

# Linear Algebra I - せんけいだいすうがく 線形代数学 I

## Overview notes

G30 Program, Nagoya University (Fall 2021)

Henrik Bachmann (Math. Building Room 457, [henrik.bachmann@math.nagoya-u.ac.jp](mailto:henrik.bachmann@math.nagoya-u.ac.jp))

Lecture notes and exercises are available at: [https://www.henrikbachmann.com/la1\\_2021.html](https://www.henrikbachmann.com/la1_2021.html)

---

These notes serve as a compact overview of the definitions, propositions, lemmas, corollaries, and theorems given in the lectures. It will be updated regularly (This is Version 3 from January 18, 2022). The proofs, examples, and explanations are provided in the handwritten notes/lectures. The reference book for this course is [B], and we will probably cover Chapters 1,2,3 and 5 during this semester. (Chapter 4 will be part of Linear Algebra II)

If you find any typos in this note, please let me know!

## Contents

<b>1</b>	<b>Linear systems</b>	<b>2</b>
<b>2</b>	<b>Matrices &amp; Vectors</b>	<b>2</b>
<b>3</b>	<b>Sets &amp; Functions</b>	<b>4</b>
<b>4</b>	<b>Linear maps</b>	<b>5</b>
<b>5</b>	<b>Linear maps in geometry</b>	<b>5</b>
<b>6</b>	<b>Composition of linear maps &amp; Matrix multiplication</b>	<b>6</b>
<b>7</b>	<b>The inverse of a linear map</b>	<b>7</b>
<b>8</b>	<b>Subspaces, Kernel &amp; Image</b>	<b>8</b>
<b>9</b>	<b>Linear independence</b>	<b>9</b>
<b>10</b>	<b>Bases &amp; dimensions</b>	<b>10</b>
<b>11</b>	<b>Coordinates</b>	<b>10</b>
<b>12</b>	<b>Orthonormal bases &amp; Gram-Schmidt algorithm</b>	<b>12</b>
<b>13</b>	<b>Orthogonal Projection &amp; Least squares</b>	<b>13</b>

## References

[B] O. Bretscher: *Linear Algebra with Applications*, 4th edition, Pearson 2009.

## 1 Linear systems

By  $\mathbb{R}$  we will denote the set of real numbers.  $\mathbb{R}$  contains all numbers you usually considered in high school, such as  $1, -1, 0, 2, 3, \frac{3}{8}, \pi, \sqrt{2}, e, \dots$ . There are rigorous definitions of the real numbers, which would be part of a pure mathematics lecture<sup>1</sup>. But in this course, we just assume that they exist and that everyone is familiar with them.

**Definition 1.1.** *i) For real numbers  $a_1, a_2, \dots, a_n, b \in \mathbb{R}$  an equation of the form*

$$a_1x_1 + \dots + a_nx_n = b$$

*is called a **linear equation**.*

*ii) A finite collection of linear equations is called a **linear system**.*

*iii) A **solution of a linear system** is a simultaneous solution for all of its equations.*

**Definition 1.2.** *The following operations on a linear system are called **elementary row operations**.*

*(R1) Add a multiple of an equation to another.*

*(R2) Multiply an equation with a non-zero number.*

*(R3) Change the order of the equations.*

**Proposition 1.3.** *Applying an elementary row operation to a linear system does not change the set of all solutions of this linear system.*

## 2 Matrices & Vectors

**Definition 2.1.** *i) A  $m \times n$ -**matrix** is given by an array ( $m$  rows,  $n$  columns) of numbers  $a_{ij} \in \mathbb{R}$*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

*Notation: We often just write  $A = (a_{ij})$  if the size of  $A$ , i.e.  $m$  and  $n$ , are known from context. By  $\mathbb{R}^{m \times n}$  we denote the set all of all  $m \times n$ -matrices.*

*ii) A (column-) **vector** of size  $n$  is a  $n \times 1$ -matrix*

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

*and the set of all vectors of size  $n$  is denoted by  $\mathbb{R}^n = \mathbb{R}^{n \times 1}$ .*

---

<sup>1</sup>See for example [https://en.wikipedia.org/wiki/Construction\\_of\\_the\\_real\\_numbers](https://en.wikipedia.org/wiki/Construction_of_the_real_numbers) for an overview of the "construction" of real numbers.

**Definition 2.2.** For matrices  $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n}$  and a real number  $\lambda \in \mathbb{R}$  we define

$$\begin{aligned} A + B &= (a_{ij} + b_{ij}) \in \mathbb{R}^{m \times n} && \text{(Sum of two matrices),} \\ \lambda A &= (\lambda a_{ij}) \in \mathbb{R}^{m \times n} && \text{(Scalar multiplication).} \end{aligned}$$

In the case  $\lambda = -1$  we write  $(-1)A = -A$  and  $A - B$  means  $A + (-1)B$ .

The matrices  $A$  and  $B$  need to be of the same size, otherwise the sum  $A + B$  is not defined. A special case of the addition of matrices is given by the addition of vectors. For  $u, v \in \mathbb{R}^n$  and  $\lambda \in \mathbb{R}$  we have

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad u + v = \begin{pmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{pmatrix}, \quad \lambda v = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix}.$$

**Definition 2.3.** The product of a matrix  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  and a vector  $v \in \mathbb{R}^n$  is defined by

$$Av = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \dots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \dots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \dots + a_{mn}v_n \end{pmatrix} \in \mathbb{R}^m.$$

We have:  $(m \times n\text{-matrix}) \cdot (\text{vector of size } n) = (\text{vector of size } m)$ .

**Proposition 2.4.** We have for  $A \in \mathbb{R}^{m \times n}$ ,  $x, y \in \mathbb{R}^n$  and  $\lambda \in \mathbb{R}$

- i)  $A(x + y) = Ax + Ay$ ,
- ii)  $A(\lambda x) = \lambda(Ax)$ .

**Definition 2.5.** For a matrix  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  and a vector  $b \in \mathbb{R}^m$  the matrix

$$(A | b) = \left( \begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array} \right) \in \mathbb{R}^{m \times (n+1)}$$

is called the **augmented matrix** of the linear system  $Ax = b$ .

The augmented matrix  $(A | b)$  is just the matrix  $A$  where we append the vector  $b$  as a column. The line  $|$  is a useful notation to distinguish between the left- and right-hand side of the corresponding linear system but it has no mathematical meaning. We will view  $(A | b)$  as a usual matrix with  $m$  rows and  $n + 1$  columns.

**Definition 2.6.** The following operations on a matrix are called **elementary row operations**.

- (R1) Add a multiple of one row to another row.
- (R2) Multiply a row with a non-zero number.

(R3) *Interchange two rows.*

Applying a row operation to a linear system (Definition 1.2) corresponds to the same row operation (Definition 2.6) on the corresponding augmented matrix of this linear system.

**Definition 2.7.** *Two matrices  $A$  and  $B$  are called **row equivalent**, if  $B$  can be obtained from  $A$  by elementary row operations. In this case we write*

$$A \sim B.$$

Notice that if  $A \sim B$ , then also  $B \sim A$ , i.e.  $A$  can be obtained from  $B$  by elementary row operations.

**Proposition 2.8.** *Let  $A, B \in \mathbb{R}^{m \times n}$  and  $b, c \in \mathbb{R}^m$ . If  $(A | b) \sim (B | c)$  then the linear systems  $Ax = b$  and  $Bx = c$  have the same solutions.*

**Definition 2.9.** *A matrix  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  is on **row-reduced echelon form** if*

- i) The first non-zero element on each row (if any) is equal to 1.*
- ii) If there is a leading 1 in a row, then all rows above contain a leading 1 further to the left.*
- iii) If  $a_{ij}$  is the first non-zero element in row  $i$ , then there are no other non-zero elements in the  $j$ -th column.*

*The first non-zero element in a row of a matrix in row-reduced echelon form is called **pivot element**.*

**Theorem 2.10.** *Every matrix  $A$  is row equivalent to a unique matrix  $B$  on row-reduced echelon form and we write*

$$B = \text{rref}(A).$$

**Definition 2.11.** *Let  $A \in \mathbb{R}^{m \times n}$  be a matrix. The **rank**  $\text{rk}(A)$  of  $A$  is the number of pivot elements in  $\text{rref}(A)$ .*

### 3 Sets & Functions

**Definition 3.1.** *Let  $X$  and  $Y$  be two sets.*

- i) A **function**  $f : X \rightarrow Y$  is a rule, assigning to each element  $x \in X$  an element  $f(x) \in Y$ . This is also denoted by*

$$\begin{aligned} f : X &\longrightarrow Y \\ x &\longmapsto f(x). \end{aligned}$$

- ii) For  $f : X \rightarrow Y$  the set  $X$  is called the **domain of  $f$**  and  $Y$  is called the **codomain of  $f$** .*

A function is also sometimes called a **map**. These two names (for the exact same mathematical object) are used interchangeably in the literature.

**Definition 3.2.** For a function  $f : X \rightarrow Y$  the **image of  $f$**  is defined by

$$\text{im}(f) = \{y \in Y \mid \exists x \in X : y = f(x)\}.$$

Another notation is  $\text{im}(f) = f(X)$ . The image is a subset of  $Y$ , i.e.  $\text{im}(f) \subset Y$ .

**Definition 3.3.** A function  $f : X \rightarrow Y$  is called

i) **injective** if  $f(x_1) \neq f(x_2)$  whenever  $x_1 \neq x_2$ . ( $x_1, x_2 \in X$ )

ii) **surjective** if  $\text{im}(f) = Y$ .

iii) **bijective** if it is injective and surjective.

## 4 Linear maps

**Definition 4.1.** A function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a **linear map** if for all  $u, v \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$  we have

i)  $F(u + v) = F(u) + F(v)$ ,

ii)  $F(\lambda u) = \lambda F(u)$ .

**Theorem 4.2.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map. Then there exists a unique matrix  $[F] \in \mathbb{R}^{m \times n}$ , such that for all  $x \in \mathbb{R}^n$  we have

$$F(x) = [F]x.$$

Here the left-hand side is the evaluation of the function  $F$  at  $x$  and the right-hand side is the multiplication of the matrix  $[F]$  with the vector  $x$ .

**Definition 4.3.** The matrix  $[F]$  in Theorem 4.2 is called **the matrix of  $F$** .

## 5 Linear maps in geometry

**Definition 5.1.** Let  $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ .

i) The **dot product** of  $u$  and  $v$  is defined by

$$u \bullet v = u_1 v_1 + \cdots + u_n v_n.$$

ii)  $u$  and  $v$  are called **orthogonal** if  $u \bullet v = 0$ .

iii) The **norm** of  $u$  is defined by

$$\|u\| = \sqrt{u \bullet u} = \sqrt{u_1^2 + \cdots + u_n^2}.$$

**Proposition 5.2.** *The dot product satisfies the following properties for all  $u, v, w \in \mathbb{R}^n$  and  $\lambda \in \mathbb{R}$ :*

- i)  $u \bullet v = v \bullet u$ ,
- ii)  $u \bullet (v + w) = u \bullet v + u \bullet w$ ,
- iii)  $u \bullet (\lambda v) = \lambda(u \bullet v)$ .

**Definition 5.3.** *Let  $u \in \mathbb{R}^n$  with  $u \neq 0$ . We define the **orthogonal projection**  $P_u$  onto the line spanned by  $u$  as*

$$P_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$x \longmapsto \frac{u \bullet x}{u \bullet u} u.$$

**Proposition 5.4.**  $P_u : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a linear map.

**Definition 5.5.** *Let  $u \in \mathbb{R}^n$  with  $u \neq 0$ . We define the **reflection**  $\rho_u$  along the line spanned by  $u$  as*

$$\rho_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$x \longmapsto 2 \frac{u \bullet x}{u \bullet u} u - x.$$

**Proposition 5.6.**  $\rho_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$  is a linear map.

**Definition 5.7.** *For  $\varphi \in \mathbb{R}$  the counterclockwise **rotation by an angle**  $\varphi$  (in  $\mathbb{R}^2$ ) is given by*

$$\text{rot}_\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$x \longmapsto \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} x.$$

## 6 Composition of linear maps & Matrix multiplication

**Theorem 6.1.** *If  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $G : \mathbb{R}^m \rightarrow \mathbb{R}^l$  are linear, then  $GF : \mathbb{R}^n \rightarrow \mathbb{R}^l$  is linear.*

**Definition 6.2.** *Let  $A \in \mathbb{R}^{l \times m}$  and  $B \in \mathbb{R}^{m \times n}$ , where  $B$  has the columns  $v_1, \dots, v_n \in \mathbb{R}^m$ , i.e.*

$$B = \left( \begin{array}{c|ccc|c} & & & & \\ & v_1 & \dots & v_n & \\ & | & & | & \end{array} \right).$$

*Then the **product of  $A$  and  $B$**  is the  $l \times n$ -matrix with columns  $Av_1, \dots, Av_n \in \mathbb{R}^l$ , i.e.*

$$A \cdot B = \left( \begin{array}{c|ccc|c} & & & & \\ & Av_1 & \dots & Av_n & \\ & | & & | & \end{array} \right) \in \mathbb{R}^{l \times n}.$$

**Theorem 6.3.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $G : \mathbb{R}^m \rightarrow \mathbb{R}^l$  be linear maps. Then the matrix of  $GF$  is given by the product of the matrices of  $G$  and  $F$ , i.e.

$$[GF] = [G] \cdot [F].$$

**Proposition 6.4.** For all  $A \in \mathbb{R}^{l \times m}$ ,  $B, D \in \mathbb{R}^{m \times n}$ ,  $C \in \mathbb{R}^{n \times p}$  and  $\lambda \in \mathbb{R}$  we have

- i)  $A \cdot I_m = I_l \cdot A = A$ , where  $I_m$  denotes the  $m \times m$ -identity matrix.
- ii)  $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ .
- iii)  $A \cdot (B + D) = A \cdot B + A \cdot D$ .
- iv)  $(B + D) \cdot C = B \cdot C + D \cdot C$ .
- v)  $\lambda(A \cdot B) = (\lambda A) \cdot B = A \cdot (\lambda B)$ .

## 7 The inverse of a linear map

The rank of a linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by the rank of its matrix, i.e.  $\text{rk}(F) := \text{rk}([F])$ .

**Theorem 7.1.** A linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is invertible if and only if  $m = n = \text{rk}(F)$ .

**Proposition 7.2.** If a linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is invertible, then its inverse  $F^{-1}$  is also linear.

**Definition 7.3.** If  $A \in \mathbb{R}^{n \times n}$  is the matrix of an invertible linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  (i.e.  $A = [F]$ ), then we define the **inverse of  $A$**  by  $A^{-1} := [F^{-1}]$ .

**Theorem 7.4.** The inverse of  $A \in \mathbb{R}^{n \times n}$  exists ( $A$  is invertible) if and only if  $\text{rref}(A) = I_n$ .

**Proposition 7.5.** If  $A, B \in \mathbb{R}^{n \times n}$  are invertible we have

- i)  $AA^{-1} = A^{-1}A = I_n$ ,
- ii)  $(BA)^{-1} = A^{-1}B^{-1}$ .

(Definition 7.6 and Theorem 7.7 are just a remark and they are not so important for the rest of this course. They will appear again in detail in Linear Algebra II)

**Definition 7.6.** For  $\lambda \in \mathbb{R}$  with  $\lambda \neq 0$  and  $1 \leq i, j \leq n$  we define the **elementary matrices**  $R_i^{\lambda, j}, R_i^\lambda, R_{i, j} \in \mathbb{R}^{n \times n}$  by

$$R_i^{\lambda, j} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & \lambda & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}, \quad R_i^\lambda = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \lambda & \\ & & & & \ddots \\ & & & & & & 1 \end{pmatrix}, \quad R_{i, j} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & & 1 \\ & & & \ddots & \\ & & 1 & & 0 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}.$$

Here the  $\lambda$  in  $R_i^{\lambda, j}$  is in the  $i$ -th row and  $j$ -th column, in  $R_i^\lambda$  it is in the  $i$ -th row, and in  $R_{i, j}$  the 0 are on the diagonal in the  $i$ -th row and  $j$ -th column.

Multiplying with an elementary matrix from the left corresponds to the elementary row operations (Definition 2.6)

(R1) Multiplying with  $R_i^{\lambda,j}$ : Add  $\lambda$ -times row  $j$  to row  $i$ .

(R2) Multiplying with  $R_i^\lambda$ : Multiply row  $j$  by  $\lambda$ . ( $\lambda \neq 0$ )

(R3) Multiplying with  $R_{i,j}$ : Change row  $i$  and  $j$ .

**Theorem 7.7.** *Every invertible matrix is a product of elementary matrices.*

## 8 Subspaces, Kernel & Image

**Definition 8.1.** *A subset  $U \subset \mathbb{R}^n$  is a subspace of  $\mathbb{R}^n$  if*

- i)  $0 \in U$ ,*
- ii) for all  $u, v \in U$  we have  $u + v \in U$ ,*
- iii) for all  $u \in U$  and  $\lambda \in \mathbb{R}$  we have  $\lambda u \in U$ .*

**Definition 8.2.** *For a linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  the kernel of  $F$  is defined by*

$$\ker(F) = \{x \in \mathbb{R}^n \mid F(x) = 0\}.$$

**Proposition 8.3.** *For any linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  we have the following:*

- i) The kernel  $\ker(F)$  is a subspace of  $\mathbb{R}^n$ .*
- ii) The image  $\text{im}(F)$  is a subspace of  $\mathbb{R}^m$ .*

**Definition 8.4.** *i) A linear combination of vectors  $v_1, \dots, v_n \in \mathbb{R}^m$  is a vector of the form*

$$u = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n \in \mathbb{R}^m$$

*for some numbers  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ .*

*ii) The span of  $v_1, \dots, v_n \in \mathbb{R}^m$  is the set of all linear combinations*

$$\text{span}\{v_1, \dots, v_n\} = \{\lambda_1 v_1 + \dots + \lambda_n v_n \in \mathbb{R}^m \mid \lambda_1, \dots, \lambda_n \in \mathbb{R}\}.$$

**Proposition 8.5.** *For  $v_1, \dots, v_n \in \mathbb{R}^m$  we have the following.*

- i)  $\text{span}\{v_1, \dots, v_n\}$  is a subspace of  $\mathbb{R}^m$ .*
- ii) If  $U \subset \mathbb{R}^m$  is a subspace and  $v_1, \dots, v_n \in U$  then  $\text{span}\{v_1, \dots, v_n\} \subset U$ .*

**Proposition 8.6.** *A linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is injective if and only if  $\ker(F) = \{0\}$ .*



**Theorem 8.7.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map.

i) We have the following equivalent statements for  $F$  being injective:

$$F \text{ is injective} \iff \ker(F) = \{0\} \iff \text{rk}([F]) = n.$$

ii) We have the following equivalent statements for  $F$  being surjective:

$$F \text{ is surjective} \iff \text{im}(F) = \mathbb{R}^m \iff \text{rk}([F]) = m.$$

iii) If  $m = n$  then the following statements are equivalent:

$$F \text{ is bijective} \iff F \text{ is injective} \iff F \text{ is surjective}.$$

## 9 Linear independence

**Lemma 9.1.** Let  $v_1, \dots, v_l \in \mathbb{R}^m$ . If  $v_l \in \text{span}\{v_1, \dots, v_{l-1}\}$  then

$$\text{span}\{v_1, \dots, v_l\} = \text{span}\{v_1, \dots, v_{l-1}\}.$$

**Definition 9.2.** i) Vectors  $v_1, \dots, v_l \in \mathbb{R}^m$  are called **linearly independent** if the equation

$$\lambda_1 v_1 + \dots + \lambda_l v_l = 0 \tag{9.1}$$

with  $\lambda_1, \dots, \lambda_l \in \mathbb{R}$  just has the unique solution  $\lambda_1 = \dots = \lambda_l = 0$ .

ii) If there exist another solution of (9.1), i.e. where at least for one  $j = 1, \dots, l$  we have  $\lambda_j \neq 0$ , then the vectors  $v_1, \dots, v_l$  are called **linearly dependent**.

**Theorem 9.3.** Let  $v_1, \dots, v_l \in \mathbb{R}^m$ . The following statements are equivalent:

i)  $v_1, \dots, v_l$  are linearly dependent.

ii) There exists a  $j = 1, \dots, l$ , such that  $v_j$  is a linear combination of the other vectors.

iii) There exists a  $j = 1, \dots, l$  with

$$\text{span}\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_l\} = \text{span}\{v_1, \dots, v_l\}.$$

**Lemma 9.4.** Let  $V \subset \mathbb{R}^n$  be a subspace,  $v_1, \dots, v_l \in V$  linearly independent and  $V = \text{span}\{w_1, \dots, w_m\}$  for some  $w_1, \dots, w_m \in \mathbb{R}^n$ . Then we have  $l \leq m$ .

**Lemma 9.5.** If  $v_1, \dots, v_l \in \mathbb{R}^n$  are linearly independent and  $w \in \mathbb{R}^n$  with  $w \notin \text{span}\{v_1, \dots, v_l\}$  then  $v_1, \dots, v_l, w$  are linearly independent.

## 10 Bases & dimensions

**Definition 10.1.** Let  $V \subset \mathbb{R}^n$  be a subspace. Vectors  $v_1, \dots, v_l \in V$  form a **basis of  $V$**  if

- i)  $V = \text{span}\{v_1, \dots, v_l\}$ ,
- ii)  $v_1, \dots, v_l$  are linearly independent.

In this case we also say that  $\{v_1, \dots, v_l\}$  is a basis of  $V$ .

Later we will also be interested in the order of the  $v_j$  and write a basis as a tuple  $(v_1, \dots, v_l)$ .

**Theorem 10.2.** For any subspace  $V \subset \mathbb{R}^n$  we have the following:

- i)  $V$  has a basis.
- ii) All bases of  $V$  have the same number of elements.
- iii) If  $v_1, \dots, v_l \in V$  are linearly independent then there exist  $u_{l+1}, \dots, u_t \in V$ , such that  $\{v_1, \dots, v_l, u_{l+1}, \dots, u_t\}$  is a basis of  $V$ .
- iv) If  $V = \text{span}\{w_1, \dots, w_m\}$  then there exists a subset  $\{u_1, \dots, u_t\} \subset \{w_1, \dots, w_m\}$  such that  $\{u_1, \dots, u_t\}$  is a basis of  $V$ .

**Definition 10.3.** Let  $V \subset \mathbb{R}^n$  be a subspace. The **dimension of  $V$** , denoted by  $\dim(V)$ , is the number of elements in a basis of  $V$ .

**Corollary 10.4.** Let  $V \subset \mathbb{R}^n$  be a subspace with  $\dim(V) = m$  and  $v_1, \dots, v_m \in V$ . Then the following statements are equivalent:

- i)  $v_1, \dots, v_m$  are linearly independent.
- ii)  $V = \text{span}\{v_1, \dots, v_m\}$ .
- iii)  $\{v_1, \dots, v_m\}$  is a basis of  $V$ .

**Theorem 10.5.** For a linear map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  we have

$$n = \dim(\ker(F)) + \dim(\text{im}(F)).$$

## 11 Coordinates

From now on we will consider ordered bases, which means that we will write  $(b_1, \dots, b_n)$  (a tuple) for a basis instead of  $\{b_1, \dots, b_n\}$  (a set). The difference is, that we care about the order now. For example, the two sets  $\{b_1, b_2\} = \{b_2, b_1\}$  are the same, but  $(b_1, b_2) \neq (b_2, b_1)$ .

**Definition 11.1.** Let  $B = (b_1, \dots, b_m)$  be a basis of a subspace  $V \subset \mathbb{R}^n$ . We define the **coordinate map** by

$$c_B : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} \longmapsto \lambda_1 b_1 + \dots + \lambda_m b_m.$$

**Theorem 11.2.** Let  $B = (b_1, \dots, b_m)$  be a basis of a subspace  $V \subset \mathbb{R}^n$ .

- i) The coordinate map  $c_B : \mathbb{R}^m \rightarrow V$  is bijective.
- ii) For all  $x \in V$  there exist unique  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  such that

$$x = \lambda_1 b_1 + \dots + \lambda_m b_m.$$

**Definition 11.3.** Let  $B = (b_1, \dots, b_m)$  be a basis of a subspace  $V \subset \mathbb{R}^n$  and  $x \in V$  with

$$x = \lambda_1 b_1 + \dots + \lambda_m b_m.$$

- i) The numbers  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  are the **coordinates of  $x$  (in the basis  $B$ )**.
- ii) The **coordinate vector** of  $x$  (with respect to  $B$ ) is given by

$$[x]_B = c_B^{-1}(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}.$$

**Proposition 11.4.** Let  $B = (b_1, \dots, b_m)$  be a basis of a subspace  $V \subset \mathbb{R}^n$ ,  $x, y \in V$  and  $\mu \in \mathbb{R}$ . Then

- i)  $[x + y]_B = [x]_B + [y]_B$ ,
- ii)  $[\mu x]_B = \mu[x]_B$ ,
- iii)  $[0]_B = 0$ .

**Definition 11.5.** Let  $B = (b_1, \dots, b_n)$  be a basis of  $\mathbb{R}^n$ . The **change-of-basis matrix** associated with  $B$  is

$$S_B = [c_B] = \begin{pmatrix} | & & | \\ b_1 & \dots & b_n \\ | & & | \end{pmatrix}.$$

**Definition 11.6.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map,  $B_1$  be a basis of  $\mathbb{R}^n$  and  $B_2$  be a basis of  $\mathbb{R}^m$ . The **matrix of  $F$  with respect to  $B_1$  and  $B_2$**  is the matrix

$$[F]_{B_1}^{B_2} := [c_{B_2}^{-1} \circ F \circ c_{B_1}].$$

In the case  $n = m$  and  $B_1 = B_2$  we just write  $[F]_{B_1} := [F]_{B_1}^{B_1}$ .

**Proposition 11.7.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map,  $B_1$  be a basis of  $\mathbb{R}^n$  and  $B_2$  be a basis of  $\mathbb{R}^m$ .

- i) We have

$$[F]_{B_1}^{B_2} = S_{B_2}^{-1} [F] S_{B_1}.$$

- ii) If  $B_1 = (b_1, \dots, b_n)$  then

$$[F]_{B_1}^{B_2} = \begin{pmatrix} | & & | \\ [F(b_1)]_{B_2} & \dots & [F(b_n)]_{B_2} \\ | & & | \end{pmatrix}.$$

## 12 Orthonormal bases & Gram-Schmidt algorithm

**Definition 12.1.** i) A vector  $u \in \mathbb{R}^n$  is called a **unit vector** if  $\|u\| = 1$ . (i.e.  $u \bullet u = 1$ )

ii) Every vector  $u \in \mathbb{R}^n$  with  $u \neq 0$  can be normalized by

$$\hat{u} = \frac{1}{\|u\|}u.$$

The vector  $\hat{u}$  is a unit vector and shows in the same direction as  $u$ .

iii) Vectors  $u_1, \dots, u_l \in \mathbb{R}^n$  are called **orthonormal** if for  $1 \leq i, j \leq l$

$$u_i \bullet u_j = \begin{cases} 1 & , \text{if } i = j \\ 0 & , \text{if } i \neq j \end{cases}.$$

**Definition 12.2.** A basis  $B = (b_1, \dots, b_m)$  of a subspace  $U$  is called an **orthonormal basis (ONB)** of  $U$  if  $b_1, \dots, b_m$  are orthonormal.

**Proposition 12.3.** i) If  $v_1, \dots, v_m \in \mathbb{R}^n$  are orthonormal (ON), then they are linearly independent.

ii) Let  $B = (v_1, \dots, v_m)$  be an ONB of  $V \subset \mathbb{R}^n$  and  $u \in V$ . Then

$$[u]_B = \begin{pmatrix} u \bullet v_1 \\ \vdots \\ u \bullet v_m \end{pmatrix} \in \mathbb{R}^m,$$

i.e.  $u = \sum_{i=1}^m (u \bullet v_i)v_i$ .

iii) If  $B = (v_1, \dots, v_m)$  is an ONB of  $V \subset \mathbb{R}^n$  and  $u, w \in V$ , then

$$u \bullet w = [u]_B \bullet [w]_B.$$

**Definition 12.4.** For a subspace  $U \subset \mathbb{R}^n$  we define the **orthogonal complement of  $U$  in  $\mathbb{R}^n$**  by

$$U^\perp = \{x \in \mathbb{R}^n \mid x \bullet u = 0 \text{ for all } u \in U\}.$$

**Lemma 12.5.** Let  $U \subset \mathbb{R}^n$  be a subspace.

i)  $U^\perp \subset \mathbb{R}^n$  is a subspace.

ii) We have  $U \cap U^\perp = \{0\}$ .

iii) If  $(u_1, \dots, u_r)$  is a basis of  $U$ ,  $x \in \mathbb{R}^n$ , then

$$x \in U^\perp \iff x \bullet u_1 = \dots = x \bullet u_r = 0.$$

*iv) Let  $(f_1, \dots, f_r)$  be an ONB of  $U$  and  $x \in \mathbb{R}^n$ . Then*

$$x = x_{\parallel} + x_{\perp},$$

*where*

$$x_{\parallel} = \sum_{i=1}^r (x \bullet f_i) f_i \in U$$

$$x_{\perp} = x - x_{\parallel} \in U^{\perp}.$$

### Gram-Schmidt algorithm (GSA)

Let  $B = (b_1, \dots, b_m)$  be an arbitrary basis of a subspace  $U \subset \mathbb{R}^n$ . The GSA constructs an orthonormal basis  $F = (f_1, \dots, f_m)$  of  $U$  out of the basis  $B$  in the following way  $m$  steps:

**Step 1:** Set  $f_1 = \hat{b}_1 = \frac{1}{\|b_1\|} b_1$ .

**Step  $l$  ( $2 \leq l \leq m$ ):** We have constructed orthonormal vectors  $f_1, \dots, f_{l-1}$  in the steps before. Now set

$$w_l = b_l - (b_l \bullet f_1) f_1 - \dots - (b_l \bullet f_{l-1}) f_{l-1} = b_l - \sum_{i=1}^{l-1} (b_l \bullet f_i) f_i$$

and define  $f_l = \frac{1}{\|w_l\|} w_l$ .

**Theorem 12.6.** *Every subspace of  $\mathbb{R}^n$  has an ONB.*

**Corollary 12.7.** *Let  $U \subset \mathbb{R}^n$  be a subspace. For all  $x \in \mathbb{R}^n$  there exist unique  $x_{\parallel} \in U$  and  $x_{\perp} \in U^{\perp}$  with*

$$x = x_{\parallel} + x_{\perp}.$$

## 13 Orthogonal Projection & Least squares

**Definition 13.1.** *Let  $U \subset \mathbb{R}^n$  be a subspace. The map*

$$P_U : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$x \longmapsto x_{\parallel}$$

*is the orthogonal projection onto  $U$ .*

**Proposition 13.2.** *Let  $U \subset \mathbb{R}^n$  be a subspace.*

*i)  $P_U$  is a linear map.*

ii)  $P_U^2 = P_U$ .

iii)  $\ker(P_U) = U^\perp$  and  $\text{im } P_U = U$ .

iv) If  $(f_1, \dots, f_m)$  is an ONB of  $U$ , then

$$P_U(x) = (x \bullet f_1)f_1 + \dots + (x \bullet f_m)f_m.$$

**Proposition 13.3.** Let  $U \subset \mathbb{R}^n$  be a subspace and  $x \in \mathbb{R}^n$ . Then for all  $u \in U$  we have

$$\|x - P_U(x)\| \leq \|x - u\|.$$

We just have equality in the case when  $u = P_U(x)$ .

**Definition 13.4.** The **transpose** of a matrix  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  is the matrix  $A^T = (a_{ji}) \in \mathbb{R}^{n \times m}$ .

**Proposition 13.5.** i) For  $A, B \in \mathbb{R}^{m \times n}$  and  $\lambda \in \mathbb{R}$  we have

$$(A + B)^T = A^T + B^T, \quad (\lambda A)^T = \lambda A^T.$$

ii) For  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times l}$  we have

$$(AB)^T = B^T A^T \in \mathbb{R}^{l \times m}.$$

iii) For  $x, y \in \mathbb{R}^n$  we have  $x \bullet y = x^T y$ .

For  $A \in \mathbb{R}^{m \times n}$  we can define the linear map  $F : x \mapsto Ax$ . With this we define the image and kernel of the matrix  $A$  by  $\text{im}(A) = \text{im}(F)$  and  $\ker(A) = \ker(F)$ .

**Proposition 13.6.** For all  $A \in \mathbb{R}^{m \times n}$  we have  $\text{im}(A)^\perp = \ker(A^T)$ .

**Corollary 13.7.** Let  $A \in \mathbb{R}^{m \times n}$ .

i) We have  $\ker(A^T A) = \ker(A)$ .

ii) We have the following equivalence

$$\ker(A) = \{0\} \iff A^T A \in \mathbb{R}^{n \times n} \text{ is invertible.}$$