Mathematics for Economists

Part I and II

Dr. Christian-Oliver Ewald

University of Kaiserslautern

Winter Term 2003/04

Working Version
March 27, 2007

Contents

Chapter 1

Mathematical Notation

Learning Mathematics from the very start consists of two equal parts. One part is learning the language of Mathematics, the other part is learning the concepts, techniques and methods. Often one can hear people complaining they don't understand Mathematics, but most time they just don't speak the language. Someone who wants to study French literature has to learn the French language and someone who wants to learn the concepts, techniques and methods of Mathematics has to learn the language of Mathematics. As learning the French language and studying the masterpieces of French literature goes hand in hand, so does learning the Mathematics language and "Mathematics". However, to start with one needs a basic vocabulary. The purpose of this first chapter is to give the student this basic vocabulary in Mathematics.

1.1 Sets

A **set** consists of an arbitrary (finite or infinite) number of objects. If an object x belongs to a set M we write

$$x \in M \tag{1.1}$$

and say x is an element of M. If x is not an element of M we write

$$x \notin M.$$

We can denote a set by listing its elements. For example

$$M = \{1, 2, 3, 4, 5, 6\} \tag{1.2}$$

If M consists of an infinite number of elements we can not write down all the elements. Sometimes the elements are in a natural order and one can indicate the set by just writing down the first three or four elements of the set such as in

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

where $\mathbb{N}$ denotes the set of **Natural Numbers**. In the same way one can write

$$\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$$

to denote the **Integer Numbers**. If however the elements of the set are in no natural order, we can usually characterize the elements of the set by some conditions such as for example in

$$\mathbb{Q} = \{\frac{p}{q}|p, q \in \mathbb{Z}, q \neq 0\} \tag{1.3}$$

Here $\mathbb{Q}$ denotes the set of **Rational Numbers** and $\mathbb{Z}$ the set of integer numbers. This is always possible. For example one could write

$$M = \{x|x \in M\}. \tag{1.4}$$

We assume furthermore that the students is familiar with the set of **Real Numbers**

$$\mathbb{R} = \{x|x \text{ is a real number }\}. \tag{1.5}$$

A very important set is the so called **empty set**. This is the set which contains no elements. We denote it with

$$\emptyset \tag{1.6}$$

If we have two sets M and N and all elements of N are also elements of M we write

$$N \subset M \tag{1.7}$$

and call N a **subset** of M. We can write

$$N \subset M \tag{1.8}$$

if we want to express that $N \subset M$ but they are not equal $N \neq M$. Furthermore given two sets M and N we can build their **union**

$$M \cup N = \{x|x \in M \ \underline{\text{or}}\ x \in N\} \tag{1.9}$$

as well as their **intersection**

$$M \cap N = \{x|x \in M \ \underline{\text{and}}\ x \in N\} \tag{1.10}$$

the **complement**

$$M \setminus N = \{x|x \in M \text{ and } \underline{\text{not}} \text{ in N }\} \tag{1.11}$$

and their **Cartesian product**

$$M \times N = \{(x, y)|x \in M, y \in N\} \tag{1.12}$$

where (x, y) just denotes the (ordered) pair which consists of x and y. Important subsets of $\mathbb{R}$ are the so called **intervals** : For $a, b \in \mathbb{R}$

$$]a, b[= \{x|x \in \mathbb{R}, a < x < b\} \tag{1.13}$$

is called the **open interval** from a to b.

$$[a, b] = \{x|x \in \mathbb{R}, a \leq x \leq b\} \tag{1.14}$$

is called the **closed interval**. Furthermore we speak of half open intervals denoted by $[a, b[$ and $]a, b]$ in case one of the inequalities is strict and the other one is not.

At the end of the section some examples :

Example 1.1.1.

$$5 \in \mathbb{N}, -2 \notin \mathbb{N}, -2 \in \mathbb{Z}, \frac{1}{2} \notin \mathbb{Z}\frac{1}{2} \in \mathbb{Q}, \sqrt{2} \notin \mathbb{Q}, \sqrt{2} \notin \mathbb{R}$$

$$\emptyset \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

$$]-3, 2[\cap \mathbb{Z} = \{-2, -1, 0, 1\}, [-3, 2] \cap \mathbb{N} = 1, 2$$

1.2 Maps

The concept of maps can be found everywhere in nature. Given two sets, than a map is something which associates to each element of the first set exactly one element of the second set. For example one can consider the price of a stock at the stock-market as a function of time. At any point in time the stock has exactly one price. It will take us some time and a lot of effort to understand how this particular function can look like, but we will finally come back to this example. For now, we have to translate the concept of maps into mathematics language.

Definition 1.2.1. *A map consist of the following data :*

1. *a set M, the so called* **domain** *of the map*
2. *a set N, the so called* **range** *of the map*
3. *for each $x \in D$ exactly one element $f(x) \in W$.*

In compact notation we write

$$\begin{aligned} f : D &\rightarrow W \\ x &\mapsto f(x). \end{aligned}$$

Sometimes if it is clear from the context we abbreviate the notation and just write $f : M \rightarrow N$ or even just f. Always though we distinguish the map f from $f(x)$ which is just an element in the range.

Example 1.2.1. *We consider the following examples :*

1. $$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto f(x) = x^2 \end{aligned}$$

2. $$\begin{aligned} g : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto g(x) = 3x + 1 \end{aligned}$$

3. $$\begin{aligned} h : \mathbb{Z} &\rightarrow \mathbb{N} \\ x &\mapsto h(x) = |x| \end{aligned}$$

4. $$\begin{aligned} id_M : M &\rightarrow M \\ x &\mapsto x \end{aligned}$$

The map id_M is called the **identity map** *on M.*

Definition 1.2.2. *Let $f : M \to N$ be a map.*

1. *For any subset $U \subset M$ the set $f(U) = \{f(m)|m \in U\}$ is called the* **image** *of U under f. The set $f(M)$ is just called the image of f.*

2. *For a subset $V \subset Y$ the set $f^{-1}(V) = \{x \in M|f(x) \in V\}$ is called the* **preimage** *of V.*

The notation f^{-1} in the definition does not stand for the inverse map, which in general does not exist. The preimage of N is always the whole set M, since any element of M is mapped into N under f.

Example 1.2.2. *Reconsider the maps of Example 1.2.1. The image of f is the interval $[0, \infty[$. Furthermore we have $f^{-1}([4, 9]) = [-3, -2] \cup [2, 3]$ and $f^{-1}(\{-1\}) = \emptyset$. The image of g is $\mathbb{R}$ an the image of h is $\mathbb{N}$. The image of the set $\{-1, 1\}$ under h is $\{1\}$*

Definition 1.2.3. *Let $f : M \to N$ be a map. Then*

1. *f is called* **surjective** *if the image of f is N, i.e. $f(M) = N$.*

2. *f is called* **injective** *if for any one element subset $\{y\} \subset N$ the preimage $f^{-1}(\{y\}$ consist of at most one element.*

3. *f is called* **bijective** *iff is both injective and surjective.*

There are a lot of equivalent definitions for surjective, injective and bijective. For example a map is surjective if and only if the preimage of any one element subset of N is not empty and it is injective if and only if different elements of M are mapped under f to different elements in N.

Example 1.2.3. *Reconsider Example 1.2.1. Then f is neither injective nor surjective. h is surjective but not injective. g is bijective.*

Definition 1.2.4. *Let $f : M \to N$ be a bijective map. Then for any $y \in N$ there is exactly one element $x \in M$ such that $f(x) = y$. We denote this element with $f^{-1}(y)$ and define the* **inverse map** *f^{-1} via*

$$\begin{aligned} f^{-1} : N &\to M \\ y &\mapsto f^{-1}(y). \end{aligned}$$

f^{-1} is also bijective and we have

$$f^{-1}(f(x)) = x \; \forall x \in M.$$

Definition 1.2.5. *Let $f : M \to N$ and $g : L \to M$ be maps. Then we define a new map $f \circ g$ called the* **composition** *of f and f as*

$$\begin{aligned} f \circ g : L &\to N \\ x &\mapsto f(g(x)). \end{aligned}$$

Example 1.2.4. *Reconsider the maps from Example 1.2.1, then we have $(f \circ g)(x) = (3x + 1)^2 = 9x^2 + 6x + 1$ for all $x \in \mathbb{R}$.*

Chapter 2

Linear Algebra

Linear Algebra is basically a tool for the study of systems of linear equations. Consider for example the following system :

$$\begin{aligned} 1x_1 + 2x_2 + 3x_3 &= 4 \\ 5x_1 + 6x_2 + 7x_3 &= 8 \\ 9x_1 + 10x_2 + 11x_3 &= 12 \end{aligned}$$

The question is : Is there a triple (x_1, x_2, x_3) of numbers which satisfies the three equations above and in case there is, how many are there and how can we find them.

2.1 Vectors, Matrices and Vectorspaces

Definition 2.1.1. *Let $x_1, ..., x_n$ be real numbers. Then*

$$\begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

is called a **column vector** *and*

$$(x_1, ..., x_n)$$

is called a **row vector**.

If the number of entries in the vector is n then we speak of an n-dimensional vector. Let us put all n dimensional column vectors together in a set :

Definition 2.1.2. *The* n-**dimensional real space** *is the set which contains all* n *dimensional column vectors*

$$\mathbb{R}^n := \{ \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} | x_1, .., x_n \in \mathbb{R} \}.$$

Sometimes we write

$$(x_1, ..., x_n)^\top = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}.$$

If we consider the set of m-tuples of n-dimensional row vector we get what is called $m \times n$ matrices.

Definition 2.1.3. *The set of* $m \times n$ *matrices is the set*

$$\mathbb{R}^{m \times n} := \{ \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdot & \cdot & a_{mn} \end{pmatrix} | a_{ij} \in \mathbb{R} \}$$

Example 2.1.1.

$$\begin{pmatrix} 1 \\ 3 \\ -3 \end{pmatrix} \in \mathbb{R}^3, \ (1, 4, \sqrt{2}, 7) \in \mathbb{R}^{1\times 4}, \ \begin{pmatrix} 2 & 2\pi \\ -4 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

We can consider n-dimensional row vectors as $1 \times n$ matrices. If we don't want to list all the entries of a matrix we just write (a_{ij}) indicating that at position i-th row, j-th column in the matrix we have the entry a_{ij}.

Definition 2.1.4. *Let* $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ *and* $B = (b_{ij}) \in \mathbb{R}^{k\times l}$ *be matrices. Then we write* $A = B$ *if*

1. $m = k$ *and* $n = l$
2. $a_{ij} = b_{ij}$ *for all pairs* i, j *such that* $1 \leq i \leq m$ *and* $1 \leq j \leq n$.

Example 2.1.2.

$$\begin{pmatrix} 2 & 2\pi \\ -4 & 1 \end{pmatrix} \neq \begin{pmatrix} 2 & 2\pi \\ -4 & 1 \\ 0 & 0 \end{pmatrix}$$

Definition 2.1.5. *Let $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n}$ and $C = (c_{ij}) \in \mathbb{R}^{n \times l}$ be matrices, $x = (x_i), y = (y_i) \in \mathbb{R}^n$ be a vector and $r \in \mathbb{R}$. Then we define*

1. **Scalar Multiplication**

$$r \cdot x \quad = \quad r \cdot \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} r \cdot x_1 \\ \cdot \\ \cdot \\ \cdot \\ r \cdot x_n \end{pmatrix}$$

2. **Vector Addition**

$$x + y \quad = \quad \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n + y_n \end{pmatrix}$$

3. **Matrix Addition**

$$\begin{aligned} A + B \quad &:= \quad \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdot & \cdot & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdot & \cdot & b_{1n} \\ b_{21} & b_{22} & \cdot & \cdot & b_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{m1} & b_{m2} & \cdot & \cdot & b_{mn} \end{pmatrix} \\ &= \quad \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdot & \cdot & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdot & \cdot & a_{2n} + b_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdot & \cdot & a_{mn} + b_{mn} \end{pmatrix} \end{aligned}$$

4. **Scalar Product**

$$< x, y > \quad := \quad x_1 \cdot y_1 + ... + x_n \cdot y_n = \sum_{i=1}^{n} x_i \cdot y_i$$

5. **Matrix/Vector Multiplication**

$$A \cdot x \;=\; \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdot & \cdot & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} \cdot x_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_{j=1}^{n} a_{mj} \cdot x_j \end{pmatrix} \in \mathbb{R}^m$$

6. **Matrix Multiplication**

$$A \cdot C \;=\; \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdot & \cdot & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} c_{11} & c_{12} & \cdot & \cdot & c_{1l} \\ c_{21} & c_{22} & \cdot & \cdot & a_{2l} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdot & \cdot & c_{nl} \end{pmatrix}$$

$$:= \begin{pmatrix} \sum_{k=1}^{n} a_{1k} \cdot c_{k1} & \sum_{k=1}^{n} a_{1k} \cdot c_{k2} & \cdot & \cdot & \sum_{k=1}^{n} a_{1k} \cdot c_{kl} \\ \sum_{k=1}^{n} a_{2k} \cdot c_{k1} & \sum_{k=1}^{n} a_{2k} \cdot c_{k2} & \cdot & \cdot & \sum_{k=1}^{n} a_{2k} \cdot c_{kl} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{k=1}^{n} a_{mk} \cdot c_{k1} & \sum_{k=1}^{n} a_{mk} \cdot c_{k2} & \cdot & \cdot & \sum_{k=1}^{n} a_{mk} \cdot c_{kl} \end{pmatrix}$$

At position (i,j) of the matrix we have the element $\sum_{k=1}^{n} a_{ik} \cdot c_{kj}$.

7. **Scalar/Matrix Multiplication**

$$r \cdot A = r \cdot \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdot & \cdot & a_{mn} \end{pmatrix} \;:=\; \begin{pmatrix} r \cdot a_{11} & r \cdot a_{12} & \cdot & \cdot & r \cdot a_{1n} \\ r \cdot a_{21} & r \cdot a_{22} & \cdot & \cdot & r \cdot a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r \cdot a_{m1} & r \cdot a_{m2} & \cdot & \cdot & r \cdot a_{mn} \end{pmatrix}$$

Note that we can only multiply matrices with matrices and matrices with vectors as well as add matrices and add vectors if the sizes fit each other. The multiplications and additions defined as above satisfy the following rules:

Proposition 2.1.1. *Let* $A_1, A_2, A_3 \in \mathbb{R}^{m\times n}$,$B_1.B_2 \in \mathbb{R}^{n\times l}$,$C_1 \in \mathbb{R}^{l\times k}$,$r, s \in \mathbb{R}$ *and* $x, y, z \in \mathbb{R}^n$. *Then we have*

1. **Commutative Laws**

   (a) $x + y = y + x$

(b) $A_1 + A_2 = A_2 + A_1$

(c) $r \cdot s = s \cdot r$

2. **Associative Laws**

   (a) $x + (y + z) = (x + y) + z$

   (b) $A_1 + (A_2 + A_3) = (A_1 + A_2) + A_3$

   (c) $A_1 \cdot (B_1 \cdot C_1) = (A_1 \cdot B_1) \cdot C_1$

   (d) $r \cdot (s \cdot x) = (r \cdot s) \cdot x$

   (e) $r \cdot (s \cdot A_1) = (r \cdot s) \cdot A_1$

3. **Distributive Laws**

   (a) $A_1 \cdot (x + y) = A_1 \cdot x + A_1 \cdot y$

   (b) $r \cdot (A_1 + A_2) = r \cdot A_1 + r \cdot A_2$

   (c) $r \cdot (x + y) = r \cdot x + r \cdot y$

   (d) $(r + s) \cdot A_1 = r \cdot A_1 + s \cdot A_1$

   (e) $(r + s) \cdot x = r \cdot x + s \cdot x$

   (f) $(A_1 + A_2) \cdot x = A_1 \cdot x + A_2 \cdot x$

Proof. Exercise ! □

Let us consider some special vectors and matrices :

Definition 2.1.6. *1. The vector* $0_n = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \in \mathbb{R}^n$ *is called the (n-dimensional)* **Zero Vector**.

2. The matrix $0_{m \times n} := \begin{pmatrix} 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}$ *is called the* $m \times n$ **Zero Matrix**.

3. *The matrix* $id_{n\times n} := \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}$ *with "1" on the diagonal and "0" elsewhere is called the* $n \times n$ **Identity Matrix**.

Proposition 2.1.2. *Let* $A \in \mathbb{R}^{m\times n}$,$B \in \mathbb{R}^{n\times m}$ *and* $x \in \mathbb{R}^n$. *Then*

1. $A \cdot id_{n\times n} = A$
2. $id_{n\times n} \cdot B = B$
3. $id_{n\times n} \cdot x = x.$
4. $A \cdot 0_{n\times l} = 0_{m\times l}$
5. $0_{l\times n} \cdot B = 0_{l\times m}$
6. $0_{l\times n} \cdot x = 0_l$
7. $A + 0_{m\times n} = A$
8. $x + 0_n = x$

In proposition 2.1.1 we saw that for computations with matrices, column-vectors and real scalars we have certain rules (or laws) at hand. There are a lot of other situations in mathematics where such rules hold. To generalize the concept one defines so called vectorspaces.

Definition 2.1.7. *A nonempty set* V *together with two maps*

$$\begin{aligned} + : V \times V &\to V \\ \cdot : \mathbb{R} \times V &\to V \end{aligned}$$

is called a **real vector-space** *if the following relations hold : We assume* $u, v, w \in V$ *and write* $+(v, w) =: v + w$ *as well as* $\cdot(r, v) =: r \cdot v$. *Then*

1. **Commutative Law**
$$v + w = w + v$$
2. **Associative Law**
$$(u + v) + w = u + (v + w)$$

3. **Distributive Laws**

$$\begin{aligned} r \cdot (v + w) &= r \cdot v + r \cdot w \\ (r + s) \cdot v &= r \cdot v + s \cdot v \end{aligned}$$

4. **Scalar Associativity**

$$(r \cdot s) \cdot v = r \cdot s \cdot v$$

5. **Unitality**

$$1 \cdot v = v$$

6. **Zero Vector** *:* $\exists 0 \in V$ *s.t.* $\forall\, v \in V$

$$0 + v = v$$

7. **Negative Vector** *: For each* $v \in V$ *there exists exactly one* $w \in V$ *s.t.*

$$v + w = 0$$

We denote $-v := w$.

The elements of V *are called* **vectors**.

Example 2.1.3. *1.* $\mathbb{R}^n$ *together with* $+ :=$ *"Vector Addition" and* $\cdot :=$ *"Scalar Multiplication".*

2. $\mathbb{R}^{m \times n}$ *together with* $+ :=$ *"Matrix Addition" and* $\cdot :=$ *"Scalar/Matrix Multiplication".*

3. *The set of all polynomials* $:= \{a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + ... + a_0 | a_i \in \mathbb{R}, n \in \mathbb{N}\}$

4. *The set of all real functions* $:= \{f : \mathbb{R} \to \mathbb{R}\}$

2.2 Linear Independence,Basis' and Subspaces

Let V be a real vectorspace and $v_1, ...v_k \in V$ be an arbitrary number of vectors.

Definition 2.2.1. *The set of vectors*

$$span(v_1, .., v_k) := \{r_1 \cdot v_1 + ... + r_k \cdot v_k | r_1, ..., r_k \in \mathbb{R}\} \tag{2.1}$$

is called the **span** *of* $v_1, .., v_k$. *We call* $v_1, .., v_k$ *a* **generating system** *if* $span(v_1, .., v_k) = V$.

In any case it is clear that $span(v_1, .., v_k) \subset V$. If $v_1, .., v_k$ is a generating system then any vector $v \in V$ can be expressed as a **linear combination**

$$v = r_1 \cdot v_1 + ... + r_k \cdot v_k$$

In general the scalar coefficients in the expression are not unique. However we will see that uniqueness holds under some extra condition.

Definition 2.2.2. *The vectors* $v_1, .., v_k$ *are called* **linear independent** *if whenever we have* $r_1 \cdot v_1 + ... + r_k \cdot v_k = 0$ *we must have* $r_1 = r_2 = ... = r_k = 0$. *If furthermore* $v_1, .., v_k$ *is a generating system of* V *then we call* $v_1, .., v_k$ *a* **basis** *of* V.

Example 2.2.1. *Consider* $\mathbb{R}^2$. *Then we have basis' given by*

1. $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{R}^2$
2. $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{R}^2$

Proposition 2.2.1. *Let* $v_1, .., v_k$ *be a basis of* V. *Then any vector* $v \in V$ *can be written uniquely as a linear combination*

$$v = r_1 \cdot v_1 + .. + r_k \cdot v_k.$$

The scalars $r_1, .., r_n$ *are called the* **coordinates** *of* v *with respect to the basis* $v_1, .., v_k$.

Proof. Since $v_1, .., v_k$ is a generating system v can be written in at least one way as a linear combination $v = r_1 \cdot v_1 + .. + r_k \cdot v_k$ of the vectors $v_1, .., v_k$. Assume now that it can also be written as $v = \tilde{r}_1 \cdot \tilde{v}_1 + .. + \tilde{r}_k \cdot v_k$. Then

$$\begin{aligned} 0 = v - v &= r_1 \cdot v_1 + .. + r_k \cdot v_k - \tilde{r}_1 \cdot v_1 + .. + \tilde{r}_k \cdot v_k \\ &\underbrace{=}_{\text{Def.2.1.7 3.)}} (r_1 - \tilde{r}_1) \cdot v_1 + .. + (r_k - \tilde{r}_k) \cdot v_k. \end{aligned}$$

By definition of linear independence this equation can only hold if $r_1 - \tilde{r}_1 = .. = r_k - \tilde{r}_k = 0$. But this is the same as $r_1 = \tilde{r}_1, .., r_k = \tilde{r}_k$. □

It is important that though the coordinates of one vector are unique with respect to one fixed basis, the same vector may have different coordinates with respect to different basis'. The following is an example.

Example 2.2.2. *The coordinates of* $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ *with respect to the first basis of* $\mathbb{R}^2$ *in Example 2.1.1 are* $r_1 = 1, r_2 = 1$ *since*

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The coordinates of the same vector with respect to the second basis however are $r_1 = 0, r_2 = 1$ *since we have*

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Because of our time constraint we won't proof the following proposition :

Proposition 2.2.2. *Let* $v_1, .., v_k, w_1, .., w_l$ *be a basis' of* V. *Then* $k = l$.

This proposition basically says that all basis' of a fixed vector space[1] have the same number of elements.

Definition 2.2.3. *If* V *has a basis consisting of* k *elements then the* **dimension** *of* V *is* k. *We write*

$$dim(V) = k.$$

Example 2.2.3. *1.* $dim(\mathbb{R}^n) = n$.

2. $dim(\mathbb{R}^{m \times n}) = m \cdot n$.

Definition 2.2.4. *Let* $U \subset V$ *be a subset. We call* U *a sub vectorspace of* V *and write* $U \leq V$ *if the following two conditions hold :*

1. If $u_1, u_2 \in U$ *then also* $u_1 + u_2 \in U$.

2. If $r \in \mathbb{R}, u \in U$ *then also* $r \cdot u \in U$.

One can summarize conditions 1 and 2 above by saying that U is closed under addition and scalar multiplication.

Example 2.2.4. *1.* $V \leq V$

2. $\{0\} \leq V$

3. $span(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}) \leq \mathbb{R}^3$.

[1] which has at least on finite basis

4. $span(v_1, .., v_k) \leq V$

The proof of the following proposition is left as an exercise :

Proposition 2.2.3. *Let* $U \leq V$ *be a sub vectorspace. Then* U *is itself a vectorspace and if* $dim(V) = k$ *then* $dim(U) \leq k$ *and* $dim(U) = k$ *if and only if* $U = V$*.*

2.3 Linear Maps

Definition 2.3.1. *Let* V *and* W *be vectorspaces. A map* $f : V \to W$ *is called* **linear** *or a* **homomorphism** *if for any* $v, v_1, v_2 \in V$*,* $r \in \mathbb{R}$ *one has*

$$\begin{aligned} f(v_1 + v_2) &= f(v_1) + f(v_2) \\ f(r \cdot v) &= r \cdot f(v). \end{aligned}$$

We denote the set of all linear maps from V *to* W *with* $Hom(V, W)$*.*

$Hom(V, W)$ is itself a vectorspace with addition defined via

$$\begin{aligned} f + g : V &\to W \\ v &\mapsto f(v) + g(v) \end{aligned}$$

for $f, g \in Hom(V, W)$ and scalar multiplication

$$\begin{aligned} r \cdot f : V &\to W \\ v &\mapsto r \cdot f(v) \end{aligned}$$

for $r \in \mathbb{R}$.

For any linear map $f \in Hom(V, W)$ we have that

$$f(0_V) = f(0 \cdot 0_V) = 0 \cdot f(0_V) = 0_W.$$

In the following let V, W, Z denote vectorspaces.

Proposition 2.3.1. *Let* $f \in Hom(V, W)$ *and* $g \in Hom(W, Z)$ *then the composition* $g \circ f$ *is a linear map and hence* $g \circ f \in Hom(V, Z)$*.*

Proof. Let $v_1, v_2 \in V$ and $r \in \mathbb{R}$. Then using first the linearity of$f$ and then the linearity of$g$ as well as Definition 1.2.5 we have

$$\begin{aligned}(g \circ f)(r \cdot v_1 + v_2) &= g(f(r \cdot v_1 + v_2)) \\ &= g(r \cdot f(v_1) + f(v_2)) \\ &= r \cdot g(f(v_1)) + g(f(v_2)) \\ &= r \cdot (g \circ f)(v_1) + (g \circ f)(v_2).\end{aligned}$$

□

Let us consider the image of a linear map $f \in Hom(V, W)$. Let $w_1, w_2 \in im(f) \subset W$. Then by Definition 1.2.2 there must exist $v_1, v_2 \in V$ such that $w_1 = f(v_1), w_2 = f(v_2)$. Then

$$w_1 + w_2 = f(v_1) + f(v_2) = f(v_1 + v_2).$$

This means that $w_1 + w_2 \in im(f)$. Similarly we have that for any $r \in \mathbb{R}$ and $w \in im(f)$ we have that $r \cdot w \in im(f)$. By Definition 2.1.4 this means that $im(f) \leq W$ is a sub vectorspace. So we have proven the following proposition :

Proposition 2.3.2. *Let $f \in Hom(V, W)$ be a linear map. Then $im(f)$ is a sub vectorspace.*

Definition 2.3.2. *Let $f \in Hom(V, W)$ be a linear map. We define the* **kernel** *of f as*

$$ker(f) := \{v \in V : f(v) = 0\}.$$

As the image, the kernel of a linear map is also a sub vectorspace as the following proposition shows.

Proposition 2.3.3. *Let $f \in Hom(V, W)$ be a linear map. Then $ker(f) \leq V$ is a sub vectorspace of V.*

Proof. Let $v_1, v_2 \in ker(f)$ and $r \in \mathbb{R}$. Then using the linearity of f we get

$$f(r \cdot v_1 + v_2) = r \cdot f(v_1) + f(v_2) = r \cdot 0_V + 0_V = 0_V.$$

This means that $r \cdot v_1 + v_2 \in ker(f)$ which proves the proposition. □

Definition 2.3.3. *Let $f \in Hom(V, W)$ be a linear map.*

1. *f is called a* **monomorphism** *if f is injective.*
2. *f is called a* **epimorphism** *if f is surjective.*
3. *f is called a* **isomorphism** *if f is bijective.*

If $f \in Hom(V, W)$ is an isomorphism, then it is linear and bijective. Therefore the inverse map $f^{-1} : W \to V$ exists. It is not apriori clear that f^{-1} is also linear but it is true as the following proposition shows.

Proposition 2.3.4. *Let $f \in Hom(V, W)$ be an isomorphism, the $f^{-1} \in Hom(W, V)$ and in fact is also an isomorphism.*

Proof. Let $w_1, w_2 \in W$ and $r \in \mathbb{R}$. Since f is bijective and in particular surjective, there exists $v_1, v_2 \in V$ such that $f(v_1) = w_1$ and $f(v_2) = w_2$. Since f is linear one also has that $f(v_1 + v_2) = w_1 + W_2$. Therefore using Definition 1.2.4 we have

$$f^{-1}(w_1 + w_2) = f^{-1}(f(v_1 + v_2)) = v_1 + v_2 = f^{-1}(w_1) + f^{-1}(w_2).$$

Furthermore we have

$$\begin{aligned} f^{-1}(r \cdot w_1) &= f^{-1}(r \cdot f(v_1)) \\ &= f^{-1}(f(r \cdot v_1)) \\ &= r \cdot v_1 \\ &= r \cdot f^{-1}(w_1). \end{aligned}$$

This shows the linearity off^{-1}. □

For linear it is much easier to check whether they are injective or not. The following proposition shows why.

Proposition 2.3.5. *Let $f \in Hom(V, W)$. Then*

$$f \text{ is injective} \quad \Leftrightarrow \quad ker(f) = \{0\}.$$

Proof. "$\Leftarrow$" : By Definition 1.2.3 we have to show that for any one element subset$\{w\} \subset W$ the set $f^{-1}(\{w\}) \subset V$ consists of at most one element. Assume $v_1, v_2 \in f^{-1}(\{w\})$- Then $f(v_1) = f(v_2)$ and furthermore

$$f(v_1 - v - 2) = f(v_1) - f(v_2) = w - w = 0_W$$

This means that $v_1 - v_2$ is a element in the kernel of f. Since by assumption 0_V is the only element in the kernel of f we must have $v_1 - v_2 = 0$ or equivalently $v_1 = v_2$ which shows that $f^{-1}(\{w\})$ contains at most one element and hence that f is injective.
"$\Rightarrow$" : We have $f(0_v) = 0_W$. Therefore $0_V \in f^{-1}(\{0_W\})$. Since f is by assumption injective there can be no other element in $f^{-1}(\{0_W\})$. However we have $ker(f) = f^{-1}(\{0_W\})$ which shows that $ker(f) = \{0_V\}$. □

Definition 2.3.4. *Let $f \in Hom(V, W)$. The dimension of $im(f)$ is called the* **rank** *of f and will be denoted with $rk(f)$.*

Proposition 2.3.6. *Let $f \in Hom(V, W)$ and let $dim(V) = n$. Then the following formula holds*

$$dim(ker(f)) + rk(f) = n.$$

Proof. Due to time constraints no proof ! □

The previous proposition has a very nice application in the following proposition.

Proposition 2.3.7. *Let $f \in Hom(V, W)$ and $dim(V) = dim(W) = n$. Then the following statements are equivalent.*

1. *f is a monomorphism.*
2. *f is an epimorphism.*
3. *f is an isomorphism.*

Proof. 1.) $\Rightarrow$ 2.) : If f is a monomorphism then f is injective. By Proposition 2.3.5 this means that $ker(f) = \{0_V\}$. Therefore $dim(ker(f)) = 0$ and by Proposition 2.3.6 we have that $rk(f) = n$. This however implies that $dim(im(f)) = n = dim(W)$ and since $im(f) \leq W$ Proposition 2.2.3 implies that $im(f) = W$. Hence f is surjective and therefore a monomorphism.
2.) $\Rightarrow$ 3.) : If f is an epimorphism, then $f$is surjective and hence $im(f) = W$. Therefore $rk(f) = dim(im(f)) = dim(W) = n$ and using again the equation in Proposition 2.3.6 we must have $dim(ker(f)) = 0$. This however means that $ker(f) = 0$ and therefore that f is injective. Hence f is a surjective and injective linear map and therefore an isomorphism.
3.) $\Rightarrow$ 1.) : Any isomorphism is also a monomorphism. For this reason this implication is clear. □

Proposition 2.3.8. *Let $v_1, .., v_n$ be a basis of V and let $w_1, .., w_n$ be n arbitrary vectors in W. Then there exists exactly one linear map $f \in Hom(V, W)$ such that $f(v_i) = w_i$ for all $i \in \{1, .., n\}$.*

Proof. We define $f : V \to W$ as follows. Let $v \in V$. Using Proposition 2.2.1 v can be written uniquely as a linear combination $v = \sum_{i=1}^{n} r_i \cdot v_i$ of the v_i. Then set $f(v) = \sum_{i=1}^{n} r_i \cdot w_i$. Clearly $f(v_i) = w_i$ for all i. Furthermore f is linear and therefore $f \in Hom(V, W)$. this proves the existence. Assume now there is another linear map $f' \in Hom(V, W)$ s.t. $f'(v_i) = w_i$ for all i. Then for arbitrary v as above we have

$$\begin{aligned} f'(v) \quad &= f'(\textstyle\sum_{i=1}^{n} r_i \cdot v_i) &= \displaystyle\sum_{i=1}^{n} r_i f'(\cdot v_i) \\ \textstyle\sum_{i=1}^{n} r_i w_i & &= \displaystyle\sum_{i=1}^{n} r_i f(\cdot v_i) = f(v) \end{aligned}$$

and therefore $f = f'$ which proves the uniqueness. □

Proposition 2.3.9. *Let $f \in Hom(V, W)$ and $v_1, ..., v_n$ be a basis of V. Then f is an isomorphism, if and only if $f(v_1), .., f(v_n)$ is a basis of W.*

Proof. Let us define $w_i := f(v_i)$ for all $i \in \{1, .., n\}$

"$\Rightarrow$" : Let us assume that f is an isomorphism. Let $w \in W$. Then since f is surjective, there exists $v \in V$ such that $f(v) = w$. Since $v_1, .., v_n$ is a basis of V there exists real numbers $r_i \in \mathbb{R}$ s.t. $v = \sum_{i=1}^{n} r_i \cdot v_i$. Then

$$w = f(v) = f(\sum_{i=1}^{n} r_i \cdot v_i) = \sum_{i=1}^{n} r_i \cdot f(v_i) = \sum_{i=1}^{n} r_i \cdot w_i$$

This shows that $w_1, .., w_n$ is a generating system for W (see Definition 2.2.1). Assume now that $\sum_{i=1}^{n} r_i \cdot w_i = 0$ for some $r_i \in \mathbb{R}$. Then for $v = \sum_{i=1}^{n} r_i \cdot v_i$ we have as before $f(v) = \sum_{i=1}^{n} r_i \cdot w_i = 0$ and therefore $v \in ker(f) = \{0_V\}$. hence $\sum_{i=1}^{n} r_i \cdot v_i = 0_V$ and since $v_1, .., v_n$ is a basis of V we must have $r_i = 0$ for all i. This proves that $w_1, .., w_n$ are linearly independent. hence $w_1, .., w_n$ is a basis of W.

"$\Leftarrow$" : Exercise ! (just a slight modification of the above) □

Corollary 2.3.1. *Let $dim(V) = n = dim(W)$. Then V and W are isomorphic and both are isomorphic to $\mathbb{R}^n$*

Proof. For the first statement let $v_1, .., v_n$ be a basis of V and $w_1, .., w, n$ be a basis of W. by Proposition 2.3.8 there exists $f \in Hom(V, W)$ such that $f(v_i) = w_i$ for all i. By proposition 2.3.9 f is an isomorphism. The second statement clearly follows from the first since $dim(\mathbb{R}^n) = n$. □

Definition 2.3.5. *Let* $e_1, .., e_n$ *with* $(0, .., 1, .., 0)^\top$ *where the entry* 1 *is at the* i*-th position be the canonical basis of* $\mathbb{R}^n$*. Let* V *be a vectorspace and* $v_1, .., v_n$ *be a basis of* V*. Then by Proposition 2.3.8 there exists* $\phi_{(v_1,..,v_n)} \in Hom(\mathbb{R}^n, V)$ *such that* $f(e_i) = v_i$*. By Proposition 2.3.9* $\phi_{(v_1,..,v_n)}$ *is an isomorphism. Furthermore by linearity we have for* $(x_1, .., x_n) \in \mathbb{R}^n$ *that*

$$\phi_{(v_1,..,v_n)}(x_1, .., x_n) = x_1 \cdot v_1 + ... + x_n \cdot v_n = \sum_{i=1}^{n} x_i \cdot v_i.$$

$\phi_{(v_1,..,v_n)}$ *is called the* **basis-isomorphism** *corresponding to the basis* $v_1, .., v_n$ *of* V*.*

2.4 Relations between Linear Maps and Matrices

In this section we will see that to any linear map $f : V \to W$ given basis' of V and W we can associate a unique matrix which describes the map f. Let us first assume that $f \in Hom(\mathbb{R}^n, \mathbb{R}^m)$. We denote the standard basis of $\mathbb{R}^n$ with e_j for $j \in \{1, .., n\}$ and the standard basis for $\mathbb{R}^m$ with e'_i for $i \in \{1, .., m\}$. Then for any $j \in \{1, .., n\}$ the vector $f(e_j) \in \mathbb{R}^m$ can be expressed uniquely as a linear combination of the vectors e'_i. Therefore there must be unique real numbers $b_{ji} \in \mathbb{R}$ for $j \in \{1, .., n\}$, $i \in \{1, .., m\}$ such that

$$f(e_j) = \sum_{i=1}^{m} b_{ji} e'_i.$$

for all $i \in \{1, .., n\}$. For an arbitrary vector $(x_1, .., x_n) = x_1 \cdot e_1 + .. + x_n \cdot e_n$ we have by linearity

$$f((x_1, .., x_n)^\top) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} b_{ji} e'_i = \sum_{i=1}^{m} (\sum_{j=1}^{n} b_{ji} x_j) \cdot e'_i = \begin{pmatrix} \sum_{j=1}^{n} b_{j1} x_j \\ . \\ . \\ \sum_{j=1}^{n} b_{jm} x_j \end{pmatrix}.$$

Let us define the real numbers a_{ij} by $a_{ij} = b_{ji}$ for $i \in \{1, .., m\}$ and $j \in \{1, .., n\}$. then the last equation can be written as

$$f((x_1,..,x_n)^\top) = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ . \\ . \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}.$$

The $n \cdot m$ numbers a_{ij} define a matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Taking a look at Definition 2.1.5 on how Matrix/Vector multiplication has been defined we can see that

$$f(\begin{pmatrix} x_1 \\ . \\ . \\ x_n \end{pmatrix}) = A \cdot \begin{pmatrix} x_1 \\ . \\ . \\ x_n \end{pmatrix}$$

Definition 2.4.1. *Let $f \in Hom(\mathbb{R}^n, \mathbb{R}^m)$. The $m \times n$ matrix constructed above is called the matrix corresponding to f and the standard-basis' of $\mathbb{R}^n$ and $\mathbb{R}^m$. We denote it with $A_f \in \mathbb{R}^{m\times n}$.*

Let us now assume that we have an arbitrary linear map $f \in Hom(V, W)$ and basis' $v_1, .., v_n$ of V as well as $w_1, .., w_m$ of W. In this case we have the basis-isomorphism $\phi_{(v_1,..,v_n)} : \mathbb{R}^n \to V$ and $\psi_{(w_1,..,w_n)} : \mathbb{R}^m \to W$.We define a linear map $\tilde{f}$ as follows :

$$\begin{aligned} \tilde{f} : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto (\psi^{-1}_{(w_1,..,w_n)} \circ f \circ \phi_{(v_1,..,v_n)})(x). \end{aligned}$$

This map makes the following diagram commuatitive :

$$\begin{tikzcd} V \arrow[r, "f"] & W \\ \mathbb{R}^n \arrow[u, "\phi_{(v_1,..,v_n)}"] \arrow[r, "\tilde{f}"] & \mathbb{R}^m \arrow[u, "\psi_{(w_1,..,w_n)}"'] \end{tikzcd}$$

Definition 2.4.2. *Let $f \in Hom(V, W)$ and $v_1, .., v_n$ resp. $w_1, .., w_m$ as above. The $m \times n$ matrix $A_{\tilde{f}}$ is called the matrix corresponding to f given basis' $v_1, .., v_n$ and $w_1, .., w_m$. instead of $A_{\tilde{f}}$ we write A_f but keep in mind that this matrix depends on the basis' chosen.*

Let A_f as above be given by the matrix (a_{ij}). Then we have

$$\begin{aligned}
f(v_j) &= \psi^{-1}_{(w_1,..,w_n)} \circ \overbrace{(\psi_{(w_1,..,w_n)} \circ f \circ \phi_{(v_1,..,v_n)})}^{\tilde{f}}(\overbrace{\phi^{-1}_{(v_1,..,v_n)}(v_i)}^{e_j}) \\
&= \psi^{-1}_{(w_1,..,w_n)}(\tilde{f}(e_j)) = \psi^{-1}_{(w_1,..,w_n)}(\sum_{j=1}^{m} b_{ji}e'_i) \\
&= \sum_{i=1}^{m} b_{ji}\underbrace{\psi^{-1}_{(w_1,..,w_n)}e'_i)}_{w_i} = \sum_{i=1}^{m} b_{ji}w_i
\end{aligned}$$

where $b_{ji} = a_{ij}$. This means we can compute the matrix of f given the basis' $v_1,..,v_n$ and $w_1,..,w_m$ as follows : First we compute for each basis vector v_j $f(v_j)$ and express it as a linear combination of the w_i as

$$f(v_j) = \sum_{i=1}^{m} b_{ji}w_i$$

Then we write the numbers b_{ji} in a matrix where we write b_{ji} in the i-th row and j-th column. This means that for the entries a_{ij} of our matrix we have $a_{ij} = b_{ji}$. One can say that the coordinates corresponding to the basis $w_1,..,w_m$ of the images of the basis vectors v_j are the columns of A_f. In the easier case where $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ one can just say the columns of the matrix are the images of the standard unit vectors e_j.

Example 2.4.1. *Let the linear map f be given as*

$$\begin{aligned}
f : \mathbb{R}^2 &\to \mathbb{R}^2 \\
\begin{pmatrix} x \\ y \end{pmatrix} &\mapsto \begin{pmatrix} x+y \\ y-x \end{pmatrix}
\end{aligned}$$

We have $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ as well as

$$f(e_1) = \begin{pmatrix} 1+0 \\ 0-1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$
$$f(e_2) = \begin{pmatrix} 0+1 \\ 1-0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

As mentioned above, to get the matrix corresponding to f and the standard-basis of $\mathbb{R}^2$ we have to write the images of the standard unit vectors in the columns of our matrix. hence

$$A_f = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Now we want to find the basis of the same linear map f given the basis $v_1 = e_2, v_2 = e_1$ and $w_1 = e_1 - e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, w_2 = e_1 + e_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then we have

$$\begin{aligned} f(v_1) = f(e_2) &= e_1 + e_2 = 1 \cdot w_1 + 0 \cdot w_2 \\ f(v_2) = f(e_1) &= e_1 - e_2 = 0 \cdot w_1 + 1 \cdot w_2. \end{aligned}$$

Therefore the matrix of f corresponding to the basis v_1, v_2 w_1, w_2 is

$$A_f = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This example already shows that the same linear map can have very different matrices corresponding to different basis'.

The following proposition (without proof) shows that the composition of linear maps is strongly related to the multiplication of matrices.

Proposition 2.4.1. *Let $f \in Hom(V, W)$, $g \in Hom(W, Z)$ and $v_1, .., v_n$ a basis of V, $w_1, .., w_m$ a basis of W and $z_1, .., z_k$ a basis of Z. Let A_f and A_g as well as $A_{g \circ f}$ the matrices of f,g and $g \circ f \in Hom(V, Z)$ corresponding to the basis' above. Then the following relation holds :*

$$A_{g \circ f} = A_g \cdot A_f.$$

Proposition 2.4.2. *Let $Id_V \in \hom(V, V)$ denote the identity. Then the matrix of Id_V corresponding to any basis $v_1, .., v_n$ of V is the identity matrix, i.e.*

$$A_{Id_V} = id_{n \times n} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Proof. We have $\tilde{Id}_V = \phi^{-1}_{(v_1,..,v_n)} \circ Id_V \circ \phi_{(v_1,..,v_n)} = Id_{\mathbb{R}^n}$ and the matrix of the identity given the standard-basis on $\mathbb{R}^n$ is clearly $id_{n \times n}$. □

2.5 The Rank of a Matrix and elementary Transformations

Definition 2.5.1. *Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Then we call*

$$rk(A) = rk(A : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto A \cdot x) = dim(im(A))$$

the **rank** *of A. Let us denote with*

$$x_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ a_{m1} \end{pmatrix}, ..., x_1 = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ a_{mn} \end{pmatrix}$$

the columns of A. Then we call $dim(span(x_1, .., x_n))$ the **column rank** *of A. Analogously denoting with $y_1 = (a_1 1, .., a_1 n), .., y_m = (a_{m1}, .., a_{mn})$ the rows of A we call $dim(span(y_1^\top, ..., y_m^\top))$ the* **row rank** *of A.*

Proposition 2.5.1. *Let $A \in \mathbb{R}^{m \times n}$. Then*

$$rg(A) = \text{ row rank of } A = \text{ column rank of } A.$$

The following proposition shows that the previously introduced concept of "rank of a linear map" is compatible with the concept "rank of a matrix".

Proposition 2.5.2. *Let $f \in Hom(V, W)$ and let $A_f \in \mathbb{R}^{m \times n}$ denote the corresponding matrix given basis' $v_1, .., v_n$ of V and $w_1, .., w_m$ of W. Then*

$$rk(A_f) = rk(f) = dim(im(f)).$$

So instead of computing the rank of a linear map, one can as well compute the rank of the corresponding matrix (given any basis'). The question remains, how to compute the rank of a matrix. For this the concept of elementary transformations has proven very useful.

One distinguishes between 3 types of elementary row transformations on a matrix $A \in \mathbb{R}^{m \times n}$:

1. Interchanging of rows/columns :

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j1} & \cdot & \cdot & \cdot & a_{jn} \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \to \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j1} & \cdot & \cdot & \cdot & a_{jn} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

2. Multiplication of a row with a scalar $\lambda \neq 0$:

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \rightarrow \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda \cdot a_{i1} & \cdot & \cdot & \cdot & \lambda \cdot a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

3. Addition of a scalar multiple of a row to another row (not the same one)

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j1} & \cdot & \cdot & \cdot & a_{jn} \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \rightarrow \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j1} + \lambda \cdot a_{i1} & \cdot & \cdot & \cdot & a_{j1} + \lambda \cdot a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Similarly there are 3 types of elementary column transformations. The following proposition shows why elementary transformations can be used to compute the rank of a matrix.

Proposition 2.5.3. *Any elementary transformation does not change the rank of a matrix.*

Proof. Let $A \in \mathbb{R}^{m \times n}$ and $y_1^\top, ..y_m^\top$ denote its column vectors as in Definition 2.5.1. Then clearly

$$\begin{aligned} span(y_1^\top, .., y_i^\top, .., y_j^\top, ..y_m^\top) &= span(y_1^\top, .., y_j^\top, .., y_i^\top, ..y_m^\top) \\ span(y_1^\top, .., y_i^\top, .., ..y_m^\top) &= span(y_1^\top, .., \lambda \cdot y_i^\top, ..y_m^\top) \\ span(y_1^\top, .., y_i^\top, .., y_j^\top, ..y_m^\top) &= span(y_1^\top, .., y_i^\top, .., y_j^\top + \lambda \cdot y_i^\top, ..y_m^\top). \end{aligned}$$

Writing dim in front of all expressions and using that $rk(A) =$ row rank of A the statement above follows. $\square$

Example 2.5.1. *1. If A is of the following form*

$$A = \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & a_{22} & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & \cdot & & & \cdot & \cdot \\ & & & a_{rr} & & \cdot & \cdot \\ & 0 & & & & & \\ & & & & 0 & & \\ & & & & & & \end{pmatrix}$$

where $a_{ii} \neq 0$, the dots denote arbitrary entries and the 0 denote 0-entries. Then $rg(A) = r$.

2. *If A can be transformed through elementary transformations to the form describes in 1.), then $r = rk(A)$.*

3. *Any matrix $A \in \mathbb{R}^{m\times n}$ can be transformed through elementary transformation into a matrix which has the form described in 1.).*

2.6 Inverting Matrices

Definition 2.6.1. *Let $A \in \mathbb{R}^{n\times n}$ be a matrix. A is called* **invertible** *if there exists matrix $B \in \mathbb{R}^{n\times n}$ such that*

$$A \cdot B = id_{n\times n} = B \cdot A.$$

We denote $B =: A^{-1}$ and call it the inverse matrix of A.

Proposition 2.6.1. *Let $f \in Hom(V, W)$ and A_f be the corresponding matrix, given basis' of V and W. Then*

$$f \text{ is an isomorphism } \Leftrightarrow A_f \text{ is invertible }.$$

Proof. Clearly under both assumptions V and W must have the same dimensions. The map f is an isomorphism if and only if there exists a map $g \in Hom(V, W)$ such that $f \circ g = id_W$ and $g \circ f = id_V$. If A_g denotes the matrix of g with respect to the chosen basis' of V and W we get by application of Proposition 2.4.1 and Proposition 2.4.2 that

$$A_f \cdot A_g = A_{f\circ g} = A_{id_W} = id_{n\times n} = A_{id_V} = A_{g\circ f} = A_g \cdot A_f$$

which shows that A_f is invertible. On the other side if A_f is invertible, then the map $A_f : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto A_f \cdot x$[2] has an inverse map, which is given by multiplication with the matrix $B = (A_f)^{-1}$. Let $v_1, .., v_n$ resp. $w_1, .., w_n$ be the chosen basis of V resp. W. Then $f = \phi_{(w_1,..,w_n)} \circ A_f \circ \phi^{-1}_{(v_1,..,v_n)}$ where $\phi_{(v_1,..,v_n)}$ and $\phi_{(v_1,..,v_n)}$ denote the basis-isomorphisms. Since the composition of isomorphisms is also an isomorphism we must have that f is an isomorphism. □

So if one has to check if a homomorphism f is an isomorphism one can as well check if the matrix is invertible. The question remains, how can one

[2] The product $\cdot$ stands here for Matrix/Vector Multiplication

check if a matrix is invertible. The following is a very easy method to do this.

Method for inverting matrices :

1. Begin by writing the matrix in question on the left side and the identity on the right side of a paper in the form

$$A|Id$$

2. Then use elementary row transformation, to transform A to the identity and for any transformation done on the left side perform exactly the same transformation on the right side. Then you end up with

$$Id|B$$

3. Then $B = A^{-1}$.

Example 2.6.1. *We illustrate the method at the following example :*

	1	0	1	1	0	0
	1	1	2	0	1	0
	0	-1	0	0	0	1
	1	0	1	1	0	0
$z_2 - z_1$	0	1	1	-1	1	0
	0	-1	0	0	0	1
	1	0	1	1	0	0
$z_3 - z_2$	0	1	1	-1	1	0
	0	0	1	-1	1	1
	1	0	0	2	-1	-1
$z_1 - z_3$	0	1	1	-1	1	0
	0	0	1	-1	1	1
	1	0	0	2	-1	-1
$z_1 - z_3$	0	1	0	0	0	-1
	0	0	1	-1	1	1

The matrix on the bottom right corner is the inverse of

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & -1 & 0 \end{pmatrix}$$

In any case one should check if the matrix B computed satisfies $A \cdot B = B \cdot A = id_{n \times n}$. We will later learn about another method how to compute the determinant of a matrix. This method needs the notion of determinants which we're about to discuss next.

2.7 The Determinant

Proposition 2.7.1. *There is exactly one map*

$$det : \mathbb{R}^{n \times n} \quad \rightarrow \quad \mathbb{R}$$

with the following properties :

1. *det is linear in each row, i.e.*

$$det \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} + \lambda \cdot b_{i1} & \cdot & \cdot & \cdot & a_{in} + \lambda \cdot b_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix} =$$

$$det \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i1} & \cdot & \cdot & \cdot & a_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix} + \lambda \cdot det \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{i1} & \cdot & \cdot & \cdot & b_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix}$$

2. *if* $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ *and* $rk(A) < n$*then* $det(A) = 0$

3. $det(id_{n \times n}) = 1$

This map is called the **determinant**.

We don't prove this result but list the major properties of this map in the following proposition.

Proposition 2.7.2. *The determinant has the following properties :*

1. *If* $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ *is upper triangular, that is* $a_{ij} = 0$ *for all* $i > j$ *then the determinant of* A *is just the product overall diagonal elements, i.e.* $det(A) = \prod_{i=1}^{n} a_{ii}$

2. *If* B *evolves from* A *through an elementary transformation of type* 1 *(row or column), then* $det(B) = -det(A)$.

3. *If* B *evolves from* A *through an elementary transformation of type* 2 *(row or column), then* $det(B) = \lambda \cdot det(A)$.

4. *If* B *evolves from* A *through an elementary transformation of type* 3 *(row or column), then* $det(B) = det(A)$.

5. *If* $A, B \in \mathbb{R}^{n\times n}$ *then* $det(A \cdot B) = det(A) \cdot det(B)$.

6. A *invertible* $\Leftrightarrow det(A) \neq 0$ *and in this case* $det(A^{-1}) = \frac{1}{det(A)}$.

In principal one can use these rules to compute the determinant of any given matrix A. Similar as in the method for inverting matrices one can transform the matrix into the identity matrix (here also elementary column transformations are allowed) and take track how the determinant changes by the rules b.),c.) and d.) above. In the end, as one knows the determinant of the identity matrix, one has the determinant of A.

Example 2.7.1. *Let us reconsider Example 2.6.1 :*

	1	0	1	
$A =$	1	1	2	
	0	-1	0	
	1	0	1	
$z_2 - z_1$	0	1	1	$det(A_1) = det(A)$
	0	-1	0	
	1	0	1	
$z_3 - z_2$	0	1	1	$det(A_2) = det(A_1)$
	0	0	1	
	1	0	0	
$z_1 - z_3$	0	1	1	$det(A_3) = det(A_2)$
	0	0	1	
	1	0	0	
$z_1 - z_3$	0	1	0	$1 = det(A_4) = det(A_3)$
	0	0	1	

This shows that $det(A) = 1$.

This example was a special case, where in fact one only had to perform elementary row transformations of type 3 where the determinant does not change at all. However most people wouldn't have gotten the idea to compute the determinant of this special matrix via the method introduced above. The following formula also computes the right result :

$$\begin{aligned} det\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & -1 & 0 \end{pmatrix} &= 1 \cdot det\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \\ &= (1 \cdot 0 - 2 \cdot (-1)) - (0 \cdot 0 - (-1) \cdot 1) + 0 \cdot (0 \cdot 2 - 1 \cdot 1) \\ &= 2 - 1 = 1. \end{aligned}$$

That this method works is a special case of the following result :

Proposition 2.7.3. *(***Cramer's Rule***) Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$. Let us denote with*

$$A^{ij} = \begin{pmatrix} a_{11} & \cdot & \cdot & a_{1i-1} & a_{1i+1} & \cdot & \cdot & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j-1,1} & \cdot & \cdot & a_{j-1,i-1} & a_{j-1,i+1} & \cdot & \cdot & a_{1n} \\ a_{j+1,1} & \cdot & \cdot & a_{j+1,i-1} & a_{j+1,i+1} & \cdot & \cdot & a_{j+1,n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & a_{ni-1} & a_{ni+1} & \cdot & \cdot & a_{nn} \end{pmatrix}$$

the matrix which evolves through A by removing the i-th column and the j-th row. Let us define

$$\tilde{a}_{ij} := (-1)^{i+j} det(A^{ij}).$$

and $\tilde{A} = (\tilde{a}_{ij}) \in \mathbb{R}^{n \times n}$ the so called **complementary matrix** *of A. Then:*

1. *computation of determinant by development with respect to i-th row :*

$$det(A) = \sum_{j=1}^{n} a_{ij} \cdot \tilde{a}_{ij}$$

2. computation of the inverse matrix :

$$A^{-1} = \frac{1}{det(A)} \tilde{A}$$

Now we know about the determinant of a matrix. How about the determinant of a homomorphism $f \in Hom(V,V)$.

Definition 2.7.1. *Let $f \in Hom(V,V)$ and let $v_1, .., v_n$ be a basis of V. Let A_f denote the matrix of f with respect to this basis. We define the determinant of f via*

$$det(f) = det(A_f).$$

That the definition above makes sense is not completely clear. If we would take a different basis $v_1', .., v_n'$ of V then the matrix A_f changes. Would we still get the same number for $det(f)$ by computing $det(A_f)$? Yes and the reason is Proposition 2.7.2 5.). We leave the details for the exercises.[3]

2.8 Gauss Algorithm

The Gauss Algorithm provides a method how one can systematically solve systems of linear equations. A system of linear equations is given by

$$\begin{array}{rcl}
a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n & = & b_1 \\
. \quad . \quad . \quad . \quad . \quad . \quad . & & . \\
. \quad . \quad . \quad . \quad . \quad . \quad . & & . \\
. \quad . \quad . \quad . \quad . \quad . \quad . & & . \\
a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n & = & b_m
\end{array}$$

where we think of the a_{ij} and b_i as given and call any $x = (x_1, ..., x_n)^\top$ which satisfies all the equations above a solution of the system. Denoting with A the matrix $A = (a_{ij})$ and x, b the vectors

$$x = \begin{pmatrix} x_1 \\ . \\ . \\ x_n \end{pmatrix} \in \mathbb{R}^n, \qquad b = \begin{pmatrix} b_1 \\ . \\ . \\ b_m \end{pmatrix} \in \mathbb{R}^m$$

[3]It is important though, that when one computes the matrix A_f via computation of $\tilde{f}$ as in definition 2.4.2, one uses the same basis on both sides.

the system of linear equations can be written as

$$A \cdot x = b.$$

The matrix A is called the **coefficient matrix**. The **extended coefficient matrix** is the matrix d

$$B = (A|b) = \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} & b_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & \cdot & \cdot & \cdot & a_{mn} & b_m \end{pmatrix}$$

The following method is known as the **Gauss Algorithm** :

1. Use elementary row transformations to bring the extended coefficient matrix B into the following form :

$$\begin{pmatrix} id_{r\times r} & C & \begin{matrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_r \end{matrix} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} & \begin{matrix} \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{matrix} \end{pmatrix}$$

   where $C = (c_{ij})$ is an $r \times (n-r)$ matrix.

2. There is at least one solution if and only if $\tilde{b}_{r+1} = ... = \tilde{b}_m = 0$ and this solution is given by

$$\tilde{x} = \begin{pmatrix} \tilde{b}_1 \\ \cdot \\ \cdot \\ \tilde{b}_r \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}.$$

3. If $r = m$ then the solution is unique and given as above.

4. If the condition in 2.) is satisfied and $r < m$ then there exist infinite many solutions and denoting

$$\tilde{x}^1 = \begin{pmatrix} -c_{11} \\ \cdot \\ \cdot \\ -c_{r1} \\ 1 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \tilde{x}^2 = \begin{pmatrix} -c_{12} \\ \cdot \\ \cdot \\ -c_{r2} \\ 0 \\ 1 \\ \cdot \\ 0 \end{pmatrix}, ..., \tilde{x}^{n-r} = \begin{pmatrix} -c_{1,n-r} \\ \cdot \\ \cdot \\ -c_{r,n-r} \\ 0 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

then any vector in the set

$$L = \{\tilde{x} + \lambda_1 \cdot \tilde{x}^1 + ... + \lambda_{n-r}\tilde{x}^{n-r} | \lambda_i \in \mathbb{R}, 1 \leq i \leq n-r\}$$

is a solution of the system of linear equations above.

2.9 Scalar Products

Definition 2.9.1. *Let V be a (real) vectorspace. A scalar-product is a map*

$$\begin{array}{rcl} V \times V & \to & \mathbb{R} \\ (x,y) & \mapsto & < x, y > \end{array}$$

with the following properties :

1. **Bilinearity** *: For each $v, u, w \in V$ and $\lambda \in \mathbb{R}$*

$$< v, \lambda \cdot u + w >= \lambda \cdot < v, u > + < v, w >$$

2. **Symmetry***: For all $v, w \in V$*

$$< v, w >=< w, v >$$

3. **Positive Definiteness***: For all* $v \in V$

$$< v, v > \geq 0$$

and $< v, v >= 0$ *if and only if* $v = 0$.

Example 2.9.1. *The* **standard scalar product** *on* $\mathbb{R}^n$ *is defined as follows:*

$$< \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} >= \sum_{i=1}^{n} x_i \cdot y_i$$

Symmetry and bilinearity of this construction is clear. For positive definiteness we have for $x^\top = (x_1, .., x_n)$ *that* $< x, x >= \sum_{i=1}^{n} x_i^2 \geq 0$ *and* $0 = \sum_{i=1}^{n} x_i^2 =< x, x >$ *if and only if all* $x_i = 0$ *and therefore* $x = 0$.

Remark 2.9.1. *The scalar-product of any vector* $v \in V$ *with the zero vector* 0_V *is zero. In fact*

$$< v, 0_V >=< v, 0 \cdot 0_V >= 0\cdot < v, 0_V >= 0.$$

Definition 2.9.2. *A pair* $(V, < \cdot, \cdot >)$ *consisting of a real vectorspace and a scalar-product is called* **Euclidean vectorspace***.*

We will see how scalar-products can be used to define distances and angles in a vectorspace.

Definition 2.9.3. *Let* $(V, < \cdot, \cdot >)$ *be a Euclidean vectorspace and let* $v \in V$ *be a vector. We define the* **norm** *of* v *as*

$$\|v\| := \sqrt{< v, v >}$$

and the **distance** *between two vectors* $v, w \in V$ *via*

$$d(v, w) := \|v - w\|.$$

Example 2.9.2. *Consider the Euclidean vectorspace given by* $\mathbb{R}^2$ *and the standard scalar product.*

To define the angle between two vectors we need an inequality which is known as the Cauchy- Schwartz inequality.

Proposition 2.9.1. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace and $v, w \in V$. Then*

$$| < v, w > | \leq \|v\| \cdot \|w\|.$$

Proof. In case that $w = 0$ both sides are equal to zero and hence the inequality is true. If $w \neq 0$ then we set $\lambda := \frac{<v,w>}{\|w\|^2} \in \mathbb{R}$. Using the positive definiteness and bilinearity of the scalar product we get that

$$\begin{aligned} 0 &\leq < v - \lambda \cdot w, v - \lambda \cdot w > \\ &= < v, v > -2\lambda \cdot < v, w > +\lambda^2 < w, w > \\ &= \|v\|^2 - 2 \cdot \frac{< v, w >^2}{\|w\|^2} + \frac{< v, w >^2}{\|w\|^2} \\ &= \|v\|^2 - \frac{< v, w >^2}{\|w\|^2} \end{aligned}$$

which implies that $\frac{<v,w>^2}{\|w\|^2} \leq \|v\|^2$. Multiplying this inequality with $\|w\|^2$ gives

$$< v, w >^2 \leq \|v\|^2 \cdot \|w\|^2.$$

Taking the square-root gives the result. □

Definition 2.9.4. *Let $v, w \neq 0$ b vectors in a Euclidean vectorspace $(V, < \cdot, \cdot >)$. We define the* **angle** *between v and w as*

$$\angle(v, w) := arccos(\frac{< v, w >}{\|v\| \cdot \|w\|})$$

where $arccos : [-1, 1] \to [0, \pi]$ denotes the arc cosine function.

The following proposition tells us about important properties of the norm map:

Proposition 2.9.2. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace. Then*

1. $\|v\| \geq 0$ *for all* $v \in V$
2. $\|v\| = 0 \Leftrightarrow v = 0$
3. $\|\lambda \cdot v\| = |\lambda| \cdot \|v\|$
4. $\|v + w\| \leq \|v\| + \|w\|$.

Proof. Statement i.) to iii.) follow immediately from the properties of the scalar product. To prove iv.) let us consider

$$\begin{aligned}(\|v\| + \|w\|)^2 &= \|v\|^2 + 2 \cdot \|v\|\|w\| + \|w\|^2 \\ &\geq \|v\|^2 + 2\cdot < v, w > + \|w\|^2 \\ &= \|v + w\|^2\end{aligned}$$

where we used the Cauchy-Schwartz inequality from Proposition 2.9.1. Taking square-roots on both sides gives the result. □

The inequality iv.) in the previous proposition is often called the triangle inequality. Working with the distance function instead of the norm one sees that the name makes sense :

Proposition 2.9.3. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace and let $d : V \times V \longrightarrow [0, \infty)$ be defined as in Definition 2.9.3. Then for $u, v, w \in V$ one has*

$$d(v, w) \leq d(v, u) + d(u, w).$$

In words this inequality says that it is less distance going directly from $v$to w than going from v to w by passing v.

Given two vectors $v, w \neq 0$ in a Euclidean vectorspace we would consider them as orthogonal if the angle between those two is $90° = \frac{\pi}{2}$. Since $arccos(x) = \frac{\pi}{2} \Leftrightarrow x = cos(\frac{\pi}{2}) = 0$ this is the case if and only if $< v, w >= 0$. This leads to the following definition :

Definition 2.9.5. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace and $v, w \in V$. Then v and w are called* **orthogonal** *to each other if $< v, w >= 0$. We write $v \perp w$. If $M \subset V$ is a subset of V then v is called orthogonal to M if $v \perp m$ for all $m \in M$. In this case we write $v \perp M$. We call*

$$M^\perp := \{v \in V | v \perp M\}$$

the **orthogonal complement**.

Proposition 2.9.4. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace and $M \subset V$. Then $M^\perp$ is a sub-vectorspace of V.*

Proof. Exercise ! □

2.10 Orthonormal Systems

Definition 2.10.1. *A set of vectors $v_1, ..., v_k$ in a Euclidean vectorspace is called an* **orthonormal system** *if*

1. $< v_i, v_j >= 0$ *for all* $i \neq j$
2. $\|v_i\| = 1$ *for all* $i = 1, .., k$.

The set is called an **orthogonal system** *if condition* 1. *and not necessarily condition* 2. *holds.*

Proposition 2.10.1. *Let $v_1, ..., v_k$ be an orthogonal system in a Euclidean vectorspace $(V, < \cdot, \cdot >)$ s.t. $v_i \neq 0$ for all i. Then $v_1, ..., v_k$ are linear independent.*

Proof. Let $r_i \in \mathbb{R}$ such that

$$0_V = r_1 \cdot v_1 + ... + r_k \cdot v_k.$$

Therefore for any i we have (see Remark 2.9.1)

$$\begin{aligned} 0 &= < v_i, 0_v >=< v_i, r_1 \cdot v_1 + ... + r_k \cdot v_k > \\ &= r_1 \cdot < v_i, v_1 > +... + r_k \cdot < v_i, v_k > \\ &= r_i \cdot \underbrace{< v_i, v_i >}_{\neq 0} \end{aligned}$$

and hence $r_i = 0$. By Definition 2.2.2 $v_1, .., v_k$ are linearly independent. □

Definition 2.10.2. *Let $(V, < \cdot, \cdot >)$ be a Euclidean vectorspace and $v_1, ..., v_n$ be a basis of V. Then $v_1, ..., v_n$ is called an orthonormal basis if $v_1, ..., v_n$ is an orthonormal system.*

Example 2.10.1. *Consider $\mathbb{R}^n$ together with the standard scalar product from example 2.9.1. Then the standard basis $e_1, ..., e_n$ from example 2.2.1 is an orthonormal basis.*

In general given a basis $v_1, ..., v_n$ of V and a vector $v \in V$ and one wants to express v as a linear combination of the vectors v_i one has to solve a system of linear equations. If the basis however is an orthonormal basis this is much easier and only involve computation of scalar products. This is the result of the following proposition :

Proposition 2.10.2. *Let $v_1, ..., v_n$ be an orthonormal basis of $(V, < \cdot, \cdot >)$ and $v \in V$. Then*

$$v = \sum_{i=1}^{n} < v, v_i > \cdot v_i.$$

Proof. Since $v_1, ..., v_n$ is a basis of V there exists $r_i \in \mathbb{R}$ s.t. $v = \sum_{i=1}^{n} r_i \cdot v_i$. Then

$$< v, v_j > = < \sum_{i=1}^{n} r_i \cdot v_i, v_j > = \sum_{i=1}^{n} r_i \cdot \underbrace{< v_i, v_j >}_{\delta_{ij}} = r_j$$

and therefore $v = \sum_{i=1}^{n} < v, v_i > \cdot v_i$. □

The following proposition provides an algorithm for finding an orthonormal basis of a Euclidean vector space.

Proposition 2.10.3. Schmidt's Orthonormalization Procedure *: Let $v_1, ..., v_n$ be a basis of the Euclidean vectorspace $(V, < \cdot, \cdot >)$. Define vectors $\tilde{v}_i \in V$ for $i = 1, .., n$ inductively as follows :*

$$\begin{aligned} \tilde{v}_1 &:= \frac{v_1}{\|v_1\|} \\ \tilde{v}_{i+1} &:= \frac{v_{i+1} - \sum_{k=1}^{i} < v_{i+1}, \tilde{v}_k > \tilde{v}_k}{\|v_{i+1} - \sum_{k=1}^{i} < v_{i+1}, \tilde{v}_k > \tilde{v}_k\|} \end{aligned}$$

Then $\tilde{v}_1, ..., \tilde{v}_n$ is an orthonormal basis of V.

2.11 Orthogonal Maps

Definition 2.11.1. *Let $(V, < \cdot, \cdot >_V)$ and $(W, < \cdot, \cdot >_W)$ be two Euclidean vectorspaces and $f \in Hom(V, W)$ be a linear map. Then f is called an orthogonal map (sometimes also isometry) if*

$$< f(u), f(v) >_W = < u, v >_V$$

for all $u, vinV$.

The following proposition shows us how one can test whether a linear map is orthogonal or not.

Proposition 2.11.1. *Let V and W be as in Definition 2.10.1,$v_1, ..., v_n$ be an orthonormal basis of V and $f \in Hom(V, W)$. Then f is an orthogonal map if and only if $f(v_1), ..., f(v_n)$ is an orthonormal system in W.*

Proof. Exercise ! $\square$

Example 2.11.1. *If $A \in \mathbb{R}^{n\times m}$ is an $n \times m$ matrix, we can consider A as a linear map*

$$\begin{aligned} A : \mathbb{R}^m &\rightarrow \mathbb{R}^n \\ x &\mapsto A \cdot x. \end{aligned}$$

On $\mathbb{R}^m$ and respectively $\mathbb{R}^n$ we consider the standard scalar products of Example 2.9.1. and the standard orthonormal basis from example 2.10.1. Since the columns of the matrix A are precisely the images $A(e_i)$ we see that A is orthogonal if and only if its columns are an orthonormal system of $\mathbb{R}^n$.

Definition 2.11.2. *Let . Then the matrix $A^\top = (\tilde{a}_{ij}) \in \mathbb{R}^{m\times n}$ where $\tilde{a}_{ij} = a_{ji}$ for all i, j is called the* **transposed matrix** *of A.*

Definition 2.11.3. *The matrix $A \in \mathbb{R}^{n\times n}$ is called* **orthogonal** *if*

$$A^\top \cdot A = id_{n\times n}$$

Remark 2.11.1. *If $n = m$ then the map defined in Example 2.11.1 is orthogonal if and only if-the matrix A is orthogonal.*

Proposition 2.11.2. *$A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be a matrix. Then $det(A^\top) = det(A)$.*

Proof. Using Proposition 2.7.2 one can see that the map $A \mapsto det(A^\top)$ satisfies the conditions in Proposition 2.7.1 and therefore must coincide with the map $A \mapsto det(A^\top)$ $\square$

Proposition 2.11.3. *Let A be an orthogonal matrix, then $|det(A)| = 1$.*

Proof. Since A is orthogonal we have $A^\top \cdot A = id$ and by Proposition 2.7.2 (5) and Proposition 2.10. we have

$$1 = det(id) = det(A^\top \cdot A) = det(A^\top) \cdot det(A) = det(A)^2$$

Therefore $det(A) = 1$ or $det(A) = -1$. $\square$

2.12 Eigenvalues and Eigenvectors

Definition 2.12.1. *Let V be a (real) vectorspace and $f \in Hom(V, V)$ be a linear map. Let $v \in V \setminus \{0\}$ be a vector s.t.*

$$f(v) = \lambda \cdot v$$

*with $\lambda \in \mathbb{R}$ then v is called an **eigenvector** to the **eigenvalue** λ. If $A \in \mathbb{R}^{m\times n}$ then $x \in \mathbb{R}^n$ is called an eigenvector of A with eigenvalue λ if it is and eigenvector with eigenvalue of the corresponding linear map $A : \mathbb{R}^n \to \mathbb{R}^m$.*

Example 2.12.1. *Let V be an arbitrary vectorspace and $\lambda \in \mathbb{R}$. Define $f : V \to V$ via $v \mapsto \lambda v$. Then any vector in V except the zero vector is an eigenvector with eigenvalue λ.*

If we just say $\lambda \in \mathbb{R}$ is an eigenvalue of $f \in Hom(V, V)$ then we mean that there exists an eigenvector v with eigenvalue λ but we don't specify v. Now given a basis of V and $f \in Hom(V, V)$ we can compute the matrix of A_f of$f$ with respect to this basis (see Definition 2.4.2). We already mentioned that by choosing the basis in a suitable way, one can achieve that the matrix A_f has a simple structure. The best case that can happen when the vectorspace V admits a basis of eigenvectors of f as the following proposition shows :

Proposition 2.12.1. *Let $v_1, ..., v_n$ be a basis of V s.t. for each i the vector v_i is an eigenvector with eigenvalue λ_i of f. Let A_f denote the matrix of f with respect to this basis. Then*

$$A_f = \begin{pmatrix} \lambda_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \lambda_{n-1} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \lambda_n \end{pmatrix}$$

Proof. By definition of the v_i we have $f(v_i) = \lambda_i \cdot v_i$. As mentioned in section 2.4. the columns of the matrix A_f are the coordinate vectors of the images of the basis-vectors with respect to the same basis $v_1, ..., v_n$. □

On the other side it is clear that if the vector space V admits a basis such that the matrix A_f of f with respect to this basis has the form as in Proposition 2.11.1 then this basis is a basis of eigenvectors.

Definition 2.12.2. *An endomorphism $f \in Hom(V, V)$ is called diagonizable if V admits a basis of **eigenvectors** of f.*

Not every $f \in Hom(V,V)$ is diagonizable, for example the linear map corresponding to the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is not diagonizable. One can show this by hand (Exercise) or wait until we have some more theory at hand to see this.

Lemma 2.12.1. *Let $f \in Hom(V,V)$. Then $v \neq 0$ is an eigenvector of f with eigenvalue λ if and only if $v \in ker(f - \lambda \cdot Id)$.*

Proof. We have : v eigenvector with eigenvalue $\lambda \Leftrightarrow f(v) = \lambda \cdot v \Leftrightarrow (f - \lambda \cdot Id)(v) = 0 \Leftrightarrow v \in ker(f - \lambda \cdot Id)$. □

Definition 2.12.3. *Let $f \in hom(V,V)$ be a linear map and $\lambda \in \mathbb{R}$ be an eigenvalue of f. We call*

$$Eig(\lambda) := kern(f - \lambda \cdot Id)$$

*the **eigenspace** of f to the eigenvalue λ. We call its dimension $dim(Eig(\lambda))$ the (geometric) multiplicity of λ.*

It follows from Proposition 2.3.3 that the eigenspaces are sub vector spaces of V.

Lemma 2.12.2. *Let $f \in Hom(V,V)$ and let $v_1, ..., v_k$ be eigenvectors with eigenvalues $\lambda_1, ..., \lambda_r$ and let $\lambda_i \neq \lambda_j$ if $i \neq j$ then $v_1, ...v_k$ are linear independent.*

Proof. Let $r_1 \cdot v_1 + ... + r_r \cdot v_k = 0$. We have to show all r_j are equal to zero. Assume there exist one which is not equal to zero, w.l.o.g. $r_k \neq 0$. Furthermore via induction we can assume that $v_1, ..., v_{k-1}$ are linear independent. Applying f on both sides gives

$$r_1 \cdot \lambda_1 \cdot v_1 + ... + r_k \cdot \lambda_k \cdot v_k = 0$$

From this equation we subtract $0 = \lambda_k \cdot (r_1 \cdot v_1 + ... + r_r \cdot v_k)$ and since $0 - 0 = 0$ we get

$$r_1 \cdot (\lambda_1 - \lambda_k) \cdot v_1 + ... + r_{k-1} \cdot (\lambda_{k-1} - \lambda_k) \cdot v_{k-1} + r_k \cdot \underbrace{(\lambda_k - \lambda_k)}_{=0} \cdot v_k = 0.$$

Since $v_1, ..., v_{k-1}$ are linear independent and $\lambda_j - \lambda_k \neq 0$ for all $j = 1, .., k-1$ we have $r_1 = ... = r_{k-1} = 0$. But then the equality which we started with is just $r_k \cdot v_k = 0$. Since $r_k \neq 0$ this would imply that $v_k = 0$. This however is a contradiction since by definition v_k is an eigenvector and eigenvectors are different from zero. □

Proposition 2.12.2. *Let $f \in Hom(V,V)$ and $\lambda_1, ..., \lambda_k$ be eigenvectors of f such that $\lambda_i \neq \lambda_j$ wheneveri $\neq 0$ and $\sum_{i=1}^k Eig(\lambda_i) = dim(V)$ then f is diagonizable. In fact choosing a basis $v_1^i, ..., v_{n_i}^i$ of $Eig(\lambda_i)$ for each i then $v_1^1, ..., v_{n_1}^1, ..., v_1^k, ..., v_{n_k}^k$ is a basis of eigenvectors of f an with respect to this basis the matrix A_f of f has the form*

$$A_f = \begin{pmatrix} \lambda_1 \cdot id_{n_1 \times n_1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2 \cdot id_{n_2 \times n_2} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \lambda_{k-1} \cdot id_{n_{k-1} \times n_{k-1}} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \lambda_k \cdot id_{n_k \times n_k} \end{pmatrix}$$

where the zeroes in the matrix above denote the corresponding zero matrices.

2.13 The characteristic Polynomial

The characteristic polynomial provides us with a method how to compute eigenvalues of an endomorphism effectively. It follows from Lemma 2.12.1 and that

$$v \text{ eigenvalue of f} \Leftrightarrow f - \lambda \cdot Id \text{ is not invertible .}$$

By Definition 2.7.1. $det(f - \lambda \cdot Id) = det(A_f - \lambda \cdot id_{n \times n})$ and using Proposition 2.7.2 as well as Proposition 2.6.1 we get

$$v \text{ eigenvalue of f} \Leftrightarrow det(A_f - \lambda \cdot id_{n \times n}) = 0.$$

Definition 2.13.1. *Let $f \in Hom(V,V)$ be a linear map, then we call*

$$\chi_f(\lambda) :)det(f - \lambda \cdot Id)$$

*the **characteristic polynomial** of f.*

Remark 2.13.1. *If $A_f = (a_{ij})$ then*

$$\chi_f(\lambda) = det \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & a_{n-1,n-1} - \lambda & a_{n-1,n} \\ a_{n1} & a_{n2} & \cdot & \cdot & a_{nn} - \lambda \end{pmatrix}$$

and one can show that the characteristic polynomial does not depend on the basis chosen to compute A_f.

Example 2.13.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the map associated to the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then*

$$\chi_f(\lambda) = det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1.$$

The following important proposition follows directly from the discussion in the beginning of this section the follow

Proposition 2.13.1. *Let $f \in Hom(V, V)$ the λ is an eigenvalue of f if and only if $\chi_f(\lambda) = 0$ i.e. λ is a zero of the characteristic polynomial of f.*

Since the characteristic polynomial of f with $A_f = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in Example 2.13.1 has no (real) zeroes we see that this map has no eigenvalues and hence no eigenvectors, in particular there exists no basis of $\mathbb{R}^2$ of eigenvectors of f so that f is clearly not diagonizable.

The following proposition gives us a simple criterium to decide when a linear map f is diagonizable.

Proposition 2.13.2. *Let $f \in Hom(V, V)$ and assume the characteristic polynomial χ_f of f has the following form :*

$$\chi_f(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i)$$

where the λ_i are pairwise disjoint. Then f is diagonizable and its eigenvalues are $\lambda_1, ..., \lambda_n$.

Proof. Clearly for any i we have $\chi_f(\lambda_i) = 0$. Therefore all λ_i are eigenvalues of f and hence there exists eigenvectors v_i corresponding to λ_i. In particular $dim(Eig(\lambda_i) \geq 1$ and therefore

$$\sum_{i=1}^{n} Eig(\lambda_i) \geq n = dim(V).$$

Since the sum on the left side can also not be strictly greater then $dim(V)$ (see Lemma 2.12.1) we have $\sum_{i=1}^{n} Eig(\lambda_i) = dim(V)$ and the result follows from Proposition 2.12.2.

□

Example 2.13.2.

2.14 Self-adjoined Maps

In this section we consider only Euclidean vectorspaces.

Definition 2.14.1. *Let* $(V, < \cdot, \cdot >)$ *be a Euclidean vectorspace and* $f \in Hom(V, V)$. *Then* f *is called* **self-adjoined** *if for all* $v, w \in V$ *one has*

$$< f(v), w >=< v, f(w) > .$$

The following proposition shows that when working with self-adjoined maps the handling of eigenvalues and eigenvectors is a little bit easier.

Proposition 2.14.1. *Let* $f \in Hom(V, V)$ *be a self-adjoined map and* $v \in Eig(\lambda)$ *as well as* $w \in Eig(\mu)$ *for* $\lambda \neq \mu$. *Then* $< v, w >= 0$ *i.e.* v *and* w *are orthogonal. In particular* $Eig(\lambda) \perp Eig(\mu)$.

Proof. We have $\lambda < v, w >=< \lambda v, w >=< f(v), w >=< v, f(w) >=< v, \mu w >= \mu < v, w >$. Therefore

$$\underbrace{(\lambda - \mu)}_{\neq 0} < v, w >= 0$$

and therefore $< v, w >= 0$.

□

Also the computation of a matrix of a linear map $f \in Hom(V, V)$ with respect to an orthonormal basis is much easier than in general.

Proposition 2.14.2. *Let* $(V, < \cdot, \cdot >)$ *be a Euclidean vectorspace and* $f \in Hom(V, V)$. *Let* $v_1, ..., v_n$ *be an orthonormal basis of* V *then the matrix* A_f *of* f *with respect to this basis is given by* $A_f = (a_{ij})$ *with*

$$a_{ij} =< v_i, f(v_j) > .$$

Proof. This follows more or less directly from the discussion after Definition 2.4.2. when taking scalar products on both sides of the equation $f(v_j) = \sum_i a_{ij} v_i$. □

Corollary 2.14.1. *Let* $v_1, ..., v_n$ *be an orthonormal basis of the Euclidean vector space* $(V, < \cdot, \cdot >)$ *and* $f \in Hom(V, V)$. *Then the matrix* $A_f = (a_{ij})$ *of* f *with respect to this basis is symmetric, i.e.* $a_{ij} = a_{ji}$ *if and only if the* f *is a self-adjoined map.*

Proof. It follows from Proposition 2.4.12 that if f is self-adjoined then

$$a_{ij} =< v_i, f(v_j) >=< f(v_i), v_j >=< v_j, f(v_i) >= a_{ji}$$

so A_f is symmetric. If on the other side A_f is symmetric, then the equality above is still true. Furthermore for arbitrary $v, w \in V$ we can find $r_i, s_j \in \mathbb{R}$ s.t.

$$\begin{aligned} v &= \sum_{i=1}^{n} r_i \cdot v_i \\ w &= \sum_{i=1}^{n} s_i \cdot w_i. \end{aligned}$$

Then by definition of A_f we have

$$\begin{aligned} < f(v), w > &= < f(\sum_{i=1}^{n} r_i \cdot v_i), \sum_{j=1}^{n} s_j \cdot v_j > \\ &= \sum_{i,j} r_i \cdot s_j < f(v_i), v_j > \\ &= \sum_{i,j} r_i \cdot s_j < v_i, f(v_j) > \\ &= < \sum_{i=1}^{n} r_i \cdot v_i, f(\sum_{j=1}^{n} s_j \cdot v_j) > \\ &= < v, f(w) > . \end{aligned}$$

□

The proof of the following proposition will be postponed until the necessary tools from Analysis are available. however the statement is easy to get and of major importance.

Proposition 2.14.3. *Let* $(V, < \cdot, \cdot >)$ *be a Euclidean vectorspace and* $f \in Hom(V, V)$ *be a self adjoined map, then there exists an orthonormal basis of eigenvectors of* f*.*

So if $f \in Hom(V, V)$ then the matrix A_f of f with respect to this basis has diagonal form, where the eigenvalues stand on the diagonal. The following method is therefore good to diagonalize a linear self-adjoined map or equivalently a symmetric matrix.

1. Compute the characteristic polynomial and its zeroes. The zeroes are the eigenvalues of f.
2. For each eigenvalue λ of f compute an orthonormal basis of $Eig(\lambda)$. This can be done by first computing an ordinary basis and then apply the Schmidt orthonormalization procedure Proposition 2.10.3 to this basis.
3. Put all these basis' together to get a basis of V. this basis is automatically and orthonormal basis and with respect to this basis the matrix of f has diagonal form.

Example 2.14.1.

2.15 The Leontief Economic Model

In this section we present how the methods from Linear Algebra can be applied to mathematical modeling of economics. The following example goes back to Wassily Leontief who was awarded the Nobel prize in economics in 1973. Assume a conglomerate consists of four industries.

$$\begin{aligned} M_1 &= \text{coal} \\ M_2 &= \text{energy} \\ M_3 &= \text{steel} \\ M_4 &= \text{transportation} \end{aligned}$$

Assume each of the industries buys all that it needs from the conglomerate and that all of its needs can be met from within the conglomerate. Let p_i b the total value in millions of Euro of the production of M_i. Let c_{ij} denote the value in Euro of M_i required to produce one Eros worth of M_j. Then

$$c_{ij} \cdot p_j = \text{ value in Euro of } M_i \text{ needed to produce value } p_j \text{ of } M_j.$$

Let d_i be the value in millions of Euro of M_i consumed outside of the conglomerate over some fixed period of time. An ideal economy produces exactly the amount needed for consumption.

Problem : Consider the economy above, what are the total values $p_1, .., p_4$

s.t. everything what is produced is also consumed.

Translating into mathematics we get

$$\begin{aligned}
c_{11} \cdot p_1 + c_{12} \cdot p_2 + c_{13} \cdot p_3 + c_{14} \cdot p_4 + d_1 &= p_1 \\
c_{21} \cdot p_1 + c_{22} \cdot p_2 + c_{23} \cdot p_3 + c_{24} \cdot p_4 + d_2 &= p_2 \\
c_{31} \cdot p_1 + c_{32} \cdot p_2 + c_{33} \cdot p_3 + c_{34} \cdot p_4 + d_3 &= p_3 \\
c_{41} \cdot p_1 + c_{42} \cdot p_2 + c_{43} \cdot p_3 + c_{44} \cdot p_4 + d_4 &= p_4.
\end{aligned}$$

The matrix (c_{ij}) is called the **consumption matrix** and the vector

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}$$

is called the **output vector**. Rearranging the terms in the system of linear equations above we get

$$\begin{aligned}
(c_{11} - 1) \cdot p_1 + c_{12} \cdot p_2 + c_{13} \cdot p_3 + c_{14} \cdot p_4 &= -d_1 \\
c_{21} \cdot p_1 + (c_{22} - 1) \cdot p_2 + c_{23} \cdot p_3 + c_{24} \cdot p_4 &= -d_2 \\
c_{31} \cdot p_1 + c_{32} \cdot p_2 + (c_{33} - 1) \cdot p_3 + c_{34} \cdot p_4 &= -d_3 \\
c_{41} \cdot p_1 + c_{42} \cdot p_2 + c_{43} \cdot p_3 + (c_{44} - 1) \cdot p_4 &= -d_4.
\end{aligned}$$

This however is the same as

$$((c_{ij}) - (-1) \cdot id_{4\times4}) \cdot p = -d$$

where

$$d = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}$$

This equation has a unique solution in the $p_1, .., p_4$ if

$$det((c_{ij}) - (-1) \cdot id_{4\times4}) \neq 0$$

in equal that -1 is not a zero of the characteristic polynomial of the consumption matrix or equivalently that -1 is not an eigenvalue of the consumption matrix.

Chapter 3

Differential Calculus

Linear Algebra is mainly concerned with the theory of linear maps. There are also non linear maps such as for example the map

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto x^2. \end{aligned}$$

The theory of differential calculus can arguably be summarized as the theory about how to approximate non linear maps by linear maps. The linear map approximating the map is (if it exists) called the differential of the map. However before we discuss how the differential of a map has to be properly defined we have to get some more background.

3.1 Sequences and Infinite Sums

Definition 3.1.1. *A map $g : \mathbb{N} \rightarrow \mathbb{R}$ is called a* **sequence of real numbers**. *We think of g as an infinite sequence $(a_1, a_2, a_3, ...)$ with $a_m = f(m)$ and in general denote it with $(a_m)_{m\in\mathbb{N}}$.*

Example 3.1.1. *The following are examples of sequences :*

1. $(1, 2, 3...)$ *i.e.* $a_m = m$
2. $(-1, 1, -1, 1, ...)$ *i.e.* $a_m = (-1)^m$
3. $(1, \frac{1}{2}, \frac{1}{3}, ...)$ *i.e.* $a_m = \frac{1}{m}$

The most important concept in the theory of sequences is that of convergence.

Definition 3.1.2. *The sequence* $(a_m)_{m\in\mathbb{N}}$ *is called* **convergent** *with* **limit** a *if for any* $\epsilon > 0$ *there exists* m_0 *s.t. for all* $m \geq m_0$ *we have*

$$|a_m - a| < \epsilon.$$

In this case we write

$$\lim_{n\to\infty} a_m := a$$

and call a *the limit of* $(a_m)_{m\in\mathbb{N}}$. *We also say that* $(a_m)_{m\in\mathbb{N}}$ *converges to* a. *The sequence* $(a_m)_{m\in\mathbb{N}}$ *is called* **convergent** *if there exists any* $a \in \mathbb{R}$ *such that it converges to this* a *else it is called* **non convergent**.

Not every sequence converges. For example the first two sequences of Example 3.1.1 do not converge. The third one though is an example of a convergent sequence.

Example 3.1.2. *Consider the sequence* $(a_m)_{m\in\mathbb{N}}$ *with* $a_m = \frac{1}{m}$. *In Definition 3.1.2 set* $a = 0$ *and let* $\epsilon > 0$. *Then there exists* m_0 *s.t.* $\frac{1}{m_0} < \epsilon$. *For all* $m \geq m_0$ *we have*

$$|a_n - a| = |\frac{1}{n} - 0| = \frac{1}{n} \leq \frac{1}{n_0} < \epsilon.$$

Therefore the limit of $(\frac{1}{n})_{n\in\mathbb{N}}$ *exists and is equal to zero.*

A simple argument shows that a convergent sequence can not have more than one limit. The following proposition gives a criterium when a sequence is convergent.

Proposition 3.1.1. Cauchy Criterium *: Let* $(a_n)_{n\in\mathbb{N}}$ *a sequence of real numbers, then* $(a_n)_{n\in\mathbb{N}}$ *is convergent if and only if it is a* **Cauchy sequence**, *i.e. for all* $\epsilon > 0$ *there exists* n_0 *s.t. for all* $n, m \geq n_0$

$$|a_n - a_m| < \epsilon.$$

The next proposition gives us some rules which can help us to do computations with sequences and to determine their limits.

Proposition 3.1.2. *Let* $(a_m)_{m\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be convergent sequences with*

$$\begin{aligned} a &= \lim_{m\to\infty} a_m \\ b &= \lim_{m\to\infty} b_m. \end{aligned}$$

1. *Let* $c_m = a_m \cdot b_m$. *Then* $\lim_{m\to\infty} c_m = a \cdot b$.

2. *If* $b \neq 0 \neq b_m$ *for all* $m \in \mathbb{N}$, *then denoting* $d_m = \frac{a_m}{b_m}$ *one has* $\lim_{m\to\infty} d_m = \frac{a}{b}$.

Let us now come to a special version of sequence so called infinite sums.

Definition 3.1.3. *An infinite sum is a sequence* $(s_m)_{m\in\mathbb{N}}$ *where* s_m *is given by*

$$s_m = \sum_{i=1}^{m} a_i$$

with $a_1 = s_1$ *and* $a_i = s_i - s_{i-1}$ *for* $i > 1$. *We denote the sequence* $(s_m)_{m\in\mathbb{N}}$ *with the symbol*

$$\sum_{i=1}^{\infty} a_i.$$

We say that the infinite sum is **convergent** *with limit* a *if and only if* $\lim_{m\to\infty} s_m = a$. *In this case we write*

$$a = \sum_{i=1}^{\infty} a_i.$$

Example 3.1.3. *For each* $x \in \mathbb{R}$ *the infinite sum*

$$\sum_{i=0}^{\infty} \frac{x^k}{k!}$$

converges to e^x *where* e *is the Euler constant* $e \approx 2.71...$ *We call the function* $x \to e^x$ *the* **exponential function**.

Let us now consider sequences in the vectorspace $\mathbb{R}^n$.

Definition 3.1.4. *A map* $g : \mathbb{N} \to \mathbb{R}^n$ *is called a sequence in* $\mathbb{R}^n$. *As before we denote sequences in* $\mathbb{R}^n$ *as* $(x_m)_{m\in\mathbb{N}}$ *where* $x_m := g(m)$. *We denote the components of* x_m *with* x_m^i *for* $i = 1, .., n$ *i.e.* $x_m = (x_m^1, ..., x_m^n)$ *and for each* i *we call the sequences of real numbers* $(x_m^i)_{m\in\mathbb{N}}$ *the* i*-th component sequence of* $(x_m)_{m\in\mathbb{N}}$.

Definition 3.1.5. *A sequence* $(x_m)_{m\in\mathbb{N}}$ *in* $\mathbb{R}^n$ *is called convergent if all its component sequences converge, i.e. for each* $i \in \{1, .., n\}$ *there exists* $x^i \in \mathbb{R}$

such that $\lim_{m\to\infty} x_m^i = x^i$. *The vector* $x = (x^1, ..., x^n)^\top$ *is called the limit of* $(x_m)_{m\in\mathbb{N}}$ *and we write*

$$\lim_{m\to\infty} x_m = x$$

3.2 Continuous Maps

Definition 3.2.1. *A subset* $U \subset \mathbb{R}^n$ *is called* **open** *if for every* $x \in U$ *there exists and* $\epsilon > 0$ *s.t.*

$$U_\epsilon(x) := \{y = (y^1, ..., y^n)^\top | d(x,y) = \sqrt{\sum_{i=1}^n (x^i - y^i)^2} < \epsilon\} \subset U$$

Example 3.2.1.
1. *The open interval* $U = (a,b) = \{x \in \mathbb{R} | a < x < b\}$ *is an open subset of* $\mathbb{R}^n$
2. *The open unit ball* $B = \{x \in \mathbb{R}^n | \sum_{i=1}^n (x^i)^2 < 1\}$ *is an open subset of* $\mathbb{R}^n$.
3. *A set which contains of only one point* $x \in \mathbb{R}^n$ *is not an open set.*
4. *Closed intervals* $[a,b]$ *or half open intervals* $[a,b)$ *or* $(a,b]$ *are not open.*

The two most important concepts in analysis are those of continuity and differentiability. The definition of continuity is next :

Definition 3.2.2. *Let* $U \subset \mathbb{R}^n$ *be open and* $f : U \to \mathbb{R}^k$ *be a map. Then* f *is called* **continuous** *in the point* $x \in U$ *if whenever* $(x_m)_{m\in\mathbb{N}}$ *is a sequence with* $x_m \in U$ *and* $\lim_{m\to\infty} x_m = x$ *one has* $\lim_{m\to\infty} f(x_m) = f(x)$, *i.e. the sequence* $(f(x_m))_{m\in\mathbb{N}}$ *in* $\mathbb{R}^k$ *converges to* $f(x) \in \mathbb{R}^k$. *The map* f *is called continuous in* U *if it is continuous in* x *for all* $x \in U$.

Example 3.2.2.
1. *For each* n *the function* $id : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous.*
2. *Polynomial functions of the form*

$$\begin{aligned} p : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \sum_{j=0}^d a_j x^j \end{aligned}$$

are continuous.

3. *The exponential function*

$$\begin{array}{rcl} exp:\mathbb{R} & \rightarrow & \mathbb{R} \\ x & \rightarrow & exp(x)=e^x \end{array}$$

is continuous.

4. *The function*

$$\begin{array}{rcl} f:\mathbb{R}^2 & \rightarrow & \mathbb{R} \\ (x^1,x^2) & \mapsto & |x^1-x^2| \end{array}$$

is continuous.

5. *The function*

$$\begin{array}{rcl} f:\mathbb{R} & \rightarrow & \mathbb{R} \\ x & \mapsto & \begin{cases} 1 & \text{if } x\geq 0 \\ 0 & \text{if } x<0 \end{cases} \end{array}$$

is continuous in all $x\neq 0$ but not continuous in $x=0$.

6. *The function*

$$\begin{array}{rcl} f:\mathbb{R} & \rightarrow & \mathbb{R} \\ x & \mapsto & \begin{cases} 1 & \text{if } x\in\mathbb{Q} \\ 0 & \text{if } x\in\mathbb{R}\setminus\mathbb{Q} \end{cases} \end{array}$$

One can do a lot of operations with continuous functions which result into new continuous functions. The following discussion will illustrate this.

Proposition 3.2.1. *Let $U\subset\mathbb{R}^n$ be open. Then the set*

$$C(U,\mathbb{R}^k):=\{f:U\rightarrow\mathbb{R}^k | f \text{ is continuous on } U\}$$

is a (real) infinite dimensional vectorspace. In particular if f,g are continuous and $\lambda\in\mathbb{R}$ then also $f+g$ and λf are continuous. Furthermore $f\cdot g$ is continuous and if $g(x)\neq 0$ for all $x\in U$ and $k=1$ then $x\mapsto\frac{f(x)}{g(x)}$ is continuous on U.

Proposition 3.2.2. *Let $U\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^k$ be open and let $f\in C(U,\mathbb{R}^k)$ and $g\in C(V,\mathbb{R}^l)$ and assume $Im(f)\subset V$ then the composition $g\circ f$ is continuous on U and hence $g\circ f\in C(U,\mathbb{R}^l)$.*

At the end of this section we give a characterization of continuity which works completely without sequences and limits.

Proposition 3.2.3. *Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^k$. Then f is continuous on U if and only if for any open subset V of $\mathbb{R}^k$ the preimage $f^{-1}(V)$ is an open subset of $\mathbb{R}^n$.*

Proposition 3.2.4. *Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^k$ and for $i = 1, .., k$ $f^i : U \to \mathbb{R}$ denote the i-th component function of f. Then f is continuous at x if and only if all f^i are continuous at x.*

Proof. This follows directly by the definition of continuity and convergence in $\mathbb{R}^k$ (convergence of the component sequences). □

3.3 The Derivative

In this section we introduce the derivative for maps $f : U \to \mathbb{R}$ where $U \subset \mathbb{R}^n$ is an open subset.

Definition 3.3.1. *Let $f : U \to \mathbb{R}$ be a map defined on an open subset $U \subset \mathbb{R}^n$. Let $x \in U$ be a point and $v \in \mathbb{R}^n$ with $\|v\| = \sqrt{\sum_{i=1}^n (v))^2} = 1$ where $v = (v^1, ..., v^n)^\top$.*

1. *f is called differentiable at x in direction of v if there exists a real number which we denote with $D_v f(x)$ s.t. for any sequence (t_k) s.t. $t_k \in \mathbb{R} \setminus 0$ and $x + t_k \cdot v \in U$ for all k we have*

$$D_v f(x) = \lim_{n\to\infty} \frac{f(x + t_k \cdot v) - f(x)}{t_n}.$$

   The real number $D_v f(x)$ is called the **directional derivative** *of f at x in direction of v.*

2. *If above $v = e_i = (0, .., 0, 1, 0, .., 0)^\top$ is the i-th standard unit vector in $\mathbb{R}^n$ then we denote $D_{e_i} f(x)$ with $D_i f(x)$ or sometimes also with $\frac{\partial f}{\partial x^i}(x)$ and call it the i-th* **partial derivative**.

3. *If all partial derivative $D_i f(x)$ of f at $x \in U$ exist, then we call the vector $\nabla f(x) = (D_1 f(x), ..., D_n f(x))$ the* **gradient** *of f at x and say the gradient of f at x exists.*

Example 3.3.1. *Let $f : U \to \mathbb{R}$ be the constant function, then at all points x all partial derivatives exist and are equal to zero, i.e. $\nabla f(x) = 0$ for all $x \in U$.*

In case the gradient of f exists in all points $x \in U$, then we can consider the function

$$\begin{aligned} \nabla f : U &\rightarrow \mathbb{R}^n \\ x &\mapsto \nabla f(x). \end{aligned}$$

Definition 3.3.2. *Let $f : U \rightarrow \mathbb{R}$ be a map defined on an open subset $U \subset \mathbb{R}^n$ s.t. the gradient exists at any $x \in U$. Then f is called continuously differentiable or C^1, if the map $\nabla f : U \rightarrow \mathbb{R}^n$ is continuous.*

It follows directly from Proposition 3.2.4 that f is C^1 if and only if all the partial derivative functions $x \mapsto \frac{\partial f}{\partial x^i}(x)$ are continuous. We will now consider some examples. We start by assuming $n = 1$ and refer to high school education. In this case we write $\frac{\partial f}{\partial x} = \frac{df}{dx}$ and $D_1 f(x) = \frac{\partial f}{\partial x} = \frac{df}{dx}(x) = f'(x)$.

Example 3.3.2. *1. Polynomials :*

$$\frac{d}{dx}(a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + ... + a_1 \cdot x + a_0) = n \cdot a_n x^{n-1} + ... + 2 \cdot a_2 x^1 + a_1$$

2. Exponential function :

$$\frac{d}{dx} exp(x) = exp(x)$$

3. Trigonometric functions :

$$\frac{d}{dx} sin(x) = cos(x), \frac{d}{dx} cos(x) = -sin(x)$$

4. Hyperbolic functions : For $x \neq 0$, and $i \geq 1$

$$\frac{d}{dx} x^{-i} = -i x^{-(i+1)}$$

The partial derivative with respect to the i-th coordinate of a function $f(x^1, ..., x^n)$ is obtained in the following way : Consider all parameters $x^1, .., x^{i-1}, x^{i+1}, .., x^n$ as constants and consider the function only as a function of the i-the variable. The differentiate in the way you learnt before :

Example 3.3.3. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = (x + y)^2 = x^2 + y^2 + 2xy$. To compute the partial derivative with respect to x, y has to be considered as a constant. Therefore*

$$\frac{\partial f}{\partial x}(x, y) = 2 \cdot x + 0 + 2y$$

To compute the partial derivative with respect to y, x has to be considered as a constant and hence

$$\frac{\partial f}{\partial y}(x, y) = 2 \cdot y + 0 + 2x.$$

Combining both we get $\nabla f(x, y) = 2(x + y)(1, 1)^\top$.

The following two propositions are very helpful tools to compute gradients:

Proposition 3.3.1. (Product-rule) *Let $f, g : U \to \mathbb{R}$ be functions defined on an open subset of $\mathbb{R}^n$, let $x \in U$ and assume the gradients $\nabla f(x)$ and $\nabla g(x)$ exists. Consider the function*

$$\begin{aligned} f \cdot g : U &\to \mathbb{R} \\ x &\mapsto f(x) \cdot g(x). \end{aligned}$$

Then $\nabla(f \cdot g)(x)$ exists and

$$\nabla(f \cdot g)(x) = f(x) \cdot \nabla g(x) + g(x) \cdot \nabla f(x).$$

Proposition 3.3.2. (Chain-rule) *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}$ be open, and $f : U \to \mathbb{R}$ and $g : V \to \mathbb{R}$ be functions s.t. $im(f) \subset V$. Assume the gradient of f at $x \in U$ exists and the derivative (one dimensional gradient) of g at $f(x) \in V$ exists. Then the gradient of the function*

$$\begin{aligned} g \circ f : U &\to \mathbb{R} \\ x &\mapsto g(f(x)) \end{aligned}$$

at x exists

$$\nabla(g \circ f)(x) = g'(f(x)) \cdot \nabla f(x).$$

Example 3.3.4. *We reconsider the function f from Example 3.3. We let $g(x, y) = x+y$ and $h(z) = z^2$ and get $h \circ g(x, y) = (x+y)^2 = f(x, y)$. Clearly $\nabla g(x, y) = (1, 1)^\top$. Applying the previous proposition we get*

$$\nabla f(x, y) = \nabla(h \circ g)(x, y) = h'(g(x, y)\nabla g(x, y) = 2 \cdot (x + y) \cdot (1, 1)^\top$$

which of course coincides with the result obtained in Example 3.3.

Chapter 4

Linear Optimization

Chapter 5

Integration Theory

Chapter 6

Probability Theory