Math 330 - Additional Material

Michael Fochler
Department of Mathematics
Binghamton University

Contents

History of Updates: *2015-04-13*

Date	*Topic*
2015-03-08	*Added new section "Basic properties of sets"*
2015-03-17	*Added new section "Maxima, suprema, limsup ..."*
2015-03-25	*Added background material: new chapters "Some Basics", "Real Functions", "Vectors and vector spaces", "Convergence and continuity"*
2015-04-13	*Major rework of "Maxima, suprema, limsup ..."*

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. It consists of two very distinct portions.

A. Material directly related to Math 330:

	Topic
1.	*All of ch.4, p.18: "Sets and Functions, direct and indirect images"*
2.	*Ch.5.3, p.29: "Maxima, suprema, limsup ..."*
3.	*Parts of Ch.7, p.51) on "Convergence and Continuity": Ch.7.1, p.51 ("Metric spaces") and the definition of neighborhoods (7.2, p.54) in ch.7.2, p.54 ("Neighborhoods and open sets")*

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.

When you do not see any comment, interpret it as "Skim this". You just want to know what's available for you.

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 are:

a. countable and countably infinite v.s. denumbrable and countable:
For us 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 notation we 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 then we often write $A \uplus B$ 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 $]0, +\infty[$, $\mathbb{R}_{\geq 0}$ or $\mathbb{R}_+$ denotes the interval $[0, +\infty[$,

c) The set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of all natural numbers excludes the number zero. We write $\mathbb{N}_0$ or $\mathbb{Z}_+$ or $\mathbb{Z}_{\geq 0}$ for $\mathbb{N} \uplus \{0\}$

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 $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$.

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

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

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

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., $.\bar{3} = 0.33333333333\dots$.

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

$$0.\bar{3} \;=\; .\bar{3} \;=\; .33\bar{3} \;=\; 0.33333333333\dots \;=\; \frac{1}{3} \;=\; \frac{2}{6} \tag{3.1}$$

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

$$-\,4 \;=\; -4.000 \;=\; -3.99\bar{9} \;=\; -\frac{12}{3} \;=\; -\frac{400}{100} \tag{3.2}$$

[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.7. 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: $A = \{1, 2.6, \text{the moon, London}\}$ 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.

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 $1.6\bar{6}$. Irrational numbers must then be those with infinitely many decimal digits without any continually repeating patterns.

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

Definition 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*** .

$\mathbb{Q} := \{\frac{n}{d} : n \in \mathbb{Z}, d \in \mathbb{N}\}$ 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: $\sqrt{2}$ is irrational.

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

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

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

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

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

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 $b \geqq 0$

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

[3] The set difference $X \setminus Y$ (see 3.11 on p.11.) is the set of all elements which belong to X but not to Y.

We shall always assume that "$\sqrt{b}$" means "$+\sqrt{b}$" and not "$-\sqrt{b}$" 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:*

$$Triangle\ Inequality : \boxed{|\boldsymbol{a} + \boldsymbol{b}| \leqq |\boldsymbol{a}| + |\boldsymbol{b}|} \tag{3.3}$$

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

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

Proposition 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

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

This is the principle: Let us assume that we know that some statement can be proven 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 Assumption. We can prove the following for any $k \in \mathbb{N}_0$ such that $k \geqq k_0$: if the property is true for k than it will also be true for $k + 1$

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

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

In other words,

$$|(a_1 + a_2 + \ldots + a_k) + a_{k+1}| \;\le\; B + |a_{k+1}| \;=\; (|a_1| + |a_2| + \ldots + |a_k|) + |a_{k+1}| \tag{3.6}$$

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 $k \ge k_0 = 2$. ■

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

Remark 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 proven 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 Sets, Functions 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, ...

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**.

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.

3.3 Basic Properties of Mappings

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. [4]

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 every $x \in \mathbb{R}$.*

(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 exactly one 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 uniquely determined 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

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

[4] If you do not know about the different kinds of numbers, review the section "Numbers" on p.4. 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 ***image*** of the function $f(\cdot)$. [5]

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

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

.

They say "f maps X into Y" and "f maps the domain value x into 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 people will only call a mapping a function if its target is a subset of the real numbers [6] I'll try to stick to the following conventions: 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.23).

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

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

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).

We write $g(y) = x$ for the mapping that assigns to any $y \in Y$ the unique element $x \in X$ whose image $f(x)$ equals y. We see that $g(\cdot)$ indeed is 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 about sets.

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:

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

[5] We distinguish the image set Y of $f(\cdot)$ from the image $f(X)$.

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

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 $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$

You should have been taught about sequences already, but here are two examples:

Example 3.1 (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 3.2 (Series (summation sequence))**.** $s_k := 1 + 2 + \dots + k \ (k = 1, 2, 3, \dots)$

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

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.

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* $(x_\beta)_{\beta \in J}$.

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.

Definition 3.8 (empty set)**.** $\emptyset$ or {} denotes the ***empty set.*** It is the one set that does not contain any elements.

Definition 3.9 (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 3.10 (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 $A_i \cap A_j = \emptyset$

Definition 3.11 (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 $A \subseteq X$.

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

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 3.12 (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 3.13 (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:

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

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

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

Example 3.3 (Complement of a set relative to the unit interval)**.** Assume we are exclusively dealing with the unit interval, i.e., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. Let $a \in [0,1]$ and $\delta > 0$ and

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

the δ–neighborhood [7] 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 (3.12) on p.11). Then the following "duality principle" holds for any indexed family $(A_\alpha)_{\alpha \in I}$ of sets:*

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

To put this in words, the complement of an arbitrary union is the intersection of the complements and the complement of an arbitrary intersection is the union of the complements.

Generally speaking 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.

[7] Neighborhoods of a point will be discussed in the chapter on the topology of $\mathbb{R}^n$ (see (7.2) on p.54) In short, the δ–neighborhood of a is the set of all points with distance less than δ from a.

Definition 3.14 (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 3.4 (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

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

Now we know that the next definition makes sense:

Definition 3.15 (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 [8]

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

[8] 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.

Two elements $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$ 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 $(x_1, x_2, \ldots, x_N) = (y_1, y_2, \ldots, y_N)$.

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

Example 3.5 (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 $\mathbb{R}^N = \{(x_1, x_2, \ldots, x_N) : x_j \in \mathbb{R}\}$ 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 3.4 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*** $(x_1, x_2, \ldots, x_N)$ is also called a vector of dimension N. You will read more about this in the chapter 6, p.36, 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 $\underbrace{(0, 0, \ldots, 0)}_{N \text{times}}$ and ends in x.

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 3.16 (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* [9]

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

3.4 Countable sets

Everyone understands what a finite set is: It's a set with a finite number of elements. You may be surprised to hear this but there are ways to classify the degree of infinity when looking at infinite sets.

The "smallest degree of infinity" is found in sets that can be compared, in a sense, to the set $\mathbb{N}$ of all natural numbers. Look back to definition (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 sets whose elements can be arranged into a sequence of that nature.

Definition 3.17 (Countable and countably infinite sets)**.** Let X an arbitrary set such that all of its elements can be arranged in a sequence

$$X \;=\; \{x_1, x_2, x_3, \dots\} \qquad \text{where all } x_j \text{ are different}$$

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 ***countable set***

and we also say that X is countable.

A set that is neither finite nor countably infinite is called ***uncountable***

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

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

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. ■

[9] Mappings or functions were briefly discussed already in paragraph 3.3 on p.8. 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.

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 (3.10) on p.11). 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

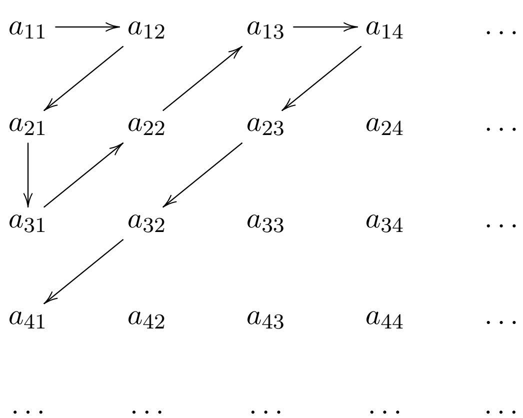
$$B_1 := A_1; \quad B_2 := A_2 \setminus B_1; \quad B_3 := A_3 \setminus B_2; \quad \dots$$

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*



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*

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

Then each set is finite and $Q_1 = \bigcup_{k \in \mathbb{N}} A_k$ *is a countable union of countably many finite sets and hence, according to the previous theorem (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 generalization of "lemma, hwk 6" which was given you in homework 6 as 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 for all $j \in J$ $X_j \subseteq Y \subseteq Z_j \subseteq W$. 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 $j \in J$, i.e., $x \in Y$ 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

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

1)

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

2)

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

Notational conveniences:

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

Proposition 4.1. *Some simple properties:*

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

$$A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2) \quad (4.5)$$

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

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

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

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

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:

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

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

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

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

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 (4.10): 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 (4.11)) - "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 (4.11)) - 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 "⊆":

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

B. Proof of "⊇":

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

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

■

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

Proposition 4.4 (Indirect image and fibers of f)**.** *We define on X the equivalence relation*

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

Then the following is true:

$$x \in X \;\Rightarrow\; \Big[\, [x]_f \;=\; \{\hat{x} \in X : f(\hat{x} = f(x))\} \;=\; f^{-1}\{f(x)\} \tag{4.33}$$

$$A \subseteq X \;\Rightarrow\; f^{-1}(f(A)) \;=\; \bigcup_{a \in A} [a]_f. \tag{4.34}$$

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*

$$f^{-1}(f(A)) = f^{-1}(\bigcup_{x \in A}\{f(x)\}) \tag{4.35}$$

$$= \bigcup_{x \in A} f^{-1}\{f(x)\} \quad (\text{see } 4.11) \tag{4.36}$$

$$= \bigcup_{x \in A} [x]_f \quad (\text{see } 4.33) \tag{4.37}$$

■

Corollary 4.1.

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

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*

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

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

Proposition 4.5.

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

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

5 Real functions (Understand this!)

5.1 Operations on real functions

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

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

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 ***sum*** $f+g$, ***difference*** $f-g$, ***product*** fg or $f \cdot g$, ***quotient*** f/g or $\frac{f}{g}$, and ***scalar product*** αf are defined by doing the operation in question with the numbers $f(x)$ and $g(x)$ for each $x \in X$.

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

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:

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

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

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

be a real function on X. The function

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

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

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

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

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

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 Measuring the distance of real functions

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 $x - y = 12.3 - 15 = -2.7$. Oops, there are situations where direction matters and a negative distance is one that goes into the opposite direction of a positive distance, but we do not worry about it 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? Here we measure distance simply as a number, without any dimension attached to it. For the above example, you get

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

It can be shown that the distance between two two–dimensional vectors $v = (v_1, v_2)$ and $w = (w_1, w_2)$ is $\sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2}$ and the distance between two three–dimensional vectors $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ is $\sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + (w_3 - v_3)^2}$.

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 5.5 (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

$$d(f, g) \ := \ d(f(\cdot), g(\cdot)) \ := \ \sup\{|f(x) - g(x)| : x \in X\} \tag{5.2}$$

The meaning of "sup" is explained in the following remark. Once you have read it you will understand that $d(f, g)$ need not necessarily be finite. For example, if

$$X =]-1, 0[= \{a \in \mathbb{R} : -1 < x < 0\}, \qquad f(x) = 0, \quad g(x) = -1/x \quad (x \in]-1, 0[)$$

then $d(f, g) = \infty$.

Remark 5.1 (Informal definition of the supremum)**.** You can read in detail about the supremum of sets and functions in def. 5.10 on p.31. To understand the concept of measuring the distance of functions you can get by with the following, not completely accurately definition: The ***supremum*** $\sup(A)$ of a set $A \in \mathbb{R}$ of real numbers is its biggest possible value.
The "supremum" is a generalization of the ***maximum*** $\max(A)$ which is defined as the element $m^\star \in A$ which is no smaller than any other $a \in A$. Note that it is a requirement that $\max(A)$ itself belong to the set A. In contrast, $\sup(A)$ might belong to A (in which case it coincides with $max(A)$. See example a). or it might not belong to A. In that case you can pick a sequence $a_1, a_2, \dots$ of members of A which will come arbitrarily close to $\sup(A)$. See example b. Even if A has no upper bound, we still say that it has a supremum, namely $+\infty$. See example c.

Here are the examples. We'll define for all three of them $f(x) := -x$ and $g(x) := x$ on various domains X_1, X_2, X_3.

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 limit 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 $\sup$ must be $+\infty$. 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.

The following picture illustrates how to measure the distance between two functions $f(\cdot)$ and $g(\cdot)$: Plot their graphs above their domain 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 $\mathbb{R}$ which corresponds to the thick portion of the x–axis.

Now that you know how to measure the distance $d(f(\cdot), g(\cdot))$ between two real functions $f(\cdot), g(\cdot)$, the next picture shows you how to visualize the δ–neighborhood

$$B_\delta(f) := \{g(\cdot) : X \to \mathbb{R} : d(f,g) < \delta\} = \{g(\cdot) : X \to \mathbb{R} : \sup_{x \in X} |f(x) - g(x)| < \delta\} \tag{5.3}$$

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 $f(\cdot)$.

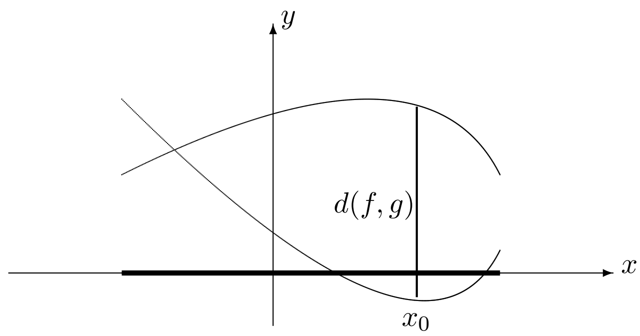


Figure 1: Distance of two real functions.

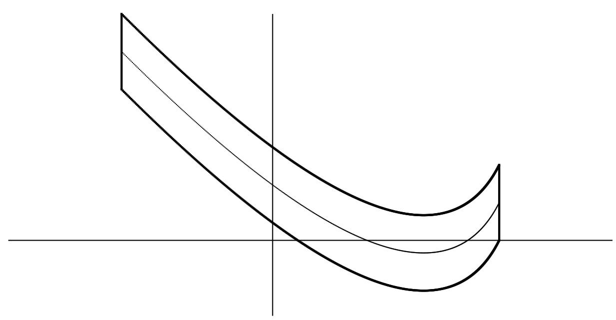


Figure 2: δ–neighborhood of a real function.

In other words assuming that the domain A is a single, connected chunk and not a collection of more than one separate intervals, the δ–neighborhood of $f(\cdot)$ is a "band" whose contours are made up on the left and right by two vertical lines and on the top and bottom by two lines that look like the graph of $f(\cdot)$ itself but have been shifted up and down by the amount of δ.

The distance of a real function $f(\cdot)$ to the zero function (see 5.3 on 23) has a special notation.

Definition 5.6 (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

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

Do you see that for any two bounded real functions $f(\cdot), g(\cdot)$ we have

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

Lemma 5.1 (Properties of the norm of a real functions)**.** *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

$$\|\cdot\| : \ \mathscr{B}(X,\mathbb{R}) \longrightarrow \mathbb{R} \qquad h(\cdot) \longmapsto \|h(\cdot)\| = \sup\{|f(x)| : x \in X\}$$

has the following three properties:

$$\text{(5.4a)} \qquad \|f\| \geqq 0 \quad \forall f \in \mathscr{B}(X,\mathbb{R}) \quad \text{and} \quad \|f\| = 0 \iff f(\cdot) = 0 \qquad \textbf{positive definite}$$

$$\text{(5.4b)} \qquad \|\alpha f(\cdot)\| = |\alpha| \cdot \|f(\cdot)\| \quad \forall f \in \mathscr{B}(X,\mathbb{R}), \forall \alpha \in \mathbb{R} \qquad \textbf{homogeneity}$$

$$\text{(5.4c)} \qquad \|f(\cdot) + g(\cdot)\| \leqq \|f(\cdot)\| + \|g(\cdot)\| \quad \forall f, g \in \mathscr{B}(X,\mathbb{R}) \qquad \textbf{triangle inequality}$$

Proof:
a. It is certainly true that $\|f\| \geqq 0$ for any function $f(\cdot)$ because the absolute value $|f(x)| \geqq 0$ for any argument x. If $0(\cdot)$ is the zero function then obviously $|0(x)| = 0$ for any argument x and hence $\|0(\cdot)\| = 0$. Conversely, let $f(\cdot)$ be a function on X such that $\|f(\cdot)\| = 0$. This means that $\sup\{|f(y)| : y \in X\} = 0$ which is only possible if there is not even one $a \in X$ where $|f(a)| > 0$. In other words, we must have $|f(x)| = 0$ for all arguments x. But this can only be if $f(x) = 0$ for all x. This in turn means that $f(\cdot)$ is the zero function $0(\cdot)$.

b. Let $f \in \mathscr{B}(X,\mathbb{R})$ and $\alpha \in \mathbb{R}$. For any $x \in X$ $f(x)$ is just an ordinary real number and we have

$$|\alpha f(x)| \ = \ |\alpha| \cdot |f(x)|$$

because it is true that $|\alpha\beta| = |\alpha| \cdot |\beta|$ for any two real numbers α and β. If you have trouble seeing that, just look separately at the four cases where α is either non–negative or negative and β is either non–negative or negative.

c. According to the triangle inequality for real numbers (see [1] B/G (Beck/Geoghegan) ch.10) we have, for any $x \in X$

$$\begin{aligned} |f(x) + g(x)| &\leqq |f(x)| + |g(x)| \\ &\leqq \sup\{|f(x_1)| : x_1 \in X\} + \sup\{|g(x_2)| : x_2 \in X\} \\ &= \|f(\cdot)\| + \|g(\cdot)\| \end{aligned}$$

In other words, the number $\|f(\cdot)\| + \|g(\cdot)\|$ is an upper bound of the set $\{|f(x) + g(x)| : x \in X\}$ It follows that it also dominates the least upper bound [10] $\sup\{|f(x) + g(x)| : x \in X\} = \|f(\cdot) + g(\cdot)\|$ of this set and we are done. ■

Lemma 5.2 (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

$$d(\cdot) : \ \mathscr{B}(X,\mathbb{R}) \times \mathscr{B}(X,\mathbb{R}) \longrightarrow \mathbb{R} \qquad (h_1, h_2) \longmapsto d(h_1, h_2) := \|h_1 - h_2\|$$

[10] see (5.10) on p.31

has the following three properties: [11]

$$(h_1, h_2) \mapsto d(h_1, h_2) := \|h_1 - h_2\|$$

has the following three properties:

(5.5a)
$d(f,g) \geqq 0 \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R}) \quad$ *and* $\quad d(f,g) = 0 \iff f(\cdot) = g(\cdot)$ ***positive definite***

(5.5b)
$d(f,g) = d(g,f) \quad \forall f(\cdot), g(\cdot) \in \mathscr{B}(X,\mathbb{R})$ ***symmetry***

(5.5c)
$d(f,h) \leqq d(f,g) + d(g,h) \quad \forall\, f,g,h \in \mathscr{B}(X,\mathbb{R})$ ***triangle inequality***

Before we prove those properties, let us quickly examine what they 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 which means that $f(x) = g(x)$ for each argument $x \in X$.
"symmetry": the distance from $f(\cdot)$ to $g(\cdot)$ is no different to that from $g(\cdot)$ to $f(\cdot)$. No such thing as a negative distance if we 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 be less than using an intermediary $g(\cdot)$ and first comparing $f(\cdot), g(\cdot)$ and then $g(\cdot), h(\cdot)$.

Proof of lemma 5.2:
Watch this proof carefully: we prove this lemma just from the properties for a norm listed in lemma 5.1: positive definiteness, homogeneity and the triangle inequality for norms.

a. It is clear that $d(f,g) = \|f - g\| \geqq 0$ because $\|h\| \geqq 0$ for all $h(\cdot)$. $d(f,g) = 0$ if and only if $\|f - g\| = 0$ if and only if $f - g = 0$ [12] *if and only if $f(x) - g(x) = 0$ for all $x \in X$* [13] *if and only if $f(\cdot) = g(\cdot)$.*

b. Clear because $|\alpha - \beta| = |\beta - \alpha|$ for any two numbers α and β, but let us find a proof that just relies on the properties of a norm instead:

$$d(f,g) = \|f - g\| = \|(-1)(g - f)\| = |-1| \cdot \|(g - f)\| = 1 \cdot \|(g - f)\| = \|(g - f)\| = d(g,f)$$

c. Because of the triangle inequality for norms:

$$d(f,h) = \|f - h\| = \|(f - g) + (g - h)\| \leqq \|f - g\| + \|g - h\| = d(f,g) + d(g,h)$$

and we are done. ■

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

Definition 5.7 (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 $l \leq a$ for all $a \in A$. We call u an ***upper bound*** of A if $u \geq a$ for all $a \in A$.

[11] 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.
[12] because of the positive definiteness for norms
[13] actually, we only needed the very general property that $f - g = 0$ is equivalent to saying $f = g$

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 $a \leq u_1$ for all $a \in A$. Hence $u_2 \leq u_1$. As u_2 is an upper bound, it follows that $u_1 \leq u_2$ and we have equality $u_1 = u_2$. The proof for minima is similar. ■

Definition 5.8. Given $A \subseteq \mathbb{R}$ we define

$$\begin{aligned} A_{\ell owb} &:= \{l \in \mathbb{R} : l \text{ is lower bound of } A\} \\ A_{uppb} &:= \{u \in \mathbb{R} : u \text{ is upper bound of } A\}. \end{aligned} \tag{5.6}$$

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

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 5.2. $A_{\ell owb}$ and/or A_{uppb} may be empty: $A = \mathbb{R}$, $A = \mathbb{R}_{>0}$, $A = \mathbb{R}_{<0}$,

Definition 5.9. Let $A \subseteq \mathbb{R}$. If A_{uppb} is not empty then $\min(A_{uppb})$ exists by axiom 5.1 and it is unique by prop. 5.5. We write $\sup(A)$ or l.u.b.(A) for $\min(A_{uppb})$ and call it the ***supremum*** or ***least upper bound*** of A.

We shall see in cor.5.1 that, if $A_{\ell owb}$ is not empty, then $\max(A_{\ell owb})$ exists and it is unique by prop. 5.5. We write $\inf(A)$ or g.l.b.(A) for $\max(A_{\ell owb})$ 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 $-A = \{-y : y \in A\}$:*

$$\begin{aligned} &-x \text{ is a lower bound of } A \iff x \text{ is an upper bound of } -A \quad \text{and vice versa}, \\ &-x \in A_{uppb} \iff x \in (-A)_{\ell owb} \quad \text{and vice versa}, \\ &-x = \sup(A) \iff x = \inf(-A) \quad \text{and vice versa}, \\ &-x = \max(A) \iff x = \min(-A) \quad \text{and vice versa}. \end{aligned} \tag{5.7}$$

Proof: A simple consequence of

$$-x \leq y \iff x \geq -y \quad \text{and} \quad -x \geq y \iff x \leq -y. \qquad ■$$

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: $\inf(A) = \sup(-A)$.

Definition 5.10. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Let

$$T_n := \{x_j : j \geq n\} = \{x_n, x_{n+1}, x_{n+2}, x_{n+3}, \dots\} \tag{5.8}$$

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 5.3. 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., $x_n \leq x_{n+1}$ for all n (see def. 2.1, p.4) and bounded above. Then $\lim\limits_{n\to\infty} x_n$ exists and coincides with $\sup\{x_n : n \in \mathbb{N}\}$ (see [1] B/G thm 10.19, p.101). And, for a sequence y_n of real numbers that is non-increasing, i.e., $y_n \geq y_{n+1}$ 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

$$\begin{aligned} \inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big) &= \lim_{n\to\infty}\big(\sup(T_n)\big) = \lim_{n\to\infty}\big(\sup\{x_j : j \in \mathbb{N}, j \geq n\}\big), \\ \sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big) &:= \lim_{n\to\infty}\big(\inf(T_n)\big) = \lim_{n\to\infty}\big(\inf\{x_j : j \in \mathbb{N}, j \geq n\}\big). \end{aligned} \tag{5.9}$$

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

In other words, (5.9) becomes

$$\begin{aligned} \inf_{n\in\mathbb{N}}\big(\sup_{j\geq n} x_j\big) &= \inf\big(\{\sup(T_n) : n \in \mathbb{N}\}\big) = \lim_{n\to\infty}\big(\sup(T_n)\big) = \lim_{n\to\infty}\big(\sup_{j\geq n} x_j\big), \\ \sup_{n\in\mathbb{N}}\big(\inf_{j\geq n} x_j\big) &= \sup\big(\{\inf(T_n) : n \in \mathbb{N}\}\big) = \lim_{n\to\infty}\big(\inf(T_n)\big) = \lim_{n\to\infty}\big(\inf_{j\geq n} x_j\big). \end{aligned} \tag{5.10}$$

The above justifies the following definition:

Definition 5.11. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and let $T_n = \{x_j : j \in \mathbb{R}, j \geq n\}$ be the tail set for x_n. Assume that T_n is bounded above for some $n_0 \in \mathbb{N}$ (and hence for all $n \geq n_0$). We call

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

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 $n \geq n_0$). We call

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

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

$$\begin{aligned}
\mathscr{U} &:= \{y \in \mathbb{R} : T_n \cap [y,\infty] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\\
\mathscr{U}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \geq y\},\\
\mathscr{U}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \cdots \in \mathbb{N} \text{ such that } x_{n_j} \geq y \text{ for all } j \in \mathbb{N}\},\\
\mathscr{U}_3 &:= \{y \in \mathbb{R} : x_n \geq y \text{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{5.11}$$

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.

$$u - \varepsilon \in \mathscr{U} \quad \text{and} \quad u + \varepsilon \notin \mathscr{U} \quad \text{for all } \varepsilon > 0. \tag{5.12}$$

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

Proof of A:

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

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

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

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 $y \leq x < \infty$, i.e., $x \geq y$. 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 $x_{n_{j_0}+k} \geq y$. 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 $x_j \geq y$. At most finitely many of those j can be less than that given n and there must be (infinitely many) $j \in J$ such that $j \geq n$ Pick any one of those, say j'. Then $x_{j'} \in T_n$ and $x_{j'} \geq y$. It follows that $y \in \mathscr{U}$

We have shown the following sequence of inclusions:

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

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

Proof of B: Let $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $y_2 \in \mathscr{U}$. It follows from $[y_2,\infty] \subseteq [y_1,\infty]$ 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 - (5.12) 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 - (5.12) part 2, $u + \varepsilon \notin \mathscr{U}$: This is trivial as $u + \varepsilon > u = \sup(\mathscr{U})$ implies that $y \leq u < u + \varepsilon$ for all $y \in \mathscr{U}$. But then $y \neq u$ for all $y \in \mathscr{U}$, i.e., $u \notin \mathscr{U}$. This proves $u + \varepsilon \notin \mathscr{U}$.

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

$$u \, - \, 1/j \, \leq x_{n_j} \, \leq \, u \, + \, 1/j. \tag{5.13}$$

Base case: We have proved as part of C.1 that $x_n \geq u + 1$ for at most finitely many indices n. Let K be the largest of those. As $u - 1 \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geq u - 1$. Infinitely many of them must exceed K. We pick one of them and that will be n_1. Clearly, n_1 satisfies (5.13) and this proves the base case.

Let us now assume that $n_1 < n_2 < \dots < n_k$ satisfying (5.13) have been constructed. $x_n \geq u + 1/(k+1)$ is possible for at most finitely many indices n. Let K be the largest of those. As $u - 1/(k+1) \in \mathscr{U}_3$, there are infinitely many n such that $x_n \geq u - 1/(k+1)$. Infinitely many of them must exceed $\max(K, n_k)$. We pick one of them and that will be n_{k+1}. Clearly, n_{k+1} satisfies (5.13) 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 (5.13) that $|x_{n_j} - u| \, \leq 1/j \, < \, 1/N < \varepsilon$ for all $j \geq n$ and this proves that $x_{n_j} \to u$ as $j \to \infty$.

We will be finished with the proof of C.2 if we can show that if $w > u$ then there is no sequence $n_1 < n_2 < \dots$ such that $x_{n_j} \to w$ as $j \to \infty$. Let $\varepsilon := (w - u)/2$. According to (5.12), $u + \varepsilon \notin \mathscr{U}$. But then, by definition of $\mathscr{U}$, there is $n \in \mathbb{N}$ such that $T_n \cap [u+\varepsilon, \infty[= \emptyset$. But $u + \varepsilon = w - \varepsilon$ and we have $T_n \cap [w - \varepsilon, \infty[= \emptyset$. This implies that $|w - x_j| \geq \varepsilon$ for all $j \geq n$ and that rules out the possibility of finding n_j such that $\lim\limits_{j\to\infty} x_{n_j} = w$. ■

Corollary 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 $\{x_j : j \geq n$ if and only if $-x$ is a lower bound for $\{y_j : j \geq n$.

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.30.

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

$$\begin{aligned}
\mathscr{L} &:= \{y \in \mathbb{R} : T_n \cap [-\infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\},\\
\mathscr{L}_1 &:= \{y \in \mathbb{R} : \text{ for all } n \in \mathbb{N} \text{ there exists } k \in \mathbb{N} \text{ such that } x_{n+k} \leq y\},\\
\mathscr{L}_2 &:= \{y \in \mathbb{R} : \exists \text{ subsequence } n_1 < n_2 < n_3 < \dots \in \mathbb{N} \text{ such that } x_{n_j} \leq y \text{ for all } j \in \mathbb{N}\},\\
\mathscr{L}_3 &:= \{y \in \mathbb{R} : x_n \leq y \text{ for infinitely many } n \in \mathbb{N}\}.
\end{aligned} \tag{5.14}$$

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.

$$l + \varepsilon \in \mathscr{L} \quad \text{and} \quad l - \varepsilon \notin \mathscr{L} \tag{5.15}$$

C2. *There exists a subsequence* $(n_j)_{j \in \mathbb{N}}$ *of integers such that* $l = \lim_{j \to \infty} x_{n_j}$ *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 Then*

$$\begin{aligned} u &= \sup(\mathscr{U}) = \sup\{y \in \mathbb{R} : T_n \cap [y, \infty[\neq \emptyset \text{ for all } n \in \mathbb{N}\}, \\ l &= \inf(\mathscr{L}) = \inf\{y \in \mathbb{R} : T_n \cap]-\infty, y] \neq \emptyset \text{ for all } n \in \mathbb{N}\}, \end{aligned} \tag{5.16}$$

Then

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

Proof that $u = \limsup_{n \to \infty} x_j$*: Let*

$$\beta_n := \sup_{j \geq n} x_j, \quad \beta := \inf_n \beta_n = \limsup_{n \to \infty} x_n. \tag{5.17}$$

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

$$b \in \mathscr{U} \text{ for } b < \beta \quad \text{and} \quad a \notin \mathscr{U} \text{ for } a > \beta. \tag{5.18}$$

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* $j \geq n_0$. *By definition,* β_j *is the least upper bound (hence an upper bound) of the tail set* T_j. *We conclude that* $x_j < a$ *for all* $j \geq n_0$. *From that we conclude that* $T_n \cap [a, \infty] = \emptyset$ *for all* $j \geq n_0$. *It follows that* $a \notin \mathscr{U}$.

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

We put everything together and see that β *has the properties listed in (5.18). This finishes the proof that* $u = \limsup_{n \to \infty} x_j$. *The proof that* $l = \liminf_{n \to \infty} x_j$ *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* $n_1 < n_2 < \dots \in \mathbb{N}$ *can be found such that* $x = \lim\limits_{j\to\infty} x_{n_j}$*.*

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

B1. $\liminf\limits_{n\to\infty} x_n$ *is the smallest of all real numbers* x *for which 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.34 that $\limsup\limits_{n\to\infty} x_n$ *is the unique number* u *described in part C of prop.5.3, p.32:* $u - \varepsilon \in \mathscr{U}$ *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* $x_n \geq u - \varepsilon$ *and* $u + \varepsilon \notin \mathscr{U} = \mathscr{U}_3$ *means that there are at most finitely many* n *such that* $x_n \geq u + \varepsilon$*. This proves A1 and A2.*

We also know from prop.5.5 that $\liminf\limits_{n\to\infty} x_n$ *is the unique number* l *described in part C of prop.5.4, p.33:* $l + \varepsilon \in \mathscr{L}$ *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* $x_n \leq l + \varepsilon$ *and* $l - \varepsilon \notin \mathscr{L} = \mathscr{L}_3$ *means that there are at most finitely many* n *such that* $x_n \leq l - \varepsilon$*. This proves B1 and B2.*
■

5.4 Bounded sets and bounded, continuous functions

Definition 5.12 (bounded sets)**.** Given is a subset A of a metric space (X, d). The ***diameter*** of A is defined as

$$diam(A) := \sup\{d(x,y) : x, y \in A\}. \tag{5.19}$$

We call A a ***bounded set*** if $diam(A) < \infty$.

Proposition 5.6 (bounded if and only if finite diameter)**.** *Given is a metric space* (X, d)*.*
A subset A *is bounded if and only if either of the following is true:*

(5.20) $\quad A.\ diam(A) < \infty$

(5.21) $\quad B.$ *There is a* $\gamma > 0$ *and* $x_0 \in X$ *such that* $A \subseteq B_\gamma(x_0)$

Proof of A: Obvious from the definition of the supremum as least upper bound (see (5.1) on p.25 for the informal definition or (5.10) on p.31 for the precise definition).

Proof of B:
$\Longrightarrow$: *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 $diam(A) \leqq 2\gamma$.

$\Longleftarrow$: *Pick an arbitrary $x_0 \in A$ and let $\gamma := diam(A)$. Then*

$$y \in A \quad \Longrightarrow \quad d(x_0,y) \leqq \sup_{x \in A} d(x,y) \leqq \sup_{x,z \in A} d(x,z) = diam(A) = \gamma$$

and it follows that $A \subseteq B_\gamma(x_0)$.
■

Definition 5.13 (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

(5.22) $$f(x) \quad < \quad \gamma_1 \qquad \text{for all arguments } x.$$

It is called ***bounded from below*** if there exists a (possibly very large) number $\gamma_2 > 0$ such that

(5.23) $$f(x) \quad > \quad -\gamma_2 \qquad \text{for all arguments } x.$$

It is called a ***bounded function*** if it is both bounded from above and below. It is obvious that if you set $\gamma := max(\gamma_1, \gamma_2)$ then bounded functions are exactly those that satisfy the inequality

(5.24) $$|f(x)| \quad < \quad \gamma \qquad \text{for all arguments } x.$$

6 Vectors and vector spaces (Understand $\mathbb{R}^N$, skip abstract vector spaces)

6.1 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.42) 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 (3.15 on p.13) 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.8? 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 $\mathbb{R}$)

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

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

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

6.2 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 $\vec{x} = (x_1, x_2, \ldots, x_N)$ and $\vec{y} = (y_1, y_2, \ldots, y_N)$ and a real number α. We define the ***sum*** $\vec{x} + \vec{y}$ of $\vec{x}$

and $\vec{y}$ as the vector $\vec{z}$ with the components

$$(6.3) \qquad z_1 = x_1 + y_1; \quad z_2 = x_2 + y_2; \quad \dots; \quad z_N = x_N + y_N;$$

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

$$(6.4) \qquad w_1 = \alpha x_1; \quad w_2 = \alpha x_2; \quad \dots; \quad w_N = \alpha x_N;$$

The following picture describes vector addition:

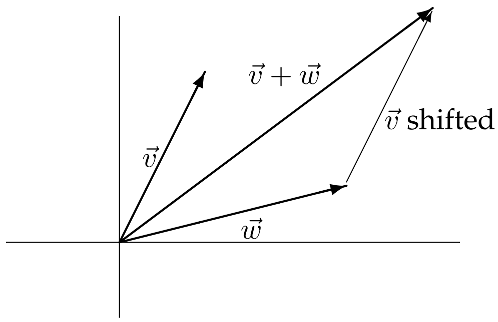


Figure 3: Adding two vectors.

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

6.3 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 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

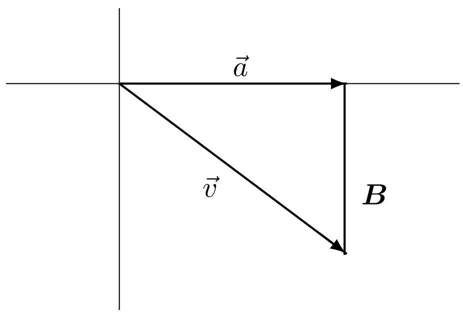


Figure 4: Length of a 2–dimensional vectors.

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* $\vec{v}$*.* [14] *We know from the previous*

[14] 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.47)

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 proven 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:

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

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

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

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

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

has the following three properties:

$$\|\vec{v}\| \geqq 0 \quad \forall \vec{v} \in \mathbb{R}^n \quad \textit{and} \quad \|\vec{v}\| = 0 \iff \vec{v} = 0 \qquad \textbf{\textit{positive definite}} \tag{6.9a}$$

$$\|\alpha\vec{v}\| = |\alpha| \cdot \|\vec{v}\| \quad \forall \vec{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R} \qquad \textbf{\textit{homogeneity}} \tag{6.9b}$$

$$\|\vec{v} + \vec{w}\| \leqq \|\vec{v}\| + \|\vec{w}\| \quad \forall\, \vec{v}, \vec{w} \in \mathbb{R}^n \qquad \textbf{\textit{triangle inequality}} \tag{6.9c}$$

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 $\vec{v} = (x_1, x_2, \ldots, x_n)$ be a vector in $\mathbb{R}^n$ such that $\|\vec{v}\| = 0$. This means that $\sqrt{\sum\limits_{j=1}^{n} {x_j}^2} = 0$ 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 $\vec{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ 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.5).

c. The proof will only be given for $n = 1, 2, 3$.
$\boldsymbol{n = 1}$: *Property (6.9.c) simply reduces to the triangle inequality for real numbers (see 3.1 on 6) and we are done.*
$\boldsymbol{n = 2, 3}$: *Look back at the picture about addition of vectors in the plane or in space (see p.38). 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.50) in the section which discusses inner products on vector spaces. ■

6.4 Vector spaces

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.37). 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.37). 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 (6.2) on p.37). 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 ((3.9) on p.10). 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 6.1 (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.37. 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.23. 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 (5.1) on p. 28) 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.43).

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 $x_1 + y_1, x_2 + y_2, \ldots, x_N + y_N,$. 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

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

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 (3.16) on p.14). 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 $y_1, y_2, \ldots$ 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 $(a_1, a_2, \ldots, a_N)$ 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 $x_1, x_2, x_3, \ldots, x_n \in L$ be a finite number of vectors in L. Let $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n \in \mathbb{R}$. 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 $\alpha_1, \alpha_2, \ldots$ 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 6.2 (Vector spaces are closed w.r.t. linear combinations). *Let L be a vector space and let $x_1, x_2, x_3, \ldots, x_n \in L$ be a finite number of vectors in L. Let $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n \in \mathbb{R}$. Then the linear combination $\sum_{j=0}^{n} \alpha_j x_j$ 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, 6). 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* $\sum_{j=0}^{4} \alpha_j x_j = \sum_{j=0}^{3} \alpha_j x_j + \alpha_4 x_4$ *can be written as the sum of two elements of L (we just saw that* $\sum_{j=0}^{3} \alpha_j x_j$ *as the sum of three elements of L belongs to L) and therefore belongs to L.*

We keep going with $n = 5, 6, 7, \dots$ *and conclude that, no matter how big an* $n \in \mathbb{N}$ *we chose,* $\sum_{j=0}^{n} \alpha_j x_j = \sum_{j=0}^{n-1} \alpha_j x_j + \alpha_n x_n$ *can be written as the sum of two elements of L (we just saw that* $\sum_{j=0}^{n-1} \alpha_j x_j$ *as the sum of* $n-1$ *elements of L belongs to L) and therefore belongs to L ....* ■

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:

(6.11a) $\quad f(x+y) \;=\; f(x) + f(y) \quad \forall x, y \in L_1$ $\quad$ ***additivity***

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

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

Note 6.1 (Note on homogeneity)**.** We encountered homogeneity when looking at the properties of norms (see (5.4) on p.28 or (6.9) on p.40) but it is defined differently there in that you had to take the absolute value $\|\alpha f(\cdot)\| = |\alpha| \cdot \|f(\cdot)\|$.

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 6.3 (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:*

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

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.6). 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 $i_1, i_2, \dots, i_M$. [15]

6.5 Normed vector spaces

The following definition of inner products and proof of the Cauchy–Schwartz inequality were taken from "Calculus of Vector Functions" (Williamson/Crowell/Trotter [10]).

[15] 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 (6.3) on p.38).

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:

(6.14a) $x \bullet x \geqq 0 \quad \forall x \in L$ and $x \bullet x = 0 \iff x = 0$ ***positive definite***

(6.14b) $x \bullet y = y \bullet x \quad \forall x, y \in L$ ***symmetry***

(6.14c) $(x+y) \bullet z = x \bullet z + y \bullet z \quad \forall\, x, y, z \in L$ ***additivity***

(6.14d) $(\lambda x) \bullet y = \lambda (x \bullet y) \quad \forall\, x, y \in L \quad \forall\, \lambda \in \mathbb{R}$ ***homogeneity***

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

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

are both linear.

Proposition 6.4 (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*

(6.15a) $$(a+b) \bullet (x+y) = a \bullet x \,+\, b \bullet x \,+\, a \bullet y \,+\, b \bullet y$$

(6.15b) $$(x+y) \bullet (x+y) = x \bullet x \,+\, 2(x \bullet y) \,+\, y \bullet y$$

(6.15c) $$(x-y) \bullet (x-y) = x \bullet x \,-\, 2(x \bullet y) \,+\, y \bullet y$$

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

$$(\vec{v}, \vec{w}) \;\mapsto\; x_1 y_1 + x_2 y_2 + \ldots + x_N y_N \;=\; \sum_{j=1}^{n} x_j y_j \tag{6.16}$$

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

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

Proposition 6.6 (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 (6.4) on p.48.
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 proven. ■

We previously discussed the sup*–norm*

$$\|f(\cdot)\|_\infty = \sup\{|f(x)| : x \in X\} \tag{6.17}$$

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

$$\|\vec{x}\|_2 = \sum_{j=1}^{n} {x_j}^2 \tag{6.18}$$

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 (5.1) on p.28 and (6.1) on p.40). 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. 51) 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:

(6.19a)	$\|x\| \geqq 0 \quad \forall x \in L$ and $\|x\| = 0 \iff x = 0$	***positive definite***
(6.19b)	$\|\alpha x\| = \|\alpha\| \cdot \|x\| \quad \forall x \in L, \forall \alpha \in \mathbb{R}$	***homogeneity***
(6.19c)	$\|x + y\| \leqq \|x\| + \|y\| \quad \forall\, x, y \in L$	***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* $\|x+y\| \leq \|x\| + \|y\|$ *and that's the triangle inequality we set out to prove.* ■

7 Convergence and Continuity (Study this!)

There is a branch of Mathematics, called topology, which deals with the concept of closeness and continous functions (those which map arbitrarily close elements of the domain into 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. A metric is a real function of two arguments which associates with any two points* $x, y \in X$ *their "distance"* $d(x, y)$.

If you worked through all material in section on measuring the distance of real functions, then you have read about the metric properties of positive definiteness, symmetry and triangle inequality (lemma (5.2) on p.28). We now will take those properties as the starting point for this chapter.

7.1 Metric spaces

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

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

$$(x,y) \mapsto d(x,y) \qquad \text{where } x, y \in X$$

with the following three properties: [16]

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

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

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

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

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

Remark 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 any vector space with a norm automatically becomes a metric space.

Theorem 7.1 (Norms define metric spaces)**.** *Let L be a vector space with a norm*

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

Then

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

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

Proof:

Positive definiteness*: For any $x,y \in X$ $d_{\|\cdot\|}(x,y) = \|y-x\| \geqq 0$ is trivial. Further,*

$$d_{\|\cdot\|}(x,y) = 0 \iff \|y-x\| = 0 \iff x = y$$

where the last equivalence is true because norms are positive definite.

Symmetry*: True because for any $x,y \in X$*

$$\begin{aligned} d_{\|\cdot\|}(x,y) &= \|x-y\| \;=\; \|(-1)(y-x)\| \;=\; |-1| \cdot \|(y-x)\| \\ &= \|(y-x)\| \;=\; d_{\|\cdot\|}(y,x) \end{aligned}$$

[16] 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.

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

$$\begin{aligned} d_{\|\cdot\|}(x,z) &= \|x-z\| \;=\; \|(x-y)+(y-z)\| \\ &\leqq \|(x-y)\| + \|(y-z)\| \;=\; d_{\|\cdot\|}(x,y) + d_{\|\cdot\|}(y,z) \end{aligned}$$

■

Here are some examples of metric spaces.

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 (5.1) on p.28). If you prefer, you can also conclude this from from (5.2) on p.28 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.40).

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

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

on $X \times X$ defines a metric.

Proof: Obviously the function is non–negative and 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. ■

7.2 Neighborhoods and open sets

A. *Given a point* $x_0 \in \mathbb{R}$ *(a real number), we can look at*

$$B_\varepsilon(x_0) = (x_0 - \varepsilon, x_0 + \varepsilon) \; = \{x \in \mathbb{R} : x_0 - \varepsilon < x < x_0 + \varepsilon\} \\ = \{x \in \mathbb{R} : d(x, x_0) = |x - x_0| < \varepsilon\} \tag{7.2}$$

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.53).*

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*

$$B_\varepsilon(\vec{x}_0) \; = \{\vec{x} \in \mathbb{R}^2 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} \\ = \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\} \tag{7.3}$$

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*

$$B_\varepsilon(\vec{x}_0) \; = \{\vec{x} \in \mathbb{R}^3 : \|\vec{x} - \vec{x}_0\| < \varepsilon\} \\ = \{(x, y, z) \in \mathbb{R}^3 : (\vec{x} - \vec{x}_0)^2 + (\vec{y} - \vec{y}_0)^2 + (\vec{z} - \vec{z}_0)^2 < \varepsilon^2\} \tag{7.4}$$

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*

$$B_\varepsilon(x_0) \; = \; \{x \in L : \|x - x_0\| < \varepsilon\} \tag{7.5}$$

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 7.2 (Neighborhood)**.** Given a metric space (X, d) and an element $x_0 \in X$, we can look at

$$B_\varepsilon(x_0) \; = \; \{x \in L : d(x, x_0) < \varepsilon\} \tag{7.6}$$

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*

$$\varepsilon - d(a, x_0) = 2\delta \tag{7.7}$$

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 (7.7).

So we have proven 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 7.3 (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 7.4 (open set)**.** Given is a metric space (X, d).
A set all of whose members are interior points is called an ***open set*** .

Proposition 7.1. *$B_\varepsilon(x_0)$ is an open set*

Proof: we showed earlier on that any $a \in B_\varepsilon(x_0)$ is an interior point of $B_\varepsilon(x_0)$. ■

Theorem 7.2 (Metric spaces are topological spaces)**.** *The following is true about open sets of a metric space (X, d):*

(7.8a) *An arbitrary union $\bigcup\limits_{i \in I} U_i$ of open sets U_i is open.*

(7.8b) *A finite intersection $U_1 \cap U_2 \cap \ldots \cap U_n$ $(n \in \mathbb{N})$ of open sets is open.*

(7.8c) *The entire set X is open and the empty set $\emptyset$ is open.*

Proof of a: Let $U := \bigcup\limits_{i \in I} U_i$ 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 $U = \bigcup\limits_{i \in I} U_i$ 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]

$$B_\varepsilon(x) \subseteq B_{\varepsilon_j}(x) \subseteq U_j \ (1 \leqq j \leqq n) \qquad \Longrightarrow \qquad B_\varepsilon(x) \subseteq \bigcap_{j=1}^{n} U_j.$$

[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.

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 $x \in X$. Choose $B_{2,123,460,708}(x)$ (you can make it bigger if you like). Then x is an inner point of X because $B_{2,123,460,708}(x) \subseteq X$. 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. If you have a problem with that, then find me an element of $\emptyset$ that is not an interior point (which is what you must do to prove that $\emptyset$ is not open) and I'll give you 100 Bucks.
■

7.3 Digression: Abstract topological spaces (Skip this!)

Theorem 7.2 on p.55 gives us a way of defining neighborhoods for sets which do not have a metric.

Definition 7.5 (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 (7.8) on p.55:

(7.9a) An arbitrary union $\bigcup_{i \in I} U_i$ of sets $U_i \in \mathfrak{U}$ belongs to $\mathfrak{U}$,

(7.9b) $U_1, U_2, \dots, U_n \in \mathfrak{U}\ (n \in \mathbb{N}) \quad \Rightarrow \quad U_1 \cap U_2 \cap \dots \cap U_n \in \mathfrak{U}$,

(7.9c) $X \in \mathfrak{U} \quad \text{and} \quad \emptyset \in \mathfrak{U}$.

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 7.6 (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 (7.3) on p.55). 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 is 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.52). Obviously, this norm defines open sets

$$\mathfrak{U}_{\|\cdot\|} \;:=\; \mathfrak{U}_{d_{\|\cdot\|}}$$

on X by means of this metric. We say that this topology is ***induced by the norm*** $\|\cdot\|$ or that it is ***generated by the norm*** $\|\cdot\|$. If there is no confusion about which norm we are talking about, we also simply speak about the ***norm topology*** .

Example 7.5 (Discrete topology)**.** Let X be non–empty. We had defined in (7.4) on p.53 the discrete metric as

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

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: $A = \bigcup_{a \in A} \{a\}$ 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 7.5 on p.56. 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.

8 Lecture Notes for Math 330 - Spring 2015

8.1 Math 330 - Notes for Lecture 37 on Mon, 2015-03-30

8.1.1 Covered material for lecture 37

Mapped out the next few weeks:

	Topic
1.	*Finish ch.12 by Tue, March 31*
2.	*Finish ch.13 by Fri, April 3 UNLESS I talk about alternate ways to construct the numbers, starting with Peano's axioms.*
3.	*After the break: revisit the definitions of liminf, limsup, indicator functions*
4.	*Topic A: continuity and uniform continuity*
5.	*Topic B: Public-Key Cryptography: very little, if any*
6.	*Topic C: $\mathbb{C}$: intend to skip*
7.	*Topic D: Groups and Graphs: must appease the computer scientists*
8.	*Topic E: Generating Functions: Skip*
9.	*Topic F: Cantor-Schröder-Bernstein Theorem (ignore ordinal numbers).*

Definition 8.1 (limits $\pm\infty$)**.** We say of a sequence $(x_n)_{j \in \mathbb{N}}$ of real numbers that x_n ***has limit*** ∞ and we write $\lim\limits_{n \to \infty} x_n = \infty$ if, for each $K^\star \in \mathbb{R}$, there is $N = N(K^\star) \in \mathbb{N}$ such that $x_n > K^\star$ for all $n \geq N(K^\star)$. In other words, the entire tail set $T_{N(K^\star)}$ must be contained within $]K^\star, \infty[$ no matter how large a $K^\star$ was chosen. The entire tail will have risen above the given upper threshold $K^\star$.

We say that x_n ***has limit*** $-\infty$. and we write $\lim\limits_{n \to \infty} x_n = -\infty$ if, for each $K_\star \in \mathbb{R}$, there is $N = N(K_\star) \in \mathbb{N}$ such that $x_n < K_\star$ for all $n \geq N(K_\star)$. In other words, the entire tail set $T_{N(K_\star)}$ must be contained within $]-\infty, K_\star[$ no matter how small a $K_\star$ was chosen. The entire tail will have dropped below the given lower threshold $K_\star$.

Definition 8.2 (limits $\pm\infty$)**.** Let $(a_j)_{j\in\mathbb{N}}$ be a sequence of real numbers. Let

$$s_n \ := \sum_{j=1}^{n} a_j, \quad s \ := \sum_{j=1}^{\infty} a_j \ := \ \lim_{n\to\infty} s_n \ = \ \lim_{n\to\infty} \sum_{j=1}^{n} a_j.$$

We call $\sum_{j=1}^{\infty} a_j$ a ***series*** or an ***infinite series*** $(a_j)_{j\in\mathbb{N}}$ is called the ***sequence of terms*** . of the series and we call s_n the n***-th partial sum*** of the series. $(s_n)_{n\in\mathbb{N}}$ is called the ***sequence of partial sums*** of the series.

If $s \ := \ \lim_{n\to\infty} s_n$ exists then we call s the ***sum*** of the series. We say the ***series converges*** to s if $s \in \mathbb{R}$, i.e., $s \neq \pm\infty$, otherwise the series diverges.

Example 8.1 (Cover page of the B/G book)**.** Geometric meaning of

$$\sum_{j=0}^{n} (1/4)^j \ = \frac{4}{3} \ = \ 1 \ + \ \sum_{j=1}^{n} (1/4)^j, \quad \text{i.e.,} \sum_{j=1}^{n} (1/4)^j; \ = \ \frac{1}{3}$$

a. each subsequent iteration has half the height (similar triangles), hence the triangles of each iteration have $1/4$ the area of the previous one. That means that if $(1/4)^j$ is the area of each of the triangles in iteration j then $(1/4) \cdot (1/4)^j$ is the area of each of the triangles in iteration $j+1$.
b. in each horizontal slice the shaded triangle has $1/3$ the area of the entire slice. That means that the limit of the sum of the shaded triangles is $1/3$ of the total area

We talked in the lecture about increasing and decreasing sequences. See def.2.1, p.4.

Proposition 8.1 (Inequalities persist in the limit)**.** *(not in B/G):*
Let x_n, $\hat{x}_n$ be two convergent sequences such that $x_n \leq \hat{x}_n$ for all $n \geq N_1$. Let $\alpha = \lim x_n$, $\beta = \lim \hat{x}_n$. Then $\alpha \leq \beta$.

Proof: Otherwise there is $\varepsilon > 0$ such that $\alpha - \beta = 3\varepsilon$. There is $N_2 = N_2(\varepsilon)$ and $N_3 = N_3(\varepsilon)$ such that $|x_n - \alpha| < \varepsilon$ for all $n \geq N_1$ and $|\hat{x}_n - \beta| < \varepsilon$ for all $n \geq N_2$. Let $n \geq N := \max(N_1, N_2, N_3)$. It follows from $\alpha = \beta + 3\varepsilon$ that $x_n > \hat{x}_n$. That inequality, together with $N \geq N_1$, contradicts the assumption that $x_n \leq \hat{x}_n$ for all $n \geq N_1$. ■

8.2 Math 330 - Notes for Lecture 38 on Tue, 2015-03-31

8.2.1 Covered material for lecture 38

Proposition 8.2 (B/G Prop.12.5, p.116)**.**

$$\begin{aligned} \textbf{a.} 0 \ \le \quad 9 \sum_{j=n}^{\infty} 10^{-j} \ &= \ \frac{1}{10^{n-1}}, \\ \textbf{b.} \quad \sum_{j=n}^{\infty} d_j 10^{-j} \ &\le \ \frac{1}{10^{n-1}}, \\ \textbf{c.} \quad \sum_{j=n}^{\infty} d_j 10^{-j} \ &= \ \frac{1}{10^{n-1}} \quad \text{iff } \ d_j = 9 \ \forall j \ge n. \end{aligned}$$

Proof of a and b:

$$\sum_{j=n}^{\infty} d_j 10^{-j} \ \le 9 \cdot \sum_{j=n}^{\infty} 10^{-j} \ = \ 9 \cdot 10^{-n} \sum_{k=0}^{\infty} 10^{-k} \ = \ \frac{9}{10^n} \frac{1}{9/10} \ = \ 10^{n-1}$$

■

Proof of c: For $k \ge n$ *let* $s_k \ = \ \sum_{j=n}^{k} d_j 10^{-j}, \quad s \ = \ \sum_{j=n}^{\infty} d_j 10^{-j}, \quad s'_k \ = \ 9 \sum_{j=n}^{k} 10^{-j}, \quad s' \ = \ 9 \sum_{j=n}^{\infty} 10^{-j}.$

If there is $j_0 \ge n$ *such that* $d_{j_0} < 9$ *then* $s'_k - s_k \ge 10^{-j_0}$ *for all* $k \ge n$*. Hence*

$$s' - s \ = \ \lim k \to \infty (s'_k - s_k) \ \ge \ 10^{-j_0}$$

because the difference of the limits is the limit of the differences and according to 8.1 on p.58. This proves the "only if" direction. The "if" direction is immediate from part a. ■

Lemma 8.1 (Uniqueness of Decimal Expansions)**.** *Given is a decimal expansion* $x \ = \ m + \sum_{j=1}^{\infty} d_j 10^{-j}$*. Then*

a. $m \le x \le m+1$.
b. $x = m$ *iff* $d_j = 0$ *for all* $j \in \mathbb{N}$.
c. $x = m+1$ *iff* $d_j = 9$ *for all* $j \in \mathbb{N}$.

Proof:

Let $s \ := \ \sum_{j=1}^{\infty} d_j 10^{-j}.$

Part a is immediate from prop.8.2.a.

Part b is clear because if $d_n > 0$ *for some* n *then* $s \ge s_n > d_n 10^{-n} > 0$.

Part c follows from prop.8.2.c. ■

Definition 8.3 (Repeating Decimals)**.** A non-negative decimal

$$x = (m, d_1, d_2, \dots) = m + \sum_{j=1}^{\infty} d_j 10^{-j} \quad (d_j \in \{0, 1, 2 \dots, 9\})$$

is ***repeating*** if there are natural numbers N and p such that

$$d_{N+n+kp} = d_{N+n} \ \forall\, 0 \leq n < p,\ k \in \mathbb{N}.$$

The above INCLUDES $p = 1$ and $d_N = 0$ (finite expansion!).

Proposition 8.3 (B/G Prop.12.11, p.119)**.** *Every repeating decimal represents a rational number.*

Proof: Let $x = (m, d_1, d_2, \dots) = m + \sum_{j=1}^{\infty}$ be a repeating decimal, i.e., there exist $N, p \in \mathbb{N}$ such that for all $0 \leq n < p$ and for all $k \in \mathbb{N}$,

$$d_{N+n+kp} = d_{N+n}.$$

Then

$$\begin{aligned}
x &= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{j=N}^{\infty} d_j 10^{-j} \\
&= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{k=0}^{\infty}\sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n+kp}} \\
&= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} d_{N+n} \sum_{k=0}^{\infty} \frac{1}{10^{N+n+kp}} \\
&= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n}} \sum_{k=0}^{\infty} \frac{1}{10^{kp}} \\
&= m + \sum_{j=1}^{N-1} d_j 10^{-j} + \sum_{n=0}^{p-1} \frac{d_{N+n}}{10^{N+n}} \frac{1}{1 - 10^{-p}}
\end{aligned}$$

is a finite sum of rational numbers and therefore rational. ■

Note 8.1 (Decimal expansions of real numbers)**.** Let $x \in \mathbb{R}$.

a. x has at most two different decimal expansions.

b. If x has two expansions then one is all zeroes except for finitely many digits and the other is all nines except for finitely many digits.

c. If x has more than one expansion then x is rational.

d. x is a repeating decimal if and only if $x \in \mathbb{Q}$.

8.3 Math 330 - Notes for Lecture 39 on Wed, 2015-04-01

8.3.1 Covered material for lecture 39

Proposition 8.4 (B/G Prop.13.3, p.122). *Prove Prop.13.3: Let $k, n \in \mathbb{N}$ such that $1 \leq k < n$. Then the function $g_k : [n-1] \longrightarrow [n] \setminus \{k\}$ defined by*

$$g_k(j) := \begin{cases} j & \text{if } j < k \\ j+1 & \text{if } j \geq k \end{cases} \tag{8.1}$$

is bijective.

$$g_k : [n-1] \longrightarrow [n] \setminus \{k\} \quad \text{defined by} \quad g_k(j) := \begin{cases} j & \text{if } j < k \\ j+1 & \text{if } j \geq k \end{cases} \tag{8.2}$$

is bijective.

Proof: Homework 17. ■

8.4 Math 330 - Notes for Lecture 40 on Fri, 2015-04-03

8.4.1 Covered material for lecture 40

Construction of the numbers systems – alternate approach: From Peano's axioms to Dedekind cuts. References for the construction of $\mathbb{Z}_{\geq 0}$: [2] Birkhoff, Mac Lane: Algebra. References for the construction of $\mathbb{R}$ with Dedekind cuts: [5] Rudin, Walter: Principles of Mathematical Analysis.

References

[1] *Matthias Beck and Ross Geoghegan.* The Art of Proof. *Springer, 1st edition, 2010.*

[2] *Garret Birkhoff and Saunders Mac Lane.* Algebra. *Macmillan, New York, 1st edition, 1968.*

[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] *Walter Rudin.* Principles of Mathematical Analysis. *McGraw-Hill, New York, San Francisco, Toronto, London, 2nd edition, 1964.*

[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] *Richard E. Williamson, Richard H. Crowell, and Hale F. Trotter.* Calculus of Vector Functions. *Prentice Hall, Englewood Cliffs, 3rd edition, 1972.*

Index