{n} \lambda_j f(x_j). \tag{6.12}$$

Proof:
First we note that $f(\lambda_j x_j) = \lambda_j f(x_j)$ for all j because linear mappings preserve scalar products. Because they also preserve the addition of any two elements, the proposition holds for $n = 2$. We prove the general case by induction (see (3.2) on p.8). Our induction assumption is

$$f(\sum_{j=0}^{n-1} \lambda_j x_j) \;=\; \sum_{j=0}^{n-1} \lambda_j f(x_j).$$

We use it in the third equality here:

$$f(\sum_{j=0}^{n} \lambda_j x_j) \;=\; f(\sum_{j=0}^{n-1} \lambda_j x_j + \lambda_n x_n) = \; f(\sum_{j=0}^{n-1} \lambda_j x_j) + f(\lambda_n x_n) = \sum_{j=0}^{n-1} \lambda_j f(x_j) + f(\lambda_n x_n) = \sum_{j=0}^{n} \lambda_j f(x_j)$$

■

Here are some examples of linear mappings.

Example 6.14 (A: Projections to any subspace)**.** Let $N \in \mathbb{N}$. The map

$$\pi_1(\cdot) : \mathbb{R}^N \to \mathbb{R} \qquad (x_1, x_2, \dots, x_N) \mapsto x_1$$

is called the **projection** on the first coordinate or the first coordinate function.

Example 6.15 (B: Projections on any coordinate)**.** More generally, let $N \in \mathbb{N}$ and $1 \leqq j \leqq N$. The map

$$\pi_j(\cdot) : \mathbb{R}^N \to \mathbb{R} \qquad (x_1, x_2, \dots, x_N) \mapsto x_j$$

is called the **projection** on the jth coordinate or the j**th coordinate function**.

It is easy to see what that means if you set $N = 2$: For the two–dimensional vector $\vec{v} := (3.5, -2) \in \mathbb{R}^2$ you get $\pi_1(\vec{v}) = 3.5$ and $\pi_2(\vec{v}) = -2$.

Example 6.16 (C: Projections to any subspace)**.** In the last two examples we projected $\mathbb{R}^N$ onto a one–dimensional space. More generally, we can project $\mathbb{R}^N$ onto a vector space $\mathbb{R}^M$ of lower dimension M (i.e., we assume $M < N$) by keeping M of the coordinates and throwing away the remaining N_M. Mathematicians express this as follows:

Let $M, N, i_1, i_2, \dots, i_M \in \mathbb{N}$ such that $M < N$ and $1 \leqq i_1 < i_2 < \dots < i_M \leqq N$. The map

(6.13) $$\pi_{i_1, i_2, \dots, i_M}(\cdot) : \mathbb{R}^N \to \mathbb{R}^M \qquad (x_1, x_2, \dots, x_N) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_M})$$

is called the **projection** on the coordinates $i_1, i_2, \dots, i_M$. [12]

Example 6.17. Let $x_0 \in A$. The mapping

(6.14) $$\varepsilon_{x_0} : \mathscr{F}(A, \mathbb{R}) \to \mathbb{R}; \qquad f(\cdot) \mapsto f(x_0)$$

which assigns to any real function on A its value at the specific point x_0 is a linear mapping because if $h(\cdot) = \sum\limits_{j=0}^{n} a_j f_j(\cdot)$ then

$$\varepsilon_{x_0}(\sum_{j=0}^{n} a_j f_j(\cdot)) = \varepsilon_{x_0}(h(\cdot)) = h(x_0) = \sum_{j=0}^{n} a_j f_j(x_0) = \sum_{j=0}^{n} a_j \varepsilon_{x_0}(f_j(\cdot))$$

and this proves the linearity of the mapping $\varepsilon_{x_0}(\cdot)$. The mapping $\varepsilon_{x_0}(\cdot)$ is called the **Radon integral** at x_0.

6.2.2 Normed vector spaces (Skip this!)

The following definition of inner products and proof of the Cauchy–Schwartz inequality were taken from "Calculus of Vector Functions" (Williamson/Crowell/Trotter [11]).

Definition 6.8 (Inner products)**.** Let L be a vector space with a function

$$\bullet(\cdot, \cdot) : L \times L \to \mathbb{R}; \qquad (x, y) \mapsto x \bullet y := \bullet(x, y)$$

which satisfies the following properties:

(6.15a) $x \bullet x \geqq 0 \quad \forall x \in L$ and $x \bullet x = 0 \iff x = 0$ **positive definite**

(6.15b) $x \bullet y = y \bullet x \quad \forall x, y \in L$ **symmetry**

(6.15c) $(x + y) \bullet z = x \bullet z + y \bullet z \quad \forall\, x, y, z \in L$ **additivity**

(6.15d) $(\lambda x) \bullet y = \lambda(x \bullet y) \quad \forall\, x, y \in L \quad \forall\, \lambda \in \mathbb{R}$ **homogeneity**

We call such a function an **inner product**.

[12] You previously encountered an example where we made use of the projection

$$\pi_{1,2}(\cdot) : \mathbb{R}^3 \to \mathbb{R}^2 \qquad (x, y, z) \mapsto (x, y).$$

This was in the course of computing the length of a 3–dimensional vector (see (6.1.3) on p.44).

Note that additivity and homogeneity of the mapping $x \mapsto x \bullet y$ for a fixed $y \in L$ imply linearity of that mapping and the symmetry property implies that the mapping $y \mapsto x \bullet y$ for a fixed $x \in L$ is linear too. In other words, an inner product is binear in the following sense:

Definition 6.9 (Bilinearity)**.** Let L be a vector space with a function

$$F(\cdot,\cdot) : L \times L \to \mathbb{R}; \qquad (x,y) \mapsto F(x,y).$$

$F(\cdot,\cdot)$ is called **bilinear** if it is linear in each component, i.e., the mappings

$$\begin{aligned} F_1 :& L \to \mathbb{R}; \quad x \mapsto F(x,y) \\ F_2 :& L \to \mathbb{R}; \quad y \mapsto F(x,y) \end{aligned}$$

are both linear.

Proposition 6.5 (Algebraic properties of the inner product)**.** *Let L be a vector space with inner product $\bullet(\cdot,\cdot)$. Let $a,b,x,y \in L$. Then*

(6.16a) $$(a+b) \bullet (x+y) = a \bullet x \;+\; b \bullet x \;+\; a \bullet y \;+\; b \bullet y$$

(6.16b) $$(x+y) \bullet (x+y) = x \bullet x \;+\; 2(x \bullet y) \;+\; y \bullet y$$

(6.16c) $$(x-y) \bullet (x-y) = x \bullet x \;-\; 2(x \bullet y) \;+\; y \bullet y$$

*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:***

$$(x+y) \bullet (x+y) \;=\; x \bullet x \;+\; y \bullet x \;+\; x \bullet y \;+\; y \bullet y$$

according to part a. Symmetry gives us $y \bullet x = x \bullet y$ and part b follows.

*Proof of **c:** Replace y with $-y$ in part b. Bilinearity gives both*

$$x \bullet -y \;=\; -(x \bullet y); \qquad -y \bullet -y \;=\; (-1)^2 y \bullet y \;=\; y \bullet y$$

and this gives c. ■

The following is the most important example of an inner product.

Proposition 6.6 (Inner product on $\mathbb{R}^N$))**.** *Let $N \in \mathbb{N}$. Then the real function*

(6.17) $$(\vec{v},\vec{w}) \;\mapsto\; x_1y_1 + x_2y_2 + \ldots + x_Ny_N \;=\; \sum_{j=1}^{n} x_jy_j$$

is an inner product on $\mathbb{R}^N \times \mathbb{R}^N$.

Proof:
a : *For $\vec{v} = \vec{w}$ we obtain $\vec{v} \bullet \vec{v}) = \|\vec{v}\|$ and positive definiteness of the inner product follows from that of the Euclidean norm.*
b : *Symmetry is clear because $x_j y_j = y_j x_j$.*
c : *Additivity follows from the fact that $(x_j + y_j)z_j = x_j z_j + y_j z_j$.*
d : *Homogeneity follows from the fact that $(\lambda x_j)y_j = \lambda(x_j y_j)$.* $\blacksquare$

Proposition 6.7 (Cauchy–Schwartz inequality for inner products)**.** *Let L be a vector space with an inner product*

$$\bullet(\cdot,\cdot) : L \times L \to \mathbb{R}; \qquad (x,y) \mapsto x \bullet y := \bullet(x,y)$$

Then

$$(x \bullet y)^2 \leqq (x \bullet x)\,(y \bullet y).$$

Proof:
Step1 **:** *We assume first that $x \bullet x = y \bullet y = 1$. Then*

$$\begin{aligned} 0 &\leqq (x - y \bullet x - y) \\ &= x \bullet x - 2x \bullet y + y \bullet y = 2 - 2x \bullet y \end{aligned}$$

where the first equality follows from proposition (6.5) on p.55.
This means $2x \bullet y \leqq 2$, i.e., $x \bullet y \leqq 1 = (x \bullet x)\,(y \bullet y)$ where the last equality is true because we had assumed $x \bullet x = y \bullet y = 1$. The Cauchy–Schwartz inequality is thus true under that special assumption.

Step2 **:** *General case: We do not assume anymore that $x \bullet x = y \bullet y = 1$. If x or y is zero then the Cauchy–Schwartz inequality is trivially true because, say if $x = 0$ then the left hand side becomes*

$$(x \bullet y)^2 = (0x \bullet y)^2 = 0(x \bullet y)^2 = 0$$

whereas the right hand side is, as the product of two non–negative numbers $x \bullet x$ and $y \bullet y$, non–negative.

So we can assume that x and y are not zero. On account of the positive definiteness we have $x \bullet x > 0$ and $y \bullet y > 0$. This allows us to define $u := x/\sqrt{x \bullet x}$ and $v := y/\sqrt{y \bullet y}$. But then

$$\begin{aligned} u \bullet u &= (x \bullet x)/\sqrt{x \bullet x}^2 = 1 \\ v \bullet v &= (y \bullet y)/\sqrt{y \bullet y}^2 = 1. \end{aligned}$$

We have already seen in step 1 that $u \bullet v \leqq 1$. It follows that

$$(x \bullet y)/(\sqrt{x \bullet x}\sqrt{y \bullet y}) = (x/\sqrt{x \bullet x}) \bullet (y/\sqrt{y \bullet y}) \leqq 1$$

We multiply both sides with $\sqrt{x \bullet x}\sqrt{y \bullet y}$,

$$x \bullet y \leqq \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We replace x by $-x$ and obtain

$$-(x \bullet y) \leqq \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

Think for a moment about the meaning of the absolute value and it is clear that the last two inequalities together prove that

$$|x \bullet y| \leqq \sqrt{x \bullet x}\sqrt{y \bullet y}.$$

We square this and obtain

$$(x \bullet y)^2 \leqq (x \bullet x)\,(y \bullet y)$$

and the Cauchy–Schwartz inequality is proved. ■

Note 6.2. We previously discussed the sup–norm

$$\|f(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\} \tag{6.18}$$

for real functions on some non–empty set X and the Euclidean norm

$$\|\vec{x}\|_2 = \sum_{j=1}^{n} {x_j}^2 \tag{6.19}$$

for n–dimensional vectors $\vec{x} = (x_1, x_2, \dots, x_n)$. You saw that either one satisfies positive definiteness, homogeneity and the triangle inequality (see (7.1) on p.63 and (6.1) on p.47). As previously mentioned, mathematicians like to define new objects that are characterized by a given set of properties. As an example we had the definition of a vector space which encompasses objects as different as finite–dimensional vectors and real functions. In chapter (7) on the topology of real numbers (p. 59) you will learn about metric spaces as a concept that generalizes the measurement of distance (or closeness, if you prefer) for the elements of a non–empty set. Now we define a norm as a real function on a vector space by demanding the three characteristics of positive definiteness, homogeneity and the triangle inequality.

Definition 6.10 (Normed vector spaces)**.** Let L be a vector space. A **norm** on L is a real function

$$\|\cdot\| : \; L \longrightarrow \mathbb{R} \qquad x \longmapsto \|x\|$$

with the following three properties:

(6.20a) $\quad \|x\| \geqq 0 \quad \forall x \in L$ and $\|x\| = 0 \iff x = 0$ **positive definite**

(6.20b) $\quad \|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in L, \forall \alpha \in \mathbb{R}$ **homogeneity**

(6.20c) $\quad \|x + y\| \leqq \|x\| + \|y\| \quad \forall\, x, y \in L$ **triangle inequality**

Theorem 6.1 (Inner products define norms)**.** *Let L be a vector space with an inner product*

$$\bullet(\cdot,\cdot) : L \times L \to \mathbb{R}; \qquad (x, y) \mapsto x \bullet y$$

Then

$$\|\cdot\| : x \;\mapsto\; \|x\| \;:=\; \sqrt{(x \bullet x)}$$

defines a norm on L

Proof:
Positive definiteness *: follows immediately from that of the inner product.*

Homogeneity *: Let $x \in L$ and $\lambda \in \mathbb{R}$. Then*

$$\|\lambda x\| = \sqrt{(\lambda x) \bullet (\lambda x)} = \sqrt{\lambda\lambda(x \bullet x))} = |\lambda|\sqrt{x \bullet x} = |\lambda|\|x\|$$

and we are done

Triangle inequality : *Let $x, y \in L$. Then*

$$\begin{aligned}
\|x+y\|^2 &= (x+y) \bullet (x+y) \\
&= x \bullet x \ + \ 2(x \bullet y) \ + \ y \bullet y \\
&\leqq x \bullet x \ + \ 2|x \bullet y| \ + \ y \bullet y \\
&\leqq x \bullet x \ + \ 2\sqrt{x \bullet x}\sqrt{y \bullet y} \ + \ y \bullet y \\
&= \|x\|^2 \ + \ 2\|x\| \, \|y\| \ + \ \|y\|^2 \\
&= \big(\|x\| \ + \ \|y\|\big)^2
\end{aligned}$$

The second equation uses bilinearity and symmetry of the inner product. The first inequality expresses the simple fact that $\alpha \leqq |\alpha|$ for any number α. The second inequality uses Cauchy–Schwartz. The next equality just substitutes the definition $\|x\| = \sqrt{(x \bullet x)}$ of the norm. The next and last equality is your beloved binomial expansion $(a+b)^2 = a^2 + 2ab + b^2$ for the ordinary real numbers $a = \|x\|$ and $b = \|y\|$. We take square roots and obtain $\|x+y\| \ \leqq \ \|x\| \ + \ \|y\|$ and that's the triangle inequality we set out to prove. ■

7 Convergence and Continuity

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: All points of the sequence must get "arbitrarily close" to its limit as $n \to \infty$. Continuity of functions also can be phrased in terms of closeness: They map arbitrarily close elements of the domain to arbitrarily close elements of the codomain. In the most general setting Topology deals with neighborhoods of a point without providing the concept of measuring the distance of two points. We won't deal with that in this document. Instead we'll deal with sets X that are equipped with a metric.

7.1 Metric spaces (Study this!)

A metric is a real function of two arguments which associates with any two points $x, y \in X$ their "distance" $d(x, y)$.

It is clear how you measure the distance (or closeness, depending on your point of view) of two numbers x and y: you plot them on an x–axis where the distance between two consecutive integers is exactly one inch, grab a ruler and see what you get. Alternate approach: you compute the difference. For example, the distance between $x = 12.3$ and $y = 15$ is $x - y = 12.3 - 15 = -2.7$. Actually, we have a problem: There are situations where direction matters and a negative distance is one that goes into the opposite direction of a positive distance, but we do want that in this context and understand the distance to be always non–negative, i.e.,

$$dist(x, y) = |y - x| = |x - y|$$

More importantly, you must forget what you learned in your in your science classes: "Never ever talk about a measure (such as distance or speed or volume) without clarifying its dimension". Is the speed measured in miles per hour our inches per second? Is the distance measured in inches or miles or micrometers? In the context of metric spaces we measure distance simply as a number, without any dimension attached to it. For the above example, you get

$$dist(12.3, 15) = |12.3 - 15| = +2.7.$$

In section 6.1.3 on p.44 it is shown in great detail that the distance between two two–dimensional vectors $\vec{v} = (v_1, v_2)$ and $\vec{w} = (w_1, w_2)$ is

$$dist(\vec{v}, \vec{w}) = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2}$$

and the distance between two three–dimensional vectors $\vec{v} = (v_1, v_2, v_3)$ and $\vec{w} = (w_1, w_2, w_3)$ is

$$dist(\vec{v}, \vec{w}) = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + (w_3 - v_3)^2}.$$

We shall see in thm 7.1 on p.60 that this distance function is a metric according to the next definition:

Definition 7.1 (Metric spaces)**.** Let X be an arbitrary, non–empty set.
A **metric** on X is a real function

$$d(\cdot, \cdot) : \quad X \times X \longrightarrow \mathbb{R} \qquad (x, y) \longmapsto d(x, y)$$

with the following three properties: [13]

(7.1a) $\quad d(x,y) \geqq 0 \quad \forall x,y \in X$ and $d(x,y) = 0 \iff x = y$ **positive definite**

(7.1b) $\quad d(x,y) = d(y,x) \quad \forall x,y \in X$ **symmetry**

(7.1c) $\quad d(x,z) \leqq d(x,y) + d(y,z) \quad \forall\, x,y,z \in X$ **triangle inequality**

The pair $(X, d(\cdot,\cdot))$, usually just written as (X,d), is called a **metric space**. We'll write X for short if it is clear which metric we are talking about.

To appreciate that last sentence, you must understand that there can be more than one metric on X. See the examples below.

Remark 7.1 (Metric properties)**.** Let us quickly examine what those properties mean:
"Positive definite": The distance is never negative and two items x and y have distance zero if and only if they are equal.
"symmetry": the distance from x to y is no different to that from y to x. That may come as a surprise to you if you have learned in Physics about the distance from point a to point b being the vector $\vec{v}$ that starts in a and ends in b and which is the opposite of the vector $\vec{w}$ that starts in b and ends in a, i.e., $\vec{v} = -\vec{w}$. In this document we care only about size and not about direction.
"Triangle inequality": If you directly walk from x to z then this will be less painful than if you must make a stopover at an intermediary y.

Before we give some examples of metric spaces, here is a theorem that tells you that a vector space with a norm, i.e., a function with the three properties of the Euclidian norm (see 6.1 on p.47), becomes a metric space as follows:

Theorem 7.1 (Norms define metric spaces)**.** *A **norm** on a vector space L is a real function* [14]

$$\|\cdot\| : L \to \mathbb{R}_+; \qquad x \mapsto \|x\|$$

such that

$$\begin{aligned}
&\|x\| \geqq 0 \quad \forall x \in L \quad \textit{and} \quad \|x\| = 0 \iff x = 0 \quad \textbf{\textit{positive definite}}\\
&\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in L, \forall \alpha \in \mathbb{R} \quad \textbf{\textit{homogeneity}} \qquad (7.2)\\
&\|x+y\| \leqq \|x\| + \|y\| \quad \forall\, x,y \in L \quad \textbf{\textit{triangle inequality}}
\end{aligned}$$

The following is true:

$$d_{\|\cdot\|}(\cdot,\cdot) : (x,y) \mapsto \|y-x\|$$

defines a metric space $(L, d_{\|\cdot\|})$

Proof The proof may be required as part of an upcoming homework and will not be given here. It is really simple, even the triangle inequality for the metric $d(x,y) = \|x-y\|$ follows easily from the triangle inequality for the norm.

Here are some examples of metric spaces.

[13] If you forgot the meaning of $X \times X$, it's time to review [1] B/G (Beck/Geoghegan) ch.5.3 on cartesian products.
[14] This definition was given in the section on abstract vector spaces (def.6.10, p.57) which is considered optional material

Example 7.1 ($\mathbb{R} : d(a,b) = |b-a|$)**.** This is a metric space because $|\cdot|$ is the Euclidean norm on $\mathbb{R} = \mathbb{R}^1$. It is obvious that if x, y are real numbers then the difference $x - y$, and hence its absolute value, is zero if and only if $x = y$ and that proves positive definiteness.
Symmetry follows from the fact that

$$d(x,y) \;=\; |x-y| \;=\; |-(y-x)| \;=\; |y-x| \;=\; d(y,x).$$

The triangle inequality follows from the one which says that

$$|a+b| \leqq |a| + |b|$$

([1] B/G (Beck/Geoghegan), prop.10.8(iv)) as follows:

$$d(x,z) \;=\; |x-z| \;=\; |(x-y)-(z-y)| \leqq |(x-y)| + |(z-y)| \;=\; d(x,y) + d(z,y) \;=\; d(x,y) + d(y,z).$$

Example 7.2 (bounded real functions with $\boldsymbol{d(f,g) =}$ **sup**–norm of $\boldsymbol{g(\cdot) - f(\cdot)}$)**.**

$$d(f,g) = \sup\{|g(x) - f(x)| : x \in X\}$$

is a metric on the set $\mathscr{B}(X, \mathbb{R})$ of all bounded real functions on X.

This follows from the fact that $f \mapsto \sup\{|f(x)| : x \in X\}$ *is a norm on the vector space* $\mathscr{B}(X,\mathbb{R})$ *(see (7.1) on p.63). If you prefer, you can also conclude this from prop.7.2 on p.64 which directly proves the metric properties of* $\sup\{|g(x) - f(x)| : x \in X\}$*.*

Example 7.3 ($\mathbb{R}^N : d(\vec{x}, \vec{y}) =$ Euclidean norm)**.**

$$d(\vec{x}, \vec{y}) \;=\; \sqrt{(y_1-x_1)^2 + (y_2-x_2)^2 + \ldots + (y_N - x_N)^2} \;=\; \sqrt{\sum_{j=1}^{N} (y_j - x_j)^2}$$

This follows from the fact that the Euclidean norm is a norm on the vector space $\mathbb{R}^N$ *(see (6.1) on p.47).*

Just in case you think that all metrics are derived from norms, this one will set you straight.

Example 7.4 (Discrete metric)**.** Let X be non–empty. Then the function

$$d(x,y) = \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y \end{cases}$$

on $X \times X$ defines a metric.

Proof: Obviously the function is non–negative and it is zero if and only if $x = y$*.*
Symmetry is obvious too. The triangle inequality $d(x,z) = d(x,y) + d(y,z)$ *is clear in the special case* $x = z$*. (Why?) So let us assume* $x \neq z$*. But then* $x \neq y$ *or* $y \neq z$ *or both must be true. (Why?) That means that*

$$d(x,z) = 1 \leqq d(x,y) + d(y,z)$$

and this proves the triangle inequality. ■

7.1.1 Measuring the distance of real functions

How do we compare two functions? Let us make our lives easier: How do we compare two real functions $f(\cdot)$ and $g(\cdot)$? One answer is to look at a picture with the graphs of $f(\cdot)$ and $g(\cdot)$ and look at the shortest distance $|f(x) - g(x)|$ as you run through all x. That means that the distance between the functions $f(x) = x$ and $g(x) = x^2$ is zero because $f(1) = g(1) = 1$. The distance between $f(x) = x + 1$ and $g(x) = 0$ (the x–axis) is also zero because $f(-1) = g(-1) = 0$. Do you really think this is a good way to measure closeness? You really do not want two items to have zero distance unless they coincide. It's a lot better to look for an argument x where the value $|f(x) - g(x)|$ is largest rather than smallest. Now we are ready for a proper definition.

Definition 7.2 (Distance between real functions)**.** Let X be an arbitrary, non-empty set and let $f(\cdot), g(\cdot) : X \to \mathbb{R}$ be two real functions on X. We define the distance between $f(\cdot)$ and $g(\cdot)$ as

$$d(f,g) \;:=\; d(f(\cdot), g(\cdot)) \;:=\; \sup\{|f(x) - g(x)| : x \in X\} \tag{7.3}$$

The following picture illustrates this 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.

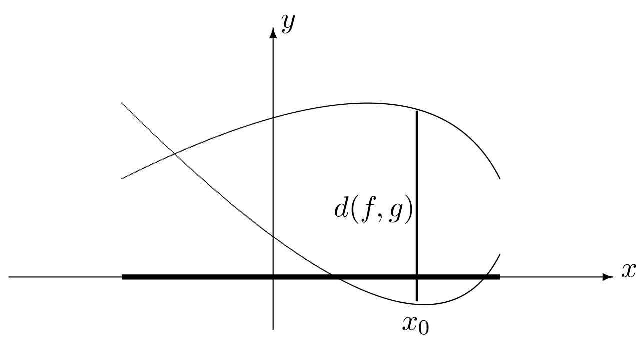

Figure 3: Distance of two real functions.

Now that you know how to measure the distance $d(f(\cdot), g(\cdot))$ between two real functions $f(\cdot), g(\cdot)$, the next picture shows you how to visualize the δ–neighborhood

$$B_\delta(f) \;:=\; \{g(\cdot) : X \to \mathbb{R} : d(f,g) < \delta\} \;=\; \{g(\cdot) : X \to \mathbb{R} : \sup_{x \in X} |f(x) - g(x)| < \delta\} \tag{7.4}$$

If X is a subset of $\mathbb{R}$, you draw the graph of $f(\cdot) + \delta$ (the graph of $f(\cdot)$ shifted up north by the amount of δ) and the graph of $f(\cdot) - \delta$ (the graph of $f(\cdot)$ shifted down south by the amount of δ). Any function $g(\cdot)$ which stays completely inside this band, without actually touching it, belongs to the δ–neighborhood of $f(\cdot)$.

In other words assuming that the domain A is a single, connected chunk and not a collection of more than one separate intervals, the δ–neighborhood of $f(\cdot)$ is a "band" whose contours are made up on the left and right

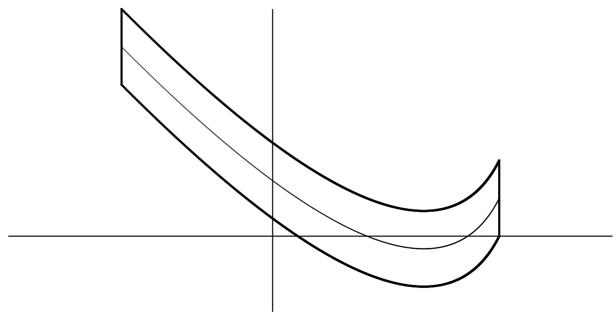

Figure 4: δ–neighborhood of a real function.

by two vertical lines and on the top and bottom by two lines that look like the graph of $f(\cdot)$ itself but have been shifted up and down by the amount of δ.

The distance of a real function $f(\cdot)$ to the zero function (see 5.3 on 33) has a special notation.

Definition 7.3 (Norm of bounded real functions)**.** Let X be an arbitrary, non-empty set. Let $f(\cdot) : X \to \mathbb{R}$ be a bounded real function on X, i.e., there exists a (possibly very large) number K such that $|f(x)| \leqq K$ for all $x \in X$. We define

$$\|f(\cdot)\| \; := \; \sup\{|f(x)| : x \in X\}$$

You should see that for any two bounded real functions $f(\cdot), g(\cdot)$ we have

$$\|f - g\| \; = \; \sup\{|f(x) - g(x)| : x \in X\} = d(f, g).$$

Proposition 7.1 (Properties of the norm of a real function)**.** *Let X be an arbitrary, non–empty set. Let*

$$\mathscr{B}(X, \mathbb{R}) := \{h(\cdot) : h(\cdot) \text{ is a bounded real function on } X\}$$

Then the norm function

$$\|\cdot\| : \; \mathscr{B}(X, \mathbb{R}) \longrightarrow \mathbb{R}_+ \qquad h(\cdot) \longmapsto \|h(\cdot)\| = \sup\{|f(x)| : x \in X\}$$

satisfies the three properties of a norm (see (7.2), p.60):

(7.5a) $\quad \|f\| \geqq 0 \quad \forall f \in \mathscr{B}(X, \mathbb{R}) \quad$ *and* $\quad \|f\| = 0 \iff f(\cdot) = 0$ ***positive definite***

(7.5b) $\quad \|\alpha f(\cdot)\| = |\alpha| \cdot \|f(\cdot)\| \quad \forall f \in \mathscr{B}(X, \mathbb{R}), \forall \alpha \in \mathbb{R}$ ***homogeneity***

(7.5c) $\quad \|f(\cdot) + g(\cdot)\| \leqq \|f(\cdot)\| + \|g(\cdot)\| \quad \forall\, f, g \in \mathscr{B}(X, \mathbb{R})$ ***triangle inequality***

Proof The proof is required as part of an upcoming homework. It is really simple, even the triangle inequality for the metric $d(x, y) = \|x - y\|$ follows easily from the triangle inequality for the norm.

Proposition 7.2 (Metric properties of the distance between real functions)**.** *Let X be an arbitrary, non–empty set.*
Let $\mathscr{B}(X,\mathbb{R}) := \{h(\cdot) : h(\cdot)$ is a bounded real function on $X\}$.
Let $f(\cdot), g(\cdot), h(\cdot) \in \mathscr{B}(X,\mathbb{R})$ Then the distance function

$$d(\cdot): \ \mathscr{B}(X,\mathbb{R}) \times \mathscr{B}(X,\mathbb{R}) \longrightarrow \mathbb{R}_+ \qquad (h_1,h_2) \longmapsto d(h_1,h_2) := \|h_1 - h_2\|$$

has the following three properties: [15]

(7.6a) $d(f,g) \geqq 0 \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R})$ *and* $d(f,g) = 0 \iff f(\cdot) = g(\cdot)$ ***positive definite***

(7.6b) $d(f,g) = d(g,f) \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R})$ ***symmetry***

(7.6c) $d(f,h) \leqq d(f,g) + d(g,h) \quad \forall\, f,g,h \in \mathscr{B}(X,\mathbb{R})$ ***triangle inequality***

We have seen in other contexts what those properties mean:
"Positive definite": The distance is never negative and two functions $f(\cdot)$ and $g(\cdot)$ have distance zero if and only if they are equal, i.e., if and only if $f(x) = g(x)$ for each argument $x \in X$.
*"symmetry": the distance from $f(\cdot)$ to $g(\cdot)$ is no different than that from $g(\cdot)$ to $f(\cdot)$. Symmetry implies that you do **not** obtain a negative distance if you walk in the opposite direction.*
"Triangle inequality": If you directly compare the maximum deviation between two functions $f(\cdot)$ and $h(\cdot)$ then this will never be more than than using an intermediary function $g(\cdot)$ and adding the distance between $f(\cdot) and g(\cdot)$ to that between $g(\cdot) and h(\cdot)$.

Proof: The proof is required as part of an upcoming homework. It is really simple, even the triangle inequality for the metric $d(x,y) = \|x - y\|$ follows easily from the triangle inequality for the norm.

7.1.2 Bounded sets and bounded functions

Definition 7.4 (bounded sets)**.** Given is a subset A of a metric space (X,d). The **diameter** of A is defined as

(7.7) $$diam(\emptyset) := 0, \qquad diam(A) := \sup\{d(x,y) : x,y \in A\} \text{ if } A \neq \emptyset.$$

We call A a **bounded set** if $diam(A) < \infty$.

Proposition 7.3 (bounded if and only if finite diameter)**.** *Given is a metric space (X,d).*
A non-empty subset A is bounded if and only if either of the following is true: [16]

(7.8) *A. $diam(A) < \infty$.*

(7.9) *B. There is a $\gamma > 0$ and $x_0 \in X$ such that $A \subseteq B_\gamma(x_0)$.*

(7.10) *C. For all $x \in X$ there is a $\gamma > 0$ such that $A \subseteq B_\gamma(x)$.*

Proof of "bounded if and only if A": Obvious from the definition of the supremum as least upper bound (see (5.7) on p.35).

[15] If you forgot the meaning of $\mathscr{B}(X,\mathbb{R}) \times \mathscr{B}(X,\mathbb{R})$, it's time to review [1] B/G (Beck/Geoghegan) ch.5.3 on cartesian products.

[16] Neighborhoods of a point will be discussed in chapter 7.1.3, p.65 about Neighborhoods and open sets. In short, the γ–neighborhood of x_0 is the set of all points with distance less than γ from x_0. (see def. 7.6 on p. 66).

Proof of "B ⇒ A": For any $x, y \in A$ we have

$$d(x,y) \leqq d(x,x_0) + d(x_0,y) \leqq 2\gamma$$

and it follows that $diam(A) \leqq 2\gamma$.

Proof of "A ⇒ B": Pick an arbitrary $x_0 \in A$ and let $\gamma := diam(A)$. Then

$$y \in A \quad \Longrightarrow \quad d(x_0,y) \leqq \sup_{x \in A} d(x,y) \leqq \sup_{x,z \in A} d(x,z) = diam(A) = \gamma.$$

It follows that $A \subseteq B_\gamma(x_0)$. ■

Proof of "C ⇒ A": We pick an arbitrary $x_0 \in A$ which is possible as A is not empty. Then there is $\gamma = \gamma(x_0)$ such that $A \subseteq B_\gamma(x_0)$. For any $y, z \in A$ we then have

$$d(y,z) \leqq d(y,x_0) + d(x_0,z) \leqq 2\gamma$$

and it follows that $diam(A) \leqq 2\gamma < \infty$.

Proof of "A ⇒ C": Given $x \in X$, pick an arbitrary $x_0 \in A$ and let $\gamma := d(x,x_0) + diam(A)$. Then

$$\begin{aligned} y \in A \quad \Longrightarrow \quad d(x,y) &\leqq d(x,x_0) + d(x_0,y) \leqq d(x,x_0) + \sup_{u \in A} d(u,y) \\ &\leqq d(x,x_0) + \sup_{u,z \in A} d(u,z) = d(x,x_0) + diam(A). = \gamma \end{aligned}$$

It follows that $A \subseteq B_\gamma(x)$.

Definition 7.5 (bounded functions)**.** Given is a metric space (X,d).
A real-valued function $f(\cdot)$ on X is called **bounded from above** if there exists a (possibly very large) number $\gamma_1 > 0$ such that

(7.11) $$f(x) < \gamma_1 \quad \text{for all arguments } x.$$

It is called **bounded from below** if there exists a (possibly very large) number $\gamma_2 > 0$ such that

(7.12) $$f(x) > -\gamma_2 \quad \text{for all arguments } x.$$

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

(7.13) $$|f(x)| < \gamma \quad \text{for all arguments } x.$$

We note that f is bounded if and only if its range $f(X)$ is a bounded subset of $\mathbb{R}$ (compare this to definition 5.9 on p.36 on supremum and infimum of functions)

7.1.3 Neighborhoods and open sets

A. *Given a point $x_0 \in \mathbb{R}$ (a real number), we can look at*

(7.14) $$\begin{aligned} B_\varepsilon(x_0) = (x_0 - \varepsilon, x_0 + \varepsilon) &= \{x \in \mathbb{R} : x_0 - \varepsilon < x < x_0 + \varepsilon\} \\ &= \{x \in \mathbb{R} : d(x,x_0) = |x - x_0| < \varepsilon\} \end{aligned}$$

which is the set of all real numbers x with a distance to x_0 of strictly less than a number ε (the open interval with end points $x_0 - \varepsilon$ and $x_0 + \varepsilon$). (see example (7.1) on p.61).

***B.** Given a point $\vec{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ (a point in the xy–plane), we can look at*

$$\begin{aligned} B_\varepsilon(\vec{x}_0) &= \{\vec{x} \in \mathbb{R}^2 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} \\ &= \{(x,y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\} \end{aligned} \tag{7.15}$$

which is the set of all points in the plane with a distance to $\vec{x}_0$ of strictly less than a number ε (the open disc around $\vec{x}_0$ with radius ε from which the points on the boundary (those with distance equal to ε) are excluded).

***C.** Given a point $\vec{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ (a point in the 3–dimensional space), we can look at*

$$\begin{aligned} B_\varepsilon(\vec{x}_0) &= \{\vec{x} \in \mathbb{R}^3 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} \\ &= \{(x,y,z) \in \mathbb{R}^3 : (\vec{x} - \vec{x}_0)^2 + (\vec{y} - \vec{y}_0)^2 + (\vec{z} - \vec{z}_0)^2 < \varepsilon^2\} \end{aligned} \tag{7.16}$$

which is the set of all points in space with a distance to $\vec{x}_0$ of strictly less than a number ε (the open ball around $\vec{x}_0$ with radius ε from which the points on the boundary (those with distance equal to ε) are excluded).

***D.** Given a normed vector space $(L, \|\cdot\|)$ and a vector $x_0 \in L$, we can look at*

$$B_\varepsilon(x_0) \;=\; \{x \in L : \|x - x_0\| < \varepsilon\} \tag{7.17}$$

which is the set of all vectors in L with a distance to x_0 of strictly less than a number ε (the open set around x_0 with "radius" ε from which the points on the boundary (those with distance equal to ε) are excluded).

There is one more item more general than neighborhoods of elements belonging to normed vector spaces, and that would be neighborhoods in metric spaces. We have arrived at the final definition:

Definition 7.6 (ε-Neighborhood)**.** Given a metric space (X, d) and an element $x_0 \in X$, we can look at

$$B_\varepsilon(x_0) \;=\; \{x \in L : d(x, x_0) < \varepsilon\} \tag{7.18}$$

which is the set of all elements of X with a distance to x_0 of strictly less than the number ε (the open set around x_0 with "radius" ε from which the points on the boundary (those with distance equal to ε) are excluded). We call $B_\varepsilon(x_0)$ the **ε–neighborhood** of x_0.

Let us not be too scientific about this, but the following should be intuitively clear: Look at any point $a \in B_\varepsilon(x_0)$. You can find $\delta > 0$ such that the entire δ–neighborhood $B_\delta(a)$ of a is contained inside $B_\varepsilon(x_0)$. Just in case you do not trust your intuition, here is the proof. It is worth while to examine it closely because you can see how the triangle inequality is put to work:

$a \in B_\varepsilon(x_0)$ *means that* $\varepsilon - d(a, x_0) > 0$ *, say*

$$\varepsilon - d(a, x_0) = 2\delta \tag{7.19}$$

where $\delta > 0$. *Let* $b \in B_\delta(a)$. *I claim that any such b is an element of $B_\varepsilon(x_0)$. How so?*

$$d(b, x_0) \leqq d(b, a) + d(a, x_0) \leqq \delta + d(a, x_0) < 2\delta + d(a, x_0) = \varepsilon$$

In the above chain, the first inequality is a consequence of the triangle inequality. The second one reflects the fact that $b \in B_\delta(a)$. *The strict inequality is trivial because we added the strictly positive number* δ. *The final equality is a consequence of* (7.19).

So we have proved that for any $b \in B_\delta(a)$ *we have* $b \in B_\varepsilon(x_0)$, *hence* $B_\delta(a) \subseteq B_\varepsilon(x_0)$.

In other words, any $a \in B_\varepsilon(x_0)$ *is an interior point of* $B_\varepsilon(x_0)$ *in the following sense:*

Definition 7.7 (Interior point)**.** Given is a metric space (X,d).
An element $a \in A \subseteq X$ is called an **interior point** of A if we can find some $\varepsilon > 0$, however small it may be, so that $B_\varepsilon(a) \subseteq A$.

Definition 7.8 (open set)**.** Given is a metric space (X,d).
A set all of whose members are interior points is called an **open set**.

Proposition 7.4. $B_\varepsilon(x_0)$ *is an open set*

Proof: we showed earlier on that any $a \in B_\varepsilon(x_0)$ *is an interior point of* $B_\varepsilon(x_0)$. ■

Definition 7.9 (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 simply called a **neighborhood** of x_0.

Remark 7.2 (Open neighborhoods are the important ones)**.** You will see that the important neighborhoods are the small ones, not the big ones. The definition above says that you can sandwich an open neighborhood U_x inbetween a point x and anyone of its neighborhoods A_x. In other words, there are many propositions and theorems where you may assume that a neighborhood you deal with is open.

Theorem 7.2 (Metric spaces are topological spaces)**.** *The following is true about open sets of a metric space* (X,d)*:*

(7.20a) *An arbitrary union* $\bigcup_{i \in I} U_i$ *of open sets* U_i *is open.*

(7.20b) *A finite intersection* $U_1 \cap U_2 \cap \ldots \cap U_n$ $(n \in \mathbb{N})$ *of open sets is open.*

(7.20c) *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* $B_\varepsilon(x) \subseteq U_{i_0}$. *But* $U_{i_0} \subseteq U$ *and we have* $B_\varepsilon(x) \subseteq U$ *and have shown that* x *is interior point of* U. *But* x *was an arbitrary point of* $U = \bigcup_{i \in I} U_i$ *which therefore is shown to be an open set.*

Proof of b: Let $x \in U := U_1 \cap U_2 \cap \ldots \cap U_n$. *Then* $x \in U_j$ *for all* $1 \leqq j \leqq n$ *according to the definition of an intersection and it is inner point of all of them because they all are open sets. Hence, for each* j *there is a suitable* $\varepsilon_j > 0$ *such that* $B_{\varepsilon_j}(x) \subseteq U_j$ *Now define*

$$\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n\}$$

Then $\varepsilon > 0$ and [17]

$$B_\varepsilon(x) \subseteq B_{\varepsilon_j}(x) \subseteq U_j \; (1 \leqq j \leqq n) \qquad \Longrightarrow \qquad B_\varepsilon(x) \subseteq \bigcap_{j=1}^{n} U_j.$$

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, B_ε is a subset of X. But then x is an inner point of X, so all members of x are inner points and this proves that X is open.
Now to the empty set $\emptyset$. You may have a hard time to accept the logic of this statement: All elements of $\emptyset$ are interior points. But remember, the premise "let $x \in X$" is always false and you may conclude from it whatever you please. ■

7.1.4 Digression: Abstract topological spaces (Skip starting at def. 7.12: Basis and neighborhood basi)!)

Theorem 7.2 on p.67 gives us a way of defining neighborhoods for sets which do not have a metric.

Definition 7.10 (Abstract topological spaces)**.** Let X be an arbitrary non-empty set and let $\mathfrak{U}$ be a set of subsets of X whose members satisfy the properties a, b and c of (7.20) on p.67:

(7.21a) An arbitrary union $\bigcup_{i \in I} U_i$ of sets $U_i \in \mathfrak{U}$ belongs to $\mathfrak{U}$,

(7.21b) $U_1, U_2, \dots, U_n \in \mathfrak{U} \; (n \in \mathbb{N}) \quad \Rightarrow \quad U_1 \cap U_2 \cap \ldots \cap U_n \in \mathfrak{U}$,

(7.21c) $X \in \mathfrak{U} \quad \text{and} \quad \emptyset \in \mathfrak{U}$.

Then $(X, \mathfrak{U})$ is called a **topological space** The members of $\mathfrak{U}$ are called "open sets" of $(X, \mathfrak{U})$ and the collection $\mathfrak{U}$ of open sets is called the **topology** of X.

Definition 7.11 (Topology induced by a metric)**.** Let (X, d) be a metric space and let $\mathfrak{U}_d$ be the set of open subsets of (X, d), i.e., all sets $U \in X$ which consist of interior points only: for each $x \in U$ there exist $\varepsilon > 0$ such that

$$B_\varepsilon(x) \;=\; \{y \in X : d(x,y) < \varepsilon\} \;\subseteq\; U$$

(see (7.7) on p.67). We have seen in theorem (7.2) that those open sets satisfy the conditions of the previous definition. In other words, $(X, \mathfrak{U}_d)$ defines a topological space. We say that its topology is **induced by the metric** $d(\cdot,\cdot)$ or that it is **generated by the metric** $d(\cdot,\cdot)$. If there is no confusion about which metric we are talking about, we also simply speak about the **metric topology**.

Let X be a vector space with a norm $\|\cdot\|$. Remember that any norm defines a metric $d_{\|\cdot\|}(\cdot,\cdot)$ via $d_{\|\cdot\|}(x,y) = \|x - y\|$ (see (7.1) on p.60). Obviously, this norm defines open sets

$$\mathfrak{U}_{\|\cdot\|} \;:=\; \mathfrak{U}_{d_{\|\cdot\|}}$$

[17] 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.

on X by means of this metric. We say that this topology is **induced by the norm** $\|\cdot\|$ or that it is **generated by the norm** $\|\cdot\|$. If there is no confusion about which norm we are talking about, we also simply speak about the **norm topology**.

Example 7.5 (Discrete topology)**.** Let X be non–empty. We had defined in (7.4) on p.61 the discrete metric as

$$d(x,y) \;=\; \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y. \end{cases}$$

The associated topology is

$$\mathfrak{U}_d \;=\; \{A : A \subseteq X\}.$$

In other words, each subset of X is open. Why? First observe that for any $x \in X$, $B_{1/2}(x) = \{x\}$. Hence, each singleton in X is open. But any subset $A \subseteq X$ is the union of it members: $A = \bigcup\limits_{a \in A} \{a\}$ and it must be open as a union of open sets. 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 definiton 7.10 on p.68. We call this topology the **discrete topology** of X.

Example 7.6 (Indiscrete topology)**.** Here is an example of a topology which is not generated by a metric. Let X be an arbitrary non–empty set and define $\mathfrak{U} := \{\emptyset, X\}$. Then $(X, \mathfrak{U})$ is a topological space. This is trivial because any intersection of members of $\mathfrak{U}$ is either $\emptyset$ (if at least one member is $\emptyset$) or X (if all members are X). Conversely, any union of members of $\mathfrak{U}$ is either $\emptyset$ (if all members are $\emptyset$) or X (if at least one member is X).

The topology $\{\emptyset, X\}$ is called the **indiscrete topology** of X. It is the smallest possible topology on X.

Definition 7.12 (Basis and neighborhood basis)**.** Let $(X, \mathfrak{U})$ be a topological space.

A subset $\mathfrak{B} \subseteq \mathfrak{U}$ of open sets is called a **basis of the topology** if any open set U can be written as a union

$$U \;=\; \bigcup_{i \in I} B_i \quad (B_i \in \mathfrak{B} \text{ for all } i \in I) \tag{7.22}$$

where I is a suitable index set.

Let $x \in X$ and $A \subseteq X$. It is not assumed that A be open. A is called a **neighborhood** of x and x is called an **interior point** of A if you can find an open set U such that

$$x \;\in\; U \;\subseteq A. \tag{7.23}$$

The set of subsets of X

$$\mathfrak{N}(x) \;:=\; \{A \subseteq X : A \text{ is a neighborhood of } X\} \tag{7.24}$$

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 basis of** x if it satisfies the following condition: For any $A \in \mathfrak{N}(x)$ you can find a $B \in \mathfrak{B}(x)$ such that $B \subseteq A$. In other words, in theorems 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 basis of x rather than having to show it for all neighborhoods of x.

Definition 7.13 (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.

Theorem 7.3 (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

(7.25) $$\mathfrak{B}(x) \;:=\; \{\, B_{1/n}(x) : n \in \mathbb{N} \,\}.$$

Then $\mathfrak{B}(x)$ is a neighborhood basis of x. ■

Definition 7.14 (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 basis for $\mathfrak{U}$.

Theorem 7.4 (Euclidean space $\mathbb{R}^N$ is second countable)**.** *Let*

(7.26) $$\mathfrak{B} \;:=\; \{\, B_{1/n}(q) : q \in \mathbb{Q}^N,\; n \in \mathbb{N} \,\}.$$

Here $\mathbb{Q}^N = \{q = (q_1, \dots, q_N) : q_j \in \mathbb{Q},\ 1 \leqq j \leqq N\}$ is the set of all points in $\mathbb{R}^N$ with rational coordinates. Then $\mathfrak{B}$ is a countable basis.

Proof (outline): You have seen that $\mathbb{Q}$ is countable (corollary 3.1 on p.21). It can be shown that $\mathbb{Q}^N$ too is countable. Let $U \in \mathfrak{U}$ be an arbitrary open set in X. Any $x \in U$ is inner point of U, hence we can find some (large) integer n_x such that the entire $3/n_x$–neighborhood $B_{3/n_x}(x)$ is contained within U. As any vector can be approximated by vectors with rational coordinates, we can find some $q = q_x \in \mathbb{Q}^N$ such that $d(x, q_x) < /n_x$. Draw a picture and you see that both $x \in B_{1/n_x}(q_x)$ and $B_{1/n_x}(q_x) \subseteq B_{3/n_x}(x)$. In other words, we have

$$x \;\in\; B_{1/n_x}(q_x) \;\subseteq\; U$$

for all $x \in U$. But then

$$U \subseteq \bigcup \left[\, B_{1/n_x}(q_x) : x \in U \,\right] \;\subseteq U$$

and it follows that U is the (countable union of the sets $B_{1/n_x}(q_x)$. ■

7.1.5 Convergence, contact points and closed sets

Definition 7.15 (convergence of sequences)**.** Given is a metric space (X, d).
We say that a sequence (x_n) of elements of X **converges** to $a \in X$ for $n \to \infty$ if almost all of the x_n will come arbitrarily close to a in the following sense:
Let δ be an arbitrarily small positive real number. Then there is a (possibly extremely large) integer n_0 such that all x_j belong to $B_\delta(a)$ just as long as $j \geqq n_0$. To say this another way: Given any number $\delta > 0$, however small, you can find an integer n_0 such that

(7.27) $$d(a, x_j) < \delta \;\text{ for all }\; j \geqq n_0$$

We write either of

(7.28) $$a = \lim_{n \to \infty} x_n \qquad \text{or} \qquad x_n \to a$$

and we call a the **limit** of the sequence (x_n)

There is yet another way of interpreting 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.

Convergence is an extremely important concept in Mathematics, but it excludes the case of sequences such as $x_n := n$ and $y_n := -n$ $(n \in \mathbb{N})$. Intuition tells us that x_n converges to ∞ and y_n converges to $-\infty$ because we think of very big numbers as being very close to $+\infty$ and very small numbers (i.e., very big ones with a minus sign) as being very close to $-\infty$.

Definition 7.16 (Limit infinity)**.** For this definition we do not deal with an arbitrary metric space but specifically with $X = \mathbb{R}$ and $d(x, y) = |b - a|$. Given a real number $K > 0$, we define

(7.29a) $$B_K(\infty) := \{x \in \mathbb{R} : x > K\}$$

(7.29b) $$B_K(-\infty) := \{x \in \mathbb{R} : x < -K\}$$

We call $B_K(\infty)$ the K-neighborhood of ∞ and $B_K(-\infty)$ the K-neighborhood of $-\infty$.
We say that a sequence (x_n) has limit ∞ and we write either of

(7.30) $$x_n \to \infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = \infty$$

if the following is true for any (big) K: There is a (possibly extremely large) integer n_0 such that all x_j belong to $B_K(\infty)$ just as long as $j \geqq n_0$.
We say that the sequence (x_n) has limit $-\infty$ and we write either of

(7.31) $$x_n \to -\infty \qquad \text{or} \qquad \lim_{n\to\infty} x_n = -\infty$$

if the following is true for any (big) K: There is a (possibly extremely large) integer n_0 such that all x_j belong to $B_K(-\infty)$ just as long as $j \geqq n_0$.

Note 7.1 (Notation for limits of monotone sequences)**.** Let (x_n) be a non-decreasing sequence of real numbers and let y_n be a non-increasing sequence. If $\xi = \lim_{k\to\infty} x_k$ (that limit might be $+\infty$) then we write suggestively

$$x_i \nearrow \xi \quad (i \to \infty)$$

If $\eta = \lim_{j\to\infty} x_j$ (that limit might be $-\infty$) then we write suggestively

$$y_j \searrow \eta \quad (j \to \infty)$$

Remark 7.3 (No convergence or divergence to infinity)**.** The majority of mathematicians does not use the expressions "convergence to ∞" or "divergence to ∞". Rather, they will use the phrase that a sequence has the limit ∞.

If you look at any ***closed interval*** *$[a, b] = \{y \in \mathbb{R} : a \leqq y \leqq b\}$, of real numbers, then all of its points are interior points, except for the end points a and b. On the other hand, a and b are contact points according to the following definition which makes sense for any metric space (X, d).*

Definition 7.17 (contact points)**.** Given is a metric space (X, d).
Let $A \subseteq X$ and $a \in X$ (a may or may not to belong to A). a is called a **contact point** of A (German: Berührungspunkt - see [10] Von Querenburg, p.21) if there exists a sequence $x_1, x_2, x_3, \dots$ of members of A which converges to a.

Proposition 7.5 (Criterion for contact points)**.** *Given is a metric space (X, d).*
Let $A \subseteq X$ and $a \in X$. Then a is a contact point of A if and only if $A \cap N \neq \emptyset$ for any neighborhood N of a.

Proof of "$\Rightarrow$": Let $x \in X$ and assume there is $(x_n)_n$such that $x_n \in A$ and $x_n \to x$. We must show that if U_x is a (open) neighborhood of x then $U_x \cap A \neq \emptyset$. Let $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U_x$. It follows from $x_n \to x$ that there is $N = N(\varepsilon)$ such that $x_n \in B_\varepsilon(x)$ for all $n \geqq N$, especially, $x_N \in B_\varepsilon(x)$. By assumption, $x_N \in A$, hence $x_N \in B_\varepsilon(x) \cap A \subseteq U_x \cap A$, hence $U_x \cap A \neq \emptyset$.

Proof of "$\Leftarrow$" Let $x \in X$ be a contact point for $A \subseteq X$. Let $x_n \in B_{1/n}(x) \cap A$. Such x_n exists: x is a contact point of A, hence $B_{1/n}(x) \cap A \neq \emptyset$. Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be chosen such that $1/\varepsilon < N$. This is possible because $\mathbb{N}$ is not bounded (above) in $\mathbb{R}$. For any $j \geqq N : d(x_j, x) < 1/j \leqq 1/N < \varepsilon$. This proves $x_n \to x$. ■

Note 7.2. We mentioned before that a contact point for a set A need not necessarily belong to A. Example: Let A be the set $]0, 1[$ of real numbers x such that $0 < x < 1$. Then 0 is a contact point because the sequence $x_n = 1/n$ converges to 0: No matter how small a δ you choose: if you set n_0 to an integer larger than $1/\delta$ then

$$n > n_0 \Rightarrow d(x_n, 0) = |x_n - 0| = |x_n| = 1/n < 1/n_0 < 1/(1/\delta) = \delta \tag{7.32}$$

and it follows that 0 is a contact point of $]0, 1[$. Similarly we can show that the sequence $x_n = 1 - 1/n$ converges to the number 1.

On the other hand, any $b \in A \subseteq (X, d)$ is a contact point of A because the constant sequence

$$x_1 = b; \quad x_2 = b; \quad x_3 = b; \cdots$$

converges to b. This means that any subset of X is contained in its closure, which we will define next.

Note 7.3 (Contact points vs Limit points)**.** Besides contact points there also is the concept of a limit point. Here is the definition (see [5] Munkres, a standard book on topology): Given is a metric space (X, d).
Let $A \subseteq X$ and $a \in X$. a is called a **limit point** or **cluster point** or **point of accumulation** of A if any neighborhood U of a intersects A in at least one point other than a. This definition excludes "isolated points" of A from being limit points of A.

Definition 7.18 (closed sets)**.** Given is a metric space (X, d)
and a subset $A \subseteq X$. We call

$$\bar{A} \quad := \{x \in X : x \text{ is a contact point of } A\}$$

the **closure** of A. A set that contains all its contact points is called a **closed set**.

Proposition 7.6. *The complement of an open set is closed.*

Proof of 7.6: Let A be an open set. Each point $a \in A$ is an interior point which can be surrounded by a δ-neighborhood $V_\delta(a)$ which, for small enough δ, will be entirely contained within A. Let $B = A^\complement = X \setminus A$ and assume $x \in X$ is a contact point of B. We want to prove that B is a closed set, so we must show that $x \in B$. We assume the opposite and show that this will lead to a contradiction. So let us assume that $x \notin B$. That means, of course, that x belongs to B's complement which is A. But A is open, so x must necessarily be an interior point of A. This means that there is an entire neighborhood $B_\delta(x)$ surrounding x which is entirely contained in A and hence has no points in common with the complement B. On the other hand we assumed that x is a contact point of B of A. That again means that there must be points in B so close to x that they also must be contained in $B_\delta(x)$ and we have reached a contradiction. ■

Proposition 7.7. *The complement of a closed set is open.*

Proof of 7.7: Let A be closed set. Let $B = A^\complement = X \setminus A$. If B is not open then there must be $b \in B$ which is not an interior point of B. We'll show now that this assumption leads to a contradiction. Because b is not an interior point of B, there is no δ-neighborhood, for whatever small δ, that entirely belongs to B. So, for each $j \in \mathbb{N}$, there is an $x_j \in B_{1/j}(b)$ which does not belong to B. In other words, we have a sequence x_j which converges to b and is entirely contained in A. The closed set A contains all its contact points and it follows that $b \in A$. But we had assumed at the outset that $b \in B$ which is the complement of A and we have a contradiction. ■

7.1.6 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 7.27 (page 70)

Definition 7.19 (Cauchy sequences)**.** Given is a metric space (X, d).
A sequence (x_n) in X is called a **Cauchy sequence** [18] 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

(7.33) $$d(x_i, x_j) < \varepsilon \qquad \text{for all } i, j \geqq n_0$$

This is called the **Cauchy criterion for convergence** of a sequence.

Example 7.7 (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

(7.34) $$|x_i - x_j| < \varepsilon \qquad \text{for all } i, j \geqq n_0$$

The following theorem of the completeness of the set of all real numbers states that any Cauchy sequence converges to a real number. To say this differently, showing that a sequence is Cauchy is all you have to do if you want to show that a sequence has a finite limit without the need to provide the actual value of that limit. This situation arises very often in Math. Matter of fact, you 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.

[18] so named after the great french mathematician Augustin–Louis Cauchy (1789–1857) who contributed massively to the most fundamental ideas of Calculus.

Theorem 7.5 (Completeness of the real numbers)**.** *The following is true for the real numbers with the metric* $d(a,b) = |b-a|$ *but will in general be false for arbitrary metric spaces: Let* (x_n) *be a Cauchy sequence in* $\mathbb{R}$*. then there exists a real number* L *such that* $L = \lim\limits_{n\to\infty} x_n$*.*

Proof: Part 1: We shall show that x_n *is bounded. There is* $N = N(1)$ *such that* $|x_i - x_j| < 1/2$ *for all* $i,j \geqq N$*. In particular,* $|x_i - x_N| < 1/2$*, hence* $|x_i| = |x_i - x_N + x_N| \leqq |x_i - x_N| + |x_N| \leqq |x_N| + 1$ *for all* $i \geqq N$*. Let* $M := \max\{|x_j| : j \leqq N\}$*. Then* $|x_j| \leqq M + 1$ *and we have proved that the sequence is bounded.*

Part 2: We shall show next that $\liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$*. Let* $\varepsilon > 0$ *and* $N \in \mathbb{N}$ *such that* $|x_i - x_j| \leqq \varepsilon$ *for all* $i,j \geqq N$*. Let* $T_n := \{x_j : j \geqq n\}$ *be the tail set of the sequence* $(x_n)_n$*. Let* $\alpha_N := \inf T_N, \beta_N := \sup T_N$*. There is some* $i \geqq \mathbb{N}$ *such that* $|x_i - \alpha_N| = x_i - \alpha_N \leqq \varepsilon$ *and there is some* $j \geqq \mathbb{N}$ *such that* $|\beta_N - x_j| = \beta_N - x_j \leqq \varepsilon$ *It follows that*

$$0 \leqq \beta_N - \alpha_N = |\beta_N - \alpha_N| \leqq |(\beta_N - x_j) + (x_j - x_i) + (x_i - \alpha_N)| \leqq 3\varepsilon.$$

Further, if $k > N$ *then* $T_k \subseteq T_N$*, hence* $\alpha_k \geqq \alpha_N$ *and* $\beta_k \leqq \beta_N$ *and it follows that*

$$\beta_k - \alpha_k \leqq \beta_N - \alpha_N \leqq 3\varepsilon.$$

But then

$$\limsup_{k\to\infty} x_k - \liminf_{k\to\infty} x_k \leq \lim_{k\to\infty} \beta_k - \lim_{k\to\infty} \alpha_k \leqq \beta_N - \alpha_N \leqq 3\varepsilon.$$

$\varepsilon > 0$ *was arbitrary, hence* $\limsup\limits_{k\to\infty} x_k = \liminf\limits_{k\to\infty} x_k$*.*

Part 3: It follows from theorem 5.2 on p.42 that the sequence $(x_n)_n$ *converges to* $L := \limsup\limits_{k\to\infty} x_k$ *and the proof is finished.* ■

Now that you have the completeness of $\mathbb{R}$ *it is not very difficult to see that it is valid for* $\mathbb{R}^N$*, too.*

Theorem 7.6 (Completeness of $\mathbb{R}^N$)**.** *The following is true for* $\mathbb{R}^N$ *with the Euclidian norm, specifically for the real numbers with* $d(a,b) = |b-a|$ *but will in general be false for arbitrary metric spaces or normed vector spaces: Let* $(\vec{x}_n)$ *be a Cauchy sequence in* $\mathbb{R}^N$*. then there exists a vector* $\vec{a} \in \mathbb{R}^N$ *such that* $\vec{a} = \lim\limits_{n\to\infty} \vec{v}_n$*.*

Proof (outline): Let $\vec{x}_n = (x_{n,1}, x_{n,2}, \dots, x_{n,N})$*. From the theorem of the completeness of the real numbers we know that there exist real numbers*

$$a_1, a_2, a_3, \dots, a_N \qquad \text{such that } a_j = \lim_{n\to\infty} x_{n,j} \ (1 \leqq j \leqq N).$$

For a given number ε *we can find natural numbers* $n_{0,1}, n_{0,2}, \dots, n_{0,N}$ *such that*

$$|x_{n,j} - a_j| < \frac{\varepsilon}{N} \qquad \text{for all } n \geqq n_{0,j}.$$

Let $n^\star := max(n_{0,1}, n_{0,2}, \dots, n_{0,N})$*. It follows that*

$$d((\vec{x}_n - \vec{a}) = \sqrt{\sum_{j=1}^{N} |x_{n,j} - a_j|^2} \leqq N \cdot \frac{\varepsilon}{N} = \varepsilon \qquad \text{for all } n \geqq n^\star.$$

■

Here is the formal definition of a complete set in a metric space.

Definition 7.20 (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.

We won't really talk about completeness in general until the chapter on compact spaces. Just to mention one of the simplest facts about completeness:

Theorem 7.7 (Complete sets are closed)**.** *Given is a metric space (X, d). Any complete subset $A \subseteq X$ is closed.*

*Proof: Let a be a contact point of A. The theorem is proved if we can show that $a \in A$. **a)** We shall employ prop.7.5 on p.72: 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 $B_{1/m}(a)$ is a neighborhood of the contact point a, hence hence $A \bigcap B_{1/m}(a) \neq \emptyset$ and we can pick a point from this intersection which we name x_m. **b)** We shall prove next that $(x_m)_m$ is Cauchy. Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be such that $N > 1/\varepsilon$. if $j \in \mathbb{N}$ and $k \in \mathbb{N}$ both exceed N then*

$$d(x_j, x_k) \leqq d(x_j, a) + d(a, x_k) \leqq \frac{1}{j} + \frac{1}{k} \leqq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

*It follows that the sequence (x_j) is Cauchy. **c)** Because A is complete, this sequence must converge to some $b \in A$. But b cannot be different from a Otherwise we could "separate" a and b by two disjoint neighborhoods: choose the open ρ–balls $B_\rho(a)$ and $B_\rho(b)$ where ρ is one half the distance between a and b (see the proof of thm.7.9 on p.91). Only finitely many of the x_n are allowed to be outside $B_\rho(a)$ and the same is true for $B_\rho(b)$. 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.* ■

Example 7.8 (Approximation of decimals)**.** The following should help to illustrate Cauchy sequences and completeness in $\mathbb{R}$. Take any real number $x \geqq 0$ and write it as a decimal. As I explained in (3.1) on (p.7), anything that can be written as a decimal number is a real number. Let's say, x starts out on the left as

$$x = 258.1408926584207531\ldots$$

If we define as x_k the leftmost part of x, truncated k digits after the decimal points:

$$x_1 = 258.1, \quad x_2 = 258.14, \quad x_3 = 258.140, \quad x_4 = 258.1408, \quad x_5 = 258.14089, \quad \ldots$$

and as y_k the leftmost part of x, truncated k digits after the decimal points, but the rightmost digit incremented by 1 (where you then might obtain a carry-over to the left when you add 1 to 9)

$$y_1 = 258.2, \quad y_2 = 258.15, \quad y_3 = 258.141, \quad y_4 = 258.1409, \quad y_5 = 258.14090, \quad \ldots$$

then the sequence (x_n) is non-decreasing: $x_{n+1} \geqq x_n$ for all n and the sequence (y_n) is non-increasing: $y_{n+1} \leqq y_n$ for all n and we have the sandwich property: $x_n \leqq x \leqq y_n$ for all n. Both sequences are Cauchy because both

$$d(x_{n+i}, x_{n+j}) = |x_{n+i} - x_{n+j}| \leqq 10^{-n} \to 0 \quad (n \to \infty)$$
$$d(y_{n+i}, y_{n+j}) = |y_{n+i} - y_{n+j}| \leqq 10^{-n} \to 0 \quad (n \to \infty)$$

It is obvious that $x = \lim_{n\to\infty} x_n = \lim_{m\to\infty} y_m$.

What just has been illustrated is that there a natural way to construct for a given $x \in \mathbb{R}$ Cauchy sequences that converge towards x. The completeness principle states that the reverse is true: For any Cauchy sequence you can find an element x against which the sequence converges.

7.1.7 Appendix: Addenda to chapter 7.1: Metric Spaces

Given a metric space (X, d)*, what is the opposite of* $\lim_{k\to\infty} x_k = L$*?*
Beware! It is NOT the statement that $\lim_{k\to\infty} x_k \neq L$ *because such a statement would mislead you to believe that such a limit exists, it just happens not to coincide with* L

The correct answer: There exists some $\varepsilon > 0$ *such that for* ***all*** $N \in \mathbb{N}$ *there exists some natural number* $j = j(N)$ *such that* $j \geq N$ *and* $d(x_j, L) \geq \varepsilon$*.*

Proposition: A sequence $(x_k)_k$ *with values in* (X, d) *does not have* $L \in X$ *as its limit iff there exists some* $\varepsilon > 0$ *and a subsequence* $n_1 < n_2 < n_3 < \dots$ *in* $\mathbb{N}$ *such that* $d(x_{n_j}, L) \geq \varepsilon$ *for* ***all*** j*.*

7.2 Continuity (Study this!)

7.2.1 Definition and characterization of continuous functions

Informally speaking a continuous function

$$f(\cdot): \ \mathbb{R} \longrightarrow \mathbb{R} \qquad x \longmapsto y = f(x)$$

is one whose graph in the xy plane is a continuous line without any disconnections or gaps. This can be stated slightly more formal by saying that if the x-values are closely together then the $f(x)$-values must be closely together too. The latter makes sense for any sets X, Y where closeness can be measured, i.e., for metric spaces (X, d_1) and (Y, d_2). Here is the formal definition:

Definition 7.21 (Continuous functions)**.** Given are two metric spaces (X, d_1) and (Y, d_2). Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. We say that $f(\cdot)$ is **continuous at $\boldsymbol{x_0}$** and we write

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{7.35}$$

if the following is true for any sequence (x_n) with values in A:

$$\text{if } x_n \to x_0 \text{ then } f(x_n) \to f(x_0). \tag{7.36}$$

We say that $f(\cdot)$ is **continuous** if $f(\cdot)$ is continuous in a for all $a \in A$.

In other words, the following must be true for any sequence (x_n) in A

$$\lim_{n\to\infty} x_n = x_0 \ \Rightarrow \ \lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x_0) \tag{7.37}$$

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. If x_0 is not an interior point of A then we ignore sequences $x_n \to x_0$ unless all its members x_n belong to A.*

Theorem 7.8 (ε-δ characterization of continuity)**.** *Let (X, d_1) and (Y, d_2) be two metric spaces. Let $A \subseteq X$, $x_0 \in A$ and let $f(\cdot) : A \to Y$ be a mapping from A to Y. Then $f(\cdot)$ is continuous at x_0 if and only if the following is true: For any (whatever small) $\varepsilon > 0$ there exists a (most likely very small) $\delta > 0$ such that*

$$f\big(B_\delta(x_0) \cap A\big) \ \subseteq \ B_\varepsilon(f(x_0)), \tag{7.38}$$

which is another way of saying that, for all $x \in A$,

$$d_1(x, x_0) < \delta \Longrightarrow d_2(f(x), f(x_0)) < \varepsilon. \tag{7.39}$$

a) **⇒***: Proof that sequence continuity implies ε-δ-continuity:*
We prove this by showing that the opposite assumption, that we have sequence continuity but not ε-δ-continuity, will lead to a contradiction.

So let us assume that there is a function f which is "sequence continuous" at x_0 but not "ε-δ-continuous". Then there exists some $\varepsilon > 0$ such that neither 7.38 nor the equivalent 7.39 is true for any $\delta > 0$.

a.1. *In other words, No matter how small a δ we choose, there is at least one $x = x(\delta) \in A$ such that $d_1(x, x_0) < \delta$ but $d_2(f(x), f(x_0)) \geqq \varepsilon$. So let us choose a whole sequence of such δ values, say $\delta := \delta(m) := 1/m(m \in \mathbb{N})$. For each such m*

(7.40) *there exists some $x_m \in B_{1/m}(x_0) \cap A$; such that; $d_2(f(x_m), f(x_0)) \geqq \varepsilon$.*

a.2. *We now show that the sequence $(x_m)_{m\in\mathbb{N}}$ converges to x_0: Let $\gamma > 0$ and pick $N := N(\gamma) \in \mathbb{N}$ so big that $N > 1/\gamma$, i.e., $1/N < \gamma$. We obtain for any $m \geqq N$ that*

$$d_1(x_m, x_0) < 1/m \leqq 1/N < \gamma$$

(we used (7.40) for the first inequality) and it is proved that $x_m \to x_0$.

a.3. *It is clear that the sequence $\big(f(x_m)\big)_{m\in\mathbb{N}}$ does not converge to $f(x_0)$ as that requires $d_2(f(x_m), f(x_0)) < \varepsilon$ for all sufficiently big m, contrary to (7.40) 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 "ε-δ-continuity" implies "sequence continuity":*
Let $x_n \to x_0$. Let $y_n := f(x_n)$ and $y := f(x_0)$. We must prove that $y_n \to y$ as $n \to \infty$.

b.1. *Let $\varepsilon > 0$. We can find $\delta > 0$ such that (7.38) and hence (7.39) are satisfied. We assumed that $x_n \to x_0$. Hence there exists $N := N(\delta) \in \mathbb{N}$ such that $d_1(x_n, x_0) < \delta$ for all $n \geqq N$.*

b.2. *It follows from (7.39) that $d_2(y_n, y) = d_2(f(x_n), f(x_0)) < \varepsilon$ for all $n \geqq N$. In other words, $y_n \to y$ as $n \to \infty$ and the proof of "$\Leftarrow$" is finished.* ■

[1] B/G: Art of Proof defines in appendix A, p.136, continuity of a function f as follows: "f^{-1}(open) $=$ open". The following proposition proves that their definition coincides with the one given here: the validity of 7.35 for all $x_0 \in X$.

a) *Note that f now is defined on all of X in the interest of avoiding additional definitions and propositions concerning "metric subspaces" A of metric spaces X and how their open sets relate to those of X.*

b) *Also note that this next proposition addresses continuity of f for **all** $x \in X$ and **not** at a specific x_0.*

Proposition 7.8 ("f^{-1}(open) $=$ open" continuity)**.** *Let (X, d_1) and (Y, d_2) be two metric spaces and let $f(\cdot) : X \to Y$ be a mapping from X to Y. Then $f(\cdot)$ is continuous if and only if the following is true: Let V be an open subset of Y. Then the inverse image $f^{-1}(V)$ is open in X.*

Proof of "$\Rightarrow$": Let V be an open set in Y. Let $U := f{-}1(V)$, $a \in U$ and $b := f(a)$. Then $b \in V$ by the definition of inverse images. b is inner point of the open set V and there is $\varepsilon > 0$ such that $B_\varepsilon(b) \subseteq V$. It follows from thm.7.8 (ε-δ characterization of continuity) that there is $\delta > 0$ such that $f(B_\delta(a)) \subseteq B_\varepsilon(b)$. It follows from the monotonicity of direct and inverse images and prop.4.1 on p.26 that

$$B_\delta(a) \subseteq f^{-1}\big(f(B_\delta(a))\big) \subseteq f^{-1}(B_\varepsilon(b)) \subseteq f^{-1}(V) = U.$$

It follows that the arbitrarily chosen $a \in U$ is an interior point of U and this proves that U is open.

Proof of "$\Leftarrow$": We now assume that all inverse images of open sets in Y are open in X. Let $a \in X, b = f(a)$ and $\varepsilon > 0$. We must find $\delta > 0$ such that $f(B_\delta(a)) \subseteq B_\varepsilon(b)$. Let $U := f^{-1}\big(B_\varepsilon(b)\big)$. Then U is open as the inverse image of the open neighborhood $B_\varepsilon(b)$ and there will be $\delta > 0$ such that $B_\delta(a) \subseteq U$. It follows from the monotonicity of direct and inverse images and prop.4.6 on p.27 that

$$f\big(B_\delta(a)\big) \subseteq f(U) = \qquad f\big(f^{-1}(f(B_\varepsilon(b)))\big) = B_\varepsilon(b) \cap f(X) \subseteq B_\varepsilon(b).$$

■

Remark 7.4 (continuity for real functions of real numbers)**.** Let $(X, d_1) = (Y, d_2) = \mathbb{R}$. In this case equation (7.39) on p.77 looks like this:

$$|x - x_0| < \delta \Longrightarrow |f(x) - f(x_0)| < \varepsilon$$

Proposition 7.9 (continuity of the identity mapping)**.** *Let $X, d)$ be a metric space and*

$$id(\cdot) : E \to E \qquad x \mapsto x$$

be the identity function on E. Then $id(\cdot)$ is continuous.

Proof: Given any $\varepsilon > 0$, let $\delta := \varepsilon$. Let $x, y \in X$. Assume that $d(x, y) < \delta$. Then

$$d(id(x), id(y)) = d(x, y) < \delta = \varepsilon$$

and we have satisfied condition (7.39) of the $\varepsilon - \delta$ characterization of continuity. This proves that the identity mapping is continuous. ■

7.2.2 Continuity of constants and sums and products

For all the following, unless stated differently, let (X, d) be a metric space and $A \subseteq X$. Let

$$f : A \longrightarrow \mathbb{R}$$
$$g : A \longrightarrow \mathbb{R}$$

be two real functions which both are continuous in a point $x_0 \in A$. Moreover, let a, b be two (constant) real numbers. You can think of any constant number a as a function on $\mathbb{R}$ as follows:

$$a(\cdot) : \quad X \longrightarrow \mathbb{R} \qquad x \longmapsto a$$

In other words, the function $a(\cdot)$ assigns to each $x \in X$ one and the same value a. We called such a function a constant function (see (5.3) on p.33).

Proposition 7.10. *Given is a metric space (X, d).*
Let $f(\cdot), g(\cdot), f_1(\cdot), f_2(\cdot), f_3(\cdot), \dots, f_n(\cdot) : A \to \mathbb{R}$ all be continuous functions in $x_0 \in A \subseteq X$. Then

a: Constant functions are continuous everywhere on $\mathbb{R}$.

b: The product $fg(\cdot) : x \mapsto f(x)g(x)$ is continuous in x_0. Especially $af(\cdot)x \mapsto a \cdot f(x)$ is continuous in x_0 and , using -1 as a constant, $-f(\cdot) : x \mapsto -f(x)$ is continuous in x_0

c: The sum $f + g(\cdot) : x \mapsto f(x) + g(x)$ is continuous in x_0

d: Any "linear combination" $\sum\limits_{j=0}^{n} a_j f_j(\cdot) : x \mapsto \sum\limits_{j=0}^{n} a_j f_j(x)$ is continuous in x_0.

Proof of a: Let $\varepsilon > 0$. We do not even have to look for a suitable δ to restrict the distance between two arguments x and x_0 because it is always true that

$$|a(x) - a(x_0)| = |a - a| = 0 < \varepsilon$$

and we are done.

Proof of b: In the following chain of calculations each inequality results from applying the triangle inequality (3.3) which states, just to remind you, that $|a + b| \leqq |a| + |b|$ for any two real numbers a and b:

$$\begin{aligned}
&|f(x_0)g(x_0) - f(x)g(x)| \\
&= |f(x_0)g(x_0) - f(x)g(x_0) + f(x)g(x_0) - f(x)g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad |f(x)| \cdot |g(x_0) - g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad |f(x) - f(x_0) + f(x_0)| \cdot |g(x_0) - g(x)| \\
&\leqq |g(x_0)| \cdot |f(x_0) - f(x)| \quad + \quad (|f(x) - f(x_0)| + |f(x_0)|) \cdot |g(x_0) - g(x)|
\end{aligned}$$

Now write x_n rather than x and assume that (x_n) is a sequence which converges to x_0 and we have just shown that

$$|f(x_0)g(x_0) - f(x_n)g(x_n)| \leqq K_1 + K_2 \tag{7.41}$$

where

$$\begin{aligned}
K_1 &= |g(x_0)| \cdot |f(x_0) - f(x_n)| \\
K_2 &= (|f(x_n) - f(x_0)| + |f(x_0)|) \cdot |g(x_0) - g(x_n)|
\end{aligned}$$

The continuity of $f(\cdot)$ and $g(\cdot)$ in x_0 and the convergence $x_n \to x_0$ for $n \to \infty$ implies that $f(x_n) \to f(x_0)$ and $g(x_n) \to g(x_0)$ (see (7.36) on p.77). So both $|f(x_0) - f(x_n)|$ and $|g(x_0) - g(x_n)|$ will converge to zero as $n \to \infty$ and the same will be true if those expressions are multiplied by the constant value $|g(x_0)|$, no matter how big it may be, or by $|f(x_n) - f(x_0)| + |f(x_0)|$ (for big n, $f(x_n)$ is very close to $f(x_0)$ so that $|f(x_n) - f(x_0)| + |f(x_0)|$ will be bounded by the constant value $1 + |f(x_0)|$) for big enough n. This means that both K_1 and K_2 will converge to zero and (7.41) shows that $fg(x_n) = f(x_n)g(x_n)$ converges to $fg(x_0)$ as $n \to \infty$. But we made no special assumption about (x_n) besides its converging against x_0 and we have proved the continuity of $(fg)(\cdot)$ in x_0. This concludes the proof of b.

Proof of c: Let $\varepsilon > 0$ and let $\tilde{\varepsilon} = \frac{\varepsilon}{2}$. Because $f(\cdot)$ and $g(\cdot)$ are both continuous in x_0, there is $\delta > 0$ such that $|f(x_0) - f(x_n)| < \tilde{\varepsilon}$ and $|g(x_0) - g(x_n)| < \tilde{\varepsilon}$ Again, we make heavy use of the triangle inequality:

$$\begin{aligned}
|f(x_0) + g(x_0) \ - \ (f(x_n) + g(x_n))| \quad &= \quad |(f(x_0) - f(x)) \ + \ (g(x_0) - g(x))| \\
&\leqq \quad |f(x_0) - f(x)| \ + \ |g(x_0) - g(x))| \\
&\leqq \quad \tilde{\varepsilon} + \tilde{\varepsilon} \quad = \quad \varepsilon
\end{aligned}$$

and we are done with the proof of c.

proof of d: For linear combinations of two functions f_1 and f_2, the proof is obvious from parts a, b and c. The proof for sums of more than two terms needs a simple (strong) induction argument: Write

$$\sum_{j=0}^{n+1} a_j f_j(x) \ = \Big(\sum_{j=0}^{n} a_j f_j(x) \Big) \ + \ a_{n+1} f_{n+1}(x) \ = \ I \ + \ II.$$

The left term "I" is continuous by the induction assumption and the entire sum $I \ + \ II$ then is continous as the sum of two continuous functions. ■

7.2.3 Function spaces (Understand this!)

Definition 7.22 (linear combinations (imprecise))**.** The following definitions were discussed in the chapter on vector spaces (see def.6.6 on p.52 and def.6.7 on p.52). As that material is optional, they are repeated here in abbreviated form for your convenience.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \dots, \boldsymbol{X}_n$ be a finite number of items for which it makes sense to multiply them with real numbers $a_1, a_2, a_3, \dots, a_n$ and to add or subtract them. We call the finite sum

$$\sum_{j=0}^{n} a_j \boldsymbol{X}_j = a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + a_3\boldsymbol{X}_3 + \ldots + a_n\boldsymbol{X}_n \tag{7.42}$$

a **linear combination** of the $\boldsymbol{X}_j$ items. The multipliers $a_1, a_2, \dots$ are called **scalars** in this context.

Definition 7.23 (linear mappings (imprecise))**.** Linear mappings also were treated in greater detail in the chapter on vector spaces Again, this is an abbreviated presentation for your convenience.

Let L_1, L_2 be two non–empty sets which contain with any elements $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ also any linear combination [19]

$$\sum_{j=0}^{n} a_j \boldsymbol{X}_j = a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + a_3\boldsymbol{X}_3 + \ldots + a_n\boldsymbol{X}_n.$$

Let $F(\cdot) : L_1 \to L_2$ be a mapping with the following properties:

$$F(x+y) = F(x) + F(y) \quad \forall x, y \in L_1 \qquad \textbf{(additivity)} \tag{7.43a}$$

$$F(\alpha x) = \alpha F(x) \quad \forall x \in L_1,\ \forall \alpha \in \mathbb{R} \qquad \textbf{homogeneity} \tag{7.43b}$$

Then we call $F(\cdot)$ a **linear mapping**.

It is easy enough to show that conditions (7.43a) and (7.43b) are equivalent to demanding that

$$F(\sum_{j=0}^{n} a_j \boldsymbol{X}_j) = \sum_{j=0}^{n} a_j F(\boldsymbol{X}_j) \tag{7.44}$$

for any linear combination in L_1. [20]

Example 7.9. It is important that you understand the following: Let $A \neq \emptyset$ and $f_j(\cdot) : A \longrightarrow \mathbb{R}$ a sequence of real functions on A. We set $\boldsymbol{X}_j := f_j(\cdot)$ and in this way create linear combinations of real-valued functions:

$$\sum_{j=0}^{n} a_j f_j(\cdot) \ :\ x \mapsto a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x) \tag{7.45}$$

is also a function which is defined on A. In other words, the set

$$\mathscr{F}(A, \mathbb{R}) = \{f(\cdot) : f(\cdot) \text{ is a real function on } A\} \tag{7.46}$$

satisfies the condition in def.7.23 that it contains all its linear combinations. In fact, $\mathscr{F}(A, \mathbb{R})$ is a vector space in the sense of def.6.4 on p.49 and so is its subset $\mathscr{B}(A, \mathbb{R})$ of all bounded functions.

[19] this "closure" with respect to linear combinations is the most important property of vector spaces
[20] see (6.4) on p.53 if you know about vector spaces.

Do not worry about the vector space property if you did not previously learn about vector spaces. Instead, review the definition (7.22) on p.81 of linear combinations. The most important aspect of vector spaces is that, with any finite number of elements, they will also contain all linear combinations you can build with them. Part d of the prop.7.10 on p.79 proves that continuous functions on any non–empty set satisfy exactly that property.

Remember though the notation $\mathscr{C}(X,\mathbb{R})$ and $\mathscr{C}_{\mathscr{B}}(X,\mathbb{R})$ for continuous and continuous bounded real functions on a set X. Besides, you will learn in the next section that if X is a bounded and closed subset of $\mathbb{R}$ then any continuous real function on X is also bounded, i.e.,

$$\mathscr{C}_{\mathscr{B}}(X,\mathbb{R}) \;=\; \mathscr{C}(X,\mathbb{R}) \qquad \textit{if} \quad X \subseteq \mathbb{R} \textit{ is closed and bounded.}$$

Example 7.10 (Vector space of continuous real functions)**.** The set

$$\mathscr{C}(X,\mathbb{R}) \;:=\; \{f(\cdot) : f(\cdot) \text{ is a continuous real function on } X\}$$

of all real continuous functions on an arbitrary non–empty set X is a vector space if you define addition and scalar multiplication as in (5.2) on p.33. The reason is that you can verify the properties A, B, C of a vector space by looking at the function values for a specific argument $x \in X$ and for each one fo those you just deal with ordinary real numbers. The "sup–norm"

$$\|f(\cdot)\| \;=\; \sup\{|f(x)| : x \in X\}$$

(see (7.3) on p.63) is **not a real function** on all of $\mathscr{C}(X,\mathbb{R})$ because $\|f(\cdot)\| = +\infty$ for any unbounded $f(\cdot) \in \mathscr{C}(X,\mathbb{R})$.

The subset

$$\mathscr{C}_{\mathscr{B}}(X,\mathbb{R}) \;:=\; \{h(\cdot) : h(\cdot) \text{ is a bounded continuous real function on } X\}$$

(see (7.1) on p. 63) is a subspace of the normed vector space of all bounded real functions on X. On this subspace the sup–norm truly is a real function in the sense that $\|f(\cdot)\| < \infty$.

7.2.4 Continuity of Polynomials (Understand this!)

Definition 7.24 (polynomials)**.** Anything that has to do with polynomials takes place in $\mathbb{R}$ and not on a metric space.

Let A be subset of the real numbers and let $p(\cdot) : A \to \mathbb{R}$ be a real function on A. $p(\cdot)$ is called a **polynomial**. if there is an integer $n \geqq 0$ and real numbers $a_1, a_2, \dots, a_n$ which are constant (they do not depend on x) so that $p(\cdot)$ can be written as a sum

$$p(x) \quad = \quad a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \tag{7.47}$$

Remember that $x^0 = 1$ and $x^1 = x$ and we have

$$p(x) \quad = \quad a_0x^0 + a_1x^1 + a_2x^2 + \ldots + a_nx^n \quad = \quad \sum_{j=0}^{n} a_jx^j \tag{7.48}$$

In other words, polynomials are linear combinations of the **monomials** $x \to x^k \quad (k \in (N)_0$.

Proposition 7.11 (All polynomials are continuous)**.**

Proof: It suffices to show that the monomials $m_j(x) := x^j$ *are continuous for all* $j = 0, 1, 2, \dots$ *because of proposition* (7.10)*, part d and because all polynomials are linear combinations of monomials.* $m_0(\cdot)$ *is continuous because it is the constant function* $x \to 1$*.* $m_1(\cdot) : x \to x$ *is continuous according to thm 7.8* [21] *(p.77) because for any given* $\varepsilon > 0$ *we choose* $\delta := \varepsilon$ *and this will ensure that* $|m_1(x) - m_1(y)| < \varepsilon$ *whenever* $|x - y| < \delta$*. But if* $m_1(\cdot)$ *is continuous then so is the product* $m_2(\cdot) = m_1(\cdot)m_1(\cdot)$*. But then so is the product* $m_3(\cdot) = m_2(\cdot)m_1(\cdot)$*. But then so is the product* $m_j(\cdot) = m_{j-1}(\cdot)m_1(\cdot)$ *for any choice of* $j > 0$*. We have shown that all monomials are continuous and so are polynomials as their linear combinations.* ■

Proposition 7.12 (Vector space property of polynomials)**.** *Sums and scalar products of polynomials are polynomials.*

Proof of a. Additivity:
Let

$$p_1(x) \;=\; a_0 + a_1 x + a_2 x^2 + \ldots + a_n x_1^n \;=\; \sum_{j=0}^{n_1} a_j x^j$$

and

$$p_2(x) \;=\; b_0 + b_1 x + b_2 x^2 + \ldots + b_n x_2^n \;=\; \sum_{j=0}^{n_2} b_j x^j$$

be two polynomials. Might as well assume that $n_1 \leqq n_2$*. Let* $a_{n_1+1} = a_{n_1+2} = \ldots = a_{n_2} = 0$*. This does not change anything and we get*

$$\begin{aligned} p_1(x) + p_1(x) &= \sum_{j=0}^{n_2} a_j x^j \;+\; \sum_{j=0}^{n_2} b_j x^j \\ &= \sum_{j=0}^{n_2} (a_j + b_j) x^j \\ &= \sum_{j=0}^{n_2} c_j x^j \qquad (c_j := a_j + b_j) \end{aligned}$$

This proves that the function $p_1(\cdot) + p_2(\cdot)$ *is of the form* (7.48) *and we have shown that it is a polynomial. The proof for the sum of more than two polynomials now follows by the principle of proof by complete induction (see* (3.2) *on p.8).*

Proof of b. Scalar product:
Let

$$p(x) \;=\; a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \;=\; \sum_{j=0}^{n} a_j x^j$$

[21] besides, $m_1(\cdot)$ is the identity mapping on $\mathbb{R}$ and we know from proposition (7.9) on p.79 that identity mappings are always continuous.

be a polynomial. Let λ be a real number. Then

$$\begin{aligned}(\lambda p)(x) &= \lambda p(x) \;=\; \lambda\sum_{j=0}^{n} a_j x^j \\ &= \sum_{j=0}^{n} \lambda a_j x^j \;=\; \sum_{j=0}^{n} c_j x^j \qquad (c_j := \lambda a_j)\end{aligned}$$

This proves that the function $\lambda p(\cdot)$ is of the form (7.48) and we are done. ■

*Polnomials may not always be given in their **normalized form** (7.48) on p.82. Here is an example:*

$$\begin{aligned} p(x) &= a_0 x^0 (1-x)^n + a_1 x^1 (1-x)^{n-1} + a_2 x^2 (1-x)^{n-2} + \ldots + a_{n-1} x^{n-1} (1-x)^1 + a_n x^n \\ &= \sum_{k=0}^{n} a_k x^k (1-x)^{n-k} \end{aligned}$$

*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***

$$p(x) \quad = \quad \sum_{k=0}^{n} \binom{n}{k} f(\frac{k}{n}) x^k (1-x)^{n-k} \qquad \text{see note}^{22}$$

are of that form.

Example 7.11 (Vector space of polynomials)**.** Let $A \subseteq \mathbb{R}$. I follows from (7.12) and (7.11) that the set

$$\{p(\cdot) : p(\cdot) \text{ is a polynomial on } A\}$$

of all polynomials on an arbitrary non–empty subset A of the real numbers is a subspace of the vector space $\mathscr{C}(A, \mathbb{R})$. (see example (7.10) on p.82. The "sup–norm"

$$\|f(\cdot)\| \;=\; \sup\{|f(x)| : x \in A\}$$

is **not a real function** on the set of all polynomials on A as its value may be ∞.. Matter of fact, it can be shown that, if the set A itself is not bounded, then the only polynomials for which $\|p(\cdot)\| < \infty$ are the constant functions on A(!)

[22] 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)

7.3 Function sequences and infinite series

7.3.1 Convergence of function sequences (Study this!)

Vectors are more complicated than numbers because an n–dimensional vector $v \in \mathbb{R}^n$ represents a grouping of a finite number n of real numbers. Matter of fact, any such vector $(x_1, x_2, x_3, \cdots, x_n)$ can be interpreted as a real function (remember: a real function is one which maps it arguments into $\mathbb{R}$)

$$f(\cdot) : \{1, 2, 3, \cdots, N\} \to \mathbb{R} \qquad j \mapsto x_j \tag{7.49}$$

(see (6.4) on p.44).

Next come sequences $(x_j)_{j\in\mathbb{N}}$ which can be interpreted as real functions

$$g(\cdot) : \mathbb{N} \to \mathbb{R} \qquad j \mapsto x_j \tag{7.50}$$

Finally we deal with any kind of real function

$$h(\cdot) : X \to \mathbb{R} \qquad x \mapsto h(x) \tag{7.51}$$

as the most general case

Now we add more complexity by not just dealing with one or two or three real functions but with an entire sequence

$$f_n(\cdot) : X \to \mathbb{R} \qquad x \mapsto f_n(x) \tag{7.52}$$

For any fixed argument x_0 we have a sequence $f_1(x_0), f_2(x_0), f_3(x_0), \cdots$ which we can examine for convergence. This sequence may converge for some or all arguments $x_0 \in X$ to a real number. Time for some definitions.

Definition 7.25 (Pointwise convergence of function sequences)**.** Let X be a non-empty set, (Y, d) a metric space and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $A \subseteq X$ be a subset of X. We say that $f_n(\cdot)$ **converges pointwise** or, simply, **converges** to $f(\cdot)$ on A and we write $f_n(\cdot) \to f(\cdot)$ if

$$f_n(x) \to f(x) \quad \text{for all } x \in A \tag{7.53}$$

Definition 7.26 (Uniform convergence of function sequences)**.** Let X be a non-empty set, (Y, d) a metric space and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $A \subseteq X$ be a subset of X. We say that $f_n(\cdot)$ **converges uniformly** to $f(\cdot)$ on A and we write [23]

$$f_n(\cdot) \stackrel{uc}{\to} f(\cdot) \tag{7.54}$$

if the following is true: For each $\varepsilon > 0$ (no matter how small) there exists a (probably huge) number n_0 which can be chosen once and for all, independently of the specific argument x, such that

$$d(f_n(x), f(x)) < \varepsilon \quad \text{for all } x \in A \quad \text{and } n \geqq n_0 \tag{7.55}$$

[23] I must confess that " $f_n(\cdot) \stackrel{uc}{\to} f(\cdot)$ " is a notation that I coined myself because it is not as tedious as writing " $f_n(\cdot)\to f(\cdot)$ uniformly"

Remark 7.5 (Uniform convergence implies pointwise convergence)**.** Look at definition (7.15) on p.70 of convergence of sequences and you should immediately see that (7.55) implies, for any given $x \in A$, ordinary convergence $f(x) = \lim_{n\to\infty} f_n(x)$ because the number $n_0 = n_0(\varepsilon)$ chosen in (7.55) will also satisfy (7.27) (p.70) for $x_n = f_n(x)$ and $a = f(x)$.

In other words, unform convergence implies pointwise convergence. But what is the difference between pointwise and uniform convergence? The difference is that, for poinwise convergence, the number n_0 will depend on both ε and x: $n_0 = n_0(\varepsilon, x)$. In the case of uniform convergence, the number n_0 will still depend on ε but can be chosen independently of the argument $x \in A$.

Example 7.12 (a. Constant sequence of functions)**.** Let X be a set and let
$f(\cdot) : X \to \mathbb{R}$ be a real function on X which may or may not be continuous anywhere. Define a sequence of functions

$$f_n(\cdot) : X \to \mathbb{R}\ (n \in \mathbb{N}) \quad \text{as} \quad f_1(\cdot) = f_2(\cdot) = \cdots = f(\cdot)$$

which is just a shorthand of writing that

$$f_1(x) = f_2(x) = \cdots = f(x)\ \forall n \in \mathbb{N},\ \forall x \in X.$$

In other words, we are looking at a constant sequence of functions (not to be confused with a sequence of constant functions – seriously!).

Then $f_n(\cdot) \overset{uc}{\to} f(\cdot)$

Proof of the example a: This is trivial. No matter how small an ε and n_0 we choose and no matter what argument $x \in X$ we are looking at, we have

$$|f_n(x) - f(x)| = 0 < \varepsilon \quad \textit{for all } x \in A \quad \textit{and } n > n_0$$

■

Example 7.13 (b. Pointwise but not uniformly convergent sequence of functions)**.** Let $X = [0,1]$, i.e., X is the closed unit interval $\{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. Let the functions $f_n(\cdot)$ be defined as follows on X:

$$f_n(x) = \begin{cases} n^2 x & \text{for } 0 \leqq x \leqq \frac{1}{n} \\ \frac{1}{x} & \text{for } \frac{1}{n} \leqq x \leqq 1 \end{cases}$$

Note that both pieces fit together in the point $a = 1/n$ because the "$\frac{1}{x}$ definition" gives $f_n(a) = \frac{1}{1/n} = n$ and the "n^2x definition" gives the same value $n = n^2\frac{1}{n}$. We do not give a formal proof that each $f_n(\cdot)$ is continuous in every point of $[0,1]$. Just accept it from the fact that the two graphs flow into each other at the "splicing point" $1/n$.

Now we define the function $f(\cdot) : [0,1] \to \mathbb{R}$ as

$$f(x) = \begin{cases} \frac{1}{x} & \text{for } 0 < x \leqq 1 \\ 0 & \text{for } x = 0 \end{cases}$$

Then the functions $f_n(\cdot)$ converge pointwise but not uniformly to $f(\cdot)$ on the entire unit interval.

Proof of example b, pointwise convergence:
first we look separately at the point $a = 0$. We have $f(0) = 0 = n^2 0 = f^n(0)$ and the constant sequence of zeroes certainly converges against zero. Now assume $a > 0$. If $n > 1/a$ then $f_n(a) = \frac{1}{a}$ for all such n. Again, we have a constant sequence $(1/a)$ except for finitely many n and it converges against $1/a = f(a)$. We have thus proved pointwise convergence.

Proof of example b, no uniform convergence:
To prove that (7.55) is not satisfied, we must find $\varepsilon > 0$ and points x_N so that for no matter how big a natural number N we choose, there will be at least one $n > N$ such that $|f_n(x) - f(x)| \geqq \varepsilon$. Let $N \in \mathbb{N}$ be any natural number. Then

$$f_N(\frac{1}{N^2}) \;=\; \frac{N^2}{N^2} = 1$$

and

$$f_{2N}(\frac{1}{N^2}) \;=\; \frac{(2N)^2}{N^2} \;= 4$$

So

$$|f_{2N}(\frac{1}{N^2}) - f_N(\frac{1}{N^2})| \;= 3$$

To recap: We found $\varepsilon > 0$ so that for each $N \in \mathbb{N}$ we were able to find an $n \geqq N$ and $x_N \in [0, 1]$ such that $|f_n(X_N) - f_N(x_N)| > \varepsilon$: we chose

$$\varepsilon = 2, \quad n = 2N, \quad x_N = \frac{1}{N^2}$$

We have thus prove that the pointwise convergence is not uniform. ■

7.3.2 Infinite Series (Understand this!)

*We start by repeating the definition of a sequence given in section 3.2 on p.9: 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 7.1 (indices of sequences)**.** Unless explicitly stated otherwise, sequences are always indexed $1, 2, 3, \ldots$, i.e., the first index is 1 and, given any index, you obtain the next one by adding 1 to it.

The simplest things that a mathematician deals with are numbers. One nice thing that is always possible with numbers, is that you can add them. Here is a very simple definition:

Definition 7.27 (Numeric Sequences and Series)**.** A sequence (a_j) of real numbers is called a **numeric sequence** if each a_j is a real number. For any such sequence, we can build another sequence (s_n) as follows:

$$s_1 := a_1; \quad s_2 := a_1 + a_2; \quad s_3 := a_1 + a_2 + a_3; \cdots \quad s_n := \sum_{k=1}^{n} a_k \tag{7.56}$$

We call (s_n) the sequence of **partial sums** associated with the sequence (a_k). We also write this more compactly as

$$a_1 + a_2 + a_3 + \cdots \quad = \quad \sum_{k=1}^{\infty} a_k \tag{7.57}$$

and we call any such object, which represents a sequence of partial sums, a **series**. Loosely speaking, a series is an infinite sum. We say that the series converges to a real number x and we write

$$x = \sum_{k=1}^{\infty} a_k \tag{7.58}$$

. if this is true for associated sequence of finite partial sums (7.56). We say that the series has limit ∞ (has limit $-\infty$) if this true for the associated partial sums and we write

$$\sum_{k=1}^{\infty} a_k = \infty \quad (\sum_{k=1}^{\infty} a_k = -\infty) \tag{7.59}$$

Proposition 7.13 (Convergence criteria for series)**.** *A series $s_n := \sum_{k=1}^{n} a_k$ of real numbers possess a limit $a \in \mathbb{R}$ if and only if either of the following two is true:*

$$\Big|\sum_{k=n}^{\infty} a_k\Big| < \varepsilon \qquad \text{for all } n \geqq n_0 \tag{7.60a}$$

$$\Big|\sum_{k=n}^{m} a_k\Big| < \varepsilon \qquad \text{for all } m, n \geqq n_0 \tag{7.60b}$$

Proof: Write

$$a := \sum_{k=1}^{\infty} a_k = \sum_{k=1}^{n} a_k + \sum_{k=n+1}^{\infty} a_k = s_n + \sum_{k=n+1}^{\infty} a_k \tag{7.61}$$

Remember the convergence criteria for numeric sequences. Convergence of a sequence (s_n) to a real number a means that for any $\varepsilon > 0$, no matter how small it may be, all but finitely many members s_n will be inside the ε–neighborhood $B_\varepsilon(a)$ of a. Written in terms of the distance to a this means there exists a suitable $n_0 \in \mathbb{N}$ such that

$$|a - s_n| < \varepsilon \qquad \text{for all } n \geqq n_0$$

(see (7.15) on p.70). According to (7.61) we can write that as

$$\Big|\sum_{k=n+1}^{\infty} a_k\Big| < \varepsilon \qquad \text{for all } n \geqq n_0$$

which is the same as (7.60.a) because it does not matter whether we we look at the sum of all terms bigger than n or $n+1$.

Alternatively, there was the Cauchy criterion

$$|x_i - x_j| < \delta \text{ for all } i, j \geqq n_0$$

(see (7.19) on p.73) which ensures convergence to some number a without specifying what it might actually be. Again we use (7.61) and obtain, assuming without loss of generality that $i < j$,

$$\Big|\sum_{k=i+1}^{j}\Big| < \delta \text{ for all } j > i \geqq n_0$$

$\blacksquare$

It is very important to understand that a series either converges to a finite number or it diverges. If it diverges it may be the case that $\sum\limits_{k=1}^{\infty} a_k = \infty$ *or* $\sum\limits_{k=1}^{\infty} a_k = -\infty$ *or there is no limit at all. As an example for a series which has no limit, look at the oscillating sequence and associated partial sums*

$$a_0 = 1; \quad a_1 = -1; \quad a_2 = 1; \quad a_3 = -1; \cdots \quad s_n = \sum_{k=0}^{n} (-1)^n \tag{7.62}$$

This also 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 saying that the limit of the series is zero because your imagination disregards the odd indices and* $s_0 = s_2 = s_4 = \cdots = s_{2j} = 0$*. Note that for any* $j \in \mathbb{N}$ *we have* $|s_j - s_{j-1}| = 1$ *because at each step we either add or subtract* 1*. This means that no matter what real number* a *and how big a number* $n_0 \in \mathbb{N}$ *we choose, it will never be true that* $|a - s_j| < 1$ *for all* $j \in \mathbb{N}$ *and* a *cannot be a limit of the series.*

Just so you understand the difference between limits and contact points (see (7.17) on p.72): Even though neither a_j *nor* s_j *has a limit, both have two contact points each.* a_j *has the contact points* $\{1, -1\}$ *and* s_j *has the contact points* $\{0, 1\}$*.*

We now turn our attention to convergence properties of series.

Definition 7.28 (Finite permutations)**.** Let $N \in \mathbb{N}$ and let $X_N := \{1, 2, 3, \ldots, N\}$ denote the set of the first N integers. A **permutation** of X_N is a mapping

$$\pi(\cdot) : X_N \to X_N; \qquad j \mapsto \pi(j)$$

which is both surjective: each element k of X_N is the image $\pi(j)$ for a suitable $j \in X_N$ and injective: different arguments $i \neq j \in X_N$ will always map to different images $\pi(i) \neq \pi(j) \in X_N$ (see (3.6) on p.11). You may remember that

$$\text{surjective} + \text{injective} = \text{bijective}$$

and that under our assumptions the inverse mapping

$$\pi^{-1}(\cdot) : X_N \to X_N; \qquad \pi(j) \mapsto \pi^{-1}\pi(j) = j,$$

which associates with each image $\pi(j)$ the unique argument j which maps into $\pi(j)$, exists (see def. 3.6 on p.11 for properties of the inverse mapping).

It is customary to write

$$i_1 \text{ instead of } \pi(1), \quad i_2 \text{ instead of } \pi(2), \quad \ldots, \quad i_j \text{ instead of } \pi(j), \quad \ldots$$

Definition 7.29 (Permutations of $\mathbb{N}$)**.** A **permutation** of $\mathbb{N}$ is a mapping

$$\pi(\cdot) : \mathbb{N} \to \mathbb{N}; \qquad j \mapsto \pi(j)$$

which is both surjective: each element k of $\mathbb{N}$ is the image $\pi(j)$ for a suitable $j \in \mathbb{N}$ and injective: different arguments $i \neq j \in \mathbb{N}$ will always map to different images $\pi(i) \neq \pi(j) \in \mathbb{N}$.

Permutations are the means of describing a reordering of the members of a finite or infinite sequence. 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 7.14 (Absolute Convergence of series with non–negative members)**.** *Let (a_n) be a sequence of non–negative members: $a_n \geqq 0$ for all $n \in \mathbb{N}$. Then one of the following will be true:*

***A**: the series $\sum\limits_{n=1}^{\infty} a_n$ converges to a (finite) number $a \in \mathbb{R}$. In that case*

$$\sum_{n=1}^{\infty} a_n \; = \; \sum_{n=1}^{\infty} a_{\pi(n)} \;\text{ for any permutation } \pi(\cdot) \text{of } \mathbb{N}.$$

***B**: the series $\sum\limits_{n=1}^{\infty} a_n$ has limit ∞. In that case it is true for any permutation $\pi(\cdot)$ of $\mathbb{N}$ that the reordered series $\sum\limits_{n=1}^{\infty} a_{\pi(n)}$ also has limit ∞.*

*Proof of **A**: Let $b_j := a_{\pi(j)}$ and, hence, $a_k = b_{\pi-1(j)}$. Let $N \in \mathbb{N}$. Let*

(7.63) $$\alpha \; := \; \max\{\pi(j) : j \leqq N\} \qquad \text{and} \qquad \beta \; := \; \max\{\pi^{-1}(k) : k \leqq N\}.$$

Note that $\alpha \geqq N$ and $\beta \geqq N$. Because all terms a_j, b_k are non–negative it follows that

$$\sum_{j=1}^{N} b_j \; = \; \sum_{j=1}^{N} a_{\pi(j)} \; \leqq \; \sum_{k=1}^{\alpha} a_k \; \leqq \; \sum_{k=1}^{\alpha} a_k \; + \; \sum_{k=\alpha+1}^{\infty} a_k \; = \; \sum_{k=1}^{\infty} a_k,$$
$$\sum_{k=1}^{N} a_k \; = \; \sum_{k=1}^{N} b_{\pi^{-1}(k)} \; \leqq \; \sum_{j=1}^{\beta} b_j \; \leqq \; \sum_{j=1}^{\beta} b_j \; + \; \sum_{j=\beta+1}^{\infty} b_j \; = \; \sum_{j=1}^{\infty} b_j.$$

We take limits as $N \to \infty$ and it follows that

$$\sum_{j=1}^{\infty} b_j \; \leq \; \sum_{k=1}^{\infty} a_k \quad \text{and} \quad \sum_{k=1}^{\infty} a_k \; \leqq \; \sum_{j=1}^{\infty} b_j.$$

This proves the lemma. ■

Definition 7.30 (absolutely convergent series)**.** A series is **absolutely convergent** if it converges and its limit is unchanged if the indices are permuted.

The last proposition then states that a convergent series of non-negative terms converges absolutely. The next proposition says that any series which converges if you replace its members by their absolute values will in fact converge absolutely.

Proposition 7.15. *Let $\sum\limits_{n=1}^{\infty} a_n$ be a series such that $\sum\limits_{n=1}^{\infty} |a_n|$ converges. Then $\sum\limits_{n=1}^{\infty} a_n$ converges absolutely.*

7.4 Appendix: Addenda to chapter 7

7.4.1 Convergence (Study this!)

Theorem 7.9 (Limits in metric spaces are uniquely determined). *Let (X,d) be a metric space . Let $(x_n)_n$ be a convergent sequence in X Then its limit is uniquely determined.*

Proof: Otherwise there would be two different points $L_1, L_2 \in X$ such that both $\lim\limits_{n\to\infty} x_n = L_1$ and $\lim\limits_{n\to\infty} x_n = L_2$ Let $\varepsilon := d(L_1,L_2)/2$. There will be $N_1, N_2 \in \mathbb{N}$ such that

$$d(x_n,L_1) < \varepsilon \; \forall n \geqq N_1 \;\; and d(x_n,L_2) < \varepsilon \; \forall n \geqq N_2.$$

It follows that, for $n \geqq N_1 + N_2$,

$$d(L_1,L_2) \;\leqq\; d(L_1,x_n) + d(x_n,L_2) \;<\; 2\varepsilon \;=\; d(L_1,L_2)$$

and we have reached a contradiction. ■

7.4.2 Completeness (Study this!)

The following is the reverse of thm.7.7.

Theorem 7.10 (Closed subsets of a complete space are complete). *Given is a complete metric space (X,d).*
Let $A \subseteq X$ be closed. Then A is complete.

Proof: Let $(x_n)_n$ be a Cauchy sequence in A. We must show that there is $a \in A$ such that $x_n \to a$.
(x_n) also is Cauchy in X. Because X is complete there exists $x \in X$ such that $x_n \to x$. If we can show that x is a contact point of A then we are done: As the set A is assumed to be closed, it contains all its contact points. It follows that $x \in A$, i.e., the arbitrary Cauchy sequence (x_n) in A converges to an element of A. We conclude that A is complete. ■

Theorem 7.11 (Convergent sequences are Cauchy). *Let $(x_{n_j})_n$ be a convergent sequence in a subset A of a metric space (X,d). Then $(x_{n_j})_n$ is a Cauchy sequence (in A).*

Proof: Let $x_n \to L$ ($L \in A$). Let $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that

$$k \geqq N \;\;\Rightarrow\;\; d(x_k,L) \;<\; \varepsilon/2. \qquad (\star) \tag{7.64}$$

It follows from $(\star)$ that, for any $i,j \geqq N$,

$$i,j \geqq N \;\;\Rightarrow\;\; d(x_i,x_j)\; leqq \; d(x_i,L) + d(L,x_j) \;<\; \varepsilon/2 + \varepsilon/2 \;=\; \varepsilon. \tag{7.65}$$

It follows that the sequence satisfies (7.33) of the definition 7.19 on p.73 of a Cauchy sequence. ■

7.4.3 Uniform convergence and metric subspaces (Study unif conv, understand subspaces!)

Theorem 7.12 (Uniform limits of continuous functions are continuous)**.** *Let (X, d_1) and (Y, d_2) be metric spaces and let $f_n(\cdot) : X \to Y$ and $f(\cdot) : X \to Y$ be functions on X $(n \in \mathbb{N})$. Let $x_0 \in X$ and let $V \subseteq X$ be a neighborhood of x_0. Assume **a)** that the functions $f_n(\cdot)$ are continuous at x_0 for all n and **b)** that $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ on V. Then f is continuous at x_0*

Proof: Let $\varepsilon > 0$.

A. Uniform convergence $f_n(\cdot) \overset{uc}{\to} f(\cdot)$ on V guarantees the existence of some $N = N(\varepsilon)$ such that

$$d_2\big(f_n(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V \text{ and } n \geqq \mathbb{N}.$$

In particular, for $n = N$,

$$d_2\big(f_N(x), f(x)\big) < \frac{\varepsilon}{3} \text{ for all } x \in V\ . \tag{7.66}$$

B. All functions f_n and in particular f_N are continuous in V. There is $\tilde\delta > 0$ such that

$$d_2(f_N(x), f_N(x_0)) < \frac{\varepsilon}{3} \text{ for all } x \in B_{\tilde\delta}(x_0). \tag{7.67}$$

C. As x_0 is an interior point of V, there exists $\hat\delta > 0$ such that $B_{\hat\delta}(x_0) \subseteq V$. Let δ be the smaller of $\hat\delta$ and $\tilde\delta$. Then (7.66) and (7.67) both hold for $x \in B_\delta(x_0)$. We note that $x_0 \in B_\delta(x_0)$ and obtain

$$d(f(x), f(x_0)) \leqq d(f(x), f_N(x)) + d(f_N(x), f_N(x_0)) + d(f_N(x_0), f(x_0)) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

The proof is finished. ■

Definition 7.31 (Restriction/Extension of a function)**.** Given are three non-empty sets $A \subseteq X$ and Y. Let $f : X \to Y$ a function with domain X. We define the **restriction of f to A** as the function

$$f\big|_A : A \to Y \text{ defined as } f\big|_A(x) := f(x)\ (x \in A) \tag{7.68}$$

Conversely let $f : A \to Y$ and $\varphi : X \to Y$ be functions such that $f = \varphi\big|_A$. We then call φ an **extension** of f to X.

Notation 7.1. As the only difference between f and $f\big|_A$ is the domain, it is customary to write f instead of $f\big|_A$ to make formulas look simpler if doing so does not give rise to confusions.

Definition 7.32 (Metric subspaces)**.** Given is a metric space (X, d) and a non–empty $A \subseteq (X, d)$. Let $d\big|_{A\times A} : A \times A \to \mathbb{R}_{\geq 0}$ be the restriction $d\big|_{A\times A}(x, y) := d(x, y)(x, y \in A)$ of the metric d to $A \times A$. It is trivial to verify that $(A, d\big|_{A\times A})$ is a metric space in the sense of def.7.1 on p.59. We call $(A, d\big|_{A\times A})$ a **metric subspace** of (X, d) and we call $d\big|_{A\times A}$ the **metric induced by d** or the **metric inherited from** (X, d).

Remark 7.6. Such a metric subspace comes with its own collections of open and closed sets, neighborhoods, ε-neighborhoods, convergent sequences, ... and you must watch out when looking at statements and their proofs whether those concepts refer to the entire space (X,d) or to the subspace $(A, d\big|_{AxA})$.

Notation 7.2. a) As the only difference between d and $d_{A\times A}$ is the domain, it is customary to write d instead of $d_{A\times A}$ to make formulas look simpler if doing so does not give rise to confusions.

b) We often shorten "open in $(A, d\big|_{A\times A})$" to "open in A", "closed in $(A, d\big|_{A\times A})$" to "closed in A", "convergent in $(A, d\big|_{A\times A})$" to "convergent in A", ...

Proposition 7.16 (ε-neighborhoods are a topology basis)**.** *Let (X,d) be a metric space. Then the set $\mathscr{B}_1 := \{B_\varepsilon(x) : x \in X, \varepsilon > 0\}$ is a basis for the topology of (X,d) (see 7.12 on p.69) and the same is true for the "thinner" set $\mathscr{B}_2 := \{B_{1/n}(x) : x \in X, n \in \mathbb{N}\}$.*

Proof: To show that $\mathscr{B}_1$ (resp., $\mathscr{B}_2$) is a basis we must prove that any open subset of X can be written as a union of (open) sets all of which belong to $\mathscr{B}_1$ (resp., $\mathscr{B}_2$). We shall prove this for $\mathscr{B}_2$.

Let $U \subseteq X$ be open. As any $x \in U$ is an interior point of U we can find some $\varepsilon = \varepsilon(x) > 0$ such that $B_{\varepsilon(x)}(x) \subseteq U$. We note that for any such $\varepsilon(x)$ there is $n(x) \in \mathbb{N}$ such that $1/n(x) \leqq \varepsilon(x)$. We observe that $U \subseteq \bigcup \big[B_{1/n(x)}(x) : x \in U\big] \subseteq U$. The first inclusion follows from the fact that $\{x\} \subseteq B_{1/n(x)}(x)$ for all $x \in U$ and the second inclusion follows from $B_{1/n(x)}(x) \subseteq U$ and the inclusion lemma (lemma 4.1 on p.22). It follows that $U = \bigcup \big[B_{1/n(x)}(x) : x \in U\big]$ and we have managed to represent our open U as a union of elements of $\mathscr{B}_2$. This proves that $\mathscr{B}_2$ is a basis for the topology of X, d. As $\mathscr{B}_2 \subseteq \mathscr{B}_1$ it follows that $\mathscr{B}_1$ also is such a basis. ■

Proposition 7.17 (Open sets in A as traces of open sets in X)**.** *Let (X,d) be a metric space and $A \subseteq X$ a nonempty subset of X, viewed as a metric subspace $(A, d\big|_{A\times A})$ of (X,d) (see def.7.32 on p.92). For $\varepsilon > 0$ and $a \in A$ let $B^A_\varepsilon(a)$ denote the ε-neighborhood of a in the subspace $(A, d\big|_{A\times A})$. In other words,*

$$B^A_\varepsilon(a) \;=\; \{x \in A : d\big|_{A\times A}(x,a) < \varepsilon\}. \tag{7.69}$$

Then

$$B^A_\varepsilon(a) \;=\; B_\varepsilon(a) \cap A. \tag{7.70}$$

More generally, a set $U \subseteq A$ is open in A if and only if there is an open $V \subseteq$ in (X,d) such that

$$U \;=\; V \cap A, \tag{7.71}$$

*i.e., U is the **trace** of an open set V in X.*

Proof: First we shall prove (7.70). As $d\big|_{A\times A}$ is the restriction of d to $A \times A$ it follows that

$$\begin{aligned} B^A_\varepsilon(a) &= \{x \in A : d(x,a) < \varepsilon\} \;=\; \{x \in A : d(x,a) < \varepsilon\} \cap A \\ &= \{x \in X : d(x,a) < \varepsilon\} \cap A \;=\; B^A_\varepsilon(a) \cap A \end{aligned}$$

Next we prove that if V is open in X then $U := V \cap A$ is open in the subspace A: Let $x \in U$. We must prove that x is an interior point of U with respect to $(A, d\big|_{A\times A})$ of (X,d). $x \in V$ and V is open in X. Hence there *is $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq V$. It follows that $B^A_\varepsilon(x) = B_\varepsilon(x) \cap A \subseteq u$ and $B^A_\varepsilon(x)$ is open in A, hence x is interior point of U.*

Finally we prove that if U is open in the subspace A then we can find $V \subseteq X$ which is open in X such that $U = V \cap A$: We can write $U = \bigcup \left[B^A_{\varepsilon(x)}(x) : x \in U\right]$ for suitable $\varepsilon(x) > 0$ (see the proof of prop.7.16 on p.93). Let $V := \bigcup \left[B_{\varepsilon(x)}(x) : x \in U\right]$ we have

$$\begin{aligned} V \cap A &= A \cap \bigcup \left[B_{\varepsilon(x)}(x) : x \in U\right] = \bigcup \left[B_{\varepsilon(x)}(x) \cap A : x \in U\right] \\ &= \bigcup \left[B^A_{\varepsilon(x)}(x) : x \in U\right] = U \end{aligned}$$

and this finishes the proof. ■

Proposition 7.18 (Distributivity of unions and intersections)**.** *Let $(A_i)_{i \in I}$ be an arbitrary family of sets and let B be a set. Then*

$$\bigcup_{i \in I} (B \cap A_i) = B \cap \bigcup_{i \in I} A_i, \tag{7.72}$$

$$\bigcap_{i \in I} (B \cup A_i) = B \cup \bigcap_{i \in I} A_i. \tag{7.73}$$

Proof: We only prove (7.72).
Proof of "$\subseteq$": It follows from $B \cap A_i \subseteq A_i$ for all i that $\bigcup_i (B \cap A_i) \subseteq \bigcup_i A_i$. Moreover, $B \cap A_i \subseteq B$ for all i implies $\bigcup_i (B \cap A_i) \subseteq \bigcup_i B$ which equals B. It follows that $\bigcup_i (B \cap A_i)$ is contained in the intersection $\left(\bigcup_i A_i\right) \cap B$.
Proof of "$\supseteq$": Let $x \in B \cap \bigcup_i A_i$. Then $x \in B$ and $x \in A_{i^\star}$ for some $i^\star \in I$, hence $x \in B \cap A_{i^\star}$, hence $x \in \bigcup_i (B \cap A_i)$. ■

7.4.4 The Hahn-Banach separation theorem [24] - Skip this!

Theorem 7.13 (Hahn-Banach)**.** *Let V be a vector space over $\mathbb{R}$ and $p : V \to \mathbb{R}$ a sublinear mapping. Suppose F is a (linear) subspace of V and $f : F \to \mathbb{R}$ is a linear mapping with $f \leqq p$ on F. Then there is an extension of f to a linear map $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f} \leqq p$ on V.*

Before proving this theorem, first we shall prove two lemmata.

Lemma 7.1. *Suppose F is a subspace of V and $a \in V \setminus F$.*
Let $k \in \mathbb{R}$ and $\tilde{f}(x + \lambda a) := f(x) + \lambda k$, i.e., $k = \tilde{f}(a)$. Then

$$k \leqq \inf_{u \in F} \{p(u + a) - f(u)\} \iff \tilde{f}(x + \lambda a) \leqq p(x + \lambda a) \text{ for all } \lambda > 0 \text{ and } x \in F, \tag{7.74}$$

$$k \geqq \sup_{v \in F} \{f(v) - p(v - a)\} \iff \tilde{f}(x + \lambda a) \leqq p(x + \lambda a) \text{ for all } \lambda < 0 \text{ and } x \in F. \tag{7.75}$$

***Proof of 7.74, $\Rightarrow$):** Let us assume that $\lambda > 0$. Then, on account of the left side of (7.74),*

$$\tilde{f}(x + \lambda a) = f(x) + \lambda k = \lambda\big(f(x/\lambda) + k\big) \leqq \lambda\Big(f(x/\lambda) + \big(p(x/\lambda + a) - f(x/\lambda)\big)\Big) = \lambda p(x/\lambda + a)$$

[24] **This chapter is optional**. The proof given here is a more detailed version of the one found in [2] Choquet.

We use the positive homogeneity of p: $\lambda p(x/\lambda + a) \;=\; p(x+\lambda a)$ to obtain $\tilde f(x+\lambda a) \;\leqq\; p(x+\lambda a)$.

Proof of 7.75, $\Rightarrow$): *Let us assume that $\lambda < 0$. Because of the left side of (7.75) and $\lambda < 0$ and positive homogeneity of p,*

$$\begin{aligned} k \;\geqq\; f(v) - p(v-a) \;&\Rightarrow \lambda k \;\leqq\; f(\lambda v) - \lambda p(v-a) \\ &\Rightarrow \;-\, f(\lambda v) + \lambda k \;\leqq (-\lambda) p(v-a) \;=\; p\big((-\lambda)(v-a)\big) \;=\; p((-\lambda)v + \lambda a). \end{aligned}$$

We substitute $v := x/\lambda \in F$:

$$-f(x) + \lambda k \;\leqq p(-x+\lambda a), \;\; \textit{hence} \;\; \tilde f(-x+\lambda a) \;=\; f(-x) + \lambda k \;\leqq\; p(-x+\lambda a)$$

We can switch from $-x$ to x as the above holds for all x in the subspace F and because $-x \in F$ iff $x \in F$. It follows that p indeed dominates $\tilde f$ for all $x \in F$ and $\lambda < 0$.

Proof of 7.74, $\Leftarrow$): *we assume $\tilde f(x+\lambda a) \;\leqq\; p(x+\lambda a)$ for all $\lambda > 0$ and $x \in F$. We shall show that $k = \tilde f(a) \;\leqq\; p(u+a) - f(u)$ for all $u \in F$.*

$$p(u+a) - f(u) \;\geqq \tilde f(u+a) - f(u) \;=\; \tilde f(u) + \tilde f(a) - f(u) \;=\; f(u) + \tilde f(a) - f(u) \;=\; \tilde f(a) \;=\; k.$$

Proof of 7.75, $\Leftarrow$): *we assume $\tilde f(x+\lambda a) \;\leqq\; p(x+\lambda a)$ for all $\lambda < 0$ and $x \in F$. We shall show that $k = \tilde f(a) \;\geqq\; f(v) - p(v-a)$ for all $v \in F$.*

$$-p(v-a) + f(v) \;\leqq\; -\,\tilde f(v-a) + f(v) \;=\; \tilde f(a-v) + f(v) \;=\; \tilde f(a) - \tilde f(v) + f(v) \;=\; \tilde f(a) \;=\; k.$$

■

Lemma 7.2. *Let $F \subset V$ be a genuine subspace of V and $a \in V \setminus F$. Let $G := span(F \uplus \{a\}$ be the subspace of all linear combinations that can be created by a and or vectors in F. Then there exists an extension $\tilde f$ of f to G.*

Proof. For $u, v \in F$ *we have*

$$f(u) + f(v) \;= f(u+v) \leqq p(u+v) \;=\; p\big((u+a) + (v-a)\big) \;\leqq\; p(u+a) + p(v-a)$$

and hence $f(v) - p(v-a) \;\leqq\; p(u+a) - f(u)$. Therefore

$$\sup_{v\in F}\{f(v) - p(v-a)\} \;\leqq\; \inf_{u\in F}\{p(u+a) - f(u)\}.$$

Now for a fixed $k \in \mathbb{R}$, we let $\tilde f(x+\lambda a) \;=\; f(x) + \lambda k$. We claim that $\tilde f \leqq p$ iff we have

$$\sup_{v\in F}\{f(v) - p(v-a)\} \;\leqq k \;\leqq\; \inf_{u\in F}\{p(u+a) - f(u)\} \tag{7.76}$$

which will conclude the proof since such a k exists by the above work. Our claim holds because $f(x) + \lambda k \;=\; \tilde f(x+\lambda a) \;\leqq\; p(x+\lambda a)$ for all λ iff

$$\begin{aligned} k \;&\leqq p(u+a) - f(u) \quad \textit{for all } u \in F \\ \textit{and} \;\; k \;&\geqq f(v) - p(v-a) \quad \textit{for all } v \in F \end{aligned}$$

(the cases $\lambda > 0$ *and* $\lambda < 0$ *respectively). This is proved above in lemma 7.1.* ■

From this we also deduce that the extension $\tilde{f}$ *is unique iff* $\sup_{v \in E}\{f(v) - p(v-a)\} = \inf_{u \in E}\{p(u+a) - f(u)\}$ *(the case in which* k *in the proof is uniquely determined (see (7.76)).*

Proof:

■

8 Compactness (Study this!)

8.1 Introduction: Closed and bounded sets in Euclidian space (Understand this!)

One of the results that are true for N-dimensional space is the "sequence compactness" of closed and bounded subsets: Any sequence that lives in such a set has a convergent subsequence. We shall discuss that next.

Theorem 8.1 (Convergent subsequences in closed and bounded sets of $\mathbb{R}$)**.** *Let A be a bounded and closed set of real numbers and let (z_n) be an arbitrary sequence in A. Then there exists $z \in A$ and a subset*

$$n_1 < n_2 < \ldots < n_j < \ldots \text{ of indices such that } z = \lim_{j\to\infty} z_{n_j}$$

i.e., the subsequence (z_{n_j}) converges to z.

Proof: Let m be the midpoint between $a := \inf(A)$ and $b := \sup(A)$. Because A is bounded, a and b must exist as finite numbers. Let

$$A_{\star 1} := A \cap [a, m]; \qquad A^{\star}{}_1 := A \cap [m, b]. \tag{8.1}$$

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.8.5 on p.106), i.e., $A_{\star 1} \cup A^{\star}{}_1 \supseteq A$. We pick such a one and call it A_1. In case both sets contain infinitely many of the z_n, it does not matter which one we pick. Do you see that $diam(A_1) \leqq diam(A)/2$?

Let m_1 be the midpoint between $a_1 := \inf(A_1)$ and $b_1 := \sup(A_1)$. Let

$$A_{\star 2} := A_1 \cap [a_1, m_1]; \qquad A^{\star}{}_2 := A \cap [m_1, b_1]. \tag{8.2}$$

Then at least one of $A_{\star 2}$, $A^{\star}{}_2$ must contain infinitely many of the z_n. We pick such a one and call it A_2. In case both sets contain infinitely many of the z_n, it does not matter which one we pick. Note that

$$diam(A_2) \;\leqq\; diam(A_1)/2 \;\leqq\; diam(A)/2^2$$

We keep picking the midpoints m_j of the sets A_j each of which has at most half the diameter of the previous one. (Why?) In other words, we have constructed a sequence

$$\begin{aligned} &A \;\supset\; A_1 \;\supset\; A_2 \;\supset\; \ldots \;\supset\; A_n \;\supset\; \ldots \qquad \text{such that} \\ &diam(A) \;\geqq\; 2diam(A_1) \;\geqq\; 2^2 diam(A_2)/ \;\ldots\; \geqq\; 2^n diam(A_n)/ \;\ldots \end{aligned} \tag{8.3}$$

which means that $diam(A_n) \leqq diam(A)/2^n \to 0$ as $n \to \infty$.

We pick a subsequence $(x_j) = (z_{n_j})$ of the original sequence (z_n) such that $z_{n_j} \in A_j$ for all $j \in \mathbb{N}$. This is not too hard because the sets A_j were picked in such a way that each one of them contains infinitely many of the z_k.

The following inequality is true because the sequence of sets (A_j) is "nested": each A_j is contained in its predecessor A_{j-1}. It follows that A_m contains all A_k for any $k > m$ and this implies that A_m contains all members $x_k = z_{n_k}$, for all $k > m$. Thus

$$|x_m - x_k| \;\leqq\; diam(A_m) \;\leqq\; \frac{diam(A)}{2^m} \qquad \text{for all } m \text{ and } k \text{ such that } \quad k > m.$$

This means that (x_n) is a Cauchy sequence (p.73). According to theorem 7.5 about the completeness of $\mathbb{R}$ (p.74) there is a contact point x such that $x_n \to x$ for $n \to \infty$.
Because A is a closed set it contains all its contact points. It follows that $x \in A$ and we have found a subsequence of the original sequence (z_n) which converges to an element of A. ■

Theorem 8.2 (Convergent subsequences in closed and bounded sets of $\mathbb{R}^N$)**.** *Let A be a bounded and closed set of $\mathbb{R}^N$ and let $(\vec{z}_n)$ be an arbitrary sequence of N–dimensional vectors in A. Then there exists $\vec{z} \in A$ and a subset*

$$n_1 < n_2 < \ldots < n_j < \ldots \text{ of indices such that } \vec{z} = \lim_{j\to\infty} \vec{z}_{n_j}$$

i.e., the subsequence $(\vec{z}_{n_j})$ converges to $\vec{z}$.

Proof (outline): We review the above proof for $\mathbb{R}$:

The base idea was to chop A in half during each step to obtain a sequence of sets A_n which become smaller and smaller in diameter but yet contain infinitely many points. of the original sequence z_n.

In higher dimensions we would still find the center point $\vec{m}_n$ which is determined by the fact that it is the center of a γ–neighborhood (N–dimensional ball) that contains A_n and does so with the smallest radius possible. We then take the minimal square (in $\mathbb{R}^2$) or the minimal N–dimensional cube (in $\mathbb{R}^N$) that is parallel to the coordinate axes and still contains that sphere or ball.

We then divide that N–dimensional cube (a square in 2 dimensions, a cube in 3 dimensions) into 2^N sectors (4 quadrants in $\mathbb{R}^2$, 8 sectors in $\mathbb{R}^3$) and partition A_n into at most 2^N pieces by intersecting it with those 2^N sectors). The set A_{n+1} would then be chosen from one of those pieces of A_n 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.8.2 on p.100. Each A_n contains infinitely many of the $(\vec{z}_k)$. Now pick $\vec{x}_k := \vec{z}_{n_k}$ where $\vec{z}_{n_k}$ is one of the infinitely many members of the original sequence $(\vec{z}_n)$ which are contained in A_k. Because $A_j \subseteq A_K$ for $j \geqq K$ and $\lim\limits_{K\to\infty} = 0$, we do the following for a given $\varepsilon > 0$: choose K so big that $diam(A_K) \leqq \varepsilon/2$. Note that

$$\text{if } i,j \geqq K \text{ then } d(\vec{x}_i,\vec{x}_j) = d(\vec{z}_{n_i},\vec{x}_{n_j}) \leqq diam(A_K) \leqq \varepsilon/2$$

because $n_i \geqq i$ (and $n_j \geqq j$), hence $\vec{x}_i, \vec{x}_j \in A_K$. It follows that the sequence $\vec{x}_j$ is Cauchy. We have seen in thm.7.6 on p.74 that $\mathbb{R}^N$ is complete, and it follows that $\vec{L} := \lim_{j\to\infty} \vec{x}_j$ exists in $\mathbb{R}^N$. The proof is complete if it can be shown that $\vec{L} \in A$. But we know that all $\vec{x}_i = \vec{z}_{n_i}$ belong to A. $\vec{L}$ must be a contact point of A because any neighborhood $B_\varepsilon(\vec{L})$ 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 8.3. *Let A be a bounded and closed set of real numbers and let $f(\cdot) : A \to \mathbb{R}$ be a continuous function on A. Then $f(\cdot)$ is a bounded function.*

Proof: Let us assume that $f(\cdot)$ is not bounded and conclude something that is impossible.
An unbounded function is not bounded from above, from below, or both. We might as well assume that $f(\cdot)$ is not bounded from above because otherwise it is not bounded from below and we can work with $-f(\cdot)$ which then is not bounded from above. This means that there must be a sequence $(z_n) \in A$ such that

$$f(z_n) > n \qquad \text{for all } n \in \mathbb{N}. \tag{8.4}$$

According to the just proved thm.8.1 on "Convergent subsequences in closed and bounded sets" there exists a subsequence $(x_j) = (z_{n_j})$ and $x_0 \in A$ such that $x_n \to x_0$ as $n \to \infty$. In particular, $f(x_0)$ exists as a finite value and $f(x_n) \to f(x_0)$ because $f(\cdot)$ is continuous in x_0. But the x_n were constructed as a subsequence of the z_j which have the property that $f(z_j) > j$ for all j and the subsequence $(f(x_n))$ cannot converge against $f(x_0)$ because $f(x_j) = f(z_{n_j}) > n_j$, i.e., $\lim\limits_{j\to\infty} f(z_{n_j}) = \infty$. We have reached a contradiction and it follows that $f(\cdot)$ is bounded. ■

Corollary 8.1. *Let $a < b$ be two real numbers and let $f(\cdot) : [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then $f(\cdot)$ is a bounded function.*

Proof: The interval interval $[a,b]$ is closed and bounded $(diam([a,b]) = b - a)$. and the proof follows from theorem 8.3. ■

8.2 Four definitions of compactness

We shall 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". There are three different ways to define compactness of a subset K of a metric space (X,d). You can say that compactness means

A.	any sequence in K has a convergent subsequence
B.	K is complete and contains only finitely many point of a grid of length ε
C.	any open covering of K has a finite subcovering
D.	K is bounded and closed - ONLY works in $\mathbb{R}^N$!

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. Instead 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.

The most important result of this chapter on compactness will be that, if you look at $\mathbb{R}^N$ with the Euclidean norm and its associated metric

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2} \qquad (\,\vec{x} = (x_1, x_2, \dots),\ \vec{y} = (y_1, y_2, \dots) \in \mathbb{R}^N\,)$$

(see (6.3) on p.46) then the first three definitions coincide. Matter of fact, all four coincide in finite dimensional Euclidian space but "covering compactness" has been moved to the sub-chapter 8.6, p.106.

8.3 ε-nets and total boundedness

We now briefly discuss ε–nets and decreasing sequences of closed sets which contract to a single point.

Definition 8.1 (ε–nets)**.** Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \in X$. let $G \subseteq A$ be a subset of A with the following property:

For each $x \in A$ there exists $g \in G$ such that $x \in B_\varepsilon(g)$.

In other words, the points of G form a "grid" or "net" fine enough so that no matter what point x of A you choose, you can always find a "grid point" g with distance less than ε to x, because that is precisely the meaning of $x \in B_\varepsilon(g)$.

We call G an ε**–net** or ε**–grid** of A and we call $g \in G$ a **grid point** of the net.

Proposition 8.1 (ε–nets and coverings)**.** *Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \in X$. Let $G \subseteq A$ be an ε–grid for A. Then $\{B_\varepsilon(g)\}_{g \in G}$ is an open covering of A in the sense of def.**8.5 on p.106**: It is a collection of open sets the union of which "covers", i.e., contains, A.*

Proof: Let $x \in A$. We can choose a point $g = g(x) \in G$ such that $x \in B_\varepsilon\big(g(x)\big)$. It follows from $\{x\} \subseteq B_\varepsilon\big(g(x)\big)$ and $g(x) \in G$ for all $x \in A$ that

$$A = \bigcup_{x \in A} \{x\} \subseteq \bigcup_{x \in A} B_\varepsilon\big(g(x)\big) \subseteq \bigcup_{g \in G} B_\varepsilon(g). \blacksquare$$

Proposition 8.2 (ε–nets in $\mathbb{R}^N$)**.** *Let (X, d) be $\mathbb{R}^N$ with the Euclidean metric.*

***A.** Let*

$$\mathbb{Z}^N = \{\vec{z} = (z_1, z_2, \dots z_N) : z \in \mathbb{Z}\}$$

In other words, $\mathbb{Z}^N$ is all points of $\mathbb{R}^N$ with integer coordinates. That's as intuitive a grid as I can think of, provided you look at the 2–dimensional plane or 3–dimensional space.

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 A will be covered by finitely many*

$$B_\varepsilon(g_1) \cup \; B_\varepsilon(g_2) \cup \; \dots, \cup B_\varepsilon(g_n) \quad (n \in \mathbb{N}, \; g_1, \dots g_n \in G_\varepsilon^{\mathbb{R}^N}).$$

***(Skip this proof!)** (all three parts A, B, C)*

*Proof of **A.***
Let $\vec{x} = (x_1, x_2, \dots x_N) \in \mathbb{R}^N$. For each x_j let $x_j^\star$ be it's integer part, i.e., we simply throw away all digits after the decimal point.

Before we continue, let's have an example, if $N = 5$ and $\vec{x} = (12.35, -12.35, 1/3, 9, -\pi)$ then its associated grid point is $\vec{x^\star} = (12, -12, 0, 9, -3)$. Let's compute the distance:

$$d(\vec{x}, \vec{x^\star}) = \sqrt{.35^2 + .35^2 + 1/3^2 + 0 + (\pi - 3)^2} \leqq \sqrt{(1 + 1 + 1 + 0 + 1)} \leqq \sqrt{N}$$

*and we see that part **A** of the lemma is true for this specific example.*

Now to the real proof. It is not really more complicated if you notice that $|x_j - x_j^\star| < 1$ *for all* $1 \leqq j \leqq N$. *We get*

$$d(\vec{x}, \vec{x^\star}) = \sqrt{\sum_{j=1}^{N} (x_j - x_j^\star)^2} \ < \ \sqrt{N \cdot 1} \ = \ \sqrt{N}$$

So, for each point you can find a grid point with integer coordinates at a distance of less than $\sqrt{N}$. *That proves that* $\mathbb{Z}^N$ *is a* $\sqrt{N}$*–net of* $\mathbb{R}^N$.

Proof of ***B.***
Let $\vec{y} \in \mathbb{R}^N$. *Let* $\vec{x} := (1/\varepsilon)\vec{y}$ *and let* $\vec{x^\star}$ *be the vector where we discarded the decimal parts. According to part* ***A.*** *we know that* $d(\vec{x^\star}, \vec{x}) < \sqrt{N}$. *Thus*

$$\begin{aligned} d(\vec{y}, \varepsilon\vec{x^\star}) \ &= d(\varepsilon\vec{x}, \varepsilon\vec{x^\star}) \ = \ \sqrt{\sum_{j=1}^{N} (\varepsilon x_j - \varepsilon x_j^\star)^2} \ = \ \sqrt{\sum_{j=1}^{N} \varepsilon^2 (x_j - x_j^\star)^2} \\ &= \varepsilon \sqrt{\sum_{j=1}^{N} (x_j - x_j^\star)^2} \ = \ \varepsilon d(\vec{x^\star}, \vec{x}) \ < \ \varepsilon\sqrt{N \cdot 1} \ = \ \varepsilon\sqrt{N} \end{aligned}$$

In other words, for any $\vec{y} \in \mathbb{R}^N$ *there is a vector* $\vec{z} \in \mathbb{Z}^N$ *such that* $d(\vec{y}, \varepsilon\vec{z}) < \varepsilon\sqrt{N}$ *(choose* $\vec{z} = \vec{x^\star}$ *). Rephrase that: For any* $\vec{y} \in \mathbb{R}^N$ *there is a vector* $\vec{g} \in G_\varepsilon^{\mathbb{R}^N} = \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^N\}$ *such that* $d(\vec{y}, \vec{g}) < \varepsilon\sqrt{N}$ *(choose* $\vec{g} = \varepsilon\vec{z} = \vec{x^\star}$ *).*

So, for each point you can find a grid point in $G_\varepsilon^{\mathbb{R}^N}$ *at a distance of less than* $\varepsilon\sqrt{N}$. *That proves 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* $A \subseteq B_{R_1}(\vec{0})$ *(see prop.7.3 on p.64).*
b. *We enlarge the radius by* ε*: Let* $R := R_1 + \varepsilon$. *The enlarged "N-dimensional ball" of radius* R $B_R(\vec{0})$ *is contained in the "N-dimensional cube"*

$$Q_R \ := \ \{\vec{x} \ = \ (x_1, x_2, \dots x_n) : -R \leqq x_j \leqq R \text{ for all } 1 \leqq j \leqq N\}.$$

c. *Let* $\vec{z} \ = \ (z_1, z_2, \dots z_n)$ *be a grid point, i.e.,* $z_j = m_j\varepsilon$ *for the* j*-th coordinate* ($m_j \in \mathbb{Z}$). *There are only finitely many integers* m*, say* K*, for which* $-R \leqq \varepsilon \cdot m \leqq R$.
d. *Hence there are only* K *possible values for the first coordinate* $z_1 = m_1\varepsilon$. *For each one of those there are only* K *possible values for* z_2*, so there are at most* K^2 *possible combinations* (z_1, z_2) *for which* $\vec{z} \in A$. *We keep going and find that there are at most* K^N *possible grid points* $\vec{z} \in Q_R$.
e. *Any point in* $\mathbb{R}^N$ *with distance less than* ε *from some point in* A *must belong to* $B_R(\vec{0})$ *(now you know why we chose augment* R_1 *by* ε*). In particular, all grid points* $g \in G_\varepsilon^{\mathbb{R}^N}$ *whose neighborhoods* $B_\varepsilon(g)$ *intersect* A *belong to* $B_R(\vec{0})$ *and hence to* Q_R. *We conclude that* $A \cap B_\varepsilon(g) = \emptyset$ *for all grid points outside* Q_R.
f. *We know from part* ***B.*** *which was already proved that* $A \subseteq \mathbb{R}^N = \bigcup[\, B_\varepsilon(g) : g \in G_\varepsilon^{\mathbb{R}^N} \,]$. *Hence,*

$$A = A \cap \bigcup[\, B_\varepsilon(g) : g \in G_\varepsilon^{\mathbb{R}^N} \,] \ = \ \bigcup[\, A \cap B_\varepsilon(g) : g \in G_\varepsilon^{\mathbb{R}^N} \,] \ = \ \bigcup[\, A \cap B_\varepsilon(g) : g \in G_\varepsilon^{\mathbb{R}^N} \cap Q_R \,].$$

It follows that $A \subseteq \bigcup[B_\varepsilon(g) : g \in G_\varepsilon^{\mathbb{R}^N} \cap Q_R \,]$ *and* ***C.*** *is proved as there are only finitely many grid points in* Q_R. ■

Remark 8.1. 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 example that shows you why this might not be possible. Look at the "standard" ε–grid $G_\varepsilon^{\mathbb{R}^N} = \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^N\}$ defined in prop.8.2, 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 ε balls belonging to grid points in B.

Definition 8.2 (Total boundedness)**.** Let (X, d) be a metric space and let A be a subset of X. We say that A is **totally bounded** if for each $\varepsilon > 0$ there is a finite collection $\mathscr{G}_\varepsilon := \{g_1, \dots g_n\}$ of points in A whose ε–balls $B_\varepsilon(g_j)$ cover A: For any $a \in A$ there is $j = j(a)$ such that $d(a, g_j) < \varepsilon$.

We shall use this definition in connection with sequence compactness which is defined in the next section.

8.4 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 8.2, p.99 on Four definitions of compactness "sequence compactness" and we shall examine that property in this chapter.

Definition 8.3 (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

$$n_1 < n_2 < \dots < n_j < \dots \quad \text{of indices such that} \quad L = \lim_{n\to\infty} x_{n_j},$$

i.e., there exists a subsequence (x_{n_j}) which converges to L.

Proposition 8.3 (Sequence compactness implies total boundedness)**.** *Let (X, d) be a metric space and let A be a sequentially compact subset of X. Then A is totally bounded.*

Proof: Nothing needs to be shown if A is empty, so we may assume that $A \neq \emptyset$. The proof will be done by contradiction.
a. *Assume that A is not totally bounded. Then there is $\varepsilon > 0$ such that for any finite collection of points $z_1, z_2, \dots z_n \in A$ the union $\bigcup_{1\leq j\leq n} B_\varepsilon(z_j)$ does not cover A: There exists $z \in A$ outside any one of those ε–neighborhoods, i.e., $z \in A \setminus \bigcup \left[B_\varepsilon(z_j) : j \leqq n\right]$. This allows us to create an infinite sequence $(x_j)_{j\in\mathbb{N}}$ such that $d(x_j, x_n) \geqq \varepsilon$ for all $j, n \in \mathbb{N}$ such that $j \neq n$ as follows: We pick*

$$x_1 \in A; \quad x_2 \in A \setminus B_\varepsilon(x_1); \quad x_3 \in A \setminus \big(B_\varepsilon(x_1) \cup B_\varepsilon(x_2)\big); \;\dots\; x_n \in A \setminus \bigcup_{j<n} B_\varepsilon(x_j); \;\dots$$

b. *The proof is done if we can show that $(x_j)_{j\in\mathbb{N}}$ does not possess a convergent subsequence. Assume to the contrary that there is $L \in A$ and $n_1 < n_2 < \dots$ such that $lim_{j\to\infty} x_{n_j} = L$. We pick the number $\varepsilon > 0$ that was used in part **a** of the proof. There exists $N = N(\varepsilon)$ such that $d(x_{n_m}, L) < \varepsilon/2$ for all $m \geqq N$. It follows for all $i, j \geqq N$ that $d(x_{n_i}, x_{n_j}) \leqq d(x_{n_i}, L) + d(L, x_{n_j}) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. But the x_n were constructed in such a fashion that $d(x_m, x_k) \geqq \varepsilon$ for **all** $m \neq k$, in particular for $m := n_i \neq k := n_j$. We have arrived at a contradiction because $n_i \neq n_j$ whenever $i \neq j$.* ■

Proposition 8.4 (Sequence compact implies complete). *Let (X,d) be a metric space and let A be a sequence compact subset of X. Then A is complete, i.e., any Cauchy sequence (x_{n_j}) in A converges to a limit $L \in A$.*

Proof: Let (x_n) be a Cauchy sequence in A and let $\varepsilon > 0$. There exists $N_1 \in \mathbb{N}$ such that

$$(8.5) \qquad k,l \geqq N_1 \;\Rightarrow\; d(x_k,x_l) \;<\; \varepsilon/2. \qquad (\star)$$

Because A is sequence compact, we can extract a subsequence $z_j := x_{n_j}$ and find $L \in A$ such that $z_j \to L$ as $j \to \infty$. It follows that for ε chosen above there exists $N_2 \in \mathbb{N}$ such that

$$(8.6) \qquad j \geqq N_2 \;\Rightarrow\; d(x_{n_j},L) \;<\; \varepsilon/2. \qquad (\star\star)$$

We observe that $n_j \geqq j$ for all j, hence $n_j \geqq N$ if $j \geqq N$. Let $N := max(N_1,N_2)$ and $j \geqq N$. Then $j \geqq N_1$ and $n_j \geqq j \geqq N \geqq N_2$ It follows from $(\star)$ that $d(x_j,x_{n_j}) < \varepsilon/2$ and from $(\star\star)$ that $d(x_{n_j},L) < \varepsilon/2$, hence $d(x_j,L) < \varepsilon$ for all $j \geqq N$. We have proved that the arbitrarily chosen Cauchy sequence (x_n) converges. ■

The last two propositions have proved that any sequence compact set in a metric space is both totally bounded and complete. The reverse is also true:

Theorem 8.4 (Sequence compact iff totally bounded and complete). *Let A be a subset of a metric space (X,d). Then A is sequence compact if and only if A is totally bounded and complete.*

Proof: We have already seen in prop.8.3 on p.102 and prop.8.4 on p.103 that if A is sequentially compact then A is totally bounded and complete. We now shall show the other direction. Let A be totally bounded and complete and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in A. All we need to show is the existence of a subsequence $z_j = x_{n_j}$ which is Cauchy: As A is complete, such a Cauchy sequence must converge to a limit $L \in A$, i.e., $x_{n_j} \to L$ as $n \to \infty$ and we have extracted a convergent subsequence $(x_{n_j})_j$ from $(x_n)_n$.

a. *Because A is totally bounded, there will be a net for $\varepsilon = 1/2$: there exists $\mathscr{G}_1 = \{g_{1,1}, g_{1,2}, \dots, g_{1,k_1}\} \subseteq A$ such that $A \subseteq U_1 := \bigcup \big[\, B_{1/2}(g_{1,j}) : j \leqq k_1 \,\big]$. It follows that $x_k \in U_1$ for each k. There are inifinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_1$. Hence there must be at least one of those which we name g_1, such that $B_1 := B_{1/2}(g_1)$ contains $x_{1,j} := x_{n_j}$ for an entire (infinite) subsequence n_j.* [25]

b. *Because A is totally bounded, there will be a net for $\varepsilon = 1/3$: there exists $\mathscr{G}_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,k_2}\} \subseteq A$ such that $A \subseteq U_2 := \bigcup \big[\, B_{1/3}(g_{2,j}) : j \leqq k_2 \,\big]$. It follows that $x_{1,k} \in U_2$ for each k. There are inifinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_2$. Hence there must be at least one of those which we name g_2, such that $B_{1/3}(g_2)$ 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 $B_2 := B_1 \cap B_{1/3}(g_2)$.*

c. *Having constructed a subsequence $(x_{n-1,j})$ of the original sequence (x_k) which lives in a set B_{n-1} contained in $B_{1/n}(g_{n-1})$ for a suitable $g_{n-1} \in A$, total boundedness of A, guarantees the existence of a net for $\varepsilon = 1/(n+1)$: there exists $\mathscr{G}_n = \{g_{n,1}, g_{n,2}, \dots, g_{n,k_n}\} \subseteq A$ such that $A \subseteq U_n := \bigcup \big[\, B_{1/(n+1)}(g_{n,j}) : j \leqq k_n \,\big]$. It follows that $x_{n,k} \in U_n$ for each k. There are inifinitely many indices k for our sequence but only finitely many points in $\mathscr{G}_n$. Hence there must be at least one of those which we name g_n, such that $B_{1/(n+1)}(g_n)$ contains $x_{n,j} := x_{n-1,n_j}$ for an entire subsequence n_j. As the entire sequence $(x_{n-1,k})$ belongs*

[25] 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 $a \in A$?

to B_{n-1}, it follows that our new subsequence $(x_{n,j})$ of $(x_{n-1,k})$ belongs to $B_n := B_{n-1} \cap B_{1/(n+1)}(g_n)$. We note that the maximal distance $d(x_{n,i}, x_{n,j})$ between any two members of that new subsequence is bounded by $2/(n+1)$ as that is the diameter of $B_{1/(n+1)}$.

***d.** Diagonalization procedure: The following trick is employed quite frequently in real analysis. We now create the "diagonal sequence" $z_1 := x_{1,1}, z_2 := x_{2,2}, \dots$ which is a subsequence of the original sequence (x_n). If we can show that it is Cauchy then the proof is complete. By construction, if $j \geqq n$ then*

$$z_j \in B_j \subseteq B_n \subseteq B_{n-1} \subseteq \cdots \subseteq B_2 \subseteq B_1 \quad \textit{and } diam(B_j) \leq \frac{2}{j+1}.$$

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 $B_j = B_{j-1} \cap B_{1/(j+1)}(g_j) \subseteq B_{1/(j+1)}(g_j)$ for a suitable $g_j \in A$, hence all its points have distance from g_j bounded by $(j+1)^{-1}$. We obtain for any $i, j \geqq N$ that

$$d(z_i, z_j) \ \leq \ d(z_i, g_N) + d(g_N, z_j) \ \leqq \ \frac{1}{N+1} + \frac{1}{N+1} < \varepsilon.$$

It follows that $(z_n)_n$ is indeed Cauchy and the proof is completed. ■

Corollary 8.2 (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 8.5 (Sequence compact sets are bounded)**.** *Let (X, d) be a metric space and let K be a sequence compact subset of X. Then K is bounded, i.e., $diam(K) = \sup\{d(x, y) : x, y \in K\} < \infty$.*

Proof: Is also given in thm.8.13 on p.114. ■

Remark 8.2. It follows from the results of this chapter and the introductory chapter on Closed and bounded sets in Euclidian space (8.1 on p.97) that, in $\mathbb{R}^N$, three of the definitions of compactness given in section 8.2 on Four definitions of compactness (p.99 are equivalent:

A subset of $\mathbb{R}^N$ is sequentially compact iff it is totally bounded and complete iff it is bounded and closed.

We shalls see later that any metric space is sequentially compact iff it is compact, i.e., covering compact (thm.8.7 on p.thm-x:compact-iff-seq-compact).

In other words, in $\mathbb{R}^{\mathbb{N}}$ all four of the definition given in section 8.2 on p.99 coincide.

8.5 Uniform continuity

Continuous real functions on the compact set $[0, 1]$ are uniformly continuous in the sense of the following definition which you should compare, for the special case of $(X, d) = (\mathbb{R}, d_{|\cdot|})$ where $d_{|\cdot|}(x, y) = |y - x|$, to [1] Beck/Geoghegan, Appendix A.3, "Uniform continuity".

Definition 8.4 (Uniform continuity of functions)**.** Let (X,d_1), (Y,d_2) be metric spaces and let A be a subset of X. A function

$$f(\cdot) : A \to Y \qquad \text{is called } \textbf{uniformly continuous}$$

if for any $\varepsilon > 0$ there exists a (possibly very small) $\delta > 0$ such that

$$d_2(f(x) - f(y)) \ < \varepsilon \quad \text{for any } x, y \in A \text{ such that } d_1(x,y) \ < \ \delta \tag{8.7}$$

Remark 8.3 (Uniform continuity vs. continuity)**.** Note the following:

***A*.** Condition (8.7) for uniform continuity looks very close to the ε–δ characterization of ordinary continuity (7.39) on p.77. Can you spot the difference? Uniform continuity is more demanding than plain continuity because when dealing with the latter you can ask for specific values of both ε and x_0 according to which you had to find a suitable δ. In other words, for plain continuity

$$\delta \ = \ \delta(\varepsilon, x_0).$$

But in the case of uniform continuity all you get is ε and you must come up with a suitable δ regardless of what arguments are thrown at you. To write that one in functional notation,

$$\delta \ = \ \delta(\varepsilon).$$

***B*.** In case you missed the point, uniform continuity implies continuity but the opposite need not be true.

Example 8.1 (Uniform continuity of the identity mapping)**.** Have another look at proposition(7.9) 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.

Theorem 8.6 (Uniform continuity on sequence compact spaces)**.** *Let (X,d_1), (Y,d_2) be metric spaces and let A be a sequence compact subset of X. Any continuous real function on A is uniformly continuous on A.*

Proof: Let us assume that $f(\cdot)$ is continuous but not uniformly continuous and find a contradiction. Because $f(\cdot)$ is not uniformly continuous, you can find $\varepsilon > 0$ such that no $\delta > 0$, however small, will satisfy (8.7) for all pairs x, y such that $d_1(x,y) \ < \ \delta$. Looking specifically at $\delta := 1/j$ for all $j \in \mathbb{N}$, we can find $x_j, x'_j \in A$ such that

$$d_1(x_j, x'_j) \ < \ \frac{1}{j} \quad \textit{but} \quad d_2(f(x), f(x')) \ \geqq \varepsilon. \tag{8.8}$$

But A is sequence compact and we can find a subsequence (x_{j_k}) of the x_j which converges to an element $x \in A$ We have

$$d_1(x'_{j_k}, x) \ \leqq \ d_1(x'_{j_k}, x_{j_k}) + d_1(x_{j_k}, x) \ \leqq \frac{1}{j_k} + d_1(x_{j_k}, x) \tag{8.9}$$

and each right hand term will converge to zero as $k \to \infty$. It's obvious for $1/j_k$ because $j_k \geqq k$ for all k and it is true for $d_1(x_{j_k}, x)$ because x_{j_k} converges to x. It follows from (8.9) that (x'_{j_k}) also converges to x. It follows from the ordinary continuity of $f(\cdot)$ that

$$f(x) \ = \ \lim_{k\to\infty} f(x'_{j_k}) \ = \ \lim_{k\to\infty} f(x_{j_k})$$

and it follows from the "ordinary" (non-uniform) convergence of sequences that there exist $N, N' \in \mathbb{N}$ *such that*

$$d_2(f(x), f(x_{j_k})) \;<\; \frac{\varepsilon}{2} \text{ for } k > N; \quad d_2(f(x), f(x'_{j_k})) \;<\; \frac{\varepsilon}{2} \text{ for } k > N'$$

and both inequalities will hold for all $k > N + N'$*. Why? It follows for all such* k *that*

$$d_2(f(x_{j_k}), f(x'_{j_k})) \;<\; d_2(f(x_{j_k}), f(x)) + d_2(f(x), f(x'_{j_k})) \;<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon \tag{8.10}$$

and we have a contradiction to (8.8)*.* ■

Corollary 8.3 (Uniform continuity on closed intervals)**.** *Let* a, b *be two real numbers such that* $a \leqq b$*. Any continuous real function on the closed interval* $[a, b]$ *is uniformly continuous on* $[a, b]$*, which means that, given any whatever small* $\varepsilon > 0$*, there exists a number* $\delta > 0$*, possibly a lot smaller, such that*

$$d(f(x) - f(y)) \;<\; \varepsilon \qquad \text{for all } x, y \in [a, b] \text{ such that } d(f(x) - f(y)) < \delta \tag{8.11}$$

Proof: This follows from the previous theorem (8.6) *because closed intervals* $[a, b]$ *are closed and bounded sets and, in* $\mathbb{R}$*, any closed and bounded set is sequence compact .* ■

8.6 Open coverings and the Heine–Borel theorem

We shall now discuss families of open sets called " open coverings" and you should review the concept of an indexed family and how it differs from that of a set (see (3.7) *on p.12).*

Definition 8.5 (Open coverings)**.** Let X be an arbitrary non–empty set and $A \subseteq X$. Let $(U_i)_{i \in I}$ be an indexed family of subsets of X such that $A \subseteq \bigcup_{i \in I} U_i$. Then we call $(U_i)_{i \in I}$ a **covering** of A.

A **finite subcovering** of a covering $(U_i)_{i \in I}$ of the set A is a finite collection

$$U_{i_1}, U_{i_2}, U_{i_3}, \dots, U_{i_n} \quad (i_j \in I \quad \text{for } 1 \leqq j \leqq n) \qquad \text{such that} \quad A \subseteq\; U_{i_1} \cup U_{i_2} \cup \ldots \cup U_{i_n}. \tag{8.12}$$

If X is a metric space and all members U_i of the family are open then $(U_i)_{i \in I}$ is called an **open covering** of A.

Definition 8.6 (Compact sets in metric spaces)**.** Let (X, d) be a metric space and $K \subseteq X$. We say that K is **compact** if it has the "extract finite subcovering" property: Given any **open** covering $(U_i)_{i \in I}$ of K, you can extract a finite subcovering. In other words, there is a (possibly very large $n \in \mathbb{N}$ and indices

$$i_1, i_2, \dots, i_n \in I \qquad \text{such that} \quad A \subseteq \bigcup_{j=1}^{n} U_{i_j}.$$

Theorem 8.7 (Sequence compact is same as compact)**.** *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 will be given in the following optional sub-chapter 8.6.1

Next comes the Heine–Borel theorem which states that, for the metric space $\mathbb{R}^N$ *with the Euclidean norm, a set is compact if and only if it is bounded and closed.*

Theorem 8.8 (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 will be given in the following optional sub-chapter 8.6.1

Theorem 8.9 (Compact sets are complete). *Let (X, d) be a metric space and let K be a compact subset of X. Then K is complete.*

Proof will be given in the following optional sub-chapter 8.6.1

8.6.1 Appendix: Proofs for (open covering) compactness (Skip this!)

This entire sub-chapter can be skipped. What you need to know about covering compactness is contained in the parent chapter.
This chapter repeats the theorems from the last chapter together with proofs. I have decided to leave the proofs in here even though they can be significantly shortened just as to not leave any gaps in this presentation. If this subject matter truly interests you then you should look at a textbook on real analysis and study the proofs in there instead.

The next lemma is complete nonsense in that its assumptions will never be valid. But still it serves its purpose to absorb most of the work to be done for the subsequent Heine Borel theorem.

Lemma 8.1 (Contracting sequences of closed sets in $\mathbb{R}^N$). *Let (X, d) be $\mathbb{R}^N$ with the Euclidean norm and its associated metric. Let $K \subseteq \mathbb{R}^N$ be a bounded and closed set. Assume that there is an open covering $(U_\alpha)_{\alpha \in I}$ of K from which you cannot extract a finite subcovering.*

Then there exists a sequence $K_1 \supset K_2 \supset \ldots \supset$ of closed subsets of K with the following properties:

(8.13a) $\quad diam(K_n) \leqq \dfrac{1}{2^{n-1}}$ *(see (7.7) on p.64)*

(8.13b) *None of the K_n can be covered by finitely many of the U_α*

(8.13c) $\quad \bigcap\limits_{j \in \mathbb{N}} K_j$ *contains exactly one element $\vec{x}^\star \in \mathbb{R}^N$*

(8.13d) *Any sequence $(\vec{x}_n)$ such that $\vec{x}_n \in K_n$ for each $n \in \mathbb{N}$ converges to $\vec{x}^\star$.*

Proof: We start with $n = 1$.
Let $\varepsilon = 1/2$. Lemma (8.2) on ε–nets shows that there are finitely many (the unproved part C of the lemma) points $\vec{g}_{1,1}, \vec{g}_{1,2}, \ldots, \vec{g}_{1,N_1}$, such that their $1/2$–neighborhoods $B_{1/2}(\vec{g}_{1,j})$ ($1 \leqq j \leqq N_1$) are a covering of K. So we have

$$B_{1/2}(\vec{g}_{1,1}) \cup B_{1/2}(\vec{g}_{1,2}) \cup \ldots \cup B_{1/2}(\vec{g}_{1,N_1}) \supset K \tag{8.14}$$

and

$$(B_{1/2}(\vec{g}_{1,1}) \cap K) \cup (B_{1/2}(\vec{g}_{1,2}) \cap K) \cup \ldots \cup (B_{1/2}(\vec{g}_{1,N_1}) \cap K) = K \tag{8.15}$$

I claim that there must be at least one j such that the set $B_{1/2}(\vec{g}_{1,j}) \cap K$ cannot be covered by finitely many U_α of the original open covering of K. Why? Well, if that was true, you would find index sets $I_1 \subseteq I$, $I_2 \subseteq I$ and finally $I_{N_1} \subseteq I$ all of which are finite such that

$$
\begin{aligned}
&\bigcup_{\alpha \in I_1} U_\alpha \supset B_{1/2}(\vec{g}_{1,1}) \cap K \\
&\bigcup_{\alpha \in I_2} U_\alpha \supset B_{1/2}(\vec{g}_{1,2}) \cap K \\
&\dots \\
&\bigcup_{\alpha \in I_{N_1}} U_\alpha \supset B_{1/2}(\vec{g}_{1,N_1}) \cap K
\end{aligned}
$$

Let us abbreviate $I^\star := I_1 \cup I_2 \cup \ldots \cup I_{N_1}$. Then $I^\star$ is finite as a finite union of finite sets. We take unions over all left sides and all right sides of the above and obtain

$$
\begin{aligned}
\bigcup_{\alpha \in I^\star} U_\alpha &= (\bigcup_{\alpha \in I_1} U_\alpha) \cup (\bigcup_{\alpha \in I_2} U_\alpha) \cup \ldots \cup (\bigcup_{\alpha \in I_{N_1}} U_\alpha) \\
&\supset (B_{1/2}(\vec{g}_{1,1}) \cap K) \cup (B_{1/2}(\vec{g}_{1,2}) \cap K) \cup \ldots \cup (B_{1/2}(\vec{g}_{1,N_1}) \cap K) \supset K
\end{aligned}
\tag{8.16}
$$

In other words, K is covered by the finitely many U_α where $\alpha \in I^\star$. But this is contrary to our original assumption (8.20) at the beginning of this proof. Now we know that there is at least one index, let's call it $j^\star$ such that the set $B_{1/2}(\vec{g}_{1,j^\star}) \cap K$ cannot be covered by finitely many U_α of the original open covering of K. I hope you understand that this set is not empty. Otherwise, how could it not be possible to find a finite subcovering for it? We define

$$
K_1 := \overline{B_{1/2}(\vec{g}_{1,j^\star})} \cap K \tag{8.17}
$$

In case you forgot, $\overline{B_{1/2}(\vec{g}_{1,j^\star})}$ is the closure of $B_{1/2}(\vec{g}_{1,j^\star})$ which is obtained by augmenting it with its contact points (see (7.18) on p.72). Note that K_1 is bounded because it is contained in the bounded set K. Matter of fact,

$$
diam(K_1) \leqq diam(\overline{B_{1/2}(\vec{g}_{1,j^\star})}) = 2 \cdot \frac{1}{2} = 1 = 2^{1-1}
$$

and K_1 is closed as the intersection of two closed sets. We finally found the first member of a sequence of sets with the properties (8.13a) and (8.13b).

Now look at $n = 2$.
Let $\varepsilon = 1/4 = 2^{-2}$. Lemma (8.1) on ε–nets shows that there are finitely many points $\vec{g}_{2,1}, \vec{g}_{2,2}, \ldots, \vec{g}_{2,N_2}$, such that their $1/4$–neighborhoods $B_{1/4}(\vec{g}_{2,j})$ ($1 \leqq j \leqq N_2$) are a covering of K_1. We use the same reasoning as we did for $n = 1$ to deduce that there is at least one index, let's call it again $j^\star$, such that the set $B_{1/4}(\vec{g}_{2,j^\star}) \cap K_1$ cannot be covered by finitely many U_α of the original open covering of K which again means that it is not empty. Now we define

$$
K_2 := \overline{B_{1/4}(\vec{g}_{2,j^\star})} \cap K_1 \tag{8.18}
$$

Clearly $K_2 \subseteq K_1$. it is bounded with diameter

$$
diam(K_2) \leqq diam(\overline{B_{1/4}(\vec{g}_{2,j^\star})}) = 2 \cdot \frac{1}{4} = \frac{1}{2} = 2^{2-1}
$$

and it is closed as the intersection of two closed sets. So we found the second member of the sequence of sets with the properties (8.13a) and (8.13b).

Now look at an arbitrary n.
We can assume that K_{n-1} has already been constructed. Let $\varepsilon = 1/2^{-n} = 2^{-n}$. Lemma (8.1) on ε–nets shows that there are finitely many points $\vec{g}_{n,1}, \vec{g}_{n,2}, \dots, \vec{g}_{n,N_n}$, with $\vec{g}_{n,j} \in G_\varepsilon = \varepsilon\mathbb{Z}$ such that their 2^{-n}–neighborhoods $B_{2^{-n}}(\vec{g}_{n,j})$ ($1 \leqq j \leqq N_n$) are a covering of K_{n-1}. We use the same reasoning as we did for $n = 1$ to deduce that there is at least one index, let's call it again $j^\star$, such that the set $B_{2^{-n}}(\vec{g}_{n,j^\star}) \cap K_{n-1}$ cannot be covered by finitely many U_α of the original open covering of K which again means that it is not empty. Now we define

$$K_n := \overline{B_{2^{-n}}(\vec{g}_{n,j^\star})} \cap K_{n-1} \tag{8.19}$$

Clearly $K_n \subseteq K_{n-1}$. It is bounded with diameter

$$diam(K_n) \leqq diam(\overline{B_{2^{-n}}(\vec{g}_{n,j^\star})}) = 2 \cdot \frac{1}{2^n} = 2^{n-1}$$

and it is closed as the intersection of two closed sets. We have found the n^{th} member of the sequence of sets with the properties (8.13a) and (8.13b).

Parts c and d will be shown together now. Lets us pick $\vec{x}_n \in K_n$ for each $n \in \mathbb{N}$. Why is the sequence $(\vec{x}_n)$ Cauchy? Actually, that's easy. Let $j, k, N_0 \in \mathbb{N}$ and assume that $j, k \geqq N_0$. Look at the members $\vec{x}_j$ and $\vec{x}_k$. Both are contained in $K_{2^{-N_0+1}}$ whose diameter does not exceed 2^{N_0-1}. In other words,

$$d(\vec{x}_j, \vec{x}_k) \leqq \frac{1}{2^{N_0-1}} \searrow 0 \quad (j, k \geqq N_0)$$

and this proves the sequence is Cauchy. But $\mathbb{R}^N$ is complete (see (7.6) on p.74) and the sequence converges against an element $\vec{x}^\star \in \mathbb{R}^N$ Given any $n \in \mathbb{N}$, all elements $\vec{x}_k$ belong to K_n for big enough k. This means $\vec{x}^\star \in K_n$ because it is a contact point of those $\vec{x}_k$ and the closed set K_n contains all its contact points. But this is true for any $n \in \mathbb{N}$ and we deduce that $\vec{x}^\star \in \bigcap\limits_{n \in \mathbb{N}} K_n$.

The last thing to show is that $\bigcap\limits_{n \in \mathbb{N}} K_n$ does not contain a second element. But if it did contain another one, say $\vec{y}$, there would be a certain distance $\delta := d(\vec{x}^\star, \vec{y}) > 0$ between them. That's kind of hard to do because

$$diam(\bigcap_{j \in \mathbb{N}} K_j) \leqq diam(K_n) \leqq 2^{n-1} \quad \text{for all } n \in \mathbb{N}$$

which means that the diameter of the intersection of all K_j is zero. This implies that $d(\vec{x}, \vec{y}) = 0$ for any two elements in that intersection and that means $\vec{x} = \vec{y}$ ■

Theorem 8.10 (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 of "$\Longleftarrow$": A closed and bounded set in $\mathbb{R}^N$ is compact:
We shall give an indirect proof. So let us assume that the set $K \subseteq \mathbb{R}^N$ is closed and bounded and

(8.20) *there is an open covering $(U_\alpha)_{\alpha \in I}$ of K from which you cannot extract a finite subcovering.*

We shall see that this leads to a contradiction.

Our assumption of not being able to obtain a finite subcovering of the U_α is precisely what we need to employ lemma (8.1) and obtain the sequence (K_n).

Now what shall we do with that sequence? First we pick an element $\vec{x}_j \in K_j$ for each $j \in \mathbb{N}$. This sequence converges to the only element $\vec{x}^\star \in \bigcap_{j \in \mathbb{N}} K_j$.

Eventually, we get back to the original open covering (U_α) of K from which we assume that no finite subcovering for K can be extracted. Because it is a covering of K and $\vec{x}^\star \in K$, there must be an index, say α_0, such that $\vec{x}^\star \in U_{\alpha_0}$. and it is an interior point of U_{α_0} because this is an open set. This means that we can find a (sufficiently small) $\delta > 0$ such that $B_\delta(\vec{x}^\star) \subseteq U_{\alpha_0}$. Let us pick $n \in \mathbb{N}$ so big that $2^{n-1} < \delta$. Pick any $\vec{y} \in K_n$. Then

$$d(\vec{y}, \vec{x}^\star) \leqq diam(K_n) \leqq 2^{n-1} \quad \Longrightarrow \quad \vec{y} \in B_\delta(\vec{x}^\star) \subseteq U_{\alpha_0}$$

That's a moment to savor. We have just shown that $K_n \subseteq U_{\alpha_0}$. What's the big deal about that? This means that the set K_n which was constructed in such a way that no finite subcovering of the U_α can cover it, is in fact covered by a single member, U_{α_0}. We can happily conclude that any closed and bounded set in $\mathbb{R}^N$ is compact.

Proof of "$\Longleftarrow$": A compact set is closed and bounded:
We needed the special properties of $\mathbb{R}^N$ with the Euclidean norm to prove that any closed and bounded set is compact. The proof that you see here to show the opposite direction needs nothing other than the properties of a metric space.

So let (X, d) be a metric space and assume that $K \subseteq X$ is compact, i.e., any open covering of K has a finite subcovering. We must show that any contact point of K belongs to K. Let $x \in X$. For $n \in \mathbb{N}$ let

$$F_n(x) := \overline{B_{1/n}(x)} = \text{closure-of}(B_{1/n}(x)) = \{y \in X : d(y, x) \leqq \frac{1}{n}.$$

The complement $U_n(x) := \complement F_n(x) = \{y \in X : d(y, x) > 1/n\}$ is open and we have $\bigcup_{j \in \mathbb{N}} U_j(x) = X \setminus \{x\}$. (Why?)

Now assume that $a \in X$ is a contact point of K. If $a \notin K$ then

$$\bigcup_{j \in \mathbb{N}} U_j(a) = X \setminus \{a\} \supset K \setminus \{a\} = K \tag{8.21}$$

In the above chain the "$\supset$" part is true simply because $X \supset K$ and the last equality follows from the definition of the set difference (see (3.12) on p.14). (8.21) shows that $(U_j(a))_{j \in \mathbb{N}}$ is an open covering of K. But K is assumed to be compact and this guarantees the existence of finitely many $j_1 < j_2 < \ldots < j_N$ such that

$$\{y \in X : d(y, a) > 1/j_N\} = U_{j_N}(a) = U_{j_1}(a) \cup U_{j_2}(a) \cup \ldots \cup U_{j_N}(a) \supset K$$

In other words, if $y \in K$ then $d(y, a) > 1/N$. This makes it impossible for a to be a contact point of K because there is an entire neighborhood $B_{1/N}(a)$ which does not contain a single element of K. But we had assumed that a is a contact point of K and we have reached a contradiction. We have proved that K contains all its contact points, i.e., K is closed.

We are not done yet. We still must prove that K is bounded. But K not being bounded means that there is no $a \in X$ and $\gamma > 0$ such that $K \subseteq B_\gamma(a)$. To phrase it differently, let us pick some arbitrary $a \in X$

and let us look at the j–neighborhoods $B_j(a)(j \in \mathbb{N})$. No matter what point $y \in X$ we choose, if j is big enough then $j > d(y,a)$ and that means $y \in B_j(a)$. In other words, $\bigcup B_j(a) = X \supset K$ and the sets $B_j(a)$ are an open covering of any set in X, so they are most certainly an open covering of the set K. But K is compact and we can extract a finite subcovering of those $B_j(a)$. This guarantees the existence of finitely many $j_1 < j_2 < \ldots < j_N$ such that

$$\{y \in X : d(y,a) < j_N\} \;=\; B_{j_N}(a) \;=\; B_{j_1}(a) \cup B_{j_2}(a) \cup \ldots \cup B_{j_N}(a) \;\supset\; K$$

Let's read that one backwards: There is a γ–neighborhood of some point a which contains all of K (set $\gamma := j_N$. But that means precisely that K is bounded. ■

Theorem 8.11 (Compact sets are complete)**.** *Let (X,d) be a metric space and let K be a compact subset of X. Then K is complete.*

Proof: We prove this indirectly and show the assumption that K is compact but not complete leads to a contradiction. Let K be compact but not complete. Then there is a Cauchy sequence (x_n) in K which does not converge to an element of K. For a given $j \in \mathbb{N}$ let

(8.22) $$F_j \;:=\; \{x_j\} \cup \{x_j + 1\} \cup \{x_j + 2\} \cup \ldots$$

Take any finite intersection of the F_j, i.e., choose finitely many

(8.23) $$j_1 < j_2 < \ldots, j_N. \qquad \textit{Then} \quad \bigcap_{k=1}^{j_N} F_{j_k} \;=\; F_{j_N} \;\neq\; \emptyset$$

For convenience, let us set $F := F_{j_N}$. Our goal is to show that F is closed, i.e., it contains all its contact points. Note that because $F \subseteq K$ and K is closed we have $\overline{F} \subseteq \overline{K} = K$ and any potential contact point of F must belong to K.

To prove that F is closed, we first rule out the case that there could be two different elements $z_1, z_2 \in X$ which both are contact points of F. Why? First, because $z_1 \neq z_2$, we can find sufficiently small ε such that

(8.24) $$d(z_1, z_2) \;>\; 3\varepsilon$$

(x_n) is Cauchy, so we can find $N_0 \in \mathbb{N}$ such that

(8.25) $$d(x_j, x_k) < \varepsilon \quad \textit{for all } j, k \geqq N_0.$$

Because z_1, z_2 are contact points of F and this set exclusively consists of members of the sequence (x_n), there must be elements of F which, indexed as a sequence, converge to z_1 and (other) elements of F which, indexed as a sequence, converge to z_2. In other words, we have sequences

$$(x_{n_j})_{j \in \mathbb{N}} \textit{ and } (x_{m_k})_{k \in \mathbb{N}} \quad \textit{such that} \quad x_{n_j} \to z_1,\; x_{m_k} \to z_2.$$

Because the full sequence $x_1, x_2, \ldots$ is Cauchy you can find $N_1 \in \mathbb{N}$ and $N_2 \in \mathbb{N}$ such that

(8.26) $$d(x_{n_j}, x_{m_k}) < \varepsilon \quad \textit{for all } j \geqq N_1,\; k \geqq N_2.$$

Set $N^\star := max(N_0, N_1, N_2)$ *Then, on account of* (8.25) *and* (8.26), *we obtain*

$$d(z_1, z_2) \leqq d(z_1, x_{n_j}) + d(x_{n_j}, x_{m_k}) + d(x_{m_k}, z_2) < 3\varepsilon$$

for all $j, k \geqq N^\star$ *and this contradicts* (8.24).

Now we know that there cannot be more than one contact point of F. *Next we show that if there is one contact point, say* z, *then* z *cannot be an element of* F. *We show that assuming otherwise leads to a contradiction too. We reuse* N_0 *from* (8.25). *Because* z *is a contact point of* F *and this set exclusively consists of members of the sequence* (x_n), *there must be elements of* F *which, indexed as a sequence, converge to* z. *In other words, we have a sequence*

$$(x_{n_j})_{j\in\mathbb{N}} \quad \textit{such that} \quad x_{n_j} \to z.$$

This in turn means that you can find $N_1 \in \mathbb{N}$ *such that*

(8.27) $$d(x_{n_j}, z) < 2\varepsilon \quad \textit{for all } j \geqq N_1.$$

Then

$$d(x_m, z) \leqq d(x_m, x_{n_j}) + d(x_{n_j}, z) < 2\varepsilon$$

for $m \geqq N_0$ *and* j *so big that both* $j \geqq N_1$ *and* $n_j \geqq j \geqq N_0$. *(think: why is* $j \leqq n_j$*?). In other words,* $z = \lim\limits_{n\to\infty} x_n$. *Again, we must have* $z \in K$ *because* $F \subseteq K$ *and* K *is closed. We have arrived at our contradiction because at the beginning of the proof we had assumed that the Cauchy sequence* (x_n) *does not have a limit in* K.

The only possibility remaining is that F *has no contact points outside* F. *Rephrase this:* F *contains all its contact point, so it is closed. Remember that because they are sets there are no duplicate elements in the sets* F_j. *But of course the original sequence* (x_n) *might have duplicate members. Question: can it have members that reoccur infinitely often? Here is why that is impossible: Let there be an* $x \in X$ *and an infinite sequence* $n_1 < n_2 < \dots$ *of indices such that* $x_{n_j} = x$ *for all* $j \in \mathbb{N}$. *Then for any* $m \in \mathbb{N}$

$$d(x_m, x) \leqq d(x_m, x_{n_j}) + d(x_{n_j}, x) = d(x_m, x_{n_j}) + 0$$

will become arbitrarily small as m, n_j *both become big because* (x_n) *is Cauchy. Hence the full sequence* (x_m) *converges to* $x \in K$ *(actually* x *even belongs to* $F_{n_1} \subseteq K$ *because* $x = x_{n_1}$ *). We have a contradiction because at the beginning of the proof we had assumed that the Cauchy sequence* (x_n) *does not have a limit in* K.

The remainder of the proof is quick: Let's go back to the original definition (8.22) *of the sets* F_j. *Obviously*

(8.28) $$\bigcap_{j\in\mathbb{N}} F_j = \emptyset$$

because the last thing we figured out is that any member of (x_n) *only occurs finitely often and cannot belong to* F_j *if* j *just is big enough. Let* $U_j := \complement F_j$. *Note that if* $m < n$ *then* $F_m \supset F_n$ *, hence* $U_m \subseteq U_n$. *According to De Morgan's law,* (8.28) *becomes*

$$\bigcup_{j\in\mathbb{N}} U_j = \complement\emptyset = X$$

and the U_j are an open covering of X, hence of the compact set K. So we can extract finitely many indices $i_1 < i_2 < \ldots < i_M$ for a suitable $M \in \mathbb{N}$ such that

$$U_{i_M} \;=\; U_{i_1} \cap U_{i_2} \cap \ldots \cap U_{i_M} \;\supset K.$$

Because any $x \in K$ belongs to U_{i_M}, it cannot belong to its complement F_{i_M}. Rephrase that: none of the $x_{i_M}, x_{i_M+1}, x_{i_M+2}, \ldots$ is an element of K. But we had assumed from the outset that all x_j belong to K and this final contradiction proves that it is impossible for a compact set to host a Cauchy sequence which does not converge to an element of K. ■

Theorem 8.12 (Sequence compact is same as compact)**.** *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 of "$\Longleftarrow$": A compact set is sequence compact:
All we need to show is that the set

$$C_1 \;:=\; \{x_1\} \cup \{x_2\} \cup \ldots$$

(the members of (x_n) with all duplicates removed) has a contact point a (i.e., any whatever small ε–neigborhood $B_\varepsilon(a)$ contains infinitely many members of (x_n) (see (7.17) on p.72). Let F_n be the closure of the tail set

$$C_n \;:=\; \{x_n\} \cup \{x_{n+1}\} \cup \ldots \qquad \textit{(see def. 5.12, p.37).}$$

Then any finite intersection of this non–increasing sequence C_j of sets is of course non–empty.

I claim that

$$F \;:=\; \bigcap_{j \in \mathbb{N}} F_j \quad \textit{must contain at least one member of } A. \tag{8.29}$$

Why? Otherwise we would have

$$A \;\subseteq\; \complement F \;=\; \bigcup_{j \in \mathbb{N}} \complement F_j$$

(De Morgan's law, (3.1) on p.15). But $U_j := \complement F_j$ is open as the complement of a closed set. Hence the U_j are an open covering of the compact set A. So there exist finitely many indices

$$n_1 < n_2 < \ldots < n_j < \ldots \quad \textit{such that} \quad U_{n_j} \;=\; \bigcup_{j = n_1, n_2, \ldots n_j} U_j \;\supset A.$$

We take complements on both sides and flip the direction of $\supset$ and obtain $F_{n_j} \subseteq \complement A$. This is impossible because all members $x_j, x_{j+1}, \ldots \in F_j$ were supposed to belong to A.

We arrived at a contradiction. Now we know that (8.29) must in fact be true. So there is $a \in F \cap A$. The way that F was constructed, this means that

$$a \in \overline{C_n} \quad \textit{where} \quad C_n \;=\; \{x_n\} \cup \{x_{n+1}\} \cup \ldots.$$

That means that any whatever small neighborhood of a contains elements of the sequence (x_j). So for $k \in \mathbb{N}$ we can pick an index n_k such that $x_{n_k} \in B_{1/k}(a)$. Obviously the x_{n_k} converge to a and we have our convergent subsequence to an element of A.

Proof of "$\Longrightarrow$": A sequence compact set is compact:
Even though this is true for arbitrary metric spaces, the proof is too hard to give here and we limit ourselves to prove the special case where (X, d) is $\mathbb{R}^N$ with the Euclidean metric. According to Heine–Borel, we only need to show that a sequence compact set is closed and bounded.

So let $A \in \mathbb{R}^N$ and assume that A is not closed, say there is a contact point a of A which does not belong to A. a can be approximated by elements of A and thus you can find for any $(1/n)$–neighborhood of a an element $x_n \in B_{1/n}(a)$, i.e., $d(x_n, a) < 1/n$. The good news is that this sequence does have a as contact point. The bad news is that it belongs to $\complement A$ and not to A, and there is nothing else that might qualify as a contact point for (x_n). This proves that a set that is not closed is not sequence compact.

Now let us assume that A is not bounded. That means that, given an arbitrary $x_0 \in \mathbb{R}^N$, none of the neighborhoods $B_j(x_0) = \{x \in \mathbb{R}^N : d(x, x_0) < j\}$ contains A and we can pick $x_j \in A$ such that $d(x_j, x_0) \geqq j$ for all $j \in \mathbb{N}$. How can there possibly a subsequence (x_{j_k}) that converges to some b anywhere in $\mathbb{R}^N$? It would have a finite distance $d(b, x_0)$ from x_0, say γ. For big enough k, all x_{j_k} would be quite close to b and we can expect that $d(x_{j_k}, b) < 1$. All together we get

$$d(x_{j_k}, x_0) \ \leqq \ d(x_{j_k}, b) + d(b, x_0) \ \leqq \ 1 + \gamma$$

for big k. On the other hand we had constructed x_{j_k} such that $d(x_{j_k}, x_0) \geqq j_k \geqq k$ where the last inequality is true because the subsequence j_k grows faster than just k It follows that $d(x_{j_k}, x_0) \to \infty$ for $k \to \infty$. We have reached a contradiction and conclude that sequence compact sets in $\mathbb{R}^N$ must be bounded. ■

8.7 Appendix: Addenda to chapter 8

8.7.1 Sequence compactness

The following theorem follows indirectly from the fact that sequence compact sets are both totally bounded and complete but here is a direct proof.

Theorem 8.13 (Sequence compact sets are closed and bounded)**.** *Let A be sequence compact subset of a metric space (X, d). Then A is a bounded and closed set.*

***a**. Proof of boundedness:*

We may assume that A is not empty because otherwise there is nothing to prove. We assume that A is not bounded, i.e., $diam(A) = \infty$. It will be proved by induction that there exists a sequence $x_n \in A$ such that $d(x_i, x_j) \geqq 1$ for any $i \neq j$.

Let $x_0 \in A$. There exists $x_1 \in A$ such that $r_1 := d(x_0, x_1) \geqq 1$. We now assume that n elements $x_1, \dots x_n$ such that $d(x_i, x_j) \geqq 1$ for any $1 \leqq i < j \leqq n$ have aready been chosen. Let $\left\{ {k:=\max \atop d(x_0, x_j) : j \leqq n} \right\}$ and $r := k + 1$. As A is not bounded, we can pick $x_{n+1} \in A \setminus B_r(x_0)$. We obtain

$$\begin{aligned} k+1 \ &\leq d(x_{n+1}, x_0) \leq d(x_{n+1}, x_j) + d(x_j, x_0) \ \leq \ d(x_{n+1}, x_j) + k, \quad \textit{i.e.,} \\ 1 \ &\leq d(x_{n+1}, x_j). \end{aligned}$$

This finishes the proof of the existence of the 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.

b. *Proof of closedness: If A was not closed then we could pick a contact point $x \in A^{\complement}$ of A. As $B_{1/m}(x) \cap A \neq \emptyset$ we can pick a sequence $x_m \in A$ such that $d(x_m, x) < 1/m$ for all $m \in \mathbb{N}$. Clearly x_m converges to $x \notin A$. Sequence compactness of A allows us to extract a subsequence $z_j = x_{n_j}$ which converges to $z \in A$. We have both z and x as limit of z_j. According to thm.7.9 on p.91, $x = z$ and we have both $x \in A^{\complement}$ and $x \in A$, a contradiction. This proves that sequence compact sets are closed.* ■

8.7.2 Continuous functions and compact spaces

Theorem 8.14 (Closed subsets of compact spaces are compact)**.** *Let A be a closed subset of a compact metric space (X, d). Then $(A, d\big|_{A \times A}$ is a compact subspace.*

Proof:

Let $(U_j)_{j \in J}$ be a family of sets open in A whose union is A. According to prop.7.17 on p.93 there are open sets V_j in X such that $U_j = V_j \cap A$. It follows that $\bigcup\limits_{j \in J} V_j \supseteq A$, hence the family $(V_j)_{j \in J}$, augmented by the (open!) set $A^{\complement}$ is an open cover of (X, d). As X is compact we can extract finitely many members from that extended family such that they still cover X. If one of them happens to be $A^{\complement}$ then we remove it and we still obtain that the remaining ones, say, $V_{i_1}, V_{i_2}, \dots, V_{i_n}$, cover A. But then the traces in A $U_{i_1} = V_{i_1} \cap A$, $U_{i_2} = V_{i_2} \cap A, \dots, U_{i_n} = V_{i_n} \cap A$, cover A and they form an open covering of the subspace A as the A-traces of open sets in X are open in A. We have proved that the given open covering in A has a finite subcover of A.
■

Theorem 8.15 (Continuous images of compact spaces are compact)**.** *Let (X, d_1) and (Y, d_2) be two metric spaces. and let $f : X \longrightarrow Y$ be continuous on X. If X is compact then the direct image $f(X)$ is compact, i.e., the metric subspace $f(X)$ of Y is compact.*

Proof: Let $(V_j)_{j \in J}$ be a family of sets open in $B := f(X)$ whose union is B. Let $U_j := f^{-1}(V_j)$.

$$\bigcup_{j \in J} U_j \;=\; \bigcup_{j \in J} f^{-1}(V_j) \;=\; f^{-1}\Big(\bigcup_{j \in J} V_j \Big) \;=\; f^{-1}(B) \;=\; f^{-1}\big(f((X)\big) \;=\; X. \tag{8.30}$$

In the above the second equation follows from prop. 4.2 (f^{-1} is compatible with all basic set ops) on p.23 and the last one follows from the fact that $f^{-1}\big(f((A)\big) \supseteq\; A$ for any subset of the domain of f (see cor. 4.1 on p. 26). According to prop.7.8 ("f^{-1}(open) $=$ open" continuity) on p.78, each U_j is open as the inverse image of an open set under the continuous function f. It follows from (8.30) 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 (4.19) on p.24) that

$$\begin{aligned} f(X) \;&= f(U_{j_1} \cup \dots \cup U_{j_n}) \;=\; f(U_{j_1}) \cup \dots \cup f(U_{j_n}) \\ &= f\big(f^{-1}(V_{j_1})\big) \cup \dots \cup f\big(f^{-1}(V_{j_n})\big) \;\subseteq\; V_{j_1} \cup \dots \cup V_{j_n}. \end{aligned}$$

The incclusion 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.4.6 on p. 27). We have proved that the arbitrary open cover $(V_j)_{j \in J}$ of $f(X)$ contains a finite subcover $V_{j_1} \cup \dots \cup V_{j_n}$ and it follows that $f(X)$ is indeed a compact metric subspace of Y. ■

Read the following remark for an easier way to prove the above theorem.

Remark 8.4. We could have proved the last two theorems more easily using sequence compactness instead of covering compactness but the proofs that were given generalize to abstract topological spaces (missing ingredient: defining topological subspaces $(A, \mathfrak{U}_A)$ of an abstract topological space $(X, \mathfrak{U})$ You do this by defining $\mathfrak{U}_A := \{U \cap A : U \in \mathfrak{U}\}$ and proving that $\mathfrak{U}_A$ satisfies the axioms for the open subsets of a topological space and the proofs above will go through with almost no alterations.

here is the outline of an alternate proof of theorem 8.15 which uses sequence compactness.

Given a sequence $y_n \in f(X)$ we construct a convergent subsequence y_{n_j} as follows: For each n there is some $x_n \in X$ such that $y_n = f(x_n)$ X is compact, hence sequence compact and it follows that there is $x \in X$ and a subsequence x_{n_j} such that x_{n_j} converges to x. We now use (sequence) continuity of f at x to conclude that $y_{n_j} = f(x_{n_j})$ converges to $f(x) \in f(X)$. ■

References

[1] *Matthias Beck and Ross Geoghegan.* The Art of Proof. *Springer, 1st edition, 2010.*

[2] *Gustave Choquet.* Lectures on Analysis, Vol. 2: Representation Theory. *Benjamin, New York, 1st edition, 1969.*

[3] *Pete L. Clark.* mazur-330-relat-func.pdf - Lecture notes on relations and functions. *1st edition.*

[4] *A.N. Kolmogorov and S.V. Fomin.* Introductory Real Analysis. *Dover, Mineola, 1st edition, 1975.*

[5] *James R. Munkres.* Topology. *Prentice-Hal, 1st edition, 2000.*

[6] *Unknown.* mazur-330-func-1.pdf - Introduction to Functions Ch.2. *1st edition.*

[7] *Unknown.* mazur-330-func-2.pdf - Properties of Functions Ch.2. *1st edition.*

[8] *Unknown.* mazur-330-sets-1.pdf - Ch.1: Introduction to Sets and Functions. *1st edition.*

[9] *Unknown.* mazur-330-sets-2.pdf - Ch.4: Applications of Methods of Proof. *1st edition.*

[10] *Boto Von Querenburg.* Mengentheoretische Topologie. *Springer, Berlin, Heidelberg, New York, 1st edition, 1973.*

[11] *Richard E. Williamson, Richard H. Crowell, and Hale F. Trotter.* Calculus of Vector Functions. *Prentice Hall, Englewood Cliffs, 3rd edition, 1972.*

List of Symbols

$A \biguplus B$ - *disjoint union*, 5
$A^{\complement}$ - *complement*, 5
$A_{\ell ow\delta}$ - *lower bounds of* A, 35
$A_{upp\delta}$ - *upper bounds of* A, 35
$f_n(\cdot) \to f(\cdot)$ *(pointwise convergence)*, 85
$f_n(\cdot) \overset{uc}{\to} f(\cdot)$ *(uniform convergence)*, 85
$\bar{A}$ *(closure of* A*)*, 72
$\complement A$ - *complement*, 5
$\inf(x_i),\ \inf(x_i)_{i\in I},\ \inf\limits_{i\in I} x_i$ *(sequences, families)*, 37
$\inf(x_n),\ \inf(x_n)_{n\in\mathbb{N}},\ \inf\limits_{n\in\mathbb{N}} x_n$ *(sequences)*, 37
$\inf(A)$ - *infimun of* A, 35
$\inf\limits_{x\in A} f(x)$ *(infimum of* $f(\cdot)$*)*, 37
$\inf_A f$ *(infimum of* $f(\cdot)$*)*, 37
$\lim\limits_{n\to\infty} x_n$, 70
$\liminf_{n\to\infty} x_j$ - *limit inferior*, 38
$\limsup_{n\to\infty} x_j$ - *limit superior*, 38
$\mathbb{N}, \mathbb{N}_0$, 5
$\mathbb{R}^+, \mathbb{R}_{>0}$, 5
$\mathbb{R}_+, \mathbb{R}_{\geqq 0}$, 5
$\mathbb{R}_{>0}, \mathbb{R}^+$, 5
$\mathbb{R}_{\geqq 0}, \mathbb{R}_+$, 5
$\mathbb{Z}_+, \mathbb{Z}_{\geqq 0}$, 5
$\max(A), \max A$ - *maximum of* A, 34
$\min(A), \min A$ - *minimum of* A, 34
$\sup(x_i),\ \sup(x_i)_{i\in I},\ \sup\limits_{i\in I} x_i$ *(sequences, families)*, 37
$\sup(x_n),\ \sup(x_n)_{n\in\mathbb{N}},\ \sup\limits_{n\in\mathbb{N}} x_n$ *(sequences)*, 37
$\sup(A)$ - *supremun of* A, 35
$\sup\limits_{x\in A} f(x)$ *(supremum of* $f(\cdot)$*)*, 37
$\sup_A f$ *(supremum of* $f(\cdot)$*)*, 37
$f(A)$ - *direct image*, 23
$f^{-1}(B)$ - *indirect image (preimage)*, 23
$x \bullet y$ - *inner product*, 54
$x_n \to -\infty$, 71
$x_n \to \infty$, 71
$x_n \to a$, 70
$(X, d(\cdot, cdot))$ *(metric space)*, 60
$(A, d_{A\times A})$ - *metric subspace*, 92
$(x_1, x_2, \dots, x_N)$ *(N–tuple)*, 17
$(x_1, x_2, x_3, \dots, x_N)$ *(N–dimensional vector)*, 43
$-f(\cdot), -f$ *(negative function)*, 34
$-x$ *(negative of* x*)*, 49
$0(\cdot)$ *(zero function)*, 33
$A \times B$ *(cartesian product of* 2 *sets)*, 16
$A^{\complement}$ *(complement of* A*)*, 15
$B_\varepsilon(x_0)$ *(*ε*–neighborhood)*, 66
$X^I = \prod\limits_{i\in I} X$ *(cartesian product)*, 18
$X_1 \times X_2 \times X_3 \times \dots \times X_N$ *(cartesian product)*, 17
$\Gamma_f, \Gamma(f)$ *(graph of* f*)*, 13
$\|x\|$ *(norm on a vector space)*, 57, 60
$\mathscr{B}(X, \mathbb{R})$ *(bounded real functions)*, 63
$\mathscr{C}(X, \mathbb{R})$ *(all continuous real functions on X)*, 82
$\mathscr{C}_{\mathscr{B}}(X, \mathbb{R})$, 82
$\mathscr{F}(X, \mathbb{R})$ *(all real functions on X)*, 50
$\mathfrak{B}$ *(basis of a topology)*, 69
$\mathfrak{N}(x)$ *(neighborhood system)*, 69
$\mathfrak{U}_{\|\cdot\|}$ *(norm topology)*, 69
$\mathfrak{U}_{d(\cdot,\cdot)}$ *(metric topology)*, 68
$\vec{x} + \vec{y}$ *(vector sum)*, 44
$\alpha\vec{x}$ *(scalar product)*, 44
αf *(scalar product of functions)*, 33
$\alpha x, \quad \alpha \cdot x$ *(scalar product)*, 49
$\complement A$ *(complement of* A*)*, 15
$\emptyset$ *(empty set)*, 14
$\mathbb{N}$ *(natural numbers)*, 6
$\mathbb{N}_0$ *(non–negative integers)*, 7
$\mathbb{Q}$ *(rational numbers)*, 6
$\mathbb{R}$ *(real numbers)*, 6
$\mathbb{R}^N$ *(all* N*–dimensional vectors)*, 43
$\mathbb{R}^\star$ *(non-zero real numbers)*, 7
$\mathbb{R}^+$ *(positive real numbers)*, 7
$\mathbb{R}_{>0}$ *(positive real numbers)*, 7
$\mathbb{R}_{\geq 0}$ *(non–negative real numbers)*, 7
$\mathbb{R}_{\neq 0}$ *(non-zero real numbers)*, 7
$\mathbb{R}_+$ *(non–negative real numbers)*, 7
$\mathbb{Z}$ *(integers)*, 6
$\mathbb{Z}_{\geq 0}$ *(non–negative integers)*, 7
$\mathbb{Z}_+$ *(non–negative integers)*, 7
$\pi_j(\cdot)$ *(*j*th coordinate function)*, 53
$\pi_{i_1,i_2,\dots,i_M}(\cdot)$ *(*M*–dim projection)*, 54
$\prod\limits_{i\in I} X_i$ *(cartesian product)*, 18
card$(X), |X|$ *(cardinality of a set)*, 18
ε_{x_0} *(Radon integral)*, 54
$d_{A\times A}$ - *induced/inherited metric*, 92

$f + g$ *(sum of functions)*, 33
$f - g$ *(difference of functions)*, 33
f/g, $\frac{f}{g}$ *(quotient of functions)*, 33
$f^{-1}(\cdot)$ *(inverse function)*, 11
f_A *- restriction of* f *to* A, 92
fg, $f \cdot g$ *(product of functions)*, 33
$id(\cdot), id_X(\cdot)$ *(identity)*, 11
$x + y$ *(vector sum)*, 49
$\|\vec{v}\|$ *(Euclidean norm)*, 46
{} (empty set), 14

$A \cap B$ *(*A *intersection* B*)*, 14
$A \setminus B$ *(*A *minus* B*)*, 14
$A \subset B$ *(*A *is strict subset of* B*)*, 14
$A \subseteq B$ *(*A *is subset of* B*)*, 14
$A \subsetneqq B$ *(*A *is strict subset of* B*)*, 14
$A \triangle B$ *(symmetric difference of* A *and* B*)*, 14
$A \uplus B$ *(*A *disjoint union* B*)*, 14
$B \supset A$ *(*B *is strict superset of* A*)*, 14
$B \supsetneqq A$ *(*B *is strict superset of* A*)*, 14
$(X, \mathfrak{U})$ *(topological space)*, 68
(x_j) *(sequence)*, 12, 87
$\sum\limits_{k=1}^{\infty} a_k$ *(series)*, 88
$\|\vec{v}\|$ *(length or norm of* $\vec{v}$*)*, 44
$\mathfrak{U}$ *(topology)*, 68
$A \cup B$ *(*A *union* B*)*, 14
$A \supseteq B$ *(*A *is superset of* B*)*, 14
$x_n \nearrow \xi \ (n \to \infty)$, 71
$x_n \searrow \xi \ (n \to \infty)$, 71

card$(X) <$ *card*(Y), 29
card$(X) =$ *card*(Y), 29
card$(X) \leqq$ *card*(Y), 29

g.l.b.(A) *- greatest lower bound of* A, 35

l.u.b.(A) *- least upper bound of* A, 35

Index

N–tuple, 17
ε–grid, 100
ε–net, 100

absolute convergence, 90
algebraic number, 32
argument, 10
associativity, 49
associativity, cartesian product, 17

basis (of a topology), 69
Bernstein polynomial, 84
bilinear, 55
bounded above, 35
bounded below, 35

cardinality, 18
    comparison of, 29
    equality, 29
cartesian product, 16, 17
cartesian product of N sets, 17
cartesian product of a family, 18
Cauchy criterion, 73
Cauchy sequence, 73
closed set, 72
closed with repect to an operation, 50
closure, 72
cluster point, 72
codomain, 10
commutativity, 49
compact, 106
    sequentially, 102
complement, 15
complete set, 75
contact point, 72
continuous function, 77
convergence, 70
convergence, uniform, 85
coordinate function, 54
countable set, 19
countably infinite set, 19
covering, 106

De Morgan's Law, 15
decreasing sequence, 5
dense set, 7

dimension, 43
direct image, 23
discrete topology, 69
distributive laws, 49
domain, 10

element of a set, 9
embed, 51
empty set, 14
equality
    arbitrary cartesian products, 18
    cartesian products, 16
    finite cartesian products, 17
Euclidean norm, 46
extension of a function, 92

family, 12
    supremum, 37
finite sets, 18
finite subcovering, 106
first axiom of countability, 70
first countable, 70
function, 10
    argument, 10
    bijective, 11
    bilinear, 55
    bounded from above, 65
    bounded from below, 65
    bounded function, 65
    codomain, 10
    constant function, 33
    continuous, 77
    convergence, 85
    difference, 33
    direct image, 23
    domain, 10
    extension, 92
    function value, 10
    identity, 11
    image, 11
    image set, 10
    independent variable, 10
    infimum, 37
    injective, 11
    inverse, 11

negative function, 34
onto, 11
pointwise convergence, 85
preimage, 10, 23
product, 33
quotient, 33
range, 11
real function, 11, 33
real valued function, 33
restriction, 92
scalar product, 33
sum, 33
supremum, 36
surjective, 11
target, 10
uniform continuity, 105
uniform convergence, 85
zero function, 33
function value, 10

graph, 13
greatest lower bound, 35
grid point, 100

identity, 11
image, 11
image set, 10
increasing sequence, 5
independent variable, 10
index set, 12
indexed family, 12
indirect image, 23
indiscrete topology, 69
induced metric, 92
infimum, 35
infimum of a family, 37
infimum of a sequence, 37
inherited metric, 92
inner product, 54
integer, 7
interior point, 67
interior point (topological space), 69
irrational number, 7
isolated point, 72

least upper bound, 35
lim inf, 38
lim sup, 38
limit, 70
limit inferior, 38
limit point, 72
limit superior, 38
linear combination, 52, 81
linear mapping, 52, 81
linear space, 49
lower bound, 34

mapping (see function), 10
maximum, 34, 36
member of a set, 9
metric, 59
induced, 92
inherited, 92
metric space, 60
metric subspace, 92
metric topology, 68
minimum, 34
monomial, 82

natural number, 7
negative, 49
neighborhood, 66, 67, 69
neighborhood (metric space), 67
neighborhood basis, 69
neighborhood system, 69
non-decreasing sequence, 5
non-increasing sequence, 5
norm
Euclidean norm, 46
norm on a vector space, 57, 60
norm topology, 69
normalized form (polynomials), 84
not countable set, 19
null vector, 49
nullspace, 49
numbers
algebraic number, 32
integers, 6
irrational number, 6
natural numbers, 6
rational numbers, 6
real numbers, 6
transcendental number, 32

open covering, 106
open neighborhood (metric space), 67

open set, 67
trace, 93
ordered pair, 16

permutation, 89
infinite, 89
point of accumulation, 72
pointwise convergence, 85
polynomial, 82
preimage, 10, 23
projection, 53, 54
projection on coordinates $i_1, i_2, \ldots, i_M$, 54

Radon integral, 54
range, 11
rational number, 7
real function, 11
real number, 7
restriction of a function, 92

scalar, 52, 81
scalar product, 49
second axiom of countability, 70
second countable, 70
sequence, 12, 87
decreasing, 5
increasing, 5
infimum, 37
non-decreasing, 5
non-increasing, 5
numeric, 87
partial sums, 87
strictly decreasing, 5
strictly increasing, 5
supremum, 37
tail sets, 37
sequence compact, 102
sequentially compact, 102
series, 12, 88
absolute convergence, 90
set, 9
bounded, 64
cardinality, 18
compact, 106
complete, 75
countable, 19
countably infinite, 19
covering, 106
dense, 7
diameter, 64
difference, 14
difference set, 14
disjoint, 14
finite, 18
finite subcovering, 106
intersection, 14
mutually disjoint, 14
not countable, 19
open covering, 106
strict subset, 14
strict superset, 14
subset, 14
superset, 14
symmetric difference, 14
uncountable, 19
union, 14
strictly decreasing sequence, 5
strictly increasing sequence, 5
subspace
metric, 92
subspace (of a vector space), 49
sum, 49
supremum, 35
supremum of a family, 37
supremum of a sequence, 37

tail sets, 37
target, 10
topological space, 68
first axiom of countability, 70
first countable, 70
second axiom of countability, 70
second countable, 70
topology, 68
discrete topology , 69
generated by metric, 68
generated by norm, 69
indiscrete topology , 69
induced by metric, 68
induced by norm, 69
metric topology, 68
norm topology , 69
totally bounded, 102
trace, 93
transcendental number, 32

triangle inequality, 7, 8

uncountable set, 19
uniform continuity, 105
uniform convergence, 85
universal set, 15
upper bound, 34

vector, 43
    length , 44
    norm , 44
    norm, *Euclidean*, 46
    scalar product, 44
    sum, 44, 48
vector (element of a vector space), 49
vector space, 43, 49
    norm, 60

zero element, 49
zero vector, 49