Notes for the course

# GEOMETRY (055715)

Davide Frapporti

Academic year 2024-2025

# Contents

# Introduction

In this preliminary chapter, we start with an informal discussion on free and applied vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, which you have probably already met in the high-school.

We will be sloppy and not completely precise: the intention is to give the gist of some concepts that will be extended and generalized during the course.

## 0.1 Geometric vectors in the plane and in the space

**Definition 0.1.** An *applied vector* is a segment with a given verse, so it is determined by an initial point $A$ and an end point $B$. It will be denoted by $\overrightarrow{AB}$.

An applied vector $\overrightarrow{AB}$ describe an absolute movement form point $A$ to point $B$.

In physics applied vectors are widely used, e.g. a force applied to body or the speed of an object, and so on.

*Remark.* An applied vector $\overrightarrow{AB}$ is determined by 4 data:

1. the application point $A$;
2. the direction (of the line passing through $A$ and $B$);
3. the verse (following which we move along the line, from $A$ towards $B$);
4. the magnitude (length of the segment with end-points $A$ and $B$).

*Remark.* Note that $\overrightarrow{BA}$ has the same direction and magnitude of $\overrightarrow{AB}$ but opposite verse!

If we get rid of the application point, we have a relative movement, which is described by a free vector.

**Definition 0.2.** A *free vector* $\overrightarrow{v}$ is determined by direction, verse, and magnitude (or length).

Each applied vector $\overrightarrow{AB}$ induces a free vector, which we denote by $[\overrightarrow{AB}]$. On the other hand, note that a free vector corresponds to infinitely many applied vectors, one for each possible application point.

### Operations with free vectors

We would like to perform operations with free vectors, and so "zero" is of crucial importance.

**Definition 0.3.** The *zero vector* $\overrightarrow{0}$ is the only vector of null magnitude. It does not have direction or verse. It is the free vector associated to $\overrightarrow{AA}$.

The first operation we consider is the multiplication of a free vector by a real number $c \in \mathbb{R}$, which we call *scalar*.

**Product with a scalar.** Let $\overrightarrow{v}$ be a free vector and let $c \in \mathbb{R}$.
If $c = 0$, then $c \cdot \overrightarrow{v} = \overrightarrow{0}$ .
If $c \neq 0$, then $c \cdot \overrightarrow{v}$ is the free vector having

(i) the same direction of $\overrightarrow{v}$;

(ii) the same verse of $\overrightarrow{v}$, if $c > 0$, and opposite verse if $c < 0$;

(iii) length equal to the length of $\overrightarrow{v}$ multiplied by $|c|$.

Note that $c \cdot \overrightarrow{v}$ is obtained by contracting/dilating and possibly flipping the vector $\overrightarrow{v}$.

**Sum of vectors.** The sum of two free vectors is determined with the parallelogram rule. Given free vectors $\overrightarrow{v}$ and $\overrightarrow{w}$, their sum $\overrightarrow{v} + \overrightarrow{w}$ is obtained in the following way: apply $\overrightarrow{v}$ at the point $O$ and then apply $\overrightarrow{w}$ in the head of $\overrightarrow{v}$. If the head of $\overrightarrow{w}$ is in $Q$, then $\overrightarrow{v} + \overrightarrow{w} = [\overrightarrow{OQ}]$.

Alternatively, apply $\overrightarrow{v}$ and $\overrightarrow{w}$ at the same point $O$ and consider the parallelogram having $\overrightarrow{v}$ and $\overrightarrow{w}$ as two consecutive sides. Then $\overrightarrow{v} + \overrightarrow{w}$ is the diagonal of the parallelogram.

These two operation satisfy the following properties.

**Properties.** *Let $\overrightarrow{v}$, $\overrightarrow{w}$ and $\overrightarrow{u}$ be free vectors, and let $c, d$ be real numbers, then*

- *the sum is associative:* $(\overrightarrow{v} + \overrightarrow{w}) + \overrightarrow{u} = \overrightarrow{v} + (\overrightarrow{w} + \overrightarrow{u})$;
- *the sum is commutative:* $\overrightarrow{v} + \overrightarrow{w} = \overrightarrow{w} + \overrightarrow{v}$;
- *the sum has a neutral element:* $\overrightarrow{v} + \overrightarrow{0} = \overrightarrow{0} + \overrightarrow{v} = \overrightarrow{v}$;
- *each vector has an additive inverse:* $-[\overrightarrow{AB}] = [\overrightarrow{BA}]$;
- *the product with a scalar is associative:* $(cd) \cdot \overrightarrow{v} = c \cdot (d \cdot \overrightarrow{v})$;
- *the product with a scalar has a neutral element:* $1 \cdot \overrightarrow{v} = \overrightarrow{v}$;
- *distributivity:* $(c + d) \cdot \overrightarrow{v} = c \cdot \overrightarrow{v} + d \cdot \overrightarrow{v}$.
- *distributivity:* $c \cdot (\overrightarrow{v} + \overrightarrow{w}) = c \cdot \overrightarrow{v} + c \cdot \overrightarrow{w}$.

## Frame

To describe a point in the plane or in three space we can use coordinates.

We consider here the plane $\mathbb{R}^2$ and the three dimensional space $\mathbb{R}^3$ with the *Cartesian coordinate system (frame)*.

In $\mathbb{R}^2$ the coordinate system is described by a point $O$ (origin), a free vector $\overrightarrow{i}$ of magnitude 1 describing the " $x$-axis", and a free vector $\overrightarrow{j}$ of magnitude 1 describing the " $y$-axis" (obtained by $\overrightarrow{i}$ with a rotation by $\frac{\pi}{2}$ counter-clockwise).

In $\mathbb{R}^3$ the coordinate system is described by a point $O$ (origin), a free vector $\overrightarrow{i}$ of magnitude 1 describing the "$x$-axis", a free vector $\overrightarrow{j}$ of magnitude 1 describing the "$y$-axis" (obtained by $\overrightarrow{i}$ with a rotation by $\frac{\pi}{2}$ counter-clockwise), and a free vector $\overrightarrow{k}$ of magnitude 1 describing the "$z$-axis" (obtained by $\overrightarrow{i}$ and $\overrightarrow{j}$ with right-hand rule).

Summing up, we have fixed the origin $O$ and two/three orthogonal, oriented axes and chosen a way to measure lengths on these axes. Then the position of each point is completely determined by taking orthogonal projections onto the axes. We can do the same with vectors, so we can identify

1. the point $P$;
2. the applied vector $\overrightarrow{OP}$;
3. and the free vector $[\overrightarrow{OP}]$.

Thus, in $\mathbb{R}^3$ (and similarly in $\mathbb{R}^2$) the free vector $\overrightarrow{v} = [\overrightarrow{OP}]$, is described through the coordinates $(x_P, y_P, z_P)$ of the end-point $P$, and indeed it can be expressed as

$$\overrightarrow{v} = x_P \cdot \overrightarrow{i} + y_P \cdot \overrightarrow{j} + z_P \cdot \overrightarrow{k} .$$

Using coordinates, operations on vectors can be easily performed, e.g. let $\overrightarrow{v} = 2\overrightarrow{i} + 3\overrightarrow{j} - 2\overrightarrow{k}$ and $\overrightarrow{w} = 3\overrightarrow{i} - 3\overrightarrow{j} + 3\overrightarrow{k}$, then

$$\begin{aligned} \overrightarrow{v} + \overrightarrow{w} &= 5\overrightarrow{i} + 0\overrightarrow{j} + \overrightarrow{k} ; \\ 2\overrightarrow{v} &= 4\overrightarrow{i} + 6\overrightarrow{j} - 4\overrightarrow{k} . \end{aligned}$$

# Chapter 1

# Matrices and Linear Systems

In this chapter we introduce, describe and manipulate two family of objects which will follow us for the entire course: matrices and linear system.

In each definition, theorem, example, exercise, etc. of this course we have to declare which numbers we are working with, so we need the concept of *field*.

**Definition 1.1.** A *field* $\mathbb{K}$ is a set endowed with two operations: $+ : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ (sum) and $\cdot : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ (product), satisfying the following properties:

F1) $\forall a, b, c \in \mathbb{K}$: $(a + b) + c = a + (b + c)$ (associativity of $+$)

F2) $\forall a, b \in \mathbb{K}$: $a + b = b + a$ (commutativity of $+$)

F3) $\exists 0 \in \mathbb{K}$: $a + 0 = 0 + a = a$ (neutral element of $+$)

F4) $\forall a \in \mathbb{K}, \exists b \in \mathbb{K}$: $a + b = b + a = 0$ (additive inverse: $b = -a$)

F5) $\forall a, b, c \in \mathbb{K}$: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (associativity of $\cdot$)

F6) $\forall a, b \in \mathbb{K}$: $a \cdot b = b \cdot a$ (commutativity of $\cdot$)

F7) $\exists 1 \in \mathbb{K}$: $a \cdot 1 = 1 \cdot a = a$ (neutral element of $\cdot$)

F8) $\forall a \in \mathbb{K}, a \neq 0, \exists b \in \mathbb{K}$: $a \cdot b = b \cdot a = 1$ (multiplicative inverse: $b = a^{-1}$)

F9) $\forall a, b \in \mathbb{K}$: $a \cdot (b + c) = a \cdot b + a \cdot c$ (Distributivity of multiplication over sum)

F10) $0 \neq 1$.

The elements of a field $\mathbb{K}$ are called *scalars*.

**Example.** i) The natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$ with the usual sum and multiplication do not form a field, since there are no additive inverses.

ii) The integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ with the usual sum and multiplication do not form a field, since there are no multiplicative inverses.

**Example.** i) The rational numbers $\mathbb{Q}$ ("fractions") and the real numbers $\mathbb{R}$ with the usual sum and multiplication form a field.

ii) The complex numbers $\mathbb{C}$ introduced in the course "Mathematical Analysis 1" form a field as well, and we have a natural chain of inclusions: $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

**Example.** There are fields with a finite numbers of elements! For example $\mathbb{K} = \{0, 1\}$ with sum and product given by the following tables:

| + | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

| · | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

## 1.1 Matrices

**Definition 1.2.** Let $m, n$ be positive integers and let $\mathbb{K}$ a field.
A $m \times n$ *matrix* (or *matrix* of type $(m, n)$) with coefficients in $\mathbb{K}$ is a table of elements of $\mathbb{K}$ arranged in $m$ rows and $n$ columns.
We denote by $\mathcal{M}_{\mathbb{K}}(m, n)$ the set of all $m \times n$ matrices with coefficients in $\mathbb{K}$.

**Notation.** Other common notations for the set of all $m \times n$ matrices with coefficients in $\mathbb{K}$ are $\mathcal{M}_{m\times n}(\mathbb{K})$, $\mathrm{Mat}_{\mathbb{K}}(m, n)$ or $\mathrm{Mat}(m \times n, \mathbb{K})$.

**Example.** $\mathbb{K} = \mathbb{R}$, $m = 2$, $n = 4$.

$$\begin{pmatrix} -1 & 0 & 2 & \pi \\ 0 & \sqrt{2} & -2/3 & 6 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(2, 4)$$

In general, to represent a matrix $A$, we use the following notation:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} = (a_{i,j})$$

So $a_{i,j}$ is the element on the $i$-th row and $j$-th column. This element can also be denoted by $(A)_{i,j}$.

**Terminology:**

- $m = n = 1$. $A = (a_{1,1})$, so we have a natural bijection $\mathcal{M}_{\mathbb{K}}(1, 1)$"="$\mathbb{K}$.
- $m = n$. A matrix of type $(n, n)$ is called a *square matrix*.
- $m = 1$, $n > 1$. $A = (a_{1,1}\ a_{1,2} \dots a_{1,n})$ is called *row vector* (of length $n$).

  A row vector depends on $n$ choices of elements in $\mathbb{K}$, so $\mathcal{M}_{\mathbb{K}}(1, n)$ is naturally in bijection with the *cartesian product*

  $$\mathbb{K}^n = \underbrace{\mathbb{K} \times \mathbb{K} \times \cdots \mathbb{K}}_{n-\text{times}} = \{(k_1, k_2, \dots, k_n) : k_i \in \mathbb{K}\}$$

- $m > 1$, $n = 1$. $A = \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix}$ is called *column vector* (of height $m$).

  As for row vectors, we have a natural bijection between the set of column vectors $\mathcal{M}_{\mathbb{K}}(m, 1)$ and $\mathbb{K}^m$, as each column vector depends on $m$ elements of $\mathbb{K}$.

### 1.1.1 Matrix operations

#### Equality

Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ and $B \in \mathcal{M}_{\mathbb{K}}(p,q)$ be two matrices. To be equal $A$ and $B$ must be of the same type: $m = p$ and $n = q$, and have the same entries: $a_{i,j} = b_{i,j}$ $\forall i = 1,\dots,m\,, \forall j = 1,\dots,n$.

#### Sum of matrices

Let $A, B \in \mathcal{M}_{\mathbb{K}}(m,n)$ be matrices of the same type, then we define their sum $A+B$ as the matrix in $\mathcal{M}_{\mathbb{K}}(m,n)$, obtained by adding componentwise the entries of $A$ and $B$. Formally

$$\begin{array}{rcl} + : \mathcal{M}_{\mathbb{K}}(m,n) \times \mathcal{M}_{\mathbb{K}}(m,n) & \longrightarrow & \mathcal{M}_{\mathbb{K}}(m,n) \\ (A,B) & \longmapsto & (A+B) = (a_{i,j} + b_{i,j}) \end{array}$$

**Example.**

$$\mathbb{K} = \mathbb{Q}, m = 2, n = 3, \quad \begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & -4 \end{pmatrix} + \begin{pmatrix} -1 & 3 & 6 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 6 & 8 \\ 1 & 2 & -4 \end{pmatrix}$$

*Remark.* The sum is defined only for matrices of the same type.

It is easy to check that the sum of matrices inherits all properties of the sum in $\mathbb{K}$:

**Properties.** *The sum of matrices satisfies the following properties.*

- *Associativity:* $\forall A, B, C \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $(A+B)+C = A+(B+C)$.
- *Commutativity:* $\forall A, B \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $A+B = B+A$.
- *The neutral element is the zero matrix* $O \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $(O)_{i,j} = 0$.
- *The additive inverse of* $A = (a_{i,j}) \in \mathcal{M}_{\mathbb{K}}(m,n)$ *is the matrix* $A' = (a'_{i,j}) \in \mathcal{M}_{\mathbb{K}}(m,n)$, *with* $A + A' = O$, *namely* $a'_{i,j} = -a_{i,j}$.

#### Product of a matrix by a scalar

Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix and $k \in \mathbb{K}$ a scalar. Multiplying each entry of $A$ by the scalar $k$ we get a new matrix $kA \in \mathcal{M}_{\mathbb{K}}(m,n)$. Formally

$$\begin{array}{rcl} \cdot : \mathbb{K} \times \mathcal{M}_{\mathbb{K}}(m,n) & \longrightarrow & \mathcal{M}_{\mathbb{K}}(m,n) \\ (k,A) & \longmapsto & (kA) = (ka_{i,j}) \end{array}$$

**Example.**

$$\mathbb{K} = \mathbb{Q}, m = 2, n = 3, \quad -2 \cdot \begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & -4 \end{pmatrix} = \begin{pmatrix} -2 & -6 & -4 \\ 0 & -4 & 8 \end{pmatrix}$$

It is straightforward to verify the following properties.

**Properties.** *The product of a matrix by a scalar satisfies the following properties.*

- *Associativity:* $\forall k_1, k_2 \in \mathbb{K}\,, \forall A \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $k_1 \cdot (k_2 \cdot A) = (k_1 \cdot k_2) \cdot A$.
- $1 \in \mathbb{K}$ *is the neutral element:* $1 \cdot A = A$.
- *Distributivity:* $\forall k_1, k_2 \in \mathbb{K}\,, \forall A \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $(k_1 + k_2) \cdot A = k_1 \cdot A + k_2 \cdot A$
- *Distributivity:* $\forall k \in \mathbb{K}\,, \forall A, B \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $k \cdot (A+B) = k \cdot A + k \cdot B$.

*Remark.* Note that $A + (-1 \cdot A) = O$, so the additive inverse of $A$ is $(-1 \cdot A) = -A$.

**Matrix multiplication (row-column)**

We define now the product of two matrices.

$$\begin{array}{rcl} \cdot : \mathcal{M}_{\mathbb{K}}(m,p) \times \mathcal{M}_{\mathbb{K}}(p,n) & \longrightarrow & \mathcal{M}_{\mathbb{K}}(m,n), \\ (A,B) & \longmapsto & A \cdot B \end{array}$$

as

$$(A \cdot B)_{i,j} = a_{i,1} \cdot b_{1,j} + a_{i,2} \cdot b_{2,j} + \cdots + a_{i,n} \cdot b_{n,j} = \sum_{l=1}^{p} (a_{i,l} \cdot b_{l,j})$$

*Remark.* Note that we do not require that $A$ and $B$ are of the same type, but we require that *the number of columns of $A$ coincides with the number of rows of $B$*, otherwise $AB$ is not defined.

The resulting matrix has as many rows as $A$ and as many columns as $B$.

**Example.** Let $A = (a_{1,1} \dots a_{1,p}) \in \mathcal{M}_{\mathbb{K}}(1,p)$ and $B = \begin{pmatrix} b_{1,1} \\ \vdots \\ b_{p_1} \end{pmatrix} \in \mathcal{M}_{\mathbb{K}}(p,1)$.

Then $C = AB \in \mathcal{M}_{\mathbb{K}}(1,1)$, and $c_{1,1} = \sum_{l=1}^{p} (a_{1,l} \cdot b_{l,1}) = a_{1,1} \cdot b_{1,1} + a_{1,2} \cdot b_{2,1} + \cdots + a_{1,n} \cdot b_{n,1}$. For example ($\mathbb{K} = \mathbb{Q}$, $p = 3$)

$$(1,2,3) \cdot \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = (0 \cdot 1 + 2 \cdot 1 + 3 \cdot (-1)) = (-1)$$

This is called row-column multiplication. In general we have to perform it $mn$ times to determine $A \cdot B$, indeed the element $(A \cdot B)_{i,j}$ is obtained multiplying the $i$-th row of $A$ with the $j$-th column of $B$.

**Example.** i) Let $A \in \mathcal{M}_{\mathbb{Q}}(2,3)$ and $B \in \mathcal{M}_{\mathbb{Q}}(3,2)$ be the matrices

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 \\ 3 & 0 \\ 6 & 0 \end{pmatrix}$$

To compute $C = AB$ we have to perform 4 row-column multiplication, e.g. $c_{1,2}$ is obtain by the row-column multiplication of the $1^{st}$ row of $A$ with the $2^{nd}$ column of $B$:

$$\begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & -4 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 \\ 3 & 0 \\ 6 & 0 \end{pmatrix} = \begin{pmatrix} 1 \cdot (-1) + 3 \cdot 3 + 2 \cdot 6 & 1 \cdot 1 + 3 \cdot 0 + 2 \cdot 0 \\ 0 \cdot (-1) + 2 \cdot 3 + (-4) \cdot 6 & 0 \cdot 1 + 2 \cdot 0 + (-4) \cdot 0 \end{pmatrix} = \begin{pmatrix} 20 & 1 \\ -18 & 0 \end{pmatrix}$$

ii) Let $A \in \mathcal{M}_{\mathbb{Q}}(2,3)$ as above and let $B' = \begin{pmatrix} -1 & 3 & 6 \\ 1 & 0 & 0 \end{pmatrix} \in \mathcal{M}_{\mathbb{Q}}(2,3)$, then the product $A \cdot B'$ is not defined, since $A$ has 3 columns, but $B'$ has only 2 rows.

**Definition 1.3.** The *identity matrix* $I_n$ is the square matrix $I_n \in \mathcal{M}_{\mathbb{K}}(n,n)$ having:

$$(I_n)_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

In a square matrix, the position with row-index equal to the column-index form the *diagonal*.

The matrix product satisfies the following properties, which are easy to verify (only the first one needs a bit of work).

**Properties.** *The matrix product satisfies the following properties:*

- *Associativity:* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n), B \in \mathcal{M}_{\mathbb{K}}(n,p), C \in \mathcal{M}_{\mathbb{K}}(p,q)$ *it holds* $(AB)C = A(BC)$.
- *Distributivity:* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n), B_1, B_2 \in \mathcal{M}_{\mathbb{K}}(n,p)$ *it holds* $A(B_1 + B_2) = AB_1 + AB_2$.

  *And* $\forall A_1, A_2 \in \mathcal{M}_{\mathbb{K}}(m,n), B \in \mathcal{M}_{\mathbb{K}}(n,p)$ *it holds* $(A_1 + A_2)B = A_1B + A_2B$.
- *Neutral element:* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n)$*, it holds* $A \cdot I_n = A$ *and* $I_m \cdot A = A$.
- *(Mixed) associativity:* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n), B \in \mathcal{M}_{\mathbb{K}}(n,p), k \in \mathbb{K}$ *it holds* $k(AB) = (kA)B = A(kB)$.

What about the commutativity?

*Remark.* In general, the matrix multiplication is not commutative!

Let $A \in \mathcal{M}_{\mathbb{K}}(m,p)$ and $B \in \mathcal{M}_{\mathbb{K}}(p,n)$.

i) $AB \in \mathcal{M}_{\mathbb{K}}(m,n)$, but $BA$ is not even defined if $m \neq n$.

ii) Assume then that $m = n$ so both $AB$ and $BA$ are defined: $AB \in \mathcal{M}_{\mathbb{K}}(m,m)$ and $BA \in \mathcal{M}_{\mathbb{K}}(p,p)$. If $p \neq m$, the two matrices are of different type, so they are different.

iii) Assume then that $m = n = p$ so $AB, BA \in \mathcal{M}_{\mathbb{K}}(n,n)$. Also in this case $AB$ and $BA$ can be different, e.g. $\mathbb{K} = \mathbb{Q}$, $m = n = p = 2$:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \text{ then } AB = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = BA$$

iv) Pay attention that in some cases, it is possible that $AB = BA$, e.g. $\mathbb{K} = \mathbb{Q}$, $m = n = p = 2$:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ then } AB = BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

*Remark.* The last example shows that it is possible that the product of 2 non-zero matrices is the zero matrix $O$, a phenomenon that cannot happen for elements in a field, for example for real numbers.

**Transpose of a matrix**

Given a matrix $A$ we can define another matrix by swapping rows and columns of $A$:

**Definition 1.4.** Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix. The *transpose of* $A$ is the matrix $B \in \mathcal{M}_{\mathbb{K}}(n,m)$ defined by $b_{i,j} = a_{j,i}$, and it is denoted by $A^T$.

**Example.** Let $A = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 2 & -4 \end{pmatrix} \in \mathcal{M}_{\mathbb{Q}}(2,3)$ then $A^T = \begin{pmatrix} 1 & 0 \\ 3 & 2 \\ 2 & -4 \end{pmatrix} \in \mathcal{M}_{\mathbb{Q}}(3,2)$.

**Properties.** *The following properties hold:*

*i)* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $(A^T)^T = A$*;*

*ii)* $\forall A, B \in \mathcal{M}_{\mathbb{K}}(m,n)$*:* $(A+B)^T = A^T + B^T$*;*

*iii)* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,n), k \in \mathbb{K}$*:* $(kA)^T = k(A^T)$*;*

*iv)* $\forall A \in \mathcal{M}_{\mathbb{K}}(m,p), B \in \mathcal{M}_{\mathbb{K}}(p,n)$*:* $(AB)^T = B^T A^T$*.*

*Proof.* Properties i)-ii)-iii) follow directly from the definitions.

iv) $(AB)_{i,j} = \sum_{l=1}^{p} a_{i,l} b_{l,j}$, so

$$((AB)^T)_{i,j} = (AB)_{j,i} = \sum_{l=1}^{p} a_{j,l} b_{l,i} = \sum_{l=1}^{p} b_{l,i} a_{j,l} = \sum_{l=1}^{p} (B^T)_{i,l} (A^T)_{l,j} = (B^T A^T)_{i,j}.$$

□

## 1.2 Linear Systems

**Definition 1.5.** Let $m, n$ be positive integers. Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix and let $b \in \mathcal{M}_{\mathbb{K}}(m,1)$ be a column vector (of height $m$). A *linear system* (or system of linear equations) is an equation of the form $Ax = b$, where $x \in \mathcal{M}_{\mathbb{K}}(n,1)$:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{1.1}$$

The matrix $A$ is called the *matrix of coefficients* and the column vector $b$ is called the *column of constant terms.*

Rewriting equation (1.1), we have a system of $m$ *linear equations* and $n$ variables:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n & = & b_2 \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots & & \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n & = & b_m \end{cases}$$

**Notation.** A linear system $Ax = b$ is usually represented through its *augmented matrix* $(A|b)$:

$$(A|b) = \left( \begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} & b_n \end{array} \right)$$

**Example.** The augmented matrix ($\mathbb{K} = \mathbb{Q}$)

$$\left(\begin{array}{rrr|r} 1 & 1 & 1 & 0 \\ 2 & -1 & 1 & 1 \end{array}\right) \text{ represents the linear system } \begin{cases} x_1 + x_2 + x_3 &= 0 \\ 2x_1 - x_2 + x_3 &= 1 \end{cases}$$

**Definition 1.6.** A *solution* of the linear system $Ax = b$, is a column vector $s \in \mathcal{M}_{\mathbb{K}}(n, 1)$ whose entries $s_1, s_2, \ldots, s_n \in \mathbb{K}$ satisfy simultaneously all $m$ equations.

We denote by $\mathrm{Sol}(A|b)$ the set of all solutions of $Ax = b$:

$$\mathrm{Sol}(A|b) = \{s \in \mathcal{M}_{\mathbb{K}}(n, 1) \text{ such that } As = b\} \subseteq \mathcal{M}_{\mathbb{K}}(n, 1)\,.$$

**Example.** Let $m = n = 1$, so that the linear system consist of a single equation in one variable: $ax = b$, with $a, b \in \mathbb{K}$.

If $a \neq 0$, we can divide both sides by $a$ and get a unique solution: $x = \frac{b}{a}$, e.g $3x = 4$ has a unique solution: $x = \frac{4}{3}$.

If $a = 0$ the equation reduces to $0x = b$ and we have 2 subcases:
i) if $b = 0$, then we have an equation of the form "$0 = 0$", so that any $k \in \mathbb{K}$ is a solution;
ii) if $b \neq 0$, then we have an equation of the form "$0 = 1$", which has no solutions at all.

**Example.** Let $m = n = 2$ ($\mathbb{K} = \mathbb{R}$) and consider the linear system

$$\begin{cases} x + y &= 2 \\ x - y &= 0 \end{cases}$$

The second equation is equivalent to $x = y$, plugging this into the first equation we get $2x = 2$, so $x = 1, y = 1$ is the unique solution of the linear system.

Consider now the linear systems

$$I) \begin{cases} x + y &= 2 \\ x + y &= 1 \end{cases} \qquad II) \begin{cases} x + y &= 2 \\ 3x + 3y &= 6 \end{cases}$$

The linear system I) has no solution since $2 \neq 1$. The linear system II) reduces to the single equation $x + y = 2$ since the second equation is 3-times the first one. So any pair $(s, 2 - s)$, $s \in \mathbb{K}$ is a solution.

Our goal is to understand if a given linear system has solutions, and, if yes, how to find all its solutions.

### 1.2.1 Echelon form

**Definition 1.7.** A matrix $A \in \mathcal{M}_{\mathbb{K}}(m, n)$ is in *echelon form* if

- every non-zero row starts with more zeroes than the previous one;
- all zero rows are below the non-zero rows.

A linear system $Ax = b$ is in *echelon form* if its augmented matrix $(A|b)$ is in echelon form.

**Definition 1.8.** Let $A \in \mathcal{M}_{\mathbb{K}}(m, n)$ be a matrix in echelon form. The first (from left) non-zero entry in each non-zero row is called *pivot*.

**Example.** The zero matrix and the identity matrix are in echelon form over any field $\mathbb{K}$. The matrices ($\mathbb{K} = \mathbb{R}$)

$$\begin{pmatrix} 2 & -1 & 3 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & -2 & 3 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

are in echelon form. The pivot of the first matrix are 2, 4 and 3; in the second matrix the pivots are 1 and 4. The following matrices ($\mathbb{K} = \mathbb{R}$) are not in echelon form:

$$\begin{pmatrix} 2 & 2 & 3 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 4 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

**Example.** Consider the following linear systems (over $\mathbb{K} = \mathbb{R}$) in echelon form

$$i)\ (A|b) = \left(\begin{array}{ccc|c} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 7 \end{array}\right) \quad \longleftrightarrow \quad \begin{cases} x_1 - x_2 + 3x_3 & = & 5 \\ x_2 + 2x_3 & = & 2 \\ x_3 & = & 7 \end{cases}$$

It is easy to solve it: the last equation returns $x_3 = 7$, we substitute this in the second equation and get $x_2 = 2 - 2 \cdot 7 = -12$; similarly $x_1 = -28$, so $\mathrm{Sol}(A|b) = \{(-28, -12, 7)^T\}$.

$$ii)\ (A|b) = \left(\begin{array}{ccc|c} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 7 \end{array}\right) \quad \longleftrightarrow \quad \begin{cases} x_1 - x_2 + 3x_3 & = & 5 \\ x_2 + 2x_3 & = & 2 \\ 0 & = & 7 \end{cases}$$

It has no solutions, since $0 = 7$ is impossible.

$$iii)\ (A|b) = \left(\begin{array}{ccc|c} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \end{array}\right) \quad \longleftrightarrow \quad \begin{cases} x_1 - x_2 + 3x_3 & = & 5 \\ x_2 + 2x_3 & = & 2 \end{cases}$$

In the last equation, 2 variables appear, so we express $x_2$ in terms of $x_3$: $x_2 = 2 - 2x_3$, by substitution in the first equation, we get $x_1 = 5 + x_2 - 3x_3 = 7 - 5x_3$, so we have one free variable ($x_3$), and $\mathrm{Sol}(A|b) = \{(7 - 5t, 2 - 2t, t)^T \mid t \in \mathbb{R}\}$.

$$iv)\ (A|b) = \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \end{array}\right) \quad \longleftrightarrow \quad x_1 + x_2 - x_3 = 1$$

There is only one equation, with 3 variables, so we express one of them, say $x_3$ in terms of the other two: $x_3 = x_1 + x_2 - 1$, so we have two free variables, and $\mathrm{Sol}(A|b) = \{(s, t, s + t - 1)^T \mid s, t \in \mathbb{R}\}$.

$$v)\ (A|b) = \left(\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right) \quad \longleftrightarrow \quad \begin{cases} x_1 - x_3 & = & 1 \\ x_3 & = & 2 \end{cases}$$

The second equation returns $x_3 = 2$, we substitute this in the first one and get $x_1 = 1 + 2 = -3$, and there is no constrains on $x_2$, which is then a free variable: $\mathrm{Sol}(A|b) = \{(3, s, 2)^T \mid s \in \mathbb{R}\}$.

Given a linear system in echelon form, we can easily check if it is solvable, and, if yes, determine all solutions by back-substitution (as above).

So our aim is now to transform an arbitrary linear system into an equivalent one (i.e. having the same set of solution), which is in echelon form. We will do it, using the *elementary row operations.*

### 1.2.2 Elementary row operations and Gauss algorithm

Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix and let $R_1, \dots, R_m$ be its rows. The *elementary row operations* are:

1. *to swap 2 rows* ($R_i \leftrightarrow R_j$):
2. *to multiply a row by a non-zero scalar* $k \in \mathbb{K}, k \neq 0$ ($R_i \to kR_i$):
3. *to add to a row a multiple of another row*: ($R_i \to R_i + kR_j$, $k \in \mathbb{K}$):

*Remark.* Considering the corresponding operations on the equations of a linear system, one easily shows that the elementary row operations do not change the set of solution of a linear system. In other words, if the matrix $(U|b')$ is obtained from $(A|b)$ through a sequence of elementary row operations, then $\mathrm{Sol}(A|b) = \mathrm{Sol}(U|b')$.

The *Gauss algorithm* takes as input a matrix $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ and returns as output a matrix $U \in \mathcal{M}_{\mathbb{K}}(m,n)$ in echelon form, obtained by $A$ via a sequence of elementary row operations. The rough idea is to "clean up" one column at a time from left to right.

**Gauss algorithm.** `Input`: $A \in \mathcal{M}_{\mathbb{K}}(m,n)$.
Step 0: If $A$ is in echelon form, then `Output`: $U = A$.

Step 1: Let $j$ be the index of the first non-zero column, and let $i$ be a row-index such that $a_{i,j} \neq 0$. Then swap the first and the $i$-th row and obtain the matrix $B$:

$$A = \begin{pmatrix} 0 & \dots & 0 & * & \dots & * \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & a_{i,j} & & \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & * & \dots & * \end{pmatrix} \overset{R_1 \leftrightarrow R_i}{\longrightarrow} B = \begin{pmatrix} 0 & \dots & 0 & p_1 & \dots & * \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & b_{i,j} & & \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & * & \dots & * \end{pmatrix}$$

Note that $p_1 = a_{i,j} \neq 0$.

Step 2: For each row $l \neq 1$ such that $b_{l,j} \neq 0$, we apply the move: $R_l \to R_l + \left(-\frac{b_{l,j}}{p_1}\right) R_1$, so that $b_{l,j}$ becomes $b_{l,j} + \left(-\frac{b_{l,j}}{p_1}\right) p_1 = 0$:

$$B = \begin{pmatrix} 0 & \dots & 0 & p_1 & \dots & * \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & b_{l,j} & & \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & * & \dots & * \end{pmatrix} \overset{R_l \to R_l + \left(-\frac{b_{l,j}}{p_1}\right) R_1}{\longrightarrow} C = \left(\begin{array}{ccccc|ccc} 0 & \dots & 0 & p_1 & * & \dots & * \\ \vdots & & \vdots & 0 & & & \\ \vdots & & \vdots & \vdots & & A' & \\ 0 & \dots & 0 & 0 & & & \end{array}\right)$$

Step 3: "Restart" the algorithm with $A' \in \mathcal{M}_{\mathbb{K}}(m-1, n')$ $(n' < n)$.

*Remark.* There are several ways to reduce a matrix $A$ in echelon form, so the output is not unique!

**Example.** Let us consider the matrix $M = \begin{pmatrix} 0 & -3 & -5 & 1 \\ 2 & -4 & 2 & 6 \\ 1 & -1 & 3 & 5 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3,4)$, we use now the elementary row operation to reduce it in echelon form:

$$\begin{pmatrix} 0 & -3 & -5 & 1 \\ 2 & -4 & 2 & 6 \\ 1 & -1 & 3 & 5 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 1 & -1 & 3 & 5 \\ 2 & -4 & 2 & 6 \\ 0 & -3 & -5 & 1 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_2 - 2R_1} \begin{pmatrix} 1 & -1 & 3 & 5 \\ 0 & -2 & -4 & -4 \\ 0 & -3 & -5 & 1 \end{pmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow -\frac{1}{2}R_2} \begin{pmatrix} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & -3 & -5 & 1 \end{pmatrix} \xrightarrow{R_3 \leftrightarrow R_3 + 3R_2} \begin{pmatrix} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

Interpreting $M$ as the augmented matrix of a linear system, we read:

$$\begin{cases} 3x_2 - 5x_3 & = & 1 \\ 2x_1 - 4x_2 + 2x_3 & = & 6 \\ x_1 - x_2 + 3x_3 & = & 5 \end{cases} \text{ is equivalent/reduces to } \begin{cases} x_1 - x_2 + 3x_3 & = & 5 \\ x_2 + 2x_3 & = & 2 \\ x_3 & = & 7 \end{cases}$$

The *Gauss-Jordan algorithm* is a refinement of the Gauss algorithm: once we have reduced our matrix in echelon form, we perform further elementary row operations to "clean out" one column at a time form right to left. The final output will be a matrix in echelon form such that every pivot is equal to 1 and every pivot is the unique non-zero element in its column.

We explain it with 2 examples ($\mathbb{K} = \mathbb{R}$).

**Example.** i) We continue the previous example :

$$\begin{pmatrix} 1 & -1 & 3 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 7 \end{pmatrix} \xrightarrow[R_1 \to R_1 - 3R_3]{R_2 \to R_2 - 2R_3} \begin{pmatrix} 1 & -1 & 0 & -16 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 7 \end{pmatrix} \xrightarrow{R_1 \to R_1 + R_3} \begin{pmatrix} 1 & 0 & 0 & -28 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

ii) Consider the linear system $\begin{cases} 2x_1 - 6x_2 + 2x_3 + 2x_4 & = & 6 \\ -3x_1 + 9x_2 + \quad\quad + 3x_4 & = & -9 \end{cases}$, its augmented matrix reduces to:

$$\left(\begin{array}{cccc|c} 2 & -6 & 2 & 2 & 6 \\ -3 & 9 & 0 & 3 & -9 \end{array}\right) \xrightarrow[R_2 \to \frac{1}{3}R_2]{R_1 \to \frac{1}{2}R_1} \left(\begin{array}{cccc|c} 1 & -3 & 1 & 1 & 3 \\ -1 & 3 & 0 & 1 & -3 \end{array}\right) \xrightarrow{R_2 \to R_2 + R_1} \left(\begin{array}{cccc|c} 1 & -3 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 & 0 \end{array}\right)$$

$$\xrightarrow{R_1 \to R_1 - R_2} \left(\begin{array}{cccc|c} 1 & -3 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & 0 \end{array}\right) \text{ which corresponds to } \begin{cases} x_1 - 3x_2 - x_4 & = & 3 \\ x_3 + 2x_4 & = & 0 \end{cases}$$

**Definition 1.9.** The variables corresponding to pivots are called *pivot variables*; the remaining ones are called *free variables*.

In the linear system $\begin{cases} x_1 - 3x_2 - x_4 & = & 3 \\ x_3 + 2x_4 & = & 0 \end{cases}$, the variables $x_2, x_4$ are free variables, while $x_1, x_3$ are pivot variables and can be expressed in term of the free variables: $x_1 = 3 + 3x_2 + x_4$ (1st equation), $x_3 = -2x_4$ (2nd equation): $\text{Sol} = \{(3+3s+t, s, -2t, t)^T \mid s, t \in \mathbb{K}\}$.

*Remark.* The Gauss(-Jordan) algorithm solves a linear system expressing the pivot variables in term of the free variables.

**Definition 1.10.** Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix, and let $U \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a reduction of $A$ in echelon form. The *rank* of $A$ (denoted $\mathrm{rk}(A)$) is the number of pivots of $U$.

**Example.** $\mathrm{rk}(O) = 0$, $\mathrm{rk}(I_n) = n$.

The matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$ reduces to $A' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, so $\mathrm{rk}(A) = 1$.

The matrix $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ reduces to $B' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$, so $\mathrm{rk}(B) = 2$.

*Remark.* i) As remarked above, there are several ways to reduce a matrix $A$ in echelon form. We will show later in the course that *all reductions of $A$ in echelon form have the same number of pivots.* In other words, the rank is independent from the reduction.
ii) Elementary row operations preserve the rank.
iii) Since the pivots are in different columns and rows, for $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, it holds $0 \le \mathrm{rk}(A) \le m$ and $0 \le \mathrm{rk}(A) \le n$, i.e. $0 \le \mathrm{rk}(A) \le \min\{m,n\}$.

### 1.2.3 Rouché-Capelli Theorem

**Example.** Consider the linear system

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 3 & 4 \\ 0 & 1 & 2 & -1 & 5 \\ 0 & -1 & -2 & 1 & \alpha \end{array}\right) \qquad \text{where } \alpha \in \mathbb{K} \text{ is a parameter.}$$

Applying the move $R_3 \to R_3 + R_2$, we get the matrix

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 3 & 4 \\ 0 & 1 & 2 & -1 & 5 \\ 0 & 0 & 0 & 0 & 5+\alpha \end{array}\right)$$

If $5 + \alpha \neq 0$, the last equation has the form "$0 = 1$" so there is no solution.

If $5 + \alpha = 0$, the last equation has the form "$0 = 0$" and we do have solutions: $x_1 = 4 - 2x_3 - 3x_4$, $x_2 = 5 - 2x_3 + x_4$:

$$\mathrm{Sol} = \left\{ \begin{pmatrix} 4-2s-3t \\ 5-2s+t \\ s \\ t \end{pmatrix} \mid s,t \in \mathbb{K} \right\} = \left\{ \begin{pmatrix} 4 \\ 5 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ -2 \\ 1 \\ 0 \end{pmatrix} s + \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix} t \mid s,t \in \mathbb{K} \right\}$$

**Theorem 1.11** (Rouché-Capelli Theorem)**.** *Let $Ax = b$ be a linear system with $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ and $b \in \mathcal{M}_{\mathbb{K}}(m,1)$.*

1. *The linear system $Ax = b$ is solvable if and only if* $\mathrm{rk}(A) = \mathrm{rk}(A|b)$.
2. *If* $\mathrm{rk}(A) = \mathrm{rk}(A|b) = r$, *then there exist $n - r + 1$ column vectors $v, w_1, \dots, w_{n-r} \in \mathcal{M}_{\mathbb{K}}(n,1)$, such that* $\mathrm{Sol}(A|b) = \{v + t_1 w_1 + \dots + t_{n-r} w_{n-r} : t_i \in \mathbb{K}\}$.

*3. In particular,* $Ax = b$ *has a unique solution if and only if* $\mathrm{rk}(A) = \mathrm{rk}(A|b) = n$

*Proof.* Applying the Gauss algorithm to $(A|b)$ we obtain a matrix $(U|b')$ in echelon form such that $\mathrm{Sol}(A|b) = \mathrm{Sol}(U|b')$. There are 2 possibilities for $(U|b')$: either all pivots are on the left (1), or there is a pivot on the right (2):

$$(1)\ \left(\begin{array}{cccccccccc|c} 0 & \dots & 0 & p_1 & & & & & & & \\ 0 & \dots & 0 & 0 & p_2 & & & & & & \\ 0 & \dots & 0 & 0 & 0 & \ddots & & & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & p_r & \dots & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & & 0 \end{array}\right) \qquad (2)\ \left(\begin{array}{cccccccc|c} 0 & \dots & 0 & p_1 & & & & & \\ 0 & \dots & 0 & 0 & p_2 & & & & \\ 0 & \dots & 0 & 0 & 0 & \ddots & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & p_r & \dots & \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & p_{r+1} \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{array}\right)$$

where the $p_j$'s are the pivots, in particular $p_j \neq 0$.

In situation (2), the $(r+1)$-th row reads "$0 = 1$" so the linear system is not solvable, and $\mathrm{rk}(A|b) = \mathrm{rk}(A) + 1$: $\mathrm{rk}(A) < \mathrm{rk}(A|b)$.

In situation (1), $\mathrm{rk}(A) = \mathrm{rk}(A|b)$ and no row is of the form "$0 = 1$", so the linear system is solvable by back-substitution: there are $r = \mathrm{rk}(A) = \mathrm{rk}(A|b)$ non-zero equations, which allow us to express the pivot variables ($r$ variables) in term of the free variables $n - r$ variables, so the solutions depend on $n - r$ parameters. □

**Example.** Consider the linear system $Ax = b$

$$\left(\begin{array}{cccc|c} -2 & 2 & -1 & 3 & -3 \\ 3 & -3 & 6 & 0 & 0 \\ 1 & -1 & 5 & 3 & -3 \end{array}\right)$$

After reducing it with the Gauss-Jordan algorithm, we obtain the equivalent linear system

$$\left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 2 \\ 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

So $\mathrm{rk}(A) = \mathrm{rk}(A|b) = 2$ and $n = 4$: there are 2 pivot variables $x_1, x_3$ and 2 free variables: $x_2$ and $x_4$:

$$\mathrm{Sol}(A|b) = \left\{ \begin{pmatrix} 2+s+2t \\ s \\ -1-t \\ t \end{pmatrix} \mid s,t \in \mathbb{K} \right\} = \left\{ \begin{pmatrix} 2 \\ 0 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 2 \\ 0 \\ -1 \\ 1 \end{pmatrix} t \mid s,t \in \mathbb{K} \right\}$$

*Remark.* a) The parametric part of the solution ($\left(\begin{smallmatrix} 1 \\ 1 \\ 0 \\ 0 \end{smallmatrix}\right) s + \left(\begin{smallmatrix} 2 \\ 0 \\ -1 \\ 1 \end{smallmatrix}\right) t$ in the last example) does not depend on the constant terms, but only on the matrix of coefficients and more precisely

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 2 \\ 0 \\ -1 \\ 1 \end{pmatrix} t \mid s,t \in \mathbb{K} \right\} = \mathrm{Sol}(A|0)$$

b) The constant terms contribute to the constant/fixed part of the solution ($\begin{pmatrix} 2 \\ 0 \\ -1 \\ 0 \end{pmatrix}$ in the last example).

**Definition 1.12.** A linear system of the form $Ax = 0$ (i.e. the constant terms is made of zeroes) is called a *homogeneous linear system.*

*Remark.* A homogeneous linear system has always a solution, namely $x = (0, \ldots, 0)^T$, called the trivial solution.

**Theorem 1.13** (Structure theorem of the solutions of a linear system)**.** *Let $Ax = b$ be a solvable linear system and let $v_0 \in \mathrm{Sol}(A|b)$. Then*

$$\mathrm{Sol}(A|b) = v_0 + \mathrm{Sol}(A|0)\,.$$

In other words: for every $v_H \in \mathrm{Sol}(A|0)$ the column vector $v_0 + v_H$ is a solution of $\mathrm{Sol}(A|b)$; conversely every solution of $Ax = b$ is of the form $v_0 + v_H$, where $v_H \in \mathrm{Sol}(A|0)$.

*Proof.* Let $v_H \in \mathrm{Sol}(A|0)$, then $A(v_0+v_H) = Av_0+Av_H = b+0 = b$, so $v_0+v_H \in \mathrm{Sol}(A|b)$.

Conversely, let $v$ be a solution of $Ax = b$, then we define $v_H = v - v_0$ and we verify that $v_H \in \mathrm{Sol}(A|0)$: $A(v - v_0) = Av - Av_0 = b - b = 0$. □

We conclude this section with two consequences of the Rouché-Capelli theorem.

**Lemma 1.14.** *Let $A \in \mathcal{M}_{\mathbb{K}}(m, n)$ be a matrix with $m < n$. Then for any $b \in \mathcal{M}_{\mathbb{K}}(m, 1)$ the linear system $Ax = b$ cannot have a unique solution.*

*Proof.* By Rouché-Capelli theorem $Ax = b$ has a unique solution if and only if $n = \mathrm{rk}(A|b) = \mathrm{rk}(A)$, but $\mathrm{rk}(A) \leq \min\{m, n\} = m < n$. □

**Theorem 1.15** (Cramer's theorem)**.** *Let $A \in \mathcal{M}_{\mathbb{K}}(n, n)$ be a square matrix. Then $\mathrm{rk}(A) = n$ if and only if the linear system $Ax = b$ has a unique solution for any $b \in \mathcal{M}_{\mathbb{K}}(n, 1)$.*

*Proof.* $\Rightarrow$] Assume $\mathrm{rk}(A) = n$ and consider a linear system $Ax = b$ for an arbitrary $b \in \mathcal{M}_{\mathbb{K}}(n, 1)$. We want to show that $Ax = b$ has a unique solution.

We consider the augmented matrix $(A|b)$ and we reduce it into echelon form using the Gauss algorithm, obtaining the matrix $(U|b')$. Since $\mathrm{rk}(A) = \mathrm{rk}(U) = n$, and $U$ is a square matrix of type $n \times n$, the pivots of $U$ appear on the diagonal:

$$(A|b) \stackrel{Gauss}{\longrightarrow} (U|b') = \left(\begin{array}{ccccc|c} p_1 & * & * & * & * & * \\ 0 & p_2 & & & & \\ 0 & 0 & \ddots & & & \\ \vdots & & & \ddots & & \\ 0 & 0 & \ldots & & p_n & * \end{array}\right)$$

So $\mathrm{rk}(A|b) = \mathrm{rk}(U|b') = n$ and by the Rouché-Capelli theorem $Ax = b$ has a unique solution.

$\Leftarrow$] Assume that the linear system $Ax = b$ has a unique solution for a $b \in \mathcal{M}_{\mathbb{K}}(n, 1)$. Then by the Rouché-Capelli theorem $\mathrm{rk}(A) = n$. □

*Remark.* Note that to check whether a square matrix $A \in \mathcal{M}_{\mathbb{K}}(n, n)$ has $\mathrm{rk}(A) = n$, it is enough to find a $b \in \mathcal{M}_{\mathbb{K}}(n, 1)$ such that $Ax = b$ has a unique solution, e.g. we check if $Ax = 0$ has a unique solution.

### 1.2.4 Inverse Matrix

For each $k \in \mathbb{K}, k \neq 0$, there exists $k' \in \mathbb{K}$ such that $k \cdot k' = 1$, $k'$ is the inverse element.
Can we do the same for matrices?

**Definition 1.16.**

- A matrix $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is *right invertible* if there exists a $B \in \mathcal{M}_{\mathbb{K}}(n,m)$ such that $AB = I_m$.
- A matrix $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is *left invertible* if there exists a $C \in \mathcal{M}_{\mathbb{K}}(n,m)$ such that $CA = I_n$.
- A matrix $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is *invertible* if it is right and left invertible.

**Proposition 1.17.** *Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be an invertible matrix, then*

1. *The right and the left inverse of $A$ (if they exist) are equal.*
2. *The inverse matrix of $A$ is unique (if it exists).*

*Proof.* 1. Assume there exist $B, C \in \mathcal{M}_{\mathbb{K}}(n,m)$ such that $AB = I_m$, $CA = I_n$ so

$$B = I_m B = (CA)B = C(AB) = CI_n = C$$

2. As 1., with $B, C$ inverse matrices of $A$. □

*Remark.* If $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is not a square matrix ($m \neq n$), then $A$ is not invertible!
We will prove this fact later in the course, so we stick to square matrices!

**Example.** Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(2,2)$. We want to determine if it is invertible, so we look for a matrix $X \in \mathcal{M}_{\mathbb{R}}(2,2)$ such that $AX = I_2$, and then we verify $XA = I_2$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix} = \begin{pmatrix} x_{1,1} + 2x_{2,1} & x_{1,2} + 2x_{2,2} \\ 3x_{1,1} + 4x_{2,1} & 3x_{1,2} + 4x_{2,2} \end{pmatrix} \longleftrightarrow \begin{cases} x_{1,1} + 2x_{2,1} &= 1 \\ 3x_{1,1} + 4x_{2,1} &= 0 \\ x_{1,2} + 2x_{2,2} &= 0 \\ 3x_{1,2} + 4x_{2,2} &= 1 \end{cases}$$

The linear system is equivalent to two linear systems (one in the variables $x_{1,1}, x_{2,1}$ and one in the variables $x_{1,2}, x_{2,2}$), both having $A$ as matrix of coefficient:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_{1,1} \\ x_{2,1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_{1,2} \\ x_{2,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

We use the Gauss-Jordan algorithm to solve them simultaneously!

$$(A|I_2) = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \xrightarrow{R_3 \to R_3 - 3R_1} \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right) \xrightarrow[R_2 \to -\frac{1}{2}R_2]{R_1 \to R_1 + R_2} \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{array}\right)$$

So $B = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$ satisfies $AB = I_2$, and it is straightforward to check that $BA = I_2$ holds true as well: $A$ is invertible and $A^{-1} = B$.

**Theorem 1.18.**
*Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix. Then the following are equivalent:*

1. *$A$ is invertible;*
2. *$\mathrm{rk}(A) = n$;*
3. *$Ax = b$ has a unique solution for any $b \in \mathcal{M}_{\mathbb{K}}(n,1)$.*

*Proof.* By Cramer's theorem we know that 2. and 3. are equivalent, so we prove the theorem by showing the following two implications: $1 \Rightarrow 3$ and $2 \Rightarrow 1$.

1. $\Rightarrow$ 3.] Assume $A$ is invertible, and consider the linear system $Ax = b$, for an arbitrary $b \in \mathcal{M}_{\mathbb{K}}(n,1)$. Then $x = A^{-1}b$ is a solution: $A(A^{-1}b) = (AA^{-1})b = I_n b = b$; and every solution has this form: $Ax = b \Rightarrow A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$. So $x = A^{-1}b$ is the unique solution.

2. $\Rightarrow$ 1.] The rank of $A$ is $n$, so by Cramer's theorem we can solve any linear system $Ax = b$ with $b \in \mathcal{M}_{\mathbb{K}}(n,1)$, and the solution is unique. Consider the $n$ linear systems:

$$A \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{n,1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad A \begin{pmatrix} x_{1,2} \\ x_{2,2} \\ \vdots \\ x_{n,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \dots, \ A \begin{pmatrix} x_{1,n} \\ x_{2,n} \\ \vdots \\ x_{n,n} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

and let $b_1, b_2, \dots, b_n \in \mathcal{M}_{\mathbb{K}}(n,1)$ be their solutions. Define $B \in \mathcal{M}_{\mathbb{K}}(n,n)$ as the matrix, whose columns are $b_1, b_2, \dots, b_n$. Then by construction $B$ is a right-inverse of $A$: $AB = I_n$; and we need to show $BA = I_n$.

Let us now consider the linear system $Bx = 0$. Applying $A$ on both sides we obtain $A(Bx) = A(0) = 0 \Rightarrow I_n x = 0 \Rightarrow x = 0$. So $Bx = 0$ has a unique solution and by Rouché-Capelli $\mathrm{rk}(B) = n$.

Arguing as for $A$, we can find a matrix $C$ such that $BC = I_n$: a right-inverse of $B$, but $A$ is a left-inverse of $B$, thus $BA = I_n = CB$. In other words $B$ is invertible, and $A = C$, so $A$ is invertible as well. □

In practice, given a square matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$, to determine whether it is invertible and find its inverse, we need to solve $n$ linear systems in $n$ variables, all having $A$ as matrix of coefficient, so it is more convenient to solve them simultaneously by using the Gauss-Jordan algorithm (as in the examples). There are 2 cases, visually:

*i)* $(A|I_n) \overset{\text{Gauss-Jordan}}{\longrightarrow} (I_n|B)$ then $A$ invertible and $A^{-1} = B$

*ii)* $(A|I_n) \overset{\text{Gauss-Jordan}}{\longrightarrow} (A'|*)$ where $A'$ has a zero row, then $A$ not invertible

**Example.** Let us determine the inverse of the following matrices:

$$i)\ A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 6 \\ 3 & 0 & 2 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3,3), \qquad ii)\ M = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(2,2)$$

$$i)\ \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 2 & 1 & 6 & 0 & 1 & 0 \\ 3 & 0 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\substack{R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 3R_1}} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 4 & -2 & 1 & 0 \\ 0 & 0 & -1 & -3 & 0 & 1 \end{array}\right) \xrightarrow{\substack{R_1 \to R_1 + R_3 \\ R_2 \to R_3 + 4R_3 \\ R_3 \to -R_3}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 0 & 1 \\ 0 & 1 & 0 & -14 & 1 & 4 \\ 0 & 0 & 1 & 3 & 0 & -1 \end{array}\right)$$

So $C = \begin{pmatrix} -2 & 0 & 1 \\ -14 & 1 & 4 \\ 3 & 0 & -1 \end{pmatrix}$ satisfies $AB = I_3 = BA = I_3$: $A$ is invertible and $A^{-1} = B$.

$$ii)\ \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{array}\right) \xrightarrow{R_2 \to R_2 - 2R_1} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right)$$

and we cannot go on, since the linear systems have no solutions. This means that $M$ is not invertible, indeed it has $\mathrm{rk}(B) = 1 < 2$.

We conclude seeing how the inverse behaves with respect to the operations of product and transpose.

**Proposition 1.19.** *Let $A, B \in \mathcal{M}_{\mathbb{K}}(n,n)$ be invertible matrices, then*

*i) $AB$ is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

*ii) $A^T$ is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

*Proof.* i) $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$.
Similarly, $(B^{-1}A^{-1})(AB) = I_n$.

ii) $A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$. Similarly, $(A^{-1})^TA^T = (AA^{-1})^T = I_n^T = I_n$. □

## 1.3 Determinant

**Definition 1.20.** The *determinant* is a function $\det : \mathcal{M}_{\mathbb{K}}(n,n) \to \mathbb{K}$ defined recursively:

$(n = 1)$ $\det(a_{1,1}) = a_{1,1}$;

$(n \geq 2)$ *Laplace expansion*: $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(\hat{A}_{i,j})$, where $i$ is a fixed row index and $\hat{A}_{i,j} \in \mathcal{M}_{\mathbb{K}}(n-1, n-1)$ is the matrix obtained from $A$ by removing the $i$-th row and the $j$-th column.

**Example.** $\mathbb{K} = \mathbb{R}, i = 1$:

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = (-1)^{1+1} 1 \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} + (-1)^{1+2} 2 \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + (-1)^{1+3} 3 \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}$$

*Remark.* i) Alternatively we can take the Laplace expansion along a fixed column (the $j$-th): $\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(\hat{A}_{i,j})$, where as above $\hat{A}_{i,j} \in \mathcal{M}_{\mathbb{K}}(n-1, n-1)$ is the matrix obtained from $A$ by removing the $i$-th row and the $j$-th column.

ii) The determinant of a matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ does not depend on the expansions.

**Example.** Let us compute the determinant of an arbitrary matrix of type $2 \times 2$, by expanding along the first column:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (-1)^{1+1} a \det(d) + (-1)^{1+2} c \det(b) = ad - bc.$$

**Example.** Let us compute the determinant of the matrix $A = \left(\begin{smallmatrix} 4 & 0 & 3 \\ 1 & 2 & 0 \\ 7 & 3 & 0 \end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{R}}(3,3)$. We expand it along the second column:

$$\begin{aligned} \det \begin{pmatrix} 4 & 0 & 3 \\ 1 & 2 & 0 \\ 7 & 3 & 0 \end{pmatrix} &= (-1)^{2+1} 0 \det \hat{A}_{2,1} + (-1)^{2+2} 2 \det \hat{A}_{2,2} + (-1)^{2+3} 3 \det \hat{A}_{2,3} \\ &= 2 \det \begin{pmatrix} 4 & 3 \\ 7 & 0 \end{pmatrix} - 3 \det \begin{pmatrix} 4 & 3 \\ 1 & 0 \end{pmatrix} = 2(4 \cdot 0 - 7 \cdot 3) - 3(4 \cdot 0 - 1 \cdot 3) = -42 + 9 = -33 \end{aligned}$$

*Remark.* Note that each time we have $a_{i,j} = 0$ we do not need to compute the corresponding hat matrix $\hat{A}_{i,j}$, so we should select the row/column with the highest number of zeroes.

**Example.** We expand the matrix $A$ along the third column:

$$\det \begin{pmatrix} 4 & 0 & 3 \\ 1 & 2 & 0 \\ 7 & 3 & 0 \end{pmatrix} = (-1)^{3+1} 3 \det \hat{A}_{3,1} + (-1)^{3+2} 0 \det \hat{A}_{3,2} + (-1)^{3+3} 0 \det \hat{A}_{3,3} = 3 \det \begin{pmatrix} 1 & 2 \\ 7 & 3 \end{pmatrix} = -33$$

Let see now some special cases:

**Definition 1.21.** Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix.
$A$ is an *upper triangular* matrix if $a_{ij} = 0$ for all $j < i$.
$A$ is a *lower triangular* matrix if $a_{ij} = 0$ for all $j > i$.
$A$ is a *diagonal* matrices: if $a_{ij} = 0$ for all $j \neq i$.

**Example.** $\left(\begin{smallmatrix} 0 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ is upper triangular, $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 4 & 0 & -1 \end{smallmatrix}\right)$ is lower triangular and $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{smallmatrix}\right)$ is diagonal. The identity matrix $I_n$ is a diagonal matrix, and every matrix in echelon form is upper triangular.

**Lemma 1.22.** *Let* $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ *be a triangular matrix.*
*Then* $\det A = a_{1,1} \cdot a_{2,2} \cdots a_{n,n}$*; in particular* $\det(I_n) = 1$.

*Proof.* Repeatedly expanding along the 1st column/row we have $\det A = a_{1,1} \det \hat{A}_{1,1} = a_{1,1} \cdot a_{2,2} \cdots a_{n,n}$. □

**Theorem 1.23.** *Let* $A, B \in \mathcal{M}_{\mathbb{K}}(n,n)$ *then*

- $\det(AB) = \det(A) \cdot \det(B)$ *(Binet's formula).*
- $\det(A) = \det(A^T)$.

**Corollary 1.24.** *Let* $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ *be an invertible matrix, then* $\det(A^{-1}) = \dfrac{1}{\det A}$.

*Proof.* $1 = \det(I_n) = \det(AA^{-1}) = \det(A) \cdot \det(A^{-1})$, so $\det(A^{-1}) = \frac{1}{\det A}$. □

### 1.3.1 Determinant and elementary row operations

**Properties.** *Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$. The elementary row operations affect the determinant as follows:*

- *Swapping to rows ($i \neq j$): $A \overset{R_i \leftrightarrow Rj}{\longrightarrow} B : \det B = -\det A$.*
- *Multiplying a row by a scalar $k \in \mathbb{K}$: $A \overset{R_i \to kR_i}{\longrightarrow} B : \det B = k \det A$.*
- *Adding to a row a multiple of another row: $A \overset{R_i \to R_i + kR_j}{\longrightarrow} B : \det B = \det A$.*

*Remark.* i) By the first property, if $A$ has 2 equal rows, then $\det A = 0$.
ii) By the second property we have: $\det(kA) = k^n \det(A)$ for any $A \in \mathcal{M}_{\mathbb{K}}(m,n)$.
iii) $\det(A) = \det(A^T)$, so the previous properties hold if we replace rows by columns.

**Example.** Let $A, B \in \mathcal{M}_{\mathbb{Q}}(3,3)$, with $\det A = -2$, $\det B = -3$, then

$$\det\left(\frac{1}{2}B^{-6}A^2B^5\right) = \left(\frac{1}{2}\right)^3 \frac{(\det A)^2}{\det B} = -\frac{1}{6}.$$

**Example.** Let us compute the determinant of the matrix $A = \begin{pmatrix} 0 & 0 & -3 & 5 \\ 1 & 8 & 6 & 7 \\ 2 & 0 & 4 & 7 \\ 3 & 0 & 6 & 9 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(4,4)$. We start expanding along the second column:

$$\begin{aligned} \det A &= 8\det\begin{pmatrix} 0 & -3 & 5 \\ 2 & 4 & 7 \\ 3 & 6 & 9 \end{pmatrix} \overset{R_3 \to \frac{1}{3}R_3}{=} 24\det\begin{pmatrix} 0 & -3 & 5 \\ 2 & 4 & 7 \\ 1 & 2 & 3 \end{pmatrix} \overset{R_2 \to R_2 - 2R_3}{=} 24\det\begin{pmatrix} 0 & -3 & 5 \\ 0 & 0 & 1 \\ 1 & 2 & 3 \end{pmatrix} \\ &= -24\det\begin{pmatrix} 0 & -3 \\ 1 & 2 \end{pmatrix} = -72 \end{aligned}$$

We conclude this chapter by giving a characterisation of invertible matrices in terms of determinant.

**Theorem 1.25** (Characterisation of invertible matrices)**.**
*Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix. Then the following are equivalent:*

1. *$A$ is invertible;*
2. *$\mathrm{rk}(A) = n$;*
3. *$Ax = b$ has a unique solution for any $b \in \mathcal{M}_{\mathbb{K}}(n,1)$;*
4. *$\det(A) \neq 0$.*

*Proof.* By Theorem 1.18, we know 1. $\Leftrightarrow$ 2. $\Leftrightarrow$ 3., so we are left to show 1. $\Leftrightarrow$ 4..
1. $\Rightarrow$ 4.] As in the proof of Corollary 1.24, if $A$ is invertible, then $\det(A) \cdot \det(A^{-1}) = 1$, thus $\det(A) \neq 0$.
4. $\Rightarrow$ 1.] Applying the Gauss algorithm to $A$ we get a matrix $U$ in echelon form with $\det(U) = c\det(A)$ for some $c \in \mathbb{K}$, $c \neq 0$, so $\det U \neq 0$. Being in echelon form $U$ is also

upper triangular: $0 \neq \det(U) = u_{1,1} \cdots u_{n,n}$, so the pivots appear on the diagonal:

$$U = \begin{pmatrix} u_{1,1} & * & * & \cdots & * \\ 0 & u_{2,2} & * & \cdots & * \\ 0 & 0 & \ddots & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \dots & & u_{n,n} \end{pmatrix}$$

We get $n = \mathrm{rk}(U) = \mathrm{rk}(A)$, so $A$ is invertible. □

# Chapter 2

# Vector spaces

In the first chapter we discussed the algebraic aspect of the objects we introduced: matrices, linear system, solutions and determinant.

In this chapter, we start to look at them from a geometric point of view which generalize what we saw in the Introduction and will explain certain choices made in the previous chapter.

## 2.1 Vector spaces

**Definition 2.1.** A *vector space* over a field $\mathbb{K}$ is a set $V$ with two operations: a *sum* $+ : V \times V \to V$ and a *product by scalar* $\cdot : \mathbb{K} \times V \to V$ satisfying:

VS1) $\forall u, v, w \in V\colon\ (u + v) + w = u + (v + w)$ (associativity of $+$)

VS2) $\forall v, w \in V\colon\ v + w = w + v$ (commutativity of $+$)

VS3) $\exists 0 \in V\colon\ v + 0 = 0 + v = v\ \forall v \in V$ (neutral element of $+$: *zero-vector* or *null-vector*)

VS4) $\forall v \in V, \exists v' \in V\colon\ v + v' = v' + v = 0$ (additive inverse)

VS5) $\forall c, d \in \mathbb{K}, v \in V\colon\ (c \cdot d) \cdot v = c \cdot (d \cdot v)$ (associativity of $\cdot$)

VS6) $\forall v \in V : 1 \cdot v = v$ (neutral element of $\cdot$)

VS7) $\forall c, d \in \mathbb{K}, v \in V\colon\ (c + d) \cdot v = c \cdot v + d \cdot v$ (distributivity)

VS8) $\forall c \in \mathbb{K}, v, w \in V\colon\ c \cdot (v + w) = c \cdot v + c \cdot w$ (distributivity)

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

**Notation.** Sometimes we call $V$ a $\mathbb{K}$-*vector space*.

**Example.**

- $\mathcal{M}_{\mathbb{K}}(m, n)$ is a vector space over $\mathbb{K}$ (with the operations defined in Chapter 1).
- free vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ form a vector space over $\mathbb{R}$ (with the operations seen in the Introduction).
- the set of continuous real functions $\mathcal{C}^0(\mathbb{R})$ is a vector space over $\mathbb{R}$.

- the set $\mathbb{K}[t] = \{\sum_{k=1}^{n} a_k t^k\}$ of polynomial with $\mathbb{K}$-coefficients is a vector space over the field $\mathbb{K}$ with respect to the usual operations:

$$\left(\sum_k a_n t^k\right) + \left(\sum_k^n b_k t^k\right) = \sum_k^n (a_k + b_k) t^k\,, \quad \lambda\left(\sum_k a_n t^k\right) = \sum_k (\lambda a_n) t^k$$

- the solution set of a homogeneous linear system $\mathrm{Sol}(A|0)$, $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is a vector space over $\mathbb{K}$ with respect to the sum and product by a scalar for matrices: let $v_1, v_2 \in \mathrm{Sol}(A|0)$, then

$$A(v_1 + v_2) = Av_1 + Av_2 = 0 + 0 = 0 \quad A(\lambda v_1) = \lambda A v_1 = \lambda 0 = 0$$

**Notation**. There are two zeroes around: the scalar zero and the zero vector. If it is not clear from the context, we will resolve this ambiguity by denoting by $0_{\mathbb{K}}$ the scalar zero and by $0_V$ the null vector of $V$.

*Remark.* i) In VS4), the additive inverse of $v \in V$ is $v' = (-1) \cdot v = -v$.
ii) $0_{\mathbb{K}} \cdot v = 0_V$ and it holds the zero-product property: $\lambda \cdot v = 0_V \Rightarrow \lambda = 0_{\mathbb{K}}$ or $v = 0_V$.
iii) The zero vector and the the additive inverse of $v \in V$ are unique.

## 2.2 Vector subspaces

**Definition 2.2.** Let $(V, +, \cdot)$ be a vector space over $\mathbb{K}$. A subset $W \subset V$ is a *vector subspace* (and one usually write $W \triangleleft V$) if the following conditions hold:

S1) $0_V \in W$;

S2) $\forall w_1, w_2 \in W$ it holds $w_1 + w_2 \in W$;

S3) $\forall w \in W$ and $\lambda \in \mathbb{K}$ it holds $\lambda \cdot w \in W$.

*Remark.* This three conditions ensure that $W$ is a vector space over $\mathbb{K}$ with the operations inherited from $V$.

**Example.**

- $W = \{0_V\} \triangleleft V$; $V \triangleleft V$.
- the set of differentiable real functions $\mathcal{C}^1(\mathbb{R})$ is a vector subspace of $\mathcal{C}^0(\mathbb{R})$.
- the set $\mathbb{K}[t]_{\leq d} = \{p(t) \in \mathbb{K}[t] \mid \deg(p) \leq d\}$ of polynomial of degree $\leq d$ is a vector subspace of $\mathbb{K}[t]$, actually:

$$\mathbb{K}[t]_{\leq d} \triangleleft \mathbb{K}[t]_{\leq d+n} \triangleleft \mathbb{K}[t]\,, \qquad \text{for } k \in \mathbb{N}.$$

- the solution set of a homogeneous linear system $\mathrm{Sol}(A|0)$, $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ is a vector subspace of $\mathcal{M}_{\mathbb{K}}(n,1)$, e.g. $W = \{(x,y,z)^T \in \mathbb{R}^3 \mid x + 2y - z = 0\} \triangleleft \mathbb{R}^3$.
- the solution set of a non-homogeneous linear system $\mathrm{Sol}(A|b)$, $b \neq 0$ is not a vector subspace of $V = \mathcal{M}_{\mathbb{K}}(n,1)$, since $b \neq A \cdot 0_V = 0$, e.g. $W = \{(x,y,z)^T \in \mathbb{R}^3 \mid x + 2y - z = 1\} \not\triangleleft \mathbb{R}^3$.
- $\{(x,y,z)^T \in \mathbb{R}^2 \mid x \geq 0, y \geq 0\}$ and $\{(x,y,z)^T \in \mathbb{R}^2 \mid xy \geq 0\}$ are not vector subspaces of $\mathbb{R}^2$.

### 2.2.1 Intersection and sum of vector subspaces

**Lemma 2.3.** *Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \triangleleft V$ be vector subspaces, then $U \cap W$ is a vector subspace too.*

*Proof.* S1) $0_V \in U$ and $0_V \in W$, so $0_V \in U \cap W$.
S2) Let $v_1, v_2 \in U \cap W$, then $v_1 + v_2 \in U$ (since $U \triangleleft V$) and $v_1 + v_2 \in W$ (since $W \triangleleft V$), so $v_1 + v_2 \in U \cap W$.
S3) Let $v \in U \cap W$ and $\lambda \in \mathbb{K}$, then $\lambda v \in U$ (since $U \triangleleft V$) and $\lambda v \in W$ (since $W \triangleleft V$), so $\lambda v \in U \cap W$. □

**Example.** Let $W = \{(x, y, z)^T \in \mathbb{R}^3 \mid x + 2y - z = 0\} \triangleleft \mathbb{R}^3$ and let $U = \{(x, y, z)^T \in \mathbb{R}^3 \mid -y + 3z = 0\} \triangleleft \mathbb{R}^3$, then

$$U \cap W = \left\{ (x, y, z)^T \in \mathbb{R}^3 \mid \left\{ \begin{array}{rcl} x + 2y - z &=& 0 \\ -y + 3z &=& 0 \end{array} \right\} \right\} = \left\{ \begin{pmatrix} -5 \\ -3 \\ 1 \end{pmatrix} t \mid t \in \mathbb{R} \right\}.$$

What about the union?

*Remark.* In general the union of two subspaces is not a subspace, for example let $U = \{(x, y)^T \in \mathbb{R}^2 \mid x + y = 0\} \triangleleft \mathbb{R}^2$ and $W = \{(x, y)^T \in \mathbb{R}^2 \mid x - y = 0\} \triangleleft \mathbb{R}^2$, then $(1, 1)$ and $(1, -1)$ belong to $U \cup W$, but $(1, 1) + (1, -1) = (2, 0) \notin U \cup W$.

**Definition 2.4.** Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \triangleleft V$ be vector subspaces. We define their *sum* $U + W$ as

$$U + W = \{v \in V | \exists u \in U, w \in W \text{ such that } v = u + w\}.$$

"$U + W$ contains all vectors that can be decomposed as sum of a vector in $U$ and a vector in $W$".

**Lemma 2.5.** *Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \triangleleft V$ be vector subspaces, then*

1. *$U + W$ is a vector subspace too.*
2. *$U \triangleleft U + W$ and $W \triangleleft U + W$*
3. *$U \cap W \triangleleft U$, $U \cap W \triangleleft W$.*

*Proof.* 1. S1) $0_V = \underbrace{0_V}_{\in U} + \underbrace{0_V}_{\in W} \in U + W$.

S2) Let $v_1 = u_1 + w_1, v_2 = u_2 + w_2 \in U + W$, with $u_1, u_2 \in U$, $w_1, w_2 \in W$.
Then $v_1 + v_2 = \underbrace{(u_1 + u_2)}_{\in U} + \underbrace{(w_1 + w_2)}_{\in W} \in U + W$.

S3) Let $\lambda \in \mathbb{K}$ and $v = u + w \in U \cap W$ with $u \in U$, $w \in W$.
Then $\lambda v = \underbrace{(\lambda u)}_{\in U} + \underbrace{(\lambda w)}_{\in W} \in U + W$.

2. and 3. are obvious. □

**Example.** a) Let $U = \{(x, y)^T \in \mathbb{R}^2 \mid x + y = 0\} \triangleleft \mathbb{R}^2$ and $W = \{(x, y)^T \in \mathbb{R}^2 \mid x - y = 0\} \triangleleft \mathbb{R}^2$, then $U + W = \mathbb{R}^2$, indeed every $(a, b) \in \mathbb{R}^2$ can be written as $(a, b) = \left(\frac{a-b}{2}, \frac{b-a}{2}\right) + \left(\frac{a+b}{2}, \frac{a+b}{2}\right) \in U + W$

b) Let $U = \left\{ \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right) t \mid t \in \mathbb{R} \right\} \triangleleft \mathbb{R}^3$ and $W = \left\{ \left(\begin{smallmatrix} 0 \\ 2 \\ -1 \end{smallmatrix}\right) s \mid s \in \mathbb{R} \right\} \triangleleft \mathbb{R}^3$, then

$$U + W = \left\{ \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right) t + \left(\begin{smallmatrix} 0 \\ 2 \\ -1 \end{smallmatrix}\right) s \mid t, s \in \mathbb{R} \right\} \triangleleft \mathbb{R}^3.$$

**Definition 2.6.** Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \lhd V$ be vector subspaces. The sum $U + W$ is called *direct* (denoted $U \oplus W$) if $U \cap W = \{0_V\}$.

The meaning of this definition is: "The sum $U + W$ is direct if all vectors can be decomposed in a unique way as sum of a vector in $U$ and a vector in $W$".

**Example.** In the last 2 examples the sum is direct, in particular in a) $U \oplus W = \mathbb{R}^2$.

## 2.3 Describing elements in a vector space

**Definition 2.7.** Let $V$ be a vector space over the field $\mathbb{K}$, and let $\{v_1, \ldots, v_n\} \subset V$ be a finite subset of vectors.

A *linear combination* of $\{v_1, \ldots, v_n\}$ is an element of $V$ of the form

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \qquad \text{with } \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{K}.$$

The set of all linear combinations of $\{v_1, \ldots, v_n\}$ is called the *span* of $\{v_1, \ldots, v_n\}$:

$$\mathrm{Span}(v_1, \ldots, v_n) = \{\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \mid \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{K}\}.$$

**Proposition 2.8.** *Let $V$ be a $\mathbb{K}$-vector space, and let $\{v_1, \ldots, v_n\} \subset V$ be a finite subset of vectors. Then $\mathrm{Span}(v_1, \ldots, v_n)$ is a vector subspace of $V$.*

*Proof.* S1) Pick $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0_{\mathbb{K}}$: $0v_1 + 0v_2 + \ldots + 0v_n = 0_V \in \mathrm{Span}(v_1, \ldots, v_n)$.

S2) Let $u = \sum_{i=1}^n \lambda_i v_i$ and $w = \sum_{i=1}^n \mu_i v_i$, then $u+w = \sum_{i=1}^n (\lambda_i+\mu_i) v_i \in \mathrm{Span}(v_1, \ldots, v_n)$.

S3) Let $u = \sum_{i=1}^n \lambda_i v_i$ and let $\gamma \in \mathbb{K}$, then $\gamma u = \sum_{i=1}^n (\gamma\lambda_i) v_i \in \mathrm{Span}(v_1, \ldots, v_n)$. □

**Example.**

1. $\mathbb{K}[t]_{\leq d} = \{p(t) \in \mathbb{K}[t] \mid \deg(p) \leq d\} = \mathrm{Span}(1, t, \ldots, t^d)$.
2. In $\mathbb{K}^3 (= \mathcal{M}_{\mathbb{K}}(3, 1))$ let $S = \left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}\right\}$, then
$$\mathrm{Span}(S) = \left\{\lambda_1\begin{pmatrix}1\\1\\0\end{pmatrix} + \lambda_2\begin{pmatrix}1\\0\\1\end{pmatrix} \mid \lambda_1, \lambda_2 \in \mathbb{K}\right\} = \left\{\begin{pmatrix}\lambda_1+\lambda_2\\\lambda_1\\\lambda_2\end{pmatrix} \mid \lambda_1, \lambda_2 \in \mathbb{K}\right\}.$$
3. Similarly, let $S = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\2\\-1\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$, then
$$\mathrm{Span}(S) = \left\{\begin{pmatrix}\lambda_1+\lambda_3\\\lambda_2+2\lambda_2\\-\lambda_3+\lambda_4\end{pmatrix} \mid \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{K}\right\}.$$

*Remark.* Note that a linear system $Ax = b$ is solvable if and only if $b$ is a linear combination of the columns of $A$! In other words, $Ax = b$ is solvable if and only if $b \in \mathrm{Span}(c_1, \ldots, c_n)$, where $c_i \in \mathcal{M}_{\mathbb{K}}(m, 1)$ are the columns of $A \in \mathcal{M}_{\mathbb{K}}(m, n)$.

For example ($n = 3$): $\begin{pmatrix}1\\1\\2\end{pmatrix} \in \mathrm{Span}\left(\begin{pmatrix}2\\0\\2\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}\right)$, indeed, the linear system $\begin{pmatrix}2&1\\0&-1\\2&0\end{pmatrix}\begin{pmatrix}\lambda_1\\\lambda_2\end{pmatrix} = \begin{pmatrix}1\\1\\2\end{pmatrix}$ is solvable ($\lambda_1 = 1, \lambda_2 = -1$).

### 2.3.1 Generators

Now we want to compare $V$ with $\mathrm{Span}(v_1, \dots, v_n)$.

**Definition 2.9.** Let $V$ be a vector space over the field $\mathbb{K}$. A finite subset $\{v_1, \dots, v_n\} \subset V$ is a *set of generators* of $V$ if $\mathrm{Span}(v_1, \dots, v_n) = V$ (i.e. if every element in $V$ is a linear combination of $\{v_1, \dots, v_n\}$).

**Notation.** Other common terminology are: $v_1, \dots, v_n$ are *generators* of $V$; $v_1, \dots, v_n$ *generate* $V$.

*Remark.* i) To be generators means that $\{v_1, \dots, v_n\}$ are enough to describe $V$!
ii) If $\{v_1, \dots, v_n\} \subseteq \{v_1, \dots, v_n, v_{n+1}, \dots, v_s\}$ and $\mathrm{Span}(v_1, \dots, v_n) = V$, then $\mathrm{Span}(v_1, \dots, v_n, \dots, v_s) = V$.

**Example.** We continue the previous examples.

1. $\{1, t, \dots, t^d\}$ is a set of generators of $\mathbb{K}[t]_{\leq d}$.
2. $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$ is not a set of generators of $\mathbb{K}^3$, e.g. the linear system
$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$
is not solvable.
3. $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a set of generators of $\mathbb{K}^3$, indeed the linear system
$$\left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & b_1 \\ 0 & 1 & 2 & 0 & b_2 \\ 0 & 0 & -1 & 1 & b_3 \end{array} \right)$$
is solvable for any $(b_1, b_2, b_3)^T \in \mathbb{K}^3$.

*Remark.* How many elements must have a set of generators of $\mathbb{K}^m$?
We need that the "corresponding" matrix (with $m$ rows) has $m$ pivots on the left (so that the linear system is solvable), thus we need at least $m$ columns: in other words, a set of generators of $\mathbb{K}^m$ has at least $m$ elements!

#### Finite generation

**Definition 2.10.** A vector space $V$ is *finitely generated* if there exist a finite subset $\{v_1, \dots, v_n\} \subset V$ such that $V = \mathrm{Span}(v_1, \dots, v_n)$.

**Example.**

- $\mathbb{K}[t]_{\leq d} = \mathrm{Span}(1, t, \dots, t^d)$ is finitely generated.
- $\mathbb{K}^n$, $\mathcal{M}_{\mathbb{K}}(m, n)$ are finitely generated:
$$\mathbb{K}^n = \mathrm{Span}\left( \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right)$$

$$\mathcal{M}_{\mathbb{K}}(m,n) = \operatorname{Span}\left( \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \right)$$

- If $V$ is finitely generated, any vector subspace $W \triangleleft V$ is finitely generated too.

**Example.**
- $\mathbb{K}[t]$ is not finitely generated: the linear combinations we may obtain from a finite set of polynomial cannot have arbitrary high degree.
- $\mathcal{C}^0(\mathbb{R})$ is not finitely generated.

*Good news!* In this course we consider only finitely generated vector spaces. From now on, *every vector space we consider will be implicitly assumed to be finitely generated.*

### 2.3.2 Linear independence

**Example.** i) $T = \{1, t, t^2, t^3\}$ is a set of generators of $\mathbb{K}[t]_{\leq 3}$, and $p(t) = (1+t)^3$ can be written in a unique way as linear combination of the vectors in $T$: $(1+t^3) = 1+3t+3t^2+t^3$.

ii) $S = \left\{ \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 2 \\ -1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right) \right\}$ is a set of generators of $\mathbb{K}^3$, and every $v \in \mathbb{K}^3$ (e.g. $\left(\begin{smallmatrix} 2 \\ -3 \\ 5 \end{smallmatrix}\right)$) can be written in several ways as linear combination of the vectors in $S$, indeed by the Rouché-Capelli theorem, the solution set of the linear system

$$\left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & 2 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & -1 & 1 & 5 \end{array} \right)$$

depends on one parameter: $\text{Sol} = \left\{ \left(\begin{smallmatrix} 2 \\ -3 \\ 0 \\ 5 \end{smallmatrix}\right) + \left(\begin{smallmatrix} -t \\ -2t \\ t \\ t \end{smallmatrix}\right) \mid t \in \mathbb{K} \right\}$. We would like to remove this ambiguity, and have a unique representation!

*Remark.* In $\mathbb{K}^m$, if we consider more than $m$ vectors, we cannot have a unique representation (if it exists): indeed, the "corresponding" matrix (with $m$ rows) would have more columns than rows (see Lemma 1.14).

In other words, in the previous example the ambiguity arises because we have to many variables/columns, so the idea is to remove vectors.

**Example.** We continue the previous example and we consider the representation of $\left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right)$, so we solve the corresponding homogeneous linear system

$$\left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \end{array} \right)$$

By Theorem 1.13 the solutions of the homogeneous linear system are $\left\{ \left(\begin{smallmatrix} -t \\ -2t \\ t \\ t \end{smallmatrix}\right) \mid t \in \mathbb{K} \right\}$ and picking $t = 1$, we get $\left(\begin{smallmatrix} 1 \\ 2 \\ -1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right) + 2\left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right) - 1\left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right) \in \operatorname{Span}(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right))$, so we may remove $\left(\begin{smallmatrix} 1 \\ 2 \\ -1 \end{smallmatrix}\right)$ from $S$ without changing the span: $\operatorname{Span}(S) = \operatorname{Span}(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right))$.

**Definition 2.11.** Let $V$ be a vector space over the field $\mathbb{K}$. A finite subset $\{v_1, \ldots, v_n\} \subset V$ is a *set of linearly independent vectors* if $\lambda_1 \cdot v_1 + \cdots \lambda_n \cdot v_n = 0_V \Rightarrow \lambda_1 = \ldots = \lambda_n = 0_{\mathbb{K}}$.

*Remark.* a) In other words, we require that $0_V$ has a unique representation as linear combination of $\{v_1, \ldots, v_n\}$.

b) To be linearly independent vectors means that we do not have redundant information.

**Notation.** If $\{v_1, \ldots, v_n\} \subset V$ is not a set of linearly independent vectors, we say that they are *linearly dependent*.
Another common terminology is: $v_1, \ldots, v_n$ are *linearly (in-)dependent*.

*Remark.* c) Two vectors $v_1, v_2 \in V$ are linearly dependent if and only if one is a multiple of the other.

d) If $S = \{v_1, \ldots, v_n\}$ is a set of linearly independent vectors, then any subset of $S$ is a set of linearly independent vectors too.

**Example.** i) $\{1, t, \ldots, t^d\}$ is a set of linearly independent vectors of $\mathbb{K}[t]_{\leq d}$.

ii) $S = \left\{ \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\2\\-1\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right) \right\}$ is a set of linearly dependent vectors, indeed

$$\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right) + 2\left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right) - 1\left(\begin{smallmatrix}1\\2\\-1\end{smallmatrix}\right) - 1\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right).$$

iii) $\left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right)$ are linearly independent, since the homogeneous linear system

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right)$$

has $(0, 0)^T$ as unique solution.

**Proposition 2.12.** *Let $V$ be a vector space over the field $\mathbb{K}$, and let $\{v_1, \ldots, v_n\} \subset V$ be a finite subset.*

1. *$\{v_1, \ldots, v_n\}$ is a set of linearly dependent vectors if and only if one of the vectors is linear combination of the others.*
2. *If $\{v_1, \ldots, v_n\}$ is a set of linearly independent vectors then $v_1, \ldots, v_n, v_{n+1}$ are linearly dependent if and only if $v_{n+1} \in \operatorname{Span}(v_1, \ldots, v_n)$.*

*Proof.* 1. $\Rightarrow]$ If $v_1, \ldots, v_n$ are linearly dependent there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ non all zero (e.g. $\lambda_j \neq 0_{\mathbb{K}}$), such that

$$\sum_{i=1}^{n} \lambda_i v_i = 0_V \Longrightarrow v_j = \sum_{i=1, i\neq j}^{n} -\left(\frac{\lambda_i}{\lambda_j}\right) v_i$$

$\Leftarrow]$ If $v_k = \sum\limits_{i=1, i\neq k}^{n} \alpha_i v_i$ $(\alpha_i \in \mathbb{K})$, then $\sum\limits_{i=1}^{k-1} \alpha_i v_i + (-1)v_k + \sum\limits_{i=k+1}^{n} \alpha_i v_i = 0$.

2. $\Leftarrow]$ The same as 1. $\Leftarrow]$.

$\Rightarrow]$ Let $\lambda_1, \ldots, \lambda_n, \lambda_{n+1} \in \mathbb{K}$ be not all zero scalars, such that $\sum\limits_{i=1}^{n+1} \lambda_i v_i = 0_V$. Since $v_1, \ldots, v_n$ are linearly independent vectors, then $\lambda_{n+1} \neq 0_{\mathbb{K}}$, and arguing as above (1. $\Rightarrow]$), we conclude $v_{n+1} \in \operatorname{Span}(v_1, \ldots, v_n)$. □

This proposition has the following consequence.

**Corollary 2.13.** *If $\{v_1, \ldots, v_n\} \subset V$ is a set of linearly dependent vectors, then $\{v_1, \ldots, v_n, v_{n+1}, \ldots, v_s\} \subset V$ is set of linearly dependent vectors as well.*

*In particular, $\{0_V, v_2, \ldots, v_n\} \subset V$ is a set of linearly dependent vectors.*

We conclude this section with a numerical relation between a set of generators and a set of linearly independent vectors

**Theorem 2.14.** *Let $V$ be a vector space over the field $\mathbb{K}$, let $\{v_1, \ldots, v_n\} \subset V$ be a set of linearly independent vectors and $\{w_1, \ldots, w_m\} \subset V$ be a set of generators. Then $n \leq m$.*

*Proof.* Aiming for a contradiction, let us assume $m < n$. Since $v_1 \in V = \operatorname{Span}(w_1, \ldots, w_m)$, there are scalars $\lambda_1, \ldots, \lambda_m \in \mathbb{K}$ , such that $v_1 = \lambda_1 w_1 + \cdots + \lambda_m w_m$. Since $v_1 \neq 0_V$, not all $\lambda_i$ are zero, without lose of generality we may assume $\lambda_1 \neq 0_{\mathbb{K}}$, so

$$\lambda_1 w_1 = v_1 - \lambda_2 w_2 - \ldots - \lambda_m w_m \Longrightarrow w_1 = \lambda_1^{-1} v_1 - \sum_{i=2}^{m} (\lambda_1^{-1}\lambda_i) w_i$$

Thus $V = \operatorname{Span}(w_1, w_2, \ldots, w_m) = \operatorname{Span}(v_1, w_2 \ldots, w_m)$.

We consider now $v_2 \in V = \operatorname{Span}(v_1, w_2, \ldots, w_m)$: there are scalars $\mu_1, \mu_2, \ldots, \mu_m \in \mathbb{K}$, such that $v_2 = \mu_1 v_1 + \mu_2 w_2 + \cdots + \mu_m w_m$. Since $v_1, v_2$ are linearly independent, not all $\mu_2, \ldots, \mu_m$ are zero and without lose of generality we may assume $\mu_2 \neq 0_{\mathbb{K}}$, so

$$w_2 = -(\mu_2^{-1}\mu_1) v_1 + \mu_2^{-1} v_2 - \sum_{i=3}^{m} (\mu_2^{-1}\mu_i) w_i$$

Thus $V = \operatorname{Span}(v_1, w_2, w_3, \ldots, w_m) = \operatorname{Span}(v_1, v_2, w_3, \ldots, w_m)$.

Repeating this process $m$-times ($m < n$), we get $V = \operatorname{Span}(w_1, \ldots, w_m) = \operatorname{Span}(v_1, \ldots, v_m)$, so $v_{m+1} \in \operatorname{Span}(v_1, \ldots, v_m)$, contradicting the fact that $v_1, \ldots, v_m, v_{m+1}$ are linearly independent. □

### 2.3.3 Bases and dimension

**Definition 2.15.** Let $V$ be a vector space over the field $\mathbb{K}$. A finite subset $\mathcal{B} = \{v_1, \ldots, v_n\} \subset V$ is a *basis* of $V$ if $\{v_1, \ldots, v_n\}$ is a set of linearly independent generators.

**Example.**

- $\mathcal{E} = \left\{ \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{e_1}, \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{e_2}, \ldots, \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{e_n} \right\}$ is a basis of $\mathbb{K}^n$ called the *canonical basis*.

- Similarly,

$$\mathcal{E} = \left\{ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \right\}$$

is a basis of $\mathcal{M}_{\mathbb{K}}(m, n)$ called the *canonical basis*.

- For any $\alpha \in \mathbb{K}$ the set $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \\ 1 \end{pmatrix} \right\}$ is a basis of $\mathbb{K}^2$.
- $\mathcal{E} = \{1, t, \ldots, t^d\}$ is a basis of $\mathbb{K}[t]_{\leq d}$ called the *canonical basis*.

To be a basis means that $\{v_1, \ldots, v_n\}$ are enough to describe $V$, and we do not carry redundant information. Indeed:

**Proposition 2.16.** *Let $V$ be a vector space over the field $\mathbb{K}$, and let $\mathcal{B} = \{v_1, \ldots, v_n\} \subset V$ be a basis of $V$. Then every element of $V$ can be written uniquely as a linear combination of $\{v_1, \ldots, v_n\}$.*

*Proof.* $V = \operatorname{Span}(v_1, \ldots, v_n)$, so for every $v \in V$ there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ such that $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$.

To prove the uniqueness, let us pick two representations of $v$:

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n \,, \qquad v = \mu_1 v_1 + \ldots + \mu_n v_n.$$

Considering the difference, we get $0_V = v - v = (\lambda_1 - \mu_1)v_1 + \ldots + (\lambda_n - \mu_n)v_n$; but $v_1, \ldots, v_n$ are linearly independent, and so $\lambda_1 - \mu_1 = 0_{\mathbb{K}}, \ldots, \lambda_n - \mu_n = 0_{\mathbb{K}}$, i.e. $\lambda_1 = \mu_1, \ldots, \lambda_n = \mu_n$. □

**Definition 2.17.** Let $V$ be a vector space over the field $\mathbb{K}$, $\mathcal{B} = \{v_1, \ldots, v_n\} \subset V$ be a basis of $V$, so that for every $v \in V$ there are unique scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ such that

$$v = \sum_{i=1}^{n} \lambda_i v_i \,.$$

The scalars $\lambda_1, \ldots, \lambda_n$ are the *coordinates* of $v$ with respect to the basis $\mathcal{B}$, and are denoted by $(v)_{\mathcal{B}} = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{K}^n$.

What do two basis of a vector space have in common? The number of elements!

**Lemma 2.18.** *Let $V$ be a vector space over the field $\mathbb{K}$, and let $\mathcal{B} = \{b_1, \ldots, b_n\} \subset V$ and $\mathcal{C} = \{c_1, \ldots, c_m\} \subset V$ be bases of $V$. Then $m = n$.*

*Proof.* Using that $b_1, \ldots, b_n$ are linearly independent and that $c_1, \ldots, c_m$ are generators, we obtain $n \leq m$ by Theorem 2.14. Switching the properties: $b_1, \ldots, b_n$ are generators and that $c_1, \ldots, c_m$ are linearly independent, we obtain $m \leq n$ by Theorem 2.14. □

**Definition 2.19.** Let $V$ be a (finitely generated) vector space over the field $\mathbb{K}$. The *dimension* of $V$ is the number of elements of a basis of $V$ and it is denoted by $\dim V$.

**Example.**

- $\dim\{0\} = 0$, indeed $\mathcal{B} = \emptyset$ is a basis of $\{0\}$.
- $\dim \mathbb{K}^n = n$.
- $\dim \mathcal{M}_{\mathbb{K}}(m, n) = m \cdot n$.
- $\dim \mathbb{K}[t]_{\leq d} = d + 1$.

- Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, then $\mathrm{Sol}(A|0) \lhd \mathbb{K}^n = \mathcal{M}_{\mathbb{K}}(n,1)$ is a vector subspace. The Rouché-Capelli Theorem tells us $\mathrm{Sol}(A|0) = \{\lambda_1 v_1 + \cdots \lambda_{n-r} v_{n-r} \mid \lambda_i \in \mathbb{K}\} = \mathrm{Span}(v_1, \ldots, v_{n-r})$ where $r = \mathrm{rk}(A)$, and it guarantees that there are no redundant generators, i.e. $v_1, \ldots, v_{n-r}$ are linearly independent, so $\dim \mathrm{Sol}(A|0) = n - \mathrm{rk}(A)$.

**Proposition 2.20.** *Let $V$ be a vector space over the field $\mathbb{K}$. The following claims hold true.*

1. *Let $W \lhd V$, then $\dim W \leq \dim V$, and $\dim W = \dim V \Leftrightarrow W = V$.*
2. *If $\dim V = n$ and $\{v_1, \ldots, v_p\}$ is a set of linearly independent vectors, then $p \leq n$.*

   *Moreover, if $p = n$ then $\{v_1, \ldots, v_p\}$ is a basis, while if $p < n$ there exist vectors $v_{p+1}, \ldots, v_n$ such that $\{v_1, \ldots, v_p, v_{p+1}, \ldots, v_n\}$ is a basis of $V$.*
   *Each set of linearly independent vectors can be completed to a basis.*
3. *If $\dim V = n$ and $\{w_1, \ldots, w_m\}$ is a set of generators of $V$, then $m \geq n$.*

   *Moreover, if $m = n$ then $\{w_1, \ldots, w_m\}$ is a basis, while if $m > n$ we can remove $m - n$ vectors in such a way that the remaining ones are a basis of $V$.*
   *From each set of generators we can extract a basis.*

*Proof.* 1. Let $\{w_1, \ldots, w_s\} \subset W \subset V$ be a basis of $W$, in particular it is a set of linearly independent vectors in $V$. By Theorem 2.14 $\dim W = s \leq \dim V$.
If $W = V$, clearly $\dim W = \dim V$.
Now, let $\dim W = \dim V = n$ and assume there exists $v \in V$, $v \notin W$. Adding $v$ to a basis $\{w_1, \ldots, w_n\}$ of $W$, we get $n + 1$ linearly independent in $V$

Then $v \notin \mathrm{Span}(w_1, \ldots, w_n)$, so $\{w_1, \ldots, w_n, v\}$ are linearly independent in $V$ (by 2. of Proposition 2.12), contradiction.

2. The first claim follows from Theorem 2.14.
For the second one, if$p = n$, pick $W = \mathrm{Span}(v_1, \ldots, v_p)$ and apply 1; else, if $p < n$, then $\mathrm{Span}(v_1, \ldots, v_p) \neq V$, so there exists $v_{p+1} \notin \mathrm{Span}(v_1, \ldots, v_p)$. By Proposition 2.12 $v_1, \ldots, v_p, v_{p+1}$ are linearly independent. Repeating $n - p$ times this step, we eventually get a basis of $V$.

3. The first claim follows from Theorem 2.14.
If $w_1, \ldots, w_m$ are linearly dependent, by 1. of Proposition 2.12 one of the vectors (say $w_m$) is linear combination of the others, so we could exclude it and still have a set of generators.

Now, if $m = n$, then $w_1, \ldots, w_m$ cannot be linearly dependent, otherwise we may exclude one vector and get a set of generators of $V$ of cardinality $n - 1$, impossible.

On the other hand, if $m > n$, then $w_1, \ldots, w_m$ are linearly dependent, by Theorem 2.14, and applying $m - n$ times the procedure above, we can exclude $m - n$ vectors and get a basis of $V$. □

From the previous proposition we deduce the following important observations.

*Remark.* i) Every (finitely generated) vector space has a basis!

ii) In a vector space $V$ of dimension $n$, a set of $n$ vectors $S = \{v_1, \ldots, v_n\}$ is a set of generators if and only if it is a set of linearly independent vectors.
So to check if $S$ is a basis, it is enough to check one condition.

**Example.**

- $\left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right)$ are linearly independent in $\mathbb{K}^3$, they are not generators of $\mathbb{K}^3$, since $2 < 3 = \dim \mathbb{K}^3$, but we can complete them to a basis, e.g. to $\left\{\left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\1\\1\end{smallmatrix}\right)\right\}$
- Let $S := \{1+t, t+t^2, 1+t^2\} \subset \mathbb{K}[t]_{\leq 2}$. Are they generators? Are they linearly independent? $\dim \mathbb{K}[t]_{\leq 2} = 3$, so we can answer both question simultaneously.

  Let us check if $1+t, t+t^2, 1+t^2$ are linearly independent:

  $$0 = \lambda_1(1+t)+\lambda_2(t+t^2)+\lambda_3(1+t^2) = (\lambda_1+\lambda_3)+(\lambda_1+\lambda_2)t+(\lambda_2+\lambda_3)t^2 \Leftrightarrow \begin{cases} \lambda_1+\lambda_3 = 0 \\ \lambda_1+\lambda_2 = 0 \\ \lambda_2+\lambda_3 = 0 \end{cases}$$

  So $1+t, t+t^2, 1+t^2$ are linearly independent if and only if the linear system has a unique solution, so if and only if its matrix of coefficients has non-zero determinant:

  $$\det \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}}_{A} = \det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1 + 1 \neq 0\,.$$

  So $\{1+t, t+t^2, 1+t^2\}$ is a basis of $\mathbb{K}[t]_{\leq 2}$.

  Note that the columns of $A$ are the coordinates of $1+t, t+t^2, 1+t^2$ with respect to the canonical basis $\{1, t, t^2\}$ of $\mathbb{K}[t]_{\leq 2}$.

*Remark.* i) By taking coordinates, we can describe a vector in $V$ via a vector in $\mathbb{K}^n$ ($n = \dim V$). So we can "translate a problem on $V$ into a problem on $\mathbb{K}^n$". We will make this more precise and concrete in the next chapter.

ii) In $\mathbb{K}^n$, to check whether $n$ vectors form a basis, it is enough to check if they are linearly independent, i.e. if the "corresponding" homogeneous linear system has a unique solution. Its matrix of coefficients $A$ is a square matrix with the $n$ vectors as columns, so the linear system has a unique solution if and only if $A$ is invertible.

In other words, $\{v_1, \ldots, v_n\} \subset \mathbb{K}^n$ is a basis of $\mathbb{K}^n$ if and only if the matrix $A = (v_1|\ldots|v_n)$ (the $v_i$ are the columns) has maximal rank (i.e. $\det A \neq 0$).

## 2.4 A comment on rank

**Definition 2.21.** Let $A \in \mathcal{M}_{\mathbb{K}}(m, n)$ be a matrix. Let $R_1, \ldots, R_m$ be its rows and let $C_1, \ldots, C_n$ be its columns.

The *row space* of $A$ is $\mathrm{Row}(A) = \mathrm{Span}(R_1, \ldots, R_m) \triangleleft \mathbb{K}^n$.

The *column space* of $A$ is $\mathrm{Col}(A) = \mathrm{Span}(C_1, \ldots, C_n) \triangleleft \mathbb{K}^m$.

**Example.** Let $A = \left(\begin{smallmatrix}1&2&3\\4&5&6\end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{K}}(2,3)$, and $B = \left(\begin{smallmatrix}1&0&1\\2&0&2\end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{K}}(2,3)$ then

$$\mathrm{Row}(A) = \mathrm{Span}((1,2,3),(4,5,6)) \triangleleft \mathbb{K}^3, \qquad \mathrm{Col}(A) = \mathrm{Span}\left(\left(\begin{smallmatrix}1\\4\end{smallmatrix}\right), \left(\begin{smallmatrix}2\\5\end{smallmatrix}\right), \left(\begin{smallmatrix}3\\6\end{smallmatrix}\right)\right) \triangleleft \mathbb{K}^2.$$
$$\mathrm{Row}(B) = \mathrm{Span}((1,0,1),(2,0,2)) \triangleleft \mathbb{K}^3, \qquad \mathrm{Col}(B) = \mathrm{Span}\left(\left(\begin{smallmatrix}1\\2\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)\right) \triangleleft \mathbb{K}^2.$$

Note that $\dim \mathrm{Row}(A) = \dim \mathrm{Col}(A) = \mathrm{rk}(A) = 2$ and $\dim \mathrm{Row}(B) = \dim \mathrm{Col}(B) = \mathrm{rk}(B) = 1$.

**Theorem 2.22.** *Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$ be a matrix. Then*

$$\operatorname{rk}(A) = \dim \operatorname{Row}(A) = \dim \operatorname{Col}(A) = \operatorname{rk}(A^T)$$

*Idea:* the elementary row operations do not change $\operatorname{Row}(A)$, so if $U$ is in echelon form and is obtained from $A$ using the Gauss algorithm, we have $\operatorname{Row}(A) = \operatorname{Row}(U)$, but $\dim \operatorname{Row}(U)$ coincides with the number of pivots, i.e. the number of non-zero rows of $U$ (they are linearly independent!)

*Attention:* In general $\operatorname{Col}(A) \neq \operatorname{Col}(U)$, but $\dim \operatorname{Col}(A) = \dim \operatorname{Col}(U) =$ number of pivots.

*Remark.* Theorem 2.22 gives us a new interpretation/proof of part 1. of Therom 1.11 (Rouché-Capelli). Indeed,

$$Ax = b \text{ is solvable } \Leftrightarrow b \in \operatorname{Col}(A) = \operatorname{Span}(C_1, \ldots, C_n) \Leftrightarrow \operatorname{Span}(C_1, \ldots, C_n) \overset{(*)}{=} \operatorname{Span}(C_1, \ldots, C_n, b)$$

But $\operatorname{Span}(C_1, \ldots, C_n) \triangleleft \operatorname{Span}(C_1, \ldots, C_n, b)$, so the equalitiy $(*)$ holds if and only if

$$\dim \operatorname{Span}(C_1, \ldots, C_n) = \dim \operatorname{Span}(C_1, \ldots, C_n, b) \Leftrightarrow \dim \operatorname{Col}(A) = \dim \operatorname{Col}(A|b) \Leftrightarrow \operatorname{rk}(A) = \operatorname{rk}(A|b).$$

## 2.5 Cartesian and parametric equations

As seen so far, there are two ways to define a vector subspace $W \triangleleft \mathbb{K}^n$:

- implicitly, as solutions of a homogeneous linear system; these are *cartesian equations*;
- explicitly, as span of some vectors; these are *parametric equations*.

*Remark.* i) If $\dim W = r$, then $W$ can be express using $r$ vectors, so $r$ parameters; or via a homogeneous linear system with $n - r$ equations (by Rouché-Capelli theorem).
ii) We may use more vectors or equations, but they would not be linearly independent, i.e. we would have redundant information.

Cartesian $\rightarrow$ parametric) Given cartesian equations (i.e. a homogeneous linear system) by solving them we easily get parametric equations.

Parametric $\rightarrow$ cartesian) given parametric equations $W = \operatorname{Span}(w_1, \ldots, w_r) \triangleleft \mathbb{K}^n$, we can interpret it as $W = \operatorname{Col}(A)$, (where $A \in \mathcal{M}_{\mathbb{K}}(n,r)$ has $w_1, \ldots, w_r$ as columns). By the remark after Theorem 2.22, we see that a vector $v = (x_1, \ldots, x_n) \in \mathbb{K}^n$ belongs to $W$ if and only if $v \in \operatorname{Col}(A)$ if and only if $\operatorname{rk}(A) = \operatorname{rk}(A|v)$. By imposing this condition we obtain cartesian equations of $W$.

**Example.** Let $W = \operatorname{Span}((2,-2,1,0)^T, (1,0,-1,1)) \triangleleft \mathbb{R}^4$, then $(x_1, x_2, x_3, x_4)^T \in W$ if and only if the linear system

$$\left(\begin{array}{cc|c} 2 & 1 & x_1 \\ -2 & 0 & x_2 \\ 1 & -1 & x_3 \\ 0 & 1 & x_4 \end{array}\right)$$

is solvable. Using elementary row operations the linear system reduces to

$$\left(\begin{array}{cc|c} 1 & -1 & x_3 \\ 0 & 1 & x_4 \\ 0 & 0 & x_1 - 2x_3 - 3x_4 \\ 0 & 0 & x_2 + 2x_3 + 2x_4 \end{array}\right) \text{ which is solvable if and only if } v \text{ a solution of} \begin{cases} x_1 - 2x_3 - 3x_4 = 0 \\ x_2 + 2x_3 + 2x_4 = 0 \end{cases}$$

These are cartesian equations of $W$.

> *Alternatively*, given parametric equations, we can recover cartesian equations by eliminating the parameters, as in the following examples.

**Example.** Let $W = \operatorname{Span}((2, -5, 1)^T) \lhd \mathbb{R}^3$, i.e.

$$W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid \begin{cases} x = 2t \\ y = -5t \\ z = t \end{cases} , t \in \mathbb{R} \right\} \overset{(*)}{=} \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid \begin{cases} x - 2z = 0 \\ y - 5z = 0 \end{cases} \right\}$$

where in $(*)$ we used the equation $z = t$ to eliminate the parameter $t$.

Let $U = \operatorname{Span}((2, -5, 1)^T, (1, 0, -1)) \lhd \mathbb{R}^3$, i.e.

$$U = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid \begin{cases} x = 2t + s \\ y = -5t \\ z = t - s \end{cases} , t, s \in \mathbb{R} \right\}$$

By using $y = -5t$ to eliminate $t$ we get $\begin{cases} 5x + 2y = 5s \\ 5z + y = -5s \end{cases}$, and then we eliminate $s$ to get $U = \{(x, y, z)^T \in \mathbb{R}^3 \mid 5x + 3y + 5z = 0\}$.

## 2.6 Grassmann's formula

We conclude this chapter discussing the relation between the dimensions of two vector subspaces and those of their sum and intersection.

**Theorem 2.23** (Grassmann's formula)**.** *Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \lhd V$ be two vector subspaces. Then*

$$\dim U + \dim W = \dim(U \cap W) + \dim(U + W)$$

*Proof.* Let us set $\dim U = p$, $\dim W = q$ and $\dim(U \cap W) = r$. By Proposition 2.20 we have $r \leq p$ and $r \leq q$, and we have to prove that $\dim(U + W) = p + q - r$.

Let $\{b_1, \ldots, b_r\}$ be a basis of $U \cap W$ and complete it to a basis of $U$: $\{b_1, \ldots, b_r, c_{r+1} \ldots, c_p\}$; and to a basis of $W$: $\{b_1, \ldots, b_r, d_{r+1} \ldots, d_q\}$.

We are going to show that $S = \{b_1, \ldots, b_r, c_{r+1} \ldots, c_p, d_{r+1} \ldots, d_q\}$ is a basis of $U + W$, so that $\dim(U + W) = r + (p - r) + (q - r) = p + q - r$.

They generate: take $v \in U + W$, so there exist $u \in U, w \in W$ such that $v = u + w$. But there exist scalars such that

$$\begin{aligned} u &= \alpha_1 b_1 + \ldots + \alpha_r b_r + \gamma_{r+1} c_{r+1} + \ldots + \gamma_p c_p, \\ w &= \beta_1 b_1 + \ldots + \beta_r b_r + \delta_{r+1} d_{r+1} + \ldots + \delta_q d_q \end{aligned}$$

and so $v = (\alpha_1 + \beta_1) b_1 + \ldots + (\alpha_r + \beta_r) b_r + \gamma_{r+1} c_{r+1} + \ldots + \gamma_p c_p + \delta_{r+1} d_{r+1} + \ldots + \delta_q d_q$.

They are <u>linearly independent</u>: pick scalars such that

$$0_V = \lambda_1 b_1 + \ldots + \lambda_r b_r + \eta_{r+1} c_{r+1} + \ldots + \eta_p c_p + \tau_{r+1} d_{r+1} + \ldots + \tau_q d_q$$

and define $u := \lambda_1 b_1 + \ldots + \lambda_r b_r + \eta_{r+1} c_{r+1} + \ldots + \eta_p c_p \in U$, so that $-u = \tau_{r+1} d_{r+1} + \ldots + \tau_q d_q \in W$ and by S3) $u \in W$. Therefore, $u \in U \cap W$ and there exists $r$ scalars $\mu_1, \ldots, \mu_r$ such that $u = \mu_1 b_1 + \ldots + \mu_r b_r$. From this it follows

$$\mu_1 b_1 + \ldots + \mu_r b_r + \tau_{r+1} d_{r+1} + \ldots + \tau_q d_q = u - u = 0_V$$

and being $b_1, \ldots, b_r, d_{r+1} \ldots, d_q$ linearly independent, we get $\mu_1 = \ldots = \mu_r = \tau_{r+1} = \ldots = \tau_q = 0_{\mathbb{K}}$. In particular,

$$0_V = u = \lambda_1 b_1 + \ldots + \lambda_r b_r + \eta_{r+1} c_{r+1} + \ldots + \eta_p c_p$$

and being $b_1, \ldots, b_r, c_{r+1} \ldots, c_p$ linearly independent, we get $\lambda_1 = \ldots = \lambda_r = \eta_{r+1} = \ldots = \eta_p = 0_{\mathbb{K}}$. □

An immediate consequence of Grassmann's formula is the following statement.

**Corollary 2.24.** *Let $V$ be a $\mathbb{K}$-vector space, and let $U, W \triangleleft V$ be two vector subspaces. Then the sum $U+W$ is direct if and only if $\dim(U \cap W) = 0$ if and only if $\dim U + \dim W = \dim(U+W)$.*

*In particular, let $\mathcal{B}_U$ be a basis of $U$ and $\mathcal{B}_W$ be a basis of $W$, then the sum $U+W$ is direct if and only if $\mathcal{B}_U \cup \mathcal{B}_W$ is a basis of $U+W$.*

**Example.** In $\mathbb{R}^4$ consider the subspaces

$$U = \left\{ (x,y,z,w)^T \in \mathbb{R}^4 \mid \left\{ \begin{array}{rcl} x+y-z &=& 0 \\ x-y+w &=& 0 \end{array} \right. \right\}, \qquad W = \operatorname{Span}\left( \left(\begin{smallmatrix}1\\1\\0\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}1\\0\\1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}2\\1\\1\\0\end{smallmatrix}\right) \right).$$

We want to determine a basis (and the dimension) of $U$, $W$, $U \cap W$, $U+W$, and determine whether the sum is direct.

Note that $U = \operatorname{Sol}(A|0)$, where $A = \left(\begin{smallmatrix}1 & 1 & -1 & 0\\ 1 & -1 & 0 & 1\end{smallmatrix}\right)$, which reduces to $\left(\begin{smallmatrix}1 & 1 & -1 & 0\\ 0 & -2 & 1 & 1\end{smallmatrix}\right)$ (by applying $R_2 \to R_2 - R_1$), so that

$$U = \left\{ \begin{pmatrix} (s-t)/2 \\ (s+t)/2 \\ s \\ t \end{pmatrix} \mid s,t \in \mathbb{R} \right\} = \operatorname{Span}\left( \begin{pmatrix}1\\1\\2\\0\end{pmatrix}, \begin{pmatrix}-1\\1\\0\\2\end{pmatrix} \right) \Rightarrow \dim U = 2$$

To find a basis of $W$, we write the generators of $W$ as columns of a matrix, and we reduce it. By the discussion in Section 2.3.2, the columns corresponding to pivots *in the original matrix* are a basis for $W$:

$$\begin{pmatrix}1 & 1 & 2\\ 1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{pmatrix} \xrightarrow{R_2 \to R_2 - R_1} \begin{pmatrix}1 & 1 & 2\\ 0 & -1 & -1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{pmatrix} \xrightarrow{R_3 \to R_3 + R_2} \begin{pmatrix}1 & 1 & 2\\ 0 & -1 & -1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$$

Thus, $W = \text{Span}\left(\begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\\0\end{pmatrix}\right)$ and $\dim W = 2$.

(*) Now we consider $U+W$. The sum $U+W$ contains all vectors that can be written as sum of a vector in $U$ and a vector in $W$, so it is generated by a basis of $U$ together with a basis of $W$:

$$U+W = \text{Span}\left(\begin{pmatrix}1\\1\\2\\0\end{pmatrix}, \begin{pmatrix}-1\\1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\\0\end{pmatrix}\right)$$

Proceeding as above, we get that those four vectors are linearly independent, so that $\dim(U+W) = 4$. By Grassmann's formula $\dim(U \cap W) = 0$, i.e. $U \cap W = \{0_V\}$.

*Alternative to (*).* We may have started by considering $U \cap W$. A vector in $W = \text{Span}\left(\left(\begin{smallmatrix}1\\1\\0\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\-1\\1\\0\end{smallmatrix}\right)\right)$ has the form $\left(\begin{smallmatrix}a\\a-b\\b\\0\end{smallmatrix}\right)$ for $a, b \in \mathbb{R}$. Such vector belongs to $U \cap W$ if it satisfies the conditions to be in $U$:

$$\begin{cases} a + (a-b) - b &= 0 \\ a - (a-b) + 0 &= 0 \end{cases} \Leftrightarrow \begin{cases} a &= b \\ b &= 0 \end{cases}$$

So there is a unique possibility: $(a,b) = (0,0)$. In other words, $U \cap W = \{0_V\}$, so that $\dim(U \cap W) = 0$. By Grassmann's formula $\dim(U+W) = 4$, i.e. $U \oplus W = \mathbb{R}^4$.

# Chapter 3

# Linear maps

In the previous chapter we have introduced vectors and vector spaces. Here, we are going to introduce the maps between vector spaces (linear maps), so we are going to see how to link two vector spaces and how to transform vectors.

Before introducing linear maps and their properties, let us briefly recall some general definitions.

**Definition 3.1.** Let $f : X \to Y$ be a function.
The function $f$ is *surjective* if "every element in the codomain can be reached":

$$\forall y \in Y \ \exists x \in X : y = f(x) .$$

The function $f$ is *injective* if "to different elements in the domain correspond different elements in the codomain":

$$\forall x_1, x_2 \in X : f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

The function $f$ is *bijective* if it is both surjective and injective:

$$\forall y \in Y \ \exists! x \in X : y = f(x)$$

"Every element in the codomain is reached by exactly one element of the domain".

If $f : X \to Y$ is bijective it is *invertible*, indeed every $y \in Y$ is associated to a single $x \in X$. So, $y \mapsto x$, where $x$ is the unique element of $X$ mapped to $y$, defines the inverse map $f^{-1} : Y \to X$. It satisfies:

$$f \circ f^{-1} = id_Y, \quad f^{-1} \circ f = id_X .$$

**Example.** The map $\mathbb{R} \to [-1, 1], t \mapsto \sin(t)$ is surjective, but not injective.
The map $\mathbb{R} \to \mathbb{R}^2, t \mapsto (2t, -t)$ is injective, but not surjective.
The map $\mathbb{R} \to \mathbb{R}, t \mapsto 3t$ is both surjective and injective, hence it is bijective.
The map $\mathbb{R} \to \mathbb{R}^2, t \mapsto (\cos(t), \sin(t))$ is neither surjective nor injective.

In this course we will focus on functions between vector spaces, which preserve the vector space structure, in particular they will map vector subspaces to vector subspaces as in the second and third examples, but not in the last one, as it sends the real line to the unit circle.

## 3.1 Linear maps: definition and basic properties

**Definition 3.2.** Let $V, W$ be vector spaces over the same field $\mathbb{K}$.
A function $f : V \to W$ is a *linear map* if

L1) $f(v_1 + v_2) = f(v_1) + f(v_2)$ for all $v_1, v_2 \in V$ (additivity)

L2) $f(\lambda \cdot v) = \lambda \cdot f(v)$ for all $v \in V$, $\lambda \in \mathbb{K}$ (homogeneity/scaling)

**Example.**

- The zero-map $V \to W, v \mapsto 0_W$ is linear.
- The identity map $id_V : V \to V, v \mapsto v$ is linear.
- Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, then the map $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m, v \mapsto Av$ is linear:

$$\mathcal{L}_A(v_1+v_2) = A(v_1+v_2) = Av_1+Av_2 = \mathcal{L}_A(v_1)+\mathcal{L}_A(v_2), \quad \mathcal{L}_A(\lambda\cdot v) = A(\lambda\cdot v) = \lambda Av = \lambda\mathcal{L}_A(v)$$

- The transposition map $\mathcal{M}_{\mathbb{K}}(m,n) \to \mathcal{M}_{\mathbb{K}}(n,m), A \to A^T$ is linear.
- The “derivative” $\mathbb{K}[t]_{\leq d} \to \mathbb{K}[t]_{\leq d}, \sum_{n=0}^{d} a_n t^n \mapsto \sum_{n=1}^{d} n a_n t^{n-1}$ is linear.
- $f : \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto x^2 + y^2$ is not linear, e.g. $f(2 \cdot (1,1)) = 8 \neq 4 = 2f(1,1)$.
- For $n > 1$, the determinant of a square matrix $\det : \mathcal{M}_{\mathbb{K}}(n,n) \to \mathbb{K}, A \mapsto \det(A)$ is not a linear map: $\det(cA) = c^n \det A$.

*Remark.* Conditions L1, L2 can be checked simultaneously by proving

$$f(\lambda v_1 + \mu v_2) = \lambda f(v_1) + \mu f(v_2) \quad \forall v_1, v_2 \in V, \lambda, \mu \in \mathbb{K}.$$

**Lemma 3.3.** *Let $f : V \to W$ be a linear map between the $\mathbb{K}$-vector spaces $V, W$. Then $f(0_V) = 0_W$.*

*Proof.* $f(0_V) = f(0_{\mathbb{K}} \cdot v) = 0_{\mathbb{K}} \cdot f(v) = 0_W$. □

In particular, if $f(0_V) \neq 0_W$, then $f : V \to W$ is not a linear map, e.g. $e^0 = 1 \neq 0$, so $\exp : \mathbb{R} \to \mathbb{R}, x \mapsto e^x$ is not linear.

Be careful, this condition is necessary, but not sufficient for the linearity: $f : \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto x^2 + y^2$ is not linear, but $f(0,0) = 0$.

**Properties.** *Let $V, W, U$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$, $g : W \to U$ be linear maps. Then*

1. *The composition $g \circ f : V \to U, v \mapsto g(f(v))$ is linear.*
2. *If $f : V \to W$ is invertible (i.e. a bijective function), then $f^{-1} : W \to V$ is linear.*

*Proof.* 1. $g(f(\lambda v_1 + \mu v_2)) = g(\lambda f(v_1) + \mu f(v_2))) = \lambda g(f(v_1)) + \mu g(f(v_2))$.

2. We need to prove $f^{-1}(\lambda w_1 + \mu w_2) = \lambda f^{-1}(w_1) + \mu f^{-1}(w_2)$. Define $v_1 = f^{-1}(w_1)$ and $v_2 = f^{-1}(w_2)$, so that $w_1 = f(v_1)$ and $w_2 = f(v_2)$. Using $f^{-1} \circ f = id_V$ and the linearity of $f$ we get

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

□

**Special case.** 1. Let $A \in \mathcal{M}_{\mathbb{K}}(m,p)$ and $B \in \mathcal{M}_{\mathbb{K}}(p,n)$. Then $\mathcal{L}_A : \mathbb{K}^p \to \mathbb{K}^m$ and $\mathcal{L}_B : \mathbb{K}^n \to \mathbb{K}^p$, so we can compose them:

$$\begin{array}{rcccc} \mathcal{L}_A \circ \mathcal{L}_B : & \mathbb{K}^n & \longrightarrow & \mathbb{K}^p & \longrightarrow & \mathbb{K}^m \\ & v & \mapsto & Bv & & \\ & & & w & \mapsto & Aw \end{array}$$

so $\mathcal{L}_A \circ \mathcal{L}_B(v) = A(Bv) = (AB)v = \mathcal{L}_{AB}(v)$, thus $\mathcal{L}_A \circ \mathcal{L}_B = \mathcal{L}_{AB}$ (note $AB \in \mathcal{M}_{\mathbb{K}}(m,n)$).

2. If we assume $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ to be invertible (i.e. $\det A \neq 0$), then $\mathcal{L}_A$ is bijective/invertible $(\mathcal{L}_A)^{-1} = \mathcal{L}_{A^{-1}}$.

"Linear maps behave well with vector subspaces" .

**Lemma 3.4.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map.*

1. *If $U \triangleleft V$, then $f(U) = \{w \in W \mid \exists u \in U \text{ s.t. } w = f(u)\}$ is a vector subspace of $W$.*
2. *If $Z \triangleleft W$, then $f^{-1}(Z) = \{v \in V \mid f(v) \in Z\}$ is a vector subspace of $V$.*

*Proof.* See Quiz 3. □

We consider now two cases which will be relevant in this course.

**Definition 3.5.** Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map.
The *kernel* of $f$ is

$$\ker(f) = f^{-1}(\{0_W\}) = \{v \in V \mid f(v) = 0_W\} \triangleleft V.$$

The *image* of $f$ is

$$\mathrm{Im}(f) = f(V) = \{w \in W \mid \exists v \in V \text{ s.t. } w = f(v)\} \triangleleft W.$$

**Notation.** Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, then we write $\ker(A)$ for $\ker(\mathcal{L}_A)$ and $\mathrm{Im}(A)$ for $\mathrm{Im}(\mathcal{L}_A)$.

**Example.** Let $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, and consider $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$, then

$$\ker(A) = \{v \in \mathbb{K}^n \mid \mathcal{L}_A(v) = 0\} = \{v \in \mathbb{K}^n \mid Av = 0\} = \mathrm{Sol}(A|0)$$

What about the image? Let $b \in \mathbb{K}^m$, then

$$b \in \mathrm{Im}(A) \Leftrightarrow \exists v \in \mathbb{K}^n \text{ s.t. } Av = b \Leftrightarrow Ax = b \text{ is solvable } \Leftrightarrow b \in \mathrm{Col}(A)$$

so $\mathrm{Im}(A) = \mathrm{Col}(A)$, in particular $\dim \mathrm{Im}(A) = \mathrm{rk}(A)$. In analogy, we define rank of a linear map.

**Definition 3.6.** Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map. The *rank* of $f$ is $\mathrm{rk}(f) := \dim \mathrm{Im}(f)$.

**Example.** Let $f : \mathbb{K}[t]_{\leq 2} \to \mathbb{K}[t]_{\leq 1}, p(t) \mapsto tp''(t) - p'(t)$.
We firstly verify that $f$ is linear using that the "derivative" is linear:

$$\begin{aligned} f(\lambda p(t) + \mu q(t)) &= t(\lambda p(t) + \mu q(t))'' - (\lambda p(t) + \mu q(t))' = t(\lambda p''(t) + \mu q''(t)) - (\lambda p'(t) + \mu q'(t)) \\ &= \lambda(tp''(t) - p'(t)) + \mu(tq''(t) - q'(t)) = \lambda f(p(t)) + \mu f(q(t)) \end{aligned}$$

Let $p(t) = a_0 + a_1 t + a_2 t^2$, then $f(p(t)) = 2a_2 t - (a_1 + 2a_2 t) = -a_1$, so

$$\ker(f) = \{p(t) \in \mathbb{K}[t]_{\leq 2} \mid tp''(t) - p'(t) = 0\} = \{a_0 + a_2 t^2 | a_0, a_2 \in \mathbb{K}\} = \mathrm{Span}(1, t^2)$$

and

$$\mathrm{Im}(f) = \{-a_1 | a_1 \in \mathbb{K}\} = \mathrm{Span}(1).$$

## 3.2 Surjectivity and injectivity of linear maps

### 3.2.1 Surjectivity

**Theorem 3.7** (Surjectivity criterion for linear maps)**.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map. Then $f$ is surjective if and only if* $\mathrm{rk}(f) = \dim W$.

*Proof.* $f$ is surjective if and only if $\mathrm{Im}(f) = W$. Since $\mathrm{Im}(f) \subset W$, the statement follows from Proposition 2.20. □

"Surjective linear maps preserve the property of being generators."

**Proposition 3.8.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map.*
*If* $V = \mathrm{Span}(v_1, \dots, v_n)$*, then* $\mathrm{Span}(f(v_1), \dots, f(v_n)) = \mathrm{Im}(f)$.
*In particular, if $f$ is surjective then* $\mathrm{Span}(f(v_1), \dots, f(v_n)) = W$.

*Proof.* Let $w \in \mathrm{Im}(f)$, so there exist $v \in V$ such that $f(v) = w$. From $V = \mathrm{Span}(v_1, \dots, v_n)$ we deduce that there exist $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ such that $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$. Thus

$$w = f(v) = f(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 f(v_1) + \ldots + \lambda_n f(v_n) \in \mathrm{Span}(f(v_1), \dots, f(v_n)) \,.$$

□

**Example.** i) Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(2,3)$. Is $\mathcal{L}_A : \mathbb{R}^3 \to \mathbb{R}^2$ surjective?
$\{e_1, e_2, e_3\}$ generates $\mathbb{R}^3$, so by the previous proposition we know

$$\mathrm{Im}(\mathcal{L}_A) = \mathrm{Span}(\mathcal{L}_A(e_1), \mathcal{L}_A(e_2), \mathcal{L}_A(e_3)) = \mathrm{Span}\left(\begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix}\right) = \mathrm{Col}(A) = \mathbb{R}^2$$

So $\mathcal{L}_A$ is surjective.

ii) Consider $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3,2)$. Is $\mathcal{L}_B : \mathbb{R}^2 \to \mathbb{R}^3$ surjective? As above, we know

$$\mathrm{Im}(\mathcal{L}_B) = \mathrm{Span}(\mathcal{L}_A(e_1), \mathcal{L}_A(e_2)) = \mathrm{Span}\left(\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}\right) = \mathrm{Col}(B)$$

Since $\dim \mathrm{Col}(B) = \mathrm{rk}(B) = 2 < 3 = \dim \mathbb{R}^3$, the map $\mathcal{L}_B$ is not surjective.

Note that, for a matrix $A \in \mathcal{M}_{\mathbb{K}}(m,n)$, the map $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$ is surjective if and only if $\dim \mathrm{Im}\, A = m$, i.e. if and only if $\mathrm{rk}(A) = m$. But $\mathrm{rk}(A) \leq \min(m,n)$, so if $\mathcal{L}_A$ is surjective, then $m \leq n$.

### 3.2.2 Injectivity

**Theorem 3.9** (Injectivity criterion for linear maps)**.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map. Then $f$ is injective if and only if* $\ker(f) = \{0_V\}$.

*Proof.* $\Rightarrow$] Assume $f$ to be injective, and let $v \in \ker f$, then $f(v) = 0_W = f(0_V)$. Since $f$ is injective, we deduce $v = 0_V$, so $\ker(f) = \{0_V\}$.

$\Leftarrow$] Assume $\ker(f) = \{0_V\}$, and let $v_1, v_2 \in V$ be vectors with $f(v_1) = f(v_2)$, then $0_W = f(v_1) - f(v_2) = f(v_1 - v_2)$, so $v_1 - v_2 \in \ker(f) = \{0_V\}$. Thus $v_1 = v_2$. □

"Injective linear maps preserve the property of being linearly independent."

**Proposition 3.10.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be an injective linear map. Then:*

*If $\{v_1, \dots, v_n\} \subset V$ are linearly independent, then $\{f(v_1), \dots, f(v_n)\} \subset V$ are linearly independent too.*

*Proof.* Write $0_W$ as linear combination of $\{f(v_1), \dots, f(v_n)\}$:

$$0_W = \lambda_1 f(v_1) + \dots \lambda_n f(v_n) = f(\lambda_1 v_1 + \dots \lambda_n v_n) \Rightarrow \lambda_1 v_1 + \dots \lambda_n v_n \in \ker(f) = \{0_V\}$$

so $\lambda_1 v_1 + \dots \lambda_n v_n = 0_V$. Using that $\{v_1, \dots, v_n\} \subset V$ are linearly independent, we get $\lambda_1 = \dots = \lambda_n = 0_{\mathbb{K}}$.

□

**Example.** i) Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(2, 3)$. The map $\mathcal{L}_A : \mathbb{R}^3 \to \mathbb{R}^2$ is injective if and only if $0 = \dim \ker(A) = \dim \mathrm{Sol}(A|0)$. By Rouché-Capelli, the solutions of the linear system $Ax = 0$ are a vector space of dimension $3 - \mathrm{rk}(A) = 3 - 2 = 1$, so $\mathcal{L}_A$ is not injective.

ii) Consider $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3, 2)$. This time the solutions of the linear system $Bx = 0$ form a vector space of dimension $2 - \mathrm{rk}(B) = 2 - 2 = 0$, so $\mathcal{L}_B : \mathbb{R}^2 \to \mathbb{R}^3$ is injective.

Note that, for a matrix $A \in \mathcal{M}_{\mathbb{K}}(m, n)$, the map $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$ is injective if and only if $\dim \ker A = 0$, i.e. if and only if $n = \mathrm{rk}(A)$. But $\mathrm{rk}(A) \leq \min(m, n)$, so if $\mathcal{L}_A$ is injective, then $n \leq m$.

### 3.2.3 Bijectivity

**Terminology**. A bijective linear map $f : V \to W$ is called *isomorphism*, and we say that $V$ and $W$ are isomorphic.

A function is bijective if it is both surjective and injective, so an isomorphism preserves both the property of being generators and the property of being linearly independent. In other words: "Bijective linear maps preserve bases".

**Proposition 3.11.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a bijective linear map. If $\{v_1, \dots, v_n\}$ is a basis of $V$, then $\{f(v_1), \dots, f(v_n)\}$ is a basis of $W$, in particular* $\dim V = \dim W$.

*Proof.* It follows immediately from Proposition 3.8 and 3.10. □

*Remark.* Let $A \in \mathcal{M}_{\mathbb{K}}(m, n)$ be a matrix. If $m \neq n$ then $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$ is not invertible, and so $A$ is also not invertible. That is why in Section 1.2.4 we stuck to *square matrices.*

## 3.3 Rank-nullity theorem and its consequences

**Theorem 3.12** (Rank-nullity theorem)**.** *Let $V, W$ be $\mathbb{K}$-vector spaces, and let $f : V \to W$ be a linear map. Then*

$$\dim \ker f + \dim \operatorname{Im}(f) = \dim V\,.$$

*Proof.* Let $\mathcal{B} = \{b_1, \ldots, b_p\}$ be a basis of $\ker f$ (recall, $\mathcal{B} = \emptyset$ if $\ker f = \{0\}$), and complete it to a basis of $V$:

$$\mathcal{B}' = \{b_1, \ldots, b_p, b_{p+1}, \ldots, b_n\}.$$

By Proposition 3.8, we know that

$$\begin{aligned}\operatorname{Im}(f) &= \operatorname{Span}(f(b_1), \ldots, f(b_p), f(b_{p+1}), \ldots, f(b_n)) = \operatorname{Span}(0_W, \ldots, 0_W, f(b_{p+1}), \ldots, f(b_n)) \\ &= \operatorname{Span}(f(b_{p+1}), \ldots, f(b_n)).\end{aligned}$$

Thus $\{f(b_{p+1}), \ldots, f(b_n)\}$ generates $\operatorname{Im}(f)$. If we show that they are linearly independent, they form a basis, so $\dim \operatorname{Im}(f) = n - p$ and we are done.

Let $\lambda_{p+1}, \ldots, \lambda_n \in \mathbb{K}$ be such that

$$0_W = \lambda_{p+1} f(v_{p+1}) + \ldots + \lambda_n f(v_n) = f(\lambda_{p+1} v_{p+1} + \ldots + \lambda_n v_n)$$

hence $\lambda_{p+1} v_{p+1} + \ldots + \lambda_n v_n \in \ker(f)$.
But $\ker(f) = \operatorname{Span}(b_1, \ldots, b_p)$, so there exists $\lambda_1, \ldots, \lambda_p \in \mathbb{K}$ such that

$$\lambda_{p+1} v_{p+1} + \ldots + \lambda_n v_n = \lambda_1 v_1 + \ldots \lambda_p v_p$$

We obtain $\lambda_1 v_1 + \ldots \lambda_p v_p - \lambda_{p+1} v_{p+1} - \ldots - \lambda_n v_n = 0_V$, but $\mathcal{B}'$ is a set of linearly independent vectors, so $\lambda_1 = \ldots = \lambda_p = \lambda_{p+1} = \ldots, \lambda_n = 0_{\mathbb{K}}$. □

**Example.** Let $f : \mathbb{K}[t]_{\leq 2} \to \mathbb{K}[t]_{\leq 1}, p(t) \mapsto tp''(t) - p'(t)$.

We have seen that $\ker(f) = \operatorname{Span}(1, t^2)$ and $\operatorname{Im}(f) = \operatorname{Span}(1)$, so $\dim \ker(f) = 2$ and $\dim \operatorname{Im}(f) = 1$.

*Remark.* Let $A \in \mathcal{M}_{\mathbb{K}}(m, n)$, then Theorem 3.12 for $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$ reads

$$\dim \operatorname{Sol}(A|0) = \dim \ker(\mathcal{L}_A) = n - \dim \operatorname{Im}(A) = n - \operatorname{rk}(A),$$

so the space of solutions of a solvable linear system depends on $n - \operatorname{rk}(A)$ parameters, as we already knew from part 2. of Therom 1.11 (Rouché-Capelli).

*Remark.* Let $f : V \to W$ be a linear map, the Rank-nullity theorem has the following consequences (1. and 2. were already deduced above in the case of $\mathcal{L}_A$, for $A \in \mathcal{M}_{\mathbb{K}}(m, n)$).

1. If $f$ is surjective, then $\dim \operatorname{Im}(f) = \dim W$, so $\dim W = \dim V - \dim \ker f \leq \dim V$. In other words, if $\dim V < \dim W$, then $f$ cannot be surjective.
2. If $f$ is injective, then $\dim \ker(f) = 0$, so $\dim V = \dim \operatorname{Im}(f) \leq \dim W$. In other words, if $\dim V > \dim W$, then $f$ cannot be injective.
3. If $f$ is bijective, then $\dim V = \dim W$. In other words, if $\dim V \neq \dim W$, then $f$ cannot be bijective (see also Proposition 3.11).
4. If $\dim V = \dim W$, then $f$ is injective if and only if $f$ is surjective.

**Example.** $\diamond$ Let $f : \mathbb{K}[t]_{\leq 3} \to \mathcal{M}_{\mathbb{K}}(2,2)$ be the linear map given by

$$f(p(t)) = \begin{pmatrix} p(0) & p'(0) \\ p''(0) & p'''(0) \end{pmatrix} \quad \text{i.e.} \quad f(a_0 + a_1 t + a_2 t^2 + a_3 t^3) = \begin{pmatrix} a_0 & a_1 \\ 2a_2 & 6a_3 \end{pmatrix}$$

$a_0 + a_1 t + a_2 t^2 + a_3 t^3 \in \ker(f) \Leftrightarrow a_0 = a_1 = 2a_2 = 6a_3 = 0$, i.e. $\ker(f) = \{0\}$, so $f$ is injective.
Since $\dim \mathbb{K}[t]_{\leq 3} = \dim \mathcal{M}_{\mathbb{K}}(2,2) = 4$, $f$ is also surjective, hence an isomorphism.

$\diamond$ Let $g : \mathbb{K}[t]_{\leq 3} \to \mathcal{M}_{\mathbb{K}}(2,2)$ be the linear map defined by $g(p(t)) = \left(\begin{smallmatrix} p(0) & p'(0) \\ -p(0) & -p''(0) \end{smallmatrix}\right)$ i.e.

$$g(a_0 + a_1 t + a_2 t^2 + a_3 t^3) = \begin{pmatrix} a_0 & a_1 \\ -a_0 & -2a_2 \end{pmatrix}$$

We get that $a_0 + a_1 t + a_2 t^2 + a_3 t^3 \in \ker(g) \Leftrightarrow a_0 = -a_0 = a_1 = -2a_2$, i.e $\ker(g) = \{a_3 t^3 \mid a_3 \in \mathbb{K}\} = \operatorname{Span}(t^3)$, so $g$ is not injective and also not surjective, because $\dim \mathbb{K}[t]_{\leq 3} = \dim \mathcal{M}_{\mathbb{K}}(2,2) = 4$.
Note that $\dim \ker(g) = 1$, so $\operatorname{rk}(g) = 4 - 1 = 3$, indeed $\operatorname{Im}(g) = \operatorname{Span}(g(1), g(t), g(t^2), g(t^3)) = \operatorname{Span}\left(\left(\begin{smallmatrix} 1 & 0 \\ -1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 0 & -2 \end{smallmatrix}\right)\right)$.

$\diamond$ Let $h : \mathbb{K}[t]_{\leq 2} \to \mathbb{K}^2$ be the linear map defined by $h(p(t)) = \left(\begin{smallmatrix} p(1) \\ p(-1) \end{smallmatrix}\right)$ i.e.

$$h(a_0 + a_1 t + a_2 t^2) = \begin{pmatrix} a_0 + a_1 + a_2 \\ a_0 - a_1 + a_2 \end{pmatrix}$$

We get that $a_0 + a_1 t + a_2 t^2 \in \ker(h) \Leftrightarrow a_0 + a_1 + a_2 = a_0 - a_1 + a_2 = 0_{\mathbb{K}}$, i.e $\ker(h) = \{a_0 + a_1 t + a_2 t^2 \mid a_1 = 0_{\mathbb{K}}, a_2 = -a_0\} = \operatorname{Span}(1 - t^2)$, so $h$ is not injective, but $\operatorname{rk}(h) = 3 - 1 = 2 = \dim \mathbb{K}^2$, so $h$ is surjective.

## 3.4 How to give a linear map?

Let $V, W$ be $\mathbb{K}$-vector spaces, how can we construct a linear map $f : V \to W$?
To define it, we need to associate to each vector $v \in V$ a vector $w \in W$ in such a way, that $f$ respects additivity (L1) and scaling (L2). Let us proceed as follows:
1. Choose a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ of $V$, so for each $v \in V$ there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ such that $v = \lambda_1 b_1 + \ldots + \lambda_n b_n$: $(v)_{\mathcal{B}} = (\lambda_1, \ldots, \lambda_n)^T$.
2. Choose the image of the $n$ vectors in $\mathcal{B}$: $w_i := f(b_i)$
3. Define $f(v)$ by linearity:

$$f(v) = f(\lambda_1 b_1 + \ldots + \lambda_n b_n) = \lambda_1 f(b_1) + \ldots + \lambda_n f(b_n) = \lambda_1 w_1 + \ldots + \lambda_n w_n$$

By construction such map is linear, so to give a linear map it is enough to give the image of the vector of a basis: "A linear map is completely determined by its behaviour on a basis".

**Example.** Let $f : \mathbb{Q}^2 \to \mathbb{Q}[t]_{\leq 2}$ be defined via

$$f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 + t, \quad f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = t^2 - t$$

then $f\left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) = f\left(-2\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) + 3\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) = -2(1+t) + 3(t^2 - t) = -2 - 5t + 3t^2$.

**Example.** Let $g : \mathbb{R}^3 \to \mathbb{R}^2$ be defined via

$$g(e_1) = \begin{pmatrix} 1 \\ 6 \end{pmatrix}, \quad g(e_2) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad g(e_3) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

then $g \begin{pmatrix} x \\ y \\ z \end{pmatrix} = g(xe_1 + ye_2 + ze_3) = x + \begin{pmatrix} 1 \\ 6 \end{pmatrix} + y \begin{pmatrix} -2 \\ 0 \end{pmatrix} + z \begin{pmatrix} 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 3 \\ 6 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

This construction easily generalize to any linear map $\mathbb{K}^n \to \mathbb{K}^m$, in other words:

**Lemma 3.13.** *Every linear map $f : \mathbb{K}^n \to \mathbb{K}^m$ is of the form $f = \mathcal{L}_A$ for some $A \in \mathcal{M}_{\mathbb{K}}(m, n)$.*

We are going to see how matrices encode linear maps between arbitrary vector spaces (and not only $\mathbb{K}^n$).

**Example.** Let $f : \mathbb{K}[t]_{\leq 3} \to \mathcal{M}_{\mathbb{K}}(2, 2)$ be the linear map defined by

$$f(a_0 + a_1 t + a_2 t^2 + a_3 t^3) = \begin{pmatrix} a_0 & a_1 \\ 2a_2 & 6a_3 \end{pmatrix}$$

Here a central role is played by the coefficients $a_0, a_1, a_2, a_3$. We would like to highlight their role.

The first step is to link any vector space with a suitable $\mathbb{K}^n$.

**Theorem 3.14** (Isomorphism theorem with $\mathbb{K}^n$)**.** *Let $V$ be a $\mathbb{K}$-vector space of dimension $\dim V = n$. Then, there exists an isomorphism $f : V \to \mathbb{K}^n$.*

*Proof.* Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis of $V$ and define the linear map $f : V \to \mathbb{K}^n$ via $f(b_j) = e_j$, where $\{e_1, \ldots, e_n\}$ is the canonical basis of $\mathbb{K}^n$.

We get immediately that $\mathrm{Im}(f) = \mathrm{Span}(f(b_1), \ldots, f(b_n)) = \mathrm{Span}(e_1, \ldots, e_n) = \mathbb{K}^n$, so $f$ is surjective. But $\dim V = n = \dim \mathbb{K}^n$, so $f$ is bijective. □

**Corollary 3.15.** *Let $V, W$ be $\mathbb{K}$-vector spaces. They are isomorphic ($V \cong W$) if and only if $\dim V = \dim W$.*

*Proof.* $\Rightarrow$] Done in Proposition 3.11.
$\Leftarrow$] By Theorem 3.14, there exists isomorphisms $f : V \to \mathbb{K}^n$ and $g : V \to \mathbb{K}^n$; hence $g^{-1} \circ f : V \to W$ is an isomorphism. □

We have then a dictionary to move from $V$ to $\mathbb{K}^n$ and back.

**Definition 3.16.** Let $V$ be a $\mathbb{K}$-vector space and let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be a basis of $V$, and let $\mathcal{E} = \{e_1, \ldots, e_n\}$ be the canonical basis of $\mathbb{K}^n$.

The linear map $X_{\mathcal{B}} : V \to \mathbb{K}^n, b_i \mapsto e_i$ is called the *coordinate map* of $\mathcal{B}$.

The linear map $P_{\mathcal{B}} : \mathbb{K}^n \to V, e_i \mapsto b_i$ is called the *parametrization map* of $\mathcal{B}$.

Let $v = \lambda_1 b_1 + \ldots + \lambda_n b_n \in V$, then $X_{\mathcal{B}}(v) = \lambda_1 e_1 + \ldots + \lambda_n e_n = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, so $X_{\mathcal{B}}$ extrapolates the coordinates of $v$ with respect to the basis $\mathcal{B}$.

Note that $P_{\mathcal{B}} = X_{\mathcal{B}}^{-1}$, so $P_{\mathcal{B}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \alpha_1 b_1 + \ldots + \alpha_n b_n$.

We have now all the ingredients to encode a linear map into a matrix.

### 3.4.1 Transformation matrix

Let $V, W$ be $\mathbb{K}$-vector spaces, let $f : V \to W$ be a linear map and let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of $V$ and $\mathcal{C} = \{c_1, \dots, c_m\}$ be a basis of $W$.

Let us consider the map $X_\mathcal{C} \circ f \circ P_\mathcal{B} : \mathbb{K}^n \to \mathbb{K}^m$. By Lemma 3.13, it coincides with the map $\mathcal{L}_A$ for a matrix $A \in \mathcal{M}_\mathbb{K}(m, n)$, and we want determine this matrix $A$.

$$X_\mathcal{C} \circ f \circ P_\mathcal{B} = \mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m, \qquad \begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ {\scriptstyle P_\mathcal{B}}\uparrow & & \downarrow{\scriptstyle X_\mathcal{C}} \\ \mathbb{K}^n & \dashrightarrow_{\mathcal{L}_A} & \mathbb{K}^m \end{array} \qquad A = ?$$

The map $X_\mathcal{C} \circ f \circ P_\mathcal{B} = \mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^m$ is determined by its behaviour on the canonical basis $\{e_1, \dots, e_n\}$ of $\mathbb{K}^n$:

$$A \cdot e_j = \mathcal{L}_A(e_j) = X_\mathcal{C}(f(P_\mathcal{B}(e_j))) = X_\mathcal{C}(f(b_j)) \quad \Longleftrightarrow \quad A = \Big( X_\mathcal{C}(f(b_1)) \Big| \dots \Big| X_\mathcal{C}(f(b_n)) \Big)$$

In words: the $j$-th column of $A$ is given by the coordinates of $f(b_j)$ with respect to the basis $\mathcal{C}$.

**Definition 3.17.** Let $V, W$ be $\mathbb{K}$-vector spaces, let $f : V \to W$ be a linear map and let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of $V$ and $\mathcal{C} = \{c_1, \dots, c_m\}$ be a basis of $W$.
The *transformation matrix* of $f$ with respect to $\mathcal{B}$ and $\mathcal{C}$ is the unique matrix $\mathcal{M}^\mathcal{B}_\mathcal{C}(f) = A \in \mathcal{M}_\mathbb{K}(m, n)$ such that

$$\mathcal{L}_A = X_\mathcal{C} \circ f \circ P_\mathcal{B},$$

i.e. whose $j$-th column is given by the coordinates of $f(b_j)$ with respect to the basis $\mathcal{C}$.

**Example.** ⋄ For a matrix $A \in \mathcal{M}_\mathbb{K}(m, n)$, we have $\mathcal{M}^\mathcal{E}_{\mathcal{E}'}(\mathcal{L}_A) = A$, where $\mathcal{E}$ and $\mathcal{E}'$ are the canonical bases of $\mathbb{K}^n$ and of $\mathbb{K}^m$.

⋄ Consider $V = \mathbb{K}[t]_{\le 2}$ and $W = \mathcal{M}_\mathbb{K}(2, 2)$, with bases $\mathcal{B} = \{1, t, t^2\}$ and $\mathcal{C} = \{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)\}$, and let $f : \mathbb{K}[t]_{\le 2} \to \mathcal{M}_\mathbb{K}(2, 2)$ be the linear map

$$f(a_0 + a_1 t + a_2 t^2) = \begin{pmatrix} a_0 & a_0 + a_1 + a_2 \\ a_0 - a_1 + a_2 & 0 \end{pmatrix}$$

Then

$$\begin{aligned} f(1) &= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = 1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + 1 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + 0 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ f(t) &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = 0 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + (-1) \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + 0 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ f(t^2) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 0 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + 1 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + 0 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \implies \mathcal{M}^\mathcal{B}_\mathcal{C}(f) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

⋄ Consider the $\mathbb{K}$-vector spaces $\mathbb{K}[t]_{\le 2}$ and $\mathbb{K}^2$, with bases $\mathcal{B} = \{1, 1 + t, 1 + t^2\}$ and $\mathcal{C} = \{e_1, e_1 + e_2\}$, and let $h : \mathbb{K}[t]_{\le 2} \to \mathbb{K}^2$ be the linear map defined by

$$h(a_0 + a_1 t + a_2 t^2) = \begin{pmatrix} a_0 + a_1 + a_2 \\ a_1 - 2a_2 \end{pmatrix}$$

Then

$$\left.\begin{aligned} h(1) &= \begin{pmatrix}1\\0\end{pmatrix} = 1\cdot e_1 + 0\cdot(e_1+e_2),\\ h(1+t) &= \begin{pmatrix}2\\1\end{pmatrix} = 1\cdot e_1 + 1\cdot(e_1+e_2),\\ h(1+t^2) &= \begin{pmatrix}2\\-2\end{pmatrix} = 4\cdot e_1 - 2\cdot(e_1+e_2)\end{aligned}\right. \implies \mathcal{M}^{\mathcal{B}}_{\mathcal{C}}(h) = \begin{pmatrix}1 & 1 & 4\\ 0 & 1 & -2\end{pmatrix}$$

*Remark.* The main utility of the transformation matrix is to provide us a very effective dictionary to translate problems on abstract vector spaces into problems on $\mathbb{K}^n$. Indeed, as explained in the following statements, via the transformation matrix we may: i) translate a problem on a linear map into a matrix problem; ii) reduce complicated computations to matrix multiplications.

**Proposition 3.18.** *Let $V, W$ be $\mathbb{K}$-vector spaces, with bases $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{C} = \{c_1, \dots, c_m\}$ respectively. Let $f : V \to W$ be a linear map and let $A := \mathcal{M}^{\mathcal{B}}_{\mathcal{C}}(f)$ be the corresponding transformation matrix. If $v = \sum_{j=1}^n \lambda_j b_j$ and $f(v) = \sum_{i=1}^m \mu_i c_i$, then*

$$A \cdot \begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\end{pmatrix} = \begin{pmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_m\end{pmatrix}$$

*In particular,* $v \in \ker f \Leftrightarrow \begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\end{pmatrix} \in \ker A$ *and* $\dim \operatorname{Im}(f) = \operatorname{rk}(f) = \operatorname{rk}(A)$.

*Proof.* By definition of $A$ we have $AX_{\mathcal{B}}(v) = X_{\mathcal{C}}(f(v))$, and the result follows by the definition of coordinate map. □

We have already seen a reason behind the definition of the matrix product (row-column), namely to translate linear systems into matrices. We see now other motivations.

**Proposition 3.19.** *Let $V, W, U$ be $\mathbb{K}$-vector spaces, with bases $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ respectively.*
*1) Let $f : V \to W$ and $g : W \to U$ be linear maps, then*

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{D}}(g \circ f) = \mathcal{M}^{\mathcal{C}}_{\mathcal{D}}(g) \cdot \mathcal{M}^{\mathcal{B}}_{\mathcal{C}}(f).$$

*2) Let $f : V \to W$ be an invertible linear map, then*

$$(\mathcal{M}^{\mathcal{B}}_{\mathcal{C}}(f))^{-1} = \mathcal{M}^{\mathcal{C}}_{\mathcal{B}}(f^{-1}).$$

*3) Let $\mathcal{B}'$, $\mathcal{C}'$ be bases of $V$ and $W$ respectively and let $f : V \to W$ be a linear map, then*

$$\mathcal{M}^{\mathcal{B}'}_{\mathcal{C}'}(f) = \mathcal{M}^{\mathcal{C}}_{\mathcal{C}'}(id_W) \cdot \mathcal{M}^{\mathcal{B}}_{\mathcal{C}}(f) \cdot \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}(id_V).$$

*Proof.* The proof of 1) is a computation using the various definitions, similar to the proof of Proposition 3.18.
2) and 3) follow from 1). □

**Definition 3.20.** Let $V$ be a $\mathbb{K}$-vector spaces, and let $\mathcal{B}, \mathcal{B}'$ be bases of $V$. The matrix

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_V),$$

is called *change of basis matrix* or *transition matrix*.

Note that $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_V) = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}(id_V)^{-1}$.

We stress that the columns of $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_V)$ are the coordinates of the vectors in $\mathcal{B}$ with respect to the basis $\mathcal{B}'$, and it holds $(v)_{\mathcal{B}'} = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_V) \cdot (v)_{\mathcal{B}}$.

**Example.** Let us consider $\mathbb{R}^2$ with bases $\mathcal{B} = \{e_1, e_2\}$ and $\mathcal{B}' = \{e_1 - 2e_2, 2e_1 - 3e_2\}$. Then

$$e_1 = -3(e_1 - 2e_2) + 2(2e_1 - 3e_2)\,, \qquad e_2 = -2(e_1 - 2e_2) + 1(2e_1 - 3e_2)\,,$$

Hence

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_{\mathbb{R}^2}) = \begin{pmatrix} -3 & -2 \\ 2 & 1 \end{pmatrix}$$

So, e.g. let $v \in \mathbb{R}^2$ be the vector having coordinates $(v)_{\mathcal{B}} = (4, -7)^T$ with respect to $\mathcal{B}$, then $v$ with respect to $\mathcal{B}'$ has coordinates

$$(v)_{\mathcal{B}'} = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(id_{\mathbb{R}^2}) = \begin{pmatrix} -3 & -2 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ -7 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}\,.$$

# Chapter 4

# Endomorphisms and diagonalization

In this chapter we focus on a certain family of linear maps, the endomorphisms: linear maps from a vector space to itself.

**Definition 4.1.** Let $V$ be a $\mathbb{K}$-vector space. A linear map $f : V \to V$ is called *endomorphism.*

We would like to understand how the choice of a basis $\mathcal{B}$ of $V$ affects the form of the transformation matrix $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f)$.

In particular, we would like to be able to find a basis $\mathcal{B}$, for which $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f)$ is as simplest as possible and for which the geometric meaning of $f$ becomes clearer.

**Example.** Let us consider the real vector space $V = \mathbb{R}^2$, and consider the linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by the reflection about a line $l \subset \mathbb{R}^2$ of slope $\theta$ passing through $0_{\mathbb{R}^2}$.

Let us consider the canonical basis $\mathcal{E} = \{e_1, e_2\}$ of $\mathbb{R}^2$ and write the transformation matrix $\mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(f)$: geometric and trigonometric considerations give:

$$f(e_1) = \begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}, \quad f(e_2) = \begin{pmatrix} \sin(2\theta) \\ -\cos(2\theta) \end{pmatrix} \implies \mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(f) = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} = A\,.$$

Let us now consider a basis $\mathcal{B} = \{b_1, b_2\}$ of $\mathbb{R}^2$, capturing the geometry of the transformation $f$. Let $b_1, b_2 \in V$ be vector such that $b_1$ lies on the line $l$ and $b_2$ is orthogonal to $l$; for example let $b_1 = \left(\begin{smallmatrix} \cos(\theta) \\ \sin(\theta) \end{smallmatrix}\right)$ and $b_2 = \left(\begin{smallmatrix} -\sin(\theta) \\ \cos(\theta) \end{smallmatrix}\right)$. In other words $f : xb_1 + yb_2 \mapsto xb_1 - yb_2$ and the transformation matrix $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f)$ is:

$$f(b_1) = b_1\,, f(b_2) = -b_2 \implies \mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = D\,.$$

Let $P$ be the change of basis matrix

$$P := \mathcal{M}^{\mathcal{B}}_{\mathcal{E}}(id_{\mathbb{R}^2}) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

By Proposition 3.19, changing basis affects the transformation matrix, as follows:

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f) = \mathcal{M}^{\mathcal{E}}_{\mathcal{B}}(id_V)\mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(f)\mathcal{M}^{\mathcal{B}}_{\mathcal{E}}(id_V) \qquad \text{i.e. } D = P^{-1}AP.$$

**Definition 4.2.** Let $A, B \in \mathcal{M}_{\mathbb{K}}(n,n)$ be matrices. They are *similar* if there exists an invertible matrix $P \in \mathcal{M}_{\mathbb{K}}(n,n)$ such that $B = P^{-1}AP$.

*Remark.* Two matrices $A, B \in \mathcal{M}_{\mathbb{K}}(n,n)$ are similar if and only if they represent the same linear map with respect to different bases.

In particular, similar matrices $A, B \in \mathcal{M}_{\mathbb{K}}(n,n)$ have the same rank and the same determinant $\det(B) = \det(P^{-1}AP) = \det(A)$ (by Binet).

**Question.** Given an endomorphism $f : V \to V$, is it possible to find a basis $\mathcal{B}$ of $V$ such that $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f)$ is diagonal?

In other words, given a matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$, are there a diagonal matrix $D \in \mathcal{M}_{\mathbb{K}}(n,n)$ and an invertible matrix $P \in \mathcal{M}_{\mathbb{K}}(n,n)$ such that $D = P^{-1}AP$?

Every endomorphism $f : V \to V$ can be encoded in a transformation matrix $A = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f) \in \mathcal{M}_{\mathbb{K}}(n,n)$, and interpreted as the endomorphism $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^n$. From now on we simplify the discussion by considering only endomorphisms of the form $\mathcal{L}_A : \mathbb{K}^n \to \mathbb{K}^n$, given by a square matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$.

All definitions and claims made for matrices can be transferred evenly to endomorphisms, by taking a transformation matrix: for example, an endomorphisms $f : V \to V$ is diagonalizable if and only if a transformation matrix $A = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(f) \in \mathcal{M}_{\mathbb{K}}(n,n)$ is diagonalizable (see Definition 4.4 and Theorems 4.5, 4.14 below).

## 4.1 Eigenvalues and eigenvectors

**Definition 4.3.** Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix.

A scalar $\lambda \in \mathbb{K}$ is an *eigenvalue* of $A$, if there exists a non-zero vector $v \in \mathbb{K}^n$, $v \neq 0$ such that $A \cdot v = \lambda \cdot v$.

A non-zero vector $v \in \mathbb{K}^n$, $v \neq 0$ is an *eigenvector* of $A$ if there exists a scalar $\lambda \in \mathbb{K}$ such that $A(v) = \lambda \cdot v$; in this case, $\lambda$ is called the *eigenvalue* of $v$.

The *spectrum* of $A$ is the set of all eigenvalues of $A$: $S(A) = \{\lambda \in \mathbb{K} \mid \lambda \text{ is an eigenvalue of } A\}$.

**Example.** 1) In the previous example, the vectors $b_1$ and $b_2$ are eigenvectors with eigenvalue 1 and $-1$ respectively.

2) Let $A = \begin{pmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \\ 2 & 4 & 1 \end{pmatrix} \in \mathcal{M}_{\mathbb{Q}}(3,3)$. The endomorphism $\mathcal{L}_A : \mathbb{Q}^3 \to \mathbb{Q}^3$ has eigenvector $u = (1,1,1)^T$ of eigenvalue 7, indeed $A \cdot u = (7,7,7) = 6 \cdot u$.

3) Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix and $v \in \ker(A), v \neq 0$, then $A(v) = 0 = 0_{\mathbb{K}} \cdot v$, so $0_{\mathbb{K}}$ is an eigenvalue of $A$.

*Remark.* a) By 3) we get that $\ker(A) \neq \{0\}$ if and only if $0_{\mathbb{K}}$ is an eigenvalue of $A$.
b) More in general, a scalar $\lambda \in \mathbb{K}$ is an eigenvalue of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ if and only if $\ker(\lambda \cdot I_n - A) \neq \{0\}$.

**Notation.** By writing a matrix containing several zeroes, we will use the convention to omit the zero entries.

We denote a diagonal matrix $\begin{pmatrix} \delta_1 & & & \\ & \delta_2 & & \\ & & \ddots & \\ & & & \delta_n \end{pmatrix}$ by the short-hand notation $\operatorname{diag}(\delta_1, \dots, \delta_n)$.

**Definition 4.4.** Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix. We say that $A$ is *diagonalizable* if there exists a basis of $\mathbb{K}^n$ made of eigenvectors of $A$.

The meaning of Definition 4.4 is that a matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ is diagonalizable if it is representable in diagonal form, i.e. $A$ is similar to a diagonal matrix.

**Theorem 4.5** ($1^{\text{st}}$ diagonalization criterion)**.** *Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix.*

*The matrix $A$ is diagonalizable if and only if there are a diagonal matrix $D \in \mathcal{M}_{\mathbb{K}}(n,n)$ and an invertible matrix $P \in \mathcal{M}_{\mathbb{K}}(n,n)$ such that $D = P^{-1}AP$.*

*Proof.* We start with the preliminary observation that $A = \mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(\mathcal{L}_A)$, where $\mathcal{E}$ is the canonical basis of $\mathbb{K}^n$.

We now prove the two implications.

$\Rightarrow$] If $A$ is diagonalizable, there exists a basis $\mathcal{C} = \{c_1, \dots, c_n\}$ of $\mathbb{K}^n$ made of eigenvectors. Let $D := \mathcal{M}^{\mathcal{C}}_{\mathcal{C}}(\mathcal{L}_A)$ and $P := \mathcal{M}^{\mathcal{C}}_{\mathcal{E}}(id)$, then $P$ is invertible (since $id$ is bijective) and $D$ is diagonal: $Ac_i = \lambda_i c_i$ and

$$D = \mathcal{M}^{\mathcal{C}}_{\mathcal{C}}(\mathcal{L}_A) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

By Proposition 3.19

$$P^{-1}AP = \mathcal{M}^{\mathcal{E}}_{\mathcal{C}}(id)\mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(\mathcal{L}_A)\mathcal{M}^{\mathcal{C}}_{\mathcal{E}}(id) = \mathcal{M}^{\mathcal{C}}_{\mathcal{C}}(\mathcal{L}_A) = D$$

$\Leftarrow$] Assume there are a diagonal matrix $D = \operatorname{diag}(\delta_1, \dots, \delta_n) \in \mathcal{M}_{\mathbb{K}}(n,n)$ and an invertible matrix $P \in \mathcal{M}_{\mathbb{K}}(n,n)$ such that $D = P^{-1}AP$. In other words, $D$ represents $\mathcal{L}_A$ with respect to another basis $\mathcal{C} = \{c_1, \dots, c_n\}$ and $P = \mathcal{M}^{\mathcal{C}}_{\mathcal{E}}(id)$.

We prove that $c_i$ are eigenvalues. Applying the definitions we get

$$Ac_i = PDP^{-1}c_i = PDe_i = P\delta_i e_i = \delta_i c_i$$ □

## 4.2 Characteristic polynomial

How do we find eigenvalues and eigenvectors?

Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix. We have the following chain of equivalences:

$$\begin{aligned} \lambda \in \mathbb{K} \text{ is an } \underline{\text{eigenvalue}} \text{ of } A \iff & \ker(\lambda \cdot I_n - A) \neq \{0_V\} \\ \iff & \operatorname{rk}(\lambda \cdot I_n - A) < n \\ \iff & \det(\lambda \cdot I_n - A) = 0 \end{aligned}$$

So to find the eigenvalues we look for scalars yielding an homogenous linear system $(\lambda \cdot I_n - A)x = 0$ having more that one solution.

The eigenvectors relative to an eigenvalue are the non-trivial solutions of this homogenous linear system.

**Example.** Consider the linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by the reflection about a line $l \subset \mathbb{R}^2$ of slope $\theta = \frac{\pi}{6}$ passing through $0_{\mathbb{R}^2}$. Its transformation matrix with respect to the canonical basis of $\mathbb{R}^2$ is:

$$A = \mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(f) = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

The eigenvalues of $A$ are real numbers such that

$$0 = \det(\lambda I_n - A) = \det \begin{pmatrix} \lambda - \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \lambda + \frac{1}{2} \end{pmatrix} = (\lambda - \frac{1}{2})(\lambda + \frac{1}{2}) - \frac{\sqrt{3}}{2} \cdot \frac{\sqrt{3}}{2} = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1),$$

so the eigenvalues are $\pm 1$.

Now, to determine the eigenvectors relative to the eigenvalue $\lambda = 1$, we have to compute $\ker(1 \cdot I_n - A)$:

$$\ker \begin{pmatrix} 1 - \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & 1 + \frac{1}{2} \end{pmatrix} = \ker \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{3}{2} \end{pmatrix} = \ker \begin{pmatrix} 1 & -\sqrt{3} \\ 0 & 0 \end{pmatrix} = \operatorname{Span}\left( \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \right)$$

Similarly, the eigenvectors relative to the eigenvalue $\lambda = -1$ are obtained by computing $\ker(-1 \cdot I_n - A)$:

$$\ker \begin{pmatrix} -1 - \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -1 + \frac{1}{2} \end{pmatrix} = \ker \begin{pmatrix} -\frac{3}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} = \ker \begin{pmatrix} \sqrt{3} & 1 \\ 0 & 0 \end{pmatrix} = \operatorname{Span}\left( \begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix} \right)$$

We got 2 eigenvalues ( $\left(\begin{smallmatrix} \sqrt{3} \\ 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 \\ -\sqrt{3} \end{smallmatrix}\right)$) which are linearly independent, so we got a basis of of $\mathbb{R}^2$: $A$ is diagonalizable (as we already knew).

Our aim is now to replicate this procedure in general and find a strategy to determine whether a square matrix is diagonalizable, and if so, how to find a basis of eigenvectors.

**Definition 4.6.** Let $A \in \mathcal{M}_{\mathbb{K}}(n, n)$ be a square matrix. The *characteristic polynomial* of $A$ is the polynomial $p_A(t) = \det(tI_n - A) \in \mathbb{K}[t]$.

**Proposition 4.7.** *Let $A \in \mathcal{M}_{\mathbb{K}}(n, n)$ be a square matrix.*

- *The eigenvalues of $A$ are the roots in $\mathbb{K}$ of its characteristic polynomial $p_A(t) \in \mathbb{K}[t]$.*
- *The characteristic polynomial $p_A(t)$ has degree $n$.*
- *In particular, the spectrum $S(A)$ is a set of cardinality at most $n$.*

**Example.** a) The characteristic polynomial of a matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{K}}(2, 2)$ is

$$\det \begin{pmatrix} t - a & -b \\ -c & t - d \end{pmatrix} = t^2 - (a + d)t + (ad - bc)$$

b) In general, the characteristic polynomial of a matrix $A \in \mathcal{M}_{\mathbb{K}}(n, n)$ has the form

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + (-1)^n \det A$$

Some authors define the characteristic polynomial to be $\det(A - tI_n)$. This definition differs from our definition $p_A(t) = \det(tI_n - A)$ by a sign $(-1)^n$, so it still have the property of having the eigenvalues of $A$ as roots; however our definition gives a *monic*[1] polynomial, whereas with the alternative definition it is monic only when $n$ is even.

[1] monic means that the coefficient of the term of highest degree is 1.

*Remark.* In the diagonalization problem, the field $\mathbb{K}$ plays an important role!

The natural chain of inclusions $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ yields the inclusions

$$\mathcal{M}_{\mathbb{Q}}(n,n) \subset \mathcal{M}_{\mathbb{R}}(n,n) \subset \mathcal{M}_{\mathbb{C}}(n,n)$$

It is possible that a matrix is diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$, and that a matrix is diagonalizable over $\mathbb{R}$ but not over $\mathbb{Q}$.

**Example.** c) The characteristic polynomial of $M = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ is $t^2 + 1$.

Over $\mathbb{R}$ the polynomial $t^2+1$ has no root, so $M$ has no eigenvalues, and it cannot be diagonalizable.

On the other hand, if $\mathbb{K} = \mathbb{C}$, then $t^2+1$ has 2 roots: $\pm i$ ($i$ is the imaginary unit: $i^2 = -1$) and we will see that $M$ is diagonalizable over $\mathbb{C}$.

d) The characteristic polynomial of $N = \left(\begin{smallmatrix} 0 & 1 \\ 2 & 0 \end{smallmatrix}\right)$ is $t^2 - 2$, which has no root in $\mathbb{Q}$ so $N$ cannot be diagonalizable over $\mathbb{Q}$. But $t^2 - 2$ has 2 real roots: $\pm\sqrt{2}$ and we will see that $N$ is diagonalizable over $\mathbb{R}$.

## 4.3 Multiplicities

Let $\lambda \in \mathbb{K}$ be an eigenvalue of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$. The eigenvectors relative to the eigenvalue $\lambda$ (together with 0) are the solutions of a certain homogeneous linear system, so they form a vector subspace of $\mathbb{K}^n$.

**Definition 4.8.** Let $\lambda \in \mathbb{K}$ be an eigenvalue of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$.

The *eigenspace* associated to $\lambda$ is the vector subspace $V_\lambda := \ker(\lambda I_n - A)$.

The *geometric multiplicity* of $\lambda$ is $g_\lambda := \dim V_\lambda = n - \operatorname{rk}(\lambda I_n - A)$.

The *algebraic multiplicity* of $\lambda$ is $a_\lambda := \max\{k \in \mathbb{N} \mid (t-\lambda)^k \text{ divides } p_A(t)\}$.

"The algebraic multiplicity of an eigenvalue $\lambda$ counts how many times $\lambda$ is a root of $p_A(t)$, i.e. to which power $(t-\lambda)$ appears in $p_A(t)$".

**Example.** Let $p_A(t) = (x-1)^3(x+2)$, then $f$ has 2 eigenvalues: 1 and $-2$, with algebraic multiplicities $a_1 = 3$, $a_{-2} = 1$.

Let $p_A(t) = (x+1)^4(x^2+1)^2$. Over $\mathbb{K} = \mathbb{Q}$ or $\mathbb{R}$, this polynomial has only $-1$ as root and $a_{-1} = 4$.

Over $\mathbb{K} = \mathbb{C}$, the polynomial factorize as $p_A(t) = (x+1)^4(x-i)^2(x+i)^2$, so it has 3 roots: $-1$, $i$, $-i$ with algebraic multiplicities $a_{-1} = 4$, $a_i = a_{-i} = 2$.

**Theorem 4.9** (Fundamental theorem of algebra)**.** *Every non-zero polynomial $p(t) \in \mathbb{C}[t]$ with complex coefficients has exactly $n$ roots, counted with multiplicity.*

*In other words, if $\{z_1, \ldots, z_r\} \subset \mathbb{C}$ are the complex roots of $p(t)$, then $\sum_{i=1}^{r} a_{z_i} = n$.*

**Proposition 4.10.** *Let $\lambda \in \mathbb{K}$ be an eigenvalue of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$. Then*

$$1 \leq g_\lambda \leq a_\lambda \leq n.$$

**Definition 4.11.** Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix and let $\lambda \in S(A)$.

The eigenvalue $\lambda$ is called *regular* if $a_\lambda = g_\lambda$.

The eigenvalue $\lambda$ is called *simple* if $a_\lambda = 1$.

Note that if $a_\lambda = 1$, then $g_\lambda = 1$; so a simple eigenvalue is automatically regular.

**Proposition 4.12.** *Let $\lambda_1, \lambda_2 \in \mathbb{K}, \lambda_1 \neq \lambda_2$ be distinct eigenvalues of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$. Then $V_{\lambda_1} \cap V_{\lambda_2} = \{0\}$*

*Proof.* Let $v \in V_{\lambda_1} \cap V_{\lambda_2}$. From $v \in V_{\lambda_1}$, we get $A(v) = \lambda_1 v$; and from $v \in V_{\lambda_2}$, we get $A(v) = \lambda_2 v$. Therefore, $\lambda_1 v = A \cdot v = \lambda_2 v$, so $(\lambda_1 - \lambda_2)v = 0$. Since $\lambda_1 - \lambda_2 \neq 0_{\mathbb{K}}$, by the zero-product property we get $v = 0$. □

In particular, two eigenvectors having distinct eigenvalues are linearly independent. More generally, "the eigenspaces carry independent information!"

**Proposition 4.13.** *Let $\lambda_1, \dots \lambda_r \in \mathbb{K}$ be pairwise distinct eigenvalues ($\lambda_i \neq \lambda_j$ for $i \neq j$) of $A \in \mathcal{M}_{\mathbb{K}}(n,n)$. For each $i = 1, \dots, r$, let $v_i$ be an eigenvalue in $V_{\lambda_i}$.*

*Then $\{v_1, \dots, v_r\} \subset \mathbb{K}^n$ is a set of linearly independent vectors.*

To diagonalize a matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ means that we can find a basis of $\mathbb{K}^n$ made of eigenvectors of $A$; in other words, we can reconstruct the effect of the transformation $A$ by looking at the effect on the single eigenenspaces. Since the eigenspaces carry independent informations, to diagonalize $A$ we need to find a basis $\mathcal{B}_{\lambda_i}$ for each eigenspace $V_{\lambda_i}$ and have $n = \sum_{i=1}^r g_{\lambda_i}$. Indeed,

$$\mathcal{B} = \mathcal{B}_{\lambda_1} \cup \cdots \mathcal{B}_{\lambda_r}$$

is a set of linearly independent vectors (by Proposition 4.13), and to generate $\mathbb{K}^n$, we need that $\mathcal{B}$ has $n$ elements.

Note that

$$\begin{array}{ccccccccc} a_{\lambda_1} & + & a_{\lambda_2} & + & \cdots & + & a_{\lambda_r} & \leq \deg p_A = n \\ \vee\!| & & \vee\!| & & & & \vee\!| & \\ g_{\lambda_1} & + & g_{\lambda_2} & + & \cdots & + & g_{\lambda_r} & \end{array}$$

so to have $g_{\lambda_1} + g_{\lambda_2} + \cdots + g_{\lambda_r} = n$ we need that all $\leq$ are equalities.

Summing up, we have deduced the following important criterion

**Theorem 4.14** (2nd diagonalization criterion)**.** *Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix. and let $S(A) = \{\lambda_1, \dots, \lambda_r\}$ be the set of eigenvalues of $A$ (roots in $\mathbb{K}$ of $p_A$). Then the following are equivalent:*

1. *$A$ is diagonalizable.*
2. *$\sum_{i=1}^{r} g_{\lambda_i} = n$.*
3. *$\sum_{i=1}^{r} a_{\lambda_i} = n$ and for every $a_{\lambda_i} = g_{\lambda_i}$ for every $i = 1, \dots, k$.*

**Corollary 4.15.** *Let $A \in \mathcal{M}_{\mathbb{K}}(n,n)$ be a square matrix.*
*If $p_A$ has all its roots in $\mathbb{K}$ and each eigenvalue is simple, then $A$ is diagonalizable.*

*Proof.* $p_A$ has all its roots in $\mathbb{K}$ means $\sum_{\lambda \in S(A)} a_\lambda = n$.

Each eigenvalue is simple means $1 \leq g_\lambda \leq a_\lambda = 1$, so $a_\lambda = g_\lambda$ for every $\lambda \in S(A)$. $\square$

**Example.** The characteristic polynomial of $M = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ is $t^2 + 1$, so $M$ is not diagonalizable over $\mathbb{R}$ (as already seen).

But over $\mathbb{C}$, the matrix $M$ has two eigenvalues: $\pm i$, both simple, so $M$ is diagonalizable over $\mathbb{C}$.

Similarly, the characteristic polynomial of $N = \left(\begin{smallmatrix} 0 & 1 \\ 2 & 0 \end{smallmatrix}\right)$ is $t^2 - 2$, so $N$ is not diagonalizable over $\mathbb{Q}$, but it is diagonalizable over $\mathbb{R}$.

**Example.** The characteristic polynomial of $B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathcal{M}_\mathbb{R}(2,2)$ is $(t-1)^2$, so it has a single eigenvalue 1 with algebraic multiplicity $a_1 = 2$.

Let us now compute the corresponding eigenspace:

$$V_1 = \ker(I_2 - B) = \ker \begin{pmatrix} 1-1 & -1 \\ 0 & 1-1 \end{pmatrix} = \ker \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} = \mathrm{Span}(\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)) \quad \Rightarrow \quad g_1 = 1 < 2$$

Thus, $B$ is not diagonalizable (neither over $\mathbb{R}$ nor over $\mathbb{C}$).

**Summary of diagonalization process**

`Input` A square matrix $A \in \mathcal{M}_\mathbb{K}(n,n)$

Step 1 Determine the characteristic polynomial $p_A(t) = \det(tI_n - A)$ of $A$ and factorize it over $\mathbb{K}$, i.e. determine the eigenvalues $\lambda_1, \ldots, \lambda_r$ of $A$ and their algebraic multiplicities $a_{\lambda_1}, \ldots, a_{\lambda_r}$.

Check 1 If $\sum_{i=1}^{r} a_{\lambda_i} < n$, then `Output`: $A$ is not diagonalizable. Else, go to Step 2:
[This check is not necessary if $\mathbb{K} = \mathbb{C}$, by Theorem 4.9]

Step 2 For each $\lambda_i$ find a basis $\mathcal{B}_{\lambda_i}$ of the eigenspace $V_{\lambda_i}$, and determine its geometric multiplicity $g_{\lambda_i}$.

Check 2 If $g_{\lambda_i} < a_{\lambda_i}$ for some $i = 1, \ldots, r$, then `Output`: $A$ is not diagonalizable. Else,

`Output` $A$ is diagonalizable: $\mathcal{B} = \mathcal{B}_{\lambda_1} \cup \ldots \cup \mathcal{B}_{\lambda_r}$ is a basis of $\mathbb{K}^n$ made of eigenvectors of $A$, and moreover:

$$\begin{array}{ccccccc} \mathcal{M}^{\mathcal{B}}_{\mathcal{E}}(id_V) & \cdot & \mathcal{M}^{\mathcal{E}}_{\mathcal{E}}(\mathcal{L}_A) & \cdot & \mathcal{M}^{\mathcal{E}}_{\mathcal{B}}(id_V) & = & \mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(\mathcal{L}_A) \\ \| & & \| & & \| & & \| \\ P^{-1} & \cdot & A & \cdot & P & = & D \end{array}$$

where $D = \mathrm{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{a_{\lambda_1}-\text{times}}, \ldots, \underbrace{\lambda_r, \ldots, \lambda_r}_{a_{\lambda_r}-\text{times}})$ and the columns of $P$ are the vectors of $\mathcal{B}_{\lambda_1} \cup \ldots \cup \mathcal{B}_{\lambda_r}$ (respecting the order!).

**Example.** • The matrix $A = \begin{pmatrix} 3 & -2 & -1 \\ 0 & 0 & 1 \\ 2 & -2 & -1 \end{pmatrix}$ has characteristic polynomial

$$\begin{aligned} p_A(t) &= \det \begin{pmatrix} t-3 & 2 & 1 \\ 0 & t & -1 \\ -2 & 2 & t+1 \end{pmatrix} = t \det \begin{pmatrix} t-3 & 1 \\ -2 & t+1 \end{pmatrix} + \det \begin{pmatrix} t-3 & 2 \\ -2 & 2 \end{pmatrix} \\ &= t(t^2 - 2t - 1) + 2t - 2 = t^3 - 2t^2 + t - 2 = (t^2+1)(t-2) \end{aligned}$$

$A$ is not diagonalizable over $\mathbb{R}$, since 2 is the only real root of $p_A(t)$ and $a_2 = 1$.

On the other hand, over $\mathbb{C}$ the matrix $A$ has 3 distinct roots, each one with algebraic multiplicity one, so over $\mathbb{C}$ the matrix $A$ is diagonalizable (Corollary 4.15).

• The matrix $A = \begin{pmatrix} 3 & -1 & -2 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$ has characteristic polynomial

$$\begin{aligned} p_A(t) &= \det \begin{pmatrix} t-3 & 1 & 2 \\ -1 & t & 1 \\ 0 & -1 & t-1 \end{pmatrix} = (t-3)\det \begin{pmatrix} t & 1 \\ -1 & t-1 \end{pmatrix} + \det \begin{pmatrix} 1 & 2 \\ -1 & t-1 \end{pmatrix} \\ &= (t-3)(t^2-t+1) + (t+1) = t^3 - 4t^2 + 5t - 2 = (t-1)^2(t-2) \end{aligned}$$

Over $\mathbb{R}$ (and also over $\mathbb{C}$) the matrix $A$ has 2 roots: 1,2 with $a_1 = 2$ and $a_2 = 1$: $1 + 2 = 3$ so Check 1 is passed.

We determine now the geometric multiplicities of the 2 eigenvalues: since 2 is a simple eigenvalue, we know $g_2 = 1$, so we proceed by computing $g_1 = \dim V_1$:

$$V_1 = \ker(1I_3 - A) = \ker \begin{pmatrix} -2 & 1 & 2 \\ -1 & 1 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & -2 & 0 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \operatorname{Span}\left( \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right)$$

so $g_1 = 1 < 2 = a_1$. Therefore, $A$ is not diagonalizable neither over $\mathbb{R}$ nor over $\mathbb{C}$.

• The matrix $A = \begin{pmatrix} 4 & -3 & -3 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$ has characteristic polynomial

$$\begin{aligned} p_A(t) &= \det \begin{pmatrix} t-4 & 3 & 3 \\ -1 & t & 1 \\ -1 & 1 & t \end{pmatrix} = -\det \begin{pmatrix} 3 & 3 \\ t & 1 \end{pmatrix} + (-1) \begin{pmatrix} t-4 & 3 \\ -1 & 1 \end{pmatrix} + t \begin{pmatrix} t-4 & 3 \\ -1 & t \end{pmatrix} \\ &= -(3-3t) - (t-1) + t(t^2 - 4t + 3) = t^3 - 4t^2 + 5t - 2 = (t-1)^2(t-2) \end{aligned}$$

The characteristic polynomial is the same of the previous example, so we proceed directly to compute $g_1 = \dim V_1$:

$$V_1 = \ker(1I_3 - A) = \ker \begin{pmatrix} -3 & 3 & 3 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \operatorname{Span}\left( \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right)$$

so $g_1 = \dim V_1 = 2 = 2 = a_1$. Therefore, $A$ is diagonalizable both over $\mathbb{R}$ and over $\mathbb{C}$.

To determine $P$ we are missing a basis of $V_2$, so we determine it:

$$V_2 = \ker(2I_3 - A) = \ker \begin{pmatrix} -2 & 3 & 3 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & -2 \\ -1 & 2 & 1 \\ -2 & 3 & 3 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & -2 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{pmatrix} = \operatorname{Span}\left( \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \right)$$

Let $P := \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ and $D := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, then $D = P^{-1}AP$.

## 4.4 Companion matrix

**Question.** Is any *monic* polynomial $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \in \mathbb{K}[t]$ the characteristic polynomial of a matrix $A \in \mathcal{M}_{\mathbb{K}}(n,n)$?

**Definition 4.16.** The *companion matrix* of the monic polynomial $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \in \mathbb{K}[t]$ is the square matrix $C(p) \in \mathcal{M}_{\mathbb{K}}(n,n)$ defined as

$$C(p) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}$$

**Example.** • The companion matrix of $t^2 - 2$ is $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$.

• The companion matrix of $t^3 - 2t^2 + 1$ is $\left(\begin{smallmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 2 \end{smallmatrix}\right)$.

**Lemma 4.17.** *The characteristic polynomial of $C(p)$ is $p(t)$.*

*Proof.* The characteristic polynomial $C(p)$ is $\det(tI_n - C(p)) =$

$$\det\begin{pmatrix} t & 0 & \dots & 0 & a_0 \\ -1 & t & \dots & 0 & a_1 \\ 0 & -1 & \dots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & t+a_{n-1} \end{pmatrix} = (-1)^{n-1}a_0 \det\underbrace{\begin{pmatrix} -1 & t & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{pmatrix}}_{=U} + t\det\underbrace{\begin{pmatrix} t & 0 & \dots & 0 & a_1 \\ -1 & t & \dots & 0 & a_2 \\ 0 & -1 & \dots & 0 & a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & t+a_{n-1} \end{pmatrix}}_{=M}$$

Since the determinant of the upper triangular matrix $U$ is $\det U = (-1)^{n-1}$, we get $\det(tI_n - C(p)) = a_0 + t\det(M)$. Expanding the determinant of $M \in \mathcal{M}_{\mathbb{K}}(n-1, n-1)$ along the first row, we achieve $\det(tI_n - C(p)) = a_0 + a_1 t + t^2 \det(N)$, where $N \in \mathcal{M}_{\mathbb{K}}(n-2, n-2)$ is obtained from $C(p)$ by removing the first two rows and columns.

Repeating this process we achieve

$$\begin{aligned} \det(tI_n - C(p)) &= a_0 + a_1 t + a_2 t^2 + \cdots + a_{n-3}t^{n-3} + t^{n-2}\det\begin{pmatrix} t & a_{n-2} \\ -1 & t + a_{n+1} \end{pmatrix} \\ &= a_0 + a_1 t + \cdots + t^{n-2}(t^2 + ta_{n+1} + a_{n+2}) = p(t) \end{aligned}$$

□

**Example.** • We would like to find a $3 \times 3$ matrix $M$ whose characteristic polynomial has three real roots, but only one is rational, e.g. $p(t) = (t-1)(t^2-2) = t^3 - t^2 - 2t + 2$.
Pick $M = C(p) = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$

• We would like to find a $3 \times 3$ matrix $N$ whose characteristic polynomial has only one real root, e.g. $q(t) = (t+3)(t^2+1) = t^3 + 3t^2 + t + 3$: pick $N = C(q) = \begin{pmatrix} 0 & 0 & -3 \\ 1 & 0 & -1 \\ 0 & 1 & -3 \end{pmatrix}$

# Chapter 5

# Real Euclidian space

In this chapter we discuss geometric properties of $\mathbb{R}^n$ related to its *euclidean* structure. We work over the real field $\mathbb{K} = \mathbb{R}$ and endow $\mathbb{R}^n$ with a "product of two vectors": the standard[1] *scalar product*. In particular, we introduce orthogonality and orthogonal projections, and discuss how these interact with the diagonalization problem.

**Definition 5.1.** The (standard[1]) *scalar product* on $\mathbb{R}^n$ is

$$\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\,, \quad \langle v, w \rangle = \left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = v_1 w_1 + \cdots + v_n w_n\,.$$

The scalar product satisfies the following properties, which are easy to verify.

**Properties.** *The scalar product is*

- *symmetric:* $\langle v, w \rangle = \langle w, v \rangle$ *holds* $\forall v, w \in \mathbb{R}^n$*;*
- *bilinear (= linear in both entries):* $\langle v + \lambda u, w \rangle = \langle v, w \rangle + \lambda \langle u, w \rangle$*, and* $\langle v, \lambda u + w \rangle = \langle v, w \rangle + \lambda \langle v, u \rangle$ *hold* $\forall v, w, u \in \mathbb{R}^n$*,* $\lambda \in \mathbb{R}$*;*
- *positive definite:* $\langle v, v \rangle \geq 0$ *holds* $\forall v \in \mathbb{R}^n$*, and* $\langle v, v \rangle = 0 \Leftrightarrow v = 0$*;*
- *non-degenerate:* $\langle v, w \rangle = 0$ *for all* $w \in \mathbb{R}^n$ *if and only if* $v = 0$.

**Example.** For example ($n = 3$): $\left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\rangle = 0 \cdot 1 + 2 \cdot 1 + 3 \cdot (-1) = -1$.

$$\left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\rangle = 1^2 + 2^2 + 3^2 = 14, \qquad \left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, -2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\rangle = -2(-1) + 14 = 16$$

**Definition 5.2.** Let $v \in \mathbb{R}^n$. The *norm* of $v$ is $||v|| = \sqrt{\langle v, v \rangle}$.

A vector $v \in \mathbb{R}^n$ is a *unit vector* if $||v|| = 1$.

Note that, if $v \in \mathbb{R}^n, v \neq 0$, then $\dfrac{v}{||v||}$ is a unit vector.

---

[1] It is also called canonical or euclidean scalar product, or dot product.
Be aware that $\mathbb{R}^n$ can be endowed with other scalar products, which will not be discussed in this course.

**Example.** For example ($n = 3$): $||(1,2,3)^T|| = \sqrt{14}$ so $(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}})^T$ is a unit vector.

**Properties.** *Let $v \in \mathbb{R}^n$. Then*

- $||v|| \geq 0$ *and* $||v|| = 0 \Leftrightarrow v = 0$*;*
- $||\lambda \cdot v|| = |\lambda| \cdot ||v||$ *for all* $\lambda \in \mathbb{R}$*;*
- $||v + w|| \leq ||v|| + ||w||$ *(Triangular inequality)*

The first 2 properties follow from the definition of norm. The third one generalize the (well known) fact that in a triangle the sum of the length of two sides always exceeds the length of the third side. This property can be proven by using the following important inequality.

**Theorem 5.3** (Cauchy-Schwartz inequality)**.** *Consider the Euclidean space $\mathbb{R}^n$ endowed with the standard scalar product $\langle \cdot, \cdot \rangle$. Then, for every $v, w \in \mathbb{R}^n$ it holds*

$$|\langle v, w \rangle| \leq ||v|| \cdot ||w|| \tag{5.1}$$

*Moreover, in* (5.1) *the equality holds if and only if $v$ and $w$ are linearly dependent.*

*Proof.* If $w = 0$, then $|\langle v, w \rangle| = 0 \leq ||v|| \cdot ||w||$; so let us assume $w \neq 0$.

The vectors $v, w$ are linearly dependent if and only if there exists $\lambda \in \mathbb{R}$ such that $v = \lambda w$, i.e. $||v - \lambda w|| = 0$. To use this observation, we introduce the function $g : \mathbb{R} \to \mathbb{R}$ defined via

$$g(t) = ||v - tw|| = \langle v - tw, v - tw \rangle = \langle v, v \rangle - 2t\langle v, w \rangle + t^2 \langle w, w \rangle$$

So, $g(t) = t^2||w||^2 - 2t\langle v, w \rangle + ||v||^2$ is a polynomial of degree 2 and $g(t) \geq 0$ for all $t \in \mathbb{R}$; moreover, there exists $\lambda \in \mathbb{R}$ such that $g(\lambda) = 0$ if and only if $v$ and $w$ are linearly dependent. Therefore, the discriminant of $g$

$$\Delta = 4\langle v, w \rangle^2 - 4||w||^2||v||^2 \leq 0 \Leftrightarrow \langle v, w \rangle^2 \leq ||w||^2||v||^2$$

and $\langle v, w \rangle^2 \leq ||w||^2||v||^2$ if and only if $v$ and $w$ are linearly dependent.

Since $||w||||v|| \geq 0$, the statement follows by taking the square root of both sides. □

Let $v, w \in \mathbb{R}^n$ be non-zero vectors: $||v||, ||w|| > 0$. By the Cauchy-Schwartz inequality we get:

$$-1 \leq \frac{\langle v, w \rangle}{||w||||v||} \leq 1$$

So we can define the "angle" between $v, w$ as follows:

**Definition 5.4.** Let $v, w \in \mathbb{R}^n$ be non-zero vectors: $||v||, ||w|| > 0$. The *angle* between $v$ and $w$ is the unique $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{\langle v, w \rangle}{||w|| \cdot ||v||}.$$

*Remark.* Form this definition, we deduce $\langle v, w \rangle = ||v|| \cdot ||w|| \cdot \cos\theta$, which coincides with the definition of scalar product you have probably seen in physics.

**Example.** The angle between $v = (1,2,3)^T$ and $w = (0,1,-1)^T$ is

$$\arccos\left(\frac{-1}{\sqrt{14}\sqrt{2}}\right) = 1.76...$$

The angle between $u_1 = (2,1)^T$ and $u_2 = (1,-2)^T$ is $\frac{\pi}{2}$, indeed $\langle u_1, u_2 \rangle = 0$.

## 5.1 Orthogonality

**Definition 5.5.** Two vectors $v, w \in \mathbb{R}^n$ are *orthogonal* if $\langle v, w\rangle = 0$ (one writes $v \perp w$). Two vector subspaces $V, W \triangleleft \mathbb{R}^n$ are *orthogonal* if $\langle v, w\rangle = 0$ for all $v \in V$ and $w \in W$.

*Remark.* The zero vector is orthogonal to any other vector: $0 \perp v$ for all $v \in \mathbb{R}^n$

**Definition 5.6.** Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$ and let $\mathcal{B} = \{b_1, \dots, b_p\}$ be a basis of $H$.

$\mathcal{B}$ is an *orthogonal basis* of $H$ if $\langle b_i, b_j\rangle = 0$ for all $i \neq j$.

$\mathcal{B}$ is an *orthonormal basis* of $H$ if $\langle b_i, b_j\rangle = 0$ for all $i \neq j$ and $||b_i|| = 1$ for all $i$.

**Example.** The canonical basis $\mathcal{E}$ of $\mathbb{R}^n$ is an orthonormal basis of $\mathbb{R}^n$.

The basis $\{(1,1)^T, (1,-1)^T\}$ of $\mathbb{R}^2$ is an orthogonal basis, but it is not an orthonormal basis. To obtain an orthonormal basis, we have to rescale each vector: $||(1,1)|| = ||(1,-1)|| = \sqrt{2}$, so $\left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T, \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^T\right\}$ is an orthonormal basis of $\mathbb{R}^2$.

**Theorem 5.7.** *Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$ and let $\mathcal{B}_H = \{b_1, \dots, b_p\}$ be an orthogonal basis of $H$. Then for each $v \in H$ we have*

$$v = \frac{\langle v, b_1\rangle}{||b_1||^2} b_1 + \cdots \frac{\langle v, b_p\rangle}{||b_p||^2} b_p$$

*If $\mathcal{B}_H = \{b_1, \dots, b_p\}$ is an orthonormal basis of $H$, then $v = \langle v, b_1\rangle b_1 + \cdots \langle v, b_p\rangle b_p$.*

*Proof.* $\mathcal{B}_H$ is a basis of $H$, so for each $v \in H$ there are scalars $\lambda_j \in \mathbb{R}$ such that $v = \lambda_1 b_1 + \ldots + \lambda_p b_p$. By computing $\langle v, b_j\rangle$ we get

$$\langle v, b_j\rangle = \langle \sum_{i=1}^{p} \lambda_i b_i, b_j\rangle = \sum_{i=1}^{p} \lambda_i \langle b_i, b_j\rangle = \lambda_j \langle b_j, b_j\rangle \Longrightarrow \lambda_j = \frac{\langle v, b_j\rangle}{||b_j||^2} \qquad \square$$

**Definition 5.8.** The coefficient $\lambda_j = \dfrac{\langle v, b_j\rangle}{||b_j||^2}$ is the *Fourier coefficients* of $v$ with respect to the vector $b_j$.

**Example.** The basis $\{(1,-2)^T, (2,1)^T\}$ is an orthogonal basis of $\mathbb{R}^2$, and $v = (4,3)^T$ decomposes as:

$$\begin{pmatrix}4\\3\end{pmatrix} = \frac{\langle (4,3)^T, (1,-2)^T\rangle}{||(1,-2)^T||^2}\begin{pmatrix}1\\-2\end{pmatrix} + \frac{\langle (4,3)^T, (2,1)^T\rangle}{||(2,1)^T||^2}\begin{pmatrix}2\\1\end{pmatrix} = -\frac{2}{5}\begin{pmatrix}1\\-2\end{pmatrix} + \frac{11}{5}\begin{pmatrix}2\\1\end{pmatrix}.$$

We see now a process to construct an orthogonal (orthonormal) basis of a vector subspace $H \triangleleft \mathbb{R}^n$ by starting from a given basis $\mathcal{B} = \{b_1, \dots, b_p\}$ of $H$.

### 5.1.1 Gram-Schmidt process

Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$ and let $\mathcal{B} = \{b_1, \dots, b_p\}$ be a basis of $H$. We construct recursively an orthogonal basis $\mathcal{C}$ of $H$, via the *Gram-Schmidt process*:

- $c_1 := b_1$;

- $c_2 := b_2 - \frac{\langle b_2, c_1\rangle}{||c_1||^2} c_1$
- $c_3 := b_3 - \frac{\langle b_3, c_1\rangle}{||c_1||^2} c_1 - \frac{\langle b_3, c_2\rangle}{||c_2||^2} c_2$
- $\ldots\ c_j := b_j - \sum_{k=1}^{j-1} \frac{\langle b_j, c_k\rangle}{||c_k||^2} c_k$

Finally, to determine an orthonormal basis $\mathcal{D}$ of $H$ we *normalize* each $c_j$: $d_j := \frac{c_j}{||c_j||}$.

Moreover, at each step we do not modify the span of the first $j$ vectors:

**Lemma 5.9.** *For all $j = 1, \ldots, p$ it holds:*

$$\mathrm{Span}(b_1, \ldots, b_j) = \mathrm{Span}(c_1, \ldots, c_j) = \mathrm{Span}(d_1, \ldots, d_j)\,.$$

*Proof.* By induction, we assume $\mathrm{Span}(b_1, \ldots, b_{j-1}) = \mathrm{Span}(c_1, \ldots, c_{j-1})$ and it follows:

$$\begin{aligned}\mathrm{Span}(c_1, \ldots, c_{j-1}, c_j)\,\mathrm{Span}(c_1, \ldots, c_{j-1}, b_j - \sum_{k=1}^{j-1} \frac{\langle b_j, c_k\rangle}{||c_k||^2} c_k) &= \mathrm{Span}(c_1, \ldots, c_{j-1}, b_j)\\ &= \mathrm{Span}(b_1, \ldots, b_{j-1}, b_j)\quad \square\end{aligned}$$

Let us now verify that the basis $\mathcal{C}$ obtained via the Gram-Schmidt process is indeed an orthogonal basis. We do it by assuming at each step that $c_1, \ldots, c_{j-1}$ are pairwise orthogonal, and by showing that the "new vector" $c_j$ is orthogonal to the previous ones:

$$\langle c_i, c_j\rangle = \langle c_i, b_j - \sum_{k=1}^{j-1} \frac{\langle b_j, c_k\rangle}{||c_k||^2} c_k\rangle = \langle c_i, b_j\rangle - \sum_{k=1}^{j-1} \frac{\langle b_j, c_k\rangle}{||c_k||^2}\langle c_i, c_k\rangle = \langle c_i, b_j\rangle - \frac{\langle b_j, c_i\rangle}{||c_i||^2}\langle c_i, c_i\rangle = 0$$

*Remark.* Every orthogonal (respectively orthonormal) basis of $H$ can be completed to a orthogonal (respectively orthonormal) basis of $\mathbb{R}^n$.

**Example.** In $\mathbb{R}^4$ consider the basis $\left\{b_1 = \begin{pmatrix}0\\1\\0\\1\end{pmatrix}, b_2 = \begin{pmatrix}2\\1\\0\\1\end{pmatrix}, b_3 = \begin{pmatrix}-1\\0\\0\\1\end{pmatrix}, b_4 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}\right\}$. We apply the Gram-Schmidt process to construct an orthogonal basis $\mathcal{C}$ of $\mathbb{R}^4$:

$$\begin{aligned}
c_1 &= b_1 = \begin{pmatrix}0\\1\\0\\1\end{pmatrix},\\
c_2 &= b_2 - \frac{\langle b_2, c_1\rangle}{||c_1||^2} c_1 = \begin{pmatrix}2\\1\\0\\1\end{pmatrix} - \frac{2}{2}\begin{pmatrix}0\\1\\0\\1\end{pmatrix} = \begin{pmatrix}2\\0\\0\\0\end{pmatrix}\\
c_3 &= b_3 - \frac{\langle b_3, c_1\rangle}{||c_1||^2} c_1 - \frac{\langle b_3, c_2\rangle}{||c_2||^2} c_2 = \begin{pmatrix}-1\\0\\0\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}0\\1\\0\\1\end{pmatrix} - \frac{-2}{4}\begin{pmatrix}2\\0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\-1/2\\0\\1/2\end{pmatrix}\\
c_4 &= b_4 - \frac{\langle b_4, c_1\rangle}{||c_1||^2} c_1 - \frac{\langle b_4, c_2\rangle}{||c_2||^2} c_2 - \frac{\langle b_4, c_3\rangle}{||c_3||^2} c_3 = b_4 - 0c_1 - 0c_2 - 0c_3 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}
\end{aligned}$$

### 5.1.2 Orthogonal projections

Let us analyse the geometric meaning of the step $c_2 := b_2 - \frac{\langle b_2, c_1\rangle}{||c_1||^2} c_1$.

Let $\theta$ be the angle between $c_1$ and $b_2$, so that

$$\langle b_2, c_1\rangle = ||b_2|| \cdot ||c_1|| \cdot \cos\theta \quad \Longrightarrow \quad c_2 := b_2 - \frac{\langle b_2, c_1\rangle}{||c_1||^2} c_1 = b_2 - \underbrace{\frac{c_1}{||c_1||} \cdot ||b_2|| \cdot \cos\theta}_{w}$$

The vector $w$ is a multiple of $c_1$ of length $||b_2|| \cdot \cos\theta$, so we are removing from $b_2$ its component lying in the direction of $c_1$ and we are left with a vector orthogonal to $c_1$. Geometric idea: $w$ is the "orthogonal projection" of $b_2$ onto the direction of $c_1$, and so $c_2$ is a vector orthogonal to $c_1$. This happens at each step, so let us formalize it.

**Definition 5.10.** Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$ and let $\mathcal{B}_H = \{b_1, \dots, b_p\}$ be an orthogonal basis of $H$. The *orthogonal projection* onto $H$ is the linear map:

$$\begin{aligned} \pi_H : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ v &\mapsto \sum_{k=1}^{p} \frac{\langle v, b_k\rangle}{||b_k||^2} b_k \end{aligned}$$

*Remark.* i) At first glance it may seem that the map depends on the choice of the of the orthogonal basis $\mathcal{B}_H$ of $H$, but Proposition 5.12 shows that the output does not depend on this choice.

ii) $\pi_H(u) = u \Leftrightarrow u \in H$.

iii) $\pi_H(v) = 0 \Leftrightarrow v \perp b_j \forall j \Leftrightarrow v \perp h \forall h \in H$.

iv) Each step of the Gram-Schmidt process can be rephrased as $c_j := b_j - \pi_{H_j}(b_j)$, where $H_j = \text{Span}(c_1, \dots, c_{j-1})$.

**Example.** Let $v = (2, 3, 0)^T \in \mathbb{R}^3$ and let $H = \text{Span}((1, 0, 1)^T, (0, 1, 0)^T) \subseteq \mathbb{R}^3$.

$(1, 0, 1)^T$ and $(0, 1, 0)^T$ are orthogonal, so the orthogonal projection of $v$ onto $H$ is:

$$\pi_H(v) = \frac{\langle (2,3,0)^T, (1,0,1)^T\rangle}{||(1,0,1)^T||^2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{\langle (2,3,0)^T, (0,1,0)^T\rangle}{||(0,1,0)^T||^2} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$$

**Definition 5.11.** Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$. The *orthogonal complement* of $H$ is

$$H^\perp = \{v \in \mathbb{R}^n \mid \langle v, u\rangle = 0 \, \forall u \in H\}$$

Note that $H^\perp = \ker(\pi_H)$; and if $\mathcal{B} = \{b_1, \dots, b_p\}$ is a basis of $H$, to check if $v \in H^\perp$ it is enough to verify $\langle v, b_j\rangle = 0$ for all $j = 1, \dots, p$.

**Proposition 5.12.** *Let $H \triangleleft \mathbb{R}^n$ be a vector subspace of $\mathbb{R}^n$, then*

$$H \oplus H^\perp = \mathbb{R}^n$$

*In particular,* $\dim H + \dim H^\perp = n$.

*Moreover, Pythagoras' Theorem holds: by writing $v \in \mathbb{R}^n$ as $v = u + w$, $u \in H$, $w \in H^\perp$, we get $||v||^2 = ||u||^2 + ||w||^2$.*

*Proof.* Let $u \in H \cap H^\perp$, then $\langle u, u\rangle = 0$, so $u = 0$, so the sum $H + H^\perp$ is direct, and we have to show that it is the whole $\mathbb{R}^n$.

Let $v \in \mathbb{R}^n$, then $v = \pi_H(v) + (v - \pi_H(v))$ and $\pi_H(v) \in H$, while $(v - \pi_H(v)) \in H^\perp$, indeed $\pi_H(v - \pi_H(v) = \pi_H(v) - \pi_H(\pi_H(v)) = \pi_H(v) - \pi_H(v) = 0$

The last claim is a simple computation using $\langle u, w\rangle = 0$: $||v||^2 = ||u + w||^2 = \langle u + w, u + w\rangle = \langle u, u\rangle + 2\langle u, w\rangle + \langle w, w\rangle = ||u||^2 + ||w||^2$ □

*Remark.* i) Because of the decomposition $\mathbb{R}^n = H \oplus H^\perp$ we have that $v = \pi_H(v) + \pi_{H^\perp}(v)$, in other words: $v - \pi_H(v) = \pi_{H^\perp}(v)$.

ii) $(H^\perp)^\perp = H$

**Example.** In $\mathbb{R}^3$ consider $H = \text{Span}((1, 2, 3)^T)$, then

$$\pi_H((x, y, z)^T) = \frac{\langle (x, y, z)^T, (1, 2, 3)^T\rangle}{||(1, 2, 3)^T||^2}\begin{pmatrix}1\\2\\3\end{pmatrix} = \frac{x + 2y + 3z}{14}\begin{pmatrix}1\\2\\3\end{pmatrix}$$

and $H^\perp = \{(x, y, z)^T \in \mathbb{R}^3 \mid x + 2y + 3z = 0\}$.

In $\mathbb{R}^3$ consider $K = \{(x, y, z)^T \in \mathbb{R}^3 \mid x + 2y + 3z = 0, 4x + 5y + 6z = 0\}$. So a vector in $K$ is orthogonal to $(1, 2, 3)^T$ and to $(4, 5, 6)^T$, hence $K^\perp = \text{Span}((1, 2, 3)^T, (4, 5, 6)^T)$.

*Remark.* If $H = \text{Sol}(A|0)$ (cartesian equations), then it is easy to have parametric equations of $H^\perp$, indeed $H^\perp = \text{Row}(A) = \text{Span}(R_1^T, \ldots, R_m^T)$, where $R_i$ are the rows of $A$.

Conversely, if $H = \text{Span}(v_1, \ldots, v_p)$ (parametric equations), then it is easy to have cartesian equations of $H^\perp$, indeed $H^\perp = \text{Sol}(M|0)$, where $M$ is the matrix having $v_1^T, \ldots, v_p^T$ as rows.

## 5.2 Orthogonality and diagonalization

**Definition 5.13.** Let $A \in \mathcal{M}_{\mathbb{R}}(n, n)$ be a square matrix.

$A$ is *symmetric* if $A^T = A$.

$A$ is *orthogonal* if $A^T = A^{-1}$ (in particular, $A$ is invertible).

$A$ is *orthogonally diagonalizable* if there exists an orthogonal basis of $\mathbb{R}^n$ made of eigenvectors of $A$.

Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be an orthogonal basis of $\mathbb{R}^n$ made of eigenvectors of $A$ and consider the corresponding orthonormal basis $\mathcal{C} = \{c_1 = \frac{b_1}{||b_1||}, \ldots, c_n = \frac{b_n}{||b_n||}\}$.

The basis $\mathcal{C}$ is a basis of eigenvectors for $A$, so by the first diagonalization theorem $P^{-1}AP = D$ is a diagonal matrix, where $P = \mathcal{M}^{\mathcal{C}}_{\mathcal{E}}(id_{\mathbb{R}^n})$: its columns are the vectors $c_1, \ldots, c_n$, and it holds

$$(P^T \cdot P)_{i,j} = c_i^T \cdot c_j = \langle c_i, c_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \Longrightarrow P^T \cdot P = I_n$$

In other words $P$ is an orthogonal matrix.

*Remark.* A matrix $P \in \mathcal{M}_{\mathbb{R}}(n, n)$ is orthogonal if and only if its columns (or rows) form an orthonormal basis of $\mathbb{R}^n$.

Reformulating: $A \in \mathcal{M}_{\mathbb{R}}(n,n)$ orthogonally diagonalizable means that there exists matrices $P, D \in \mathcal{M}_{\mathbb{R}}(n,n)$, $P$ orthogonal, $D$ invertible such that $D = P^{-1}AP = P^{T}AP$, i.e. $A = PDP^{T}$. Considering the transpose of both sides, we get

$$A^{T} = (PDP^{T})^{T} = (P^{T})^{T} D^{T} P^{T} = PDP^{T} = A$$

so an orthogonally diagonalizable matrix is symmetric. The converse holds true as well:

**Theorem 5.14** (Spectral theorem). *Let $A \in \mathcal{M}_{\mathbb{R}}(n,n)$ be a square matrix. $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.*

The proof of this result is out of the scope of these notes, but we can understand how to construct $P$ and $D$ starting from a symmetric matrix $A$.

The idea is to use the “standard” diagonalization process (see page 55) and adapt it.

`Input` A symmetric square matrix $A \in \mathcal{M}_{\mathbb{R}}(n,n)$

Step 1 Determine the characteristic polynomial $p_A(t) = \det(tI_n - A)$ of and the eigenvalues $\lambda_1, \ldots, \lambda_r$ of $A$.

Step 2 For each $\lambda_i$ find a basis $\mathcal{B}_{\lambda_i}$ of the eigenspace $V_{\lambda_i}$.

Note that Check 1 and 2 are not necessary, since $A$ is diagonalizable by the spectral theorem: so we have to find $n$ real roots (counted with multiplicity) and each root has geometric multiplicity equal to the algebraic one.

So we got an `intermedite output`: $\mathcal{B} = \mathcal{B}_{\lambda_1} \cup \ldots \cup \mathcal{B}_{\lambda_r}$ is a basis of $\mathbb{R}^n$ made of eigenvectors of $A$, but in general it is not orthogonal. Nevertheless, the orthogonality between eigenvector of distinct eigenspaces is guaranteed by the following lemma:

**Lemma 5.15.** *Let $v, w$ be eigenvectors of the symmetric matrix $A \in \mathcal{M}_{\mathbb{R}}(n,n)$ belonging to distinct eigenspaces: $v \in V_\lambda$, $w \in V_\mu$ with $\lambda \neq \mu$. Then $\langle v, w\rangle = 0$.*

*Proof.* Let us consider $\langle v, Aw\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle$. On the other hand $\langle v, Aw\rangle = v^T \cdot (A \cdot w) \overset{A=A^T}{=} v^T \cdot A^T \cdot w = (A \cdot v)^T \cdot w = \langle Av, w\rangle = \langle \lambda v, w\rangle = \lambda\langle v, w\rangle$.

Thus, $(\lambda - \mu)\langle v, w\rangle = 0$, but $\lambda \neq \mu$, so $\langle v, w\rangle = 0$. □

We need the orthogonality between eigenvectors of the same eigenspaece: we use the Gram-Schmidt process:

Step 3 Use the Gram-Schmidt process on each basis $\mathcal{B}_{\lambda_i}$, to get an orthonormal basis $\mathcal{B}^{on}_{\lambda_i}$ of the eigenspace $V_{\lambda_i}$.

`Output` $\mathcal{B}^{on} = \mathcal{B}^{on}_{\lambda_1} \cup \ldots \cup \mathcal{B}^{on}_{\lambda_r}$ is an orthonormal basis of $\mathbb{R}^n$ made of eigenvectors of $A$. By taking $P$ having for columns the vectors of $\mathcal{B}^{on}$, we get that $P^{-1}AP = P^{T}AP$ is a diagonal matrix

**Example.** a) The matrix $A = \begin{pmatrix} 1 & 3 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3,3)$ is symmetric, so it is orthogonally diagonalizable. Let us determine an orthogonal basis of $\mathbb{R}^3$ made eigenvectors for $A$. The characteristic polynomial of $A$ is

$$p_A(t) = \det \begin{pmatrix} t-1 & -3 & 0 \\ -3 & t-1 & 0 \\ 0 & 0 & t+2 \end{pmatrix} = (t+2)((t-1)^2-9) = (t+2)(t^2-2t-8) = (t+2)^2(t-4)$$

Let us now determine a basis for both eigenspaces:

$$V_4 = \ker \begin{pmatrix} 3 & -3 & 0 \\ -3 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \text{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right)$$

$$V_{-2} = \ker \begin{pmatrix} -3 & -3 & 0 \\ -3 & -3 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \ker \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span}\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$$

Since $\left(\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right)$ are orthogonal, to get an orthonormal basis we just have to normalize the 3 vectors:

$$\left\{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$$

is an orthonormal basis of $\mathbb{R}^3$ made of eigenvectors for $A$ and

$$P^T A P = \begin{pmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix} \qquad \text{where} \qquad P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

b) The matrix $A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(3,3)$ is symmetric, so it is orthogonally diagonalizable. Its characteristic polynomial is

$$p_A(t) = \det \begin{pmatrix} t-2 & 1 & 1 \\ 1 & t-2 & 1 \\ 1 & 1 & t-2 \end{pmatrix} = (t+2)((t-2)^2-1)-(t-2-1)+(1-(t-2)) = t^3-6t^2+9t = t(t-3)^2$$

Let us now determine a basis for both eigenspaces:

$$V_0 = \ker \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} = \ker \begin{pmatrix} 1 & 1 & -2 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{pmatrix} = \ker \begin{pmatrix} 1 & 1 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$$

$$V_3 = \ker \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \ker \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span}\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right)$$

Since $\langle \left(\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right) \rangle = 1$, this two vectors are not orthogonal, so we apply the Gram-Schmidt process to get an orthogonal basis of $V_3$:

$$c_1 := \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad c_2 := \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \frac{\langle (1,0,-1)^T, (1,-1,0)^T \rangle}{||(1,-1,0)^T||^2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix}$$

After normalizing, we get an orthonormal basis of $\mathbb{R}^3$ made eigenvectors for $A$:

$$\left\{ \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix} \right\}$$

and moreover it holds

$$P^T A P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \qquad \text{where} \qquad P = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}$$

# Chapter 6

# Linear analytic geometry

In the previous chapters we have focused on vector spaces and vector subspaces. In particular we have seen that any vector subspace of $\mathbb{R}^n$ can be written as $\mathrm{Sol}(A|0)$ for some $A \in \mathcal{M}_{\mathbb{R}}(m,n)$.

In this final chapter we focus on the 3-dimensional space $\mathbb{R}^3$ ($n = 3$), but we consider a more general setting, namely solution sets of solvable linear systems $\mathrm{Sol}(A|b)$, where $A \in \mathcal{M}_{\mathbb{R}}(m,3)$: the *affine subspaces*. We will use the euclidean structure of $\mathbb{R}^3$ to discuss orthogonality, reciprocal positions and distances between affine subsets of $\mathbb{R}^3$.

## 6.1 Cross product

We begin by defining another type of product: the *cross product* (or *vector product*). This is defined only on $\mathbb{R}^3$, and it will be useful in the next sections.

**Definition 6.1.** Let $v, w \in \mathbb{R}^3$. Their *cross product* (or *vector product*) is the vector

$$v \times w = ||v|| \cdot ||w|| \cdot \sin(\theta) \cdot n$$

where $\theta$ is the angle[1] between $v$ and $w$, and $n$ is the unit vector orthogonal to $v$ and $w$ and determined by the right-hand rule ($v$: thumb, $w$: index finger, $n$: middle finger).

The cartesian coordinates of $\mathbb{R}^3$ defined in the Introduction are naturally identified with the canonical basis of $\mathbb{R}^3$: $\overrightarrow{i} = e_1$, $\overrightarrow{j} = e_2$ and $\overrightarrow{k} = e_3$.

**Example.** $\overrightarrow{i} \times \overrightarrow{j} = \overrightarrow{k}$; $\quad \overrightarrow{j} \times \overrightarrow{k} = \overrightarrow{i}$; $\quad \overrightarrow{i} \times \overrightarrow{k} = -\overrightarrow{j}$.

There is an explicit formula to compute the cross product of two vectors as "mixed determinant" (i.e. the matrix contains both vectors and scalars).
Let $v = (v_1, v_2, v_3)^T, w = (w_1, w_2, w_3)^T \in \mathbb{R}^3$ then

$$v \times w = \det \begin{pmatrix} \overrightarrow{i} & v_1 & w_1 \\ \overrightarrow{j} & v_2 & w_2 \\ \overrightarrow{k} & v_3 & w_3 \end{pmatrix} = \left( \det \begin{pmatrix} v_2 & w_2 \\ v_3 & w_3 \end{pmatrix}, -\det \begin{pmatrix} v_1 & w_1 \\ v_3 & w_3 \end{pmatrix}, \det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix} \right)^T \tag{6.1}$$

where the first "determinant" is expanded *only* along the first column.

[1] By definition of angle between vectors $\theta \in [0, \pi]$, so $\sin\theta \geq 0$.

**Example.**

$$\begin{pmatrix}1\\1\\0\end{pmatrix} \times \begin{pmatrix}0\\2\\2\end{pmatrix} = \det\begin{pmatrix}\overrightarrow{i} & 1 & 0\\ \overrightarrow{j} & 1 & 2\\ \overrightarrow{k} & 0 & 2\end{pmatrix} = \begin{pmatrix}2-0\\-(2-0)\\2-0\end{pmatrix} = \begin{pmatrix}2\\-2\\2\end{pmatrix}$$

$$\begin{pmatrix}1\\2\\3\end{pmatrix} \times \begin{pmatrix}4\\5\\6\end{pmatrix} = \det\begin{pmatrix}\overrightarrow{i} & 1 & 4\\ \overrightarrow{j} & 2 & 5\\ \overrightarrow{k} & 3 & 6\end{pmatrix} = \begin{pmatrix}12-15\\-(6-12)\\5-8\end{pmatrix} = \begin{pmatrix}-3\\6\\-3\end{pmatrix}$$

**Properties.** *Let $v, w \in \mathbb{R}^3$, then*

1. *$v \times w$ is orthogonal to both $v$ and $w$, so to the vector subspace generated by them.*
2. *$||v \times w|| = ||v|| \cdot ||w|| \cdot \sin(\theta)$ is the area of the parallelogram of vertices $0$, $v$, $w$, $v + w$.*
3. *$v \times w = 0$ if and only if $v, w$ are linearly dependent.*
4. *The map $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is bilinear (= linear in both entries), and $v \times w = -w \times v$.*

*Remark.* i) The cross product is not associative; for example

$$e_1 \times (e_2 \times e_2) = e_1 \times 0 = 0\,, \text{ but } \quad (e_1 \times e_2) \times e_2 = e_3 \times e_2 = -e_1\,.$$

ii) Let $u = (u_1, u_2, u_3)^T, v = (v_1, v_2, v_3)^T, w = (w_1, w_2, w_3)^T \in \mathbb{R}^3$.
From the definition of scalar and cross product it follows:

$$\langle u, v \times w\rangle = \det\begin{pmatrix}u_1 & v_1 & w_1\\ u_2 & v_2 & w_2\\ u_3 & v_3 & w_3\end{pmatrix}\,.$$

## 6.2 Affine subspaces of $\mathbb{R}^3$

**Definition 6.2.** Let $Ax = b$ be a solvable linear system, where $A \in \mathcal{M}_{\mathbb{R}}(m, n)$ and $b \in \mathbb{R}^m$. The solution set $S := \mathrm{Sol}(A|b)$ is an *affine subspace* of $\mathbb{R}^n$.

The corresponding vector subspace $S_0 := \mathrm{Sol}(A|0)$ is called the *direction* of $S$.

*Remark.* Let $Ax = b$ be a solvable linear system, and let $P \in \mathrm{Sol}(A|b)$ be a solution. Recall that by Theorem 1.13, it holds $\mathrm{Sol}(A|b) = P + \mathrm{Sol}(A|0)$; in particular, if $P_1, P_2 \in \mathrm{Sol}(A|b)$ then $P_1 - P_2 \in \mathrm{Sol}(A|0)$.

From now on we restrict to affine subspaces of $\mathbb{R}^3$, namely we consider $A \in \mathcal{M}_{\mathbb{R}}(m, 3)$ and $b \in \mathbb{R}^m$, defining a solvable linear system $Ax = b$, in particular $\mathrm{rk}(A|b) = \mathrm{rk}(A) \leq 3$.

Note that if $m > \mathrm{rk}(A)$, then we can apply the the Gauss algorithm to reduce $(A|b)$ into echelon form, and discard the $m - \mathrm{rk}(A)$ equations corresponding to the zero rows. Thus, we can always assume $m = \mathrm{rk}(A) = \mathrm{rk}(A|b) \leq 3$.

Case $m = 0$. $(A|b)$ is the zero matrix, so $S = \mathbb{R}^3$; this case is not of much interest and we will not discuss it further.

Case $m = 1$. $Ax = b$ consists of a single equation:

$$S = \{(x, y, z)^T \in \mathbb{R}^3 \mid \alpha x + \beta y + \gamma z = \delta\},$$

and $S_0$ is a vector subspace of dimension $3 - 1 = 2$: $S$ is a *plane*.

Case $m = 2$. $Ax = b$ consists of two (independent) equations:

$$S = \{(x, y, z)^T \in \mathbb{R}^3 \mid \begin{cases} \alpha_1 x + \beta_1 y + \gamma_1 z = \delta_1 \\ \alpha_2 x + \beta_2 y + \gamma_2 z = \delta_2 \end{cases}\},$$

and $S_0$ is a vector subspace of dimension $3 - 2 = 1$: $S$ is a *line*.

Case $m = 3$. $Ax = b$ consists of three (independent) equations, so $S_0$ is a vector subspace of dimension $3 - 3 = 0$: $S_0 = \{0_{\mathbb{R}^3}\}$: $S$ is a *point*.

### 6.2.1 Cartesian and parametric equations

In the previous section we defined an affine subspace as the solution set of a solvable linear system $Ax = b$, $A \in \mathcal{M}_{\mathbb{R}}(m, 3)$ and $b \in \mathbb{R}^m$; these are *cartesian equations*.

As for vector subspaces (see Section 2.5), by solving the linear system $Ax = b$, we can describe the affine subspace $S := \mathrm{Sol}(A|b)$ more explicitly through *parametric equations*: by Theorem 1.13 (see also Theorem 1.11 (Rouché-Capelli)) it holds

$$S = \mathrm{Sol}(A|b) = P + \mathrm{Sol}(A|0) = P + \mathrm{Span}(w_1, \dots, w_{3-m})$$

where $P = (x_P, y_P, z_P)^T \in \mathbb{R}^3$ is a solution of $Ax = b$ and $\{w_1, \dots, w_{3-m}\}$ is a basis of $\mathrm{Sol}(A|0)$. More explicitly:

if $m = 3$, $S$ is a point: $S = \left\{\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix}\right\}$;

if $m = 2$, $S$ is a line : $S = \left\{\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} + t\begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \mid t \in \mathbb{R}\right\} = \begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} + \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}\mathbb{R}$;

if $m = 1$, $S$ is a plane: $S = \left\{\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} + t_1\begin{pmatrix} w_{1,x} \\ w_{1,y} \\ w_{1,z} \end{pmatrix} + t_2\begin{pmatrix} w_{2,x} \\ w_{2,y} \\ w_{2,z} \end{pmatrix} \mid t_1, t_2 \in \mathbb{R}\right\} = \begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} + \begin{pmatrix} w_{1,x} \\ w_{1,y} \\ w_{1,z} \end{pmatrix}\mathbb{R} + \begin{pmatrix} w_{2,x} \\ w_{2,y} \\ w_{2,z} \end{pmatrix}\mathbb{R}$.

Given parametric equations $T = \{P + w_1 t_1 + \cdots + w_r t_r\} = P + \mathrm{Span}(w_1, \dots, w_r)$, we would like to recover cartesian equations.

Interpreting $W = \mathrm{Span}(w_1, \dots, w_r) \triangleleft \mathbb{R}^3$ as $W = \mathrm{Col}(M)$, (where $M \in \mathcal{M}_{\mathbb{R}}(3, r)$ has $w_1, \dots, w_r$ as columns), we see that a point $Q = (x, y, z)^T \in \mathbb{R}^3$ belongs to $T$ if and only if $P - Q \in \mathrm{Span}(w_1, \dots, w_r) = W$, if and only if $\mathrm{rk}(M) = \mathrm{rk}(M|P - Q)$. By reducing $(M|P - Q)$ into echelon form and imposing this condition we obtain cartesian equations of $T$, as in the following examples.

**Example.** • Let $T = \begin{pmatrix} 1 \\ -2 \\ -3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\mathbb{R} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\mathbb{R}$, then $(x, y, z)^T \in T$ if and only if the linear system

$$\left(\begin{array}{cc|c} 1 & 0 & x - 1 \\ 0 & 1 & y + 2 \\ 1 & 1 & z + 3 \end{array}\right)$$

is solvable. Using elementary row operations the linear system reduces to

$$\left(\begin{array}{cc|c} 1 & 0 & x-1 \\ 0 & 1 & y+2 \\ 0 & 0 & z+3-x+1-y-2 \end{array}\right)$$

which is solvable if and only if $(x,y,z)^T$ is a solution of $x+y-z=2$: this is a cartesian equations of $T$.

• Let $T = \left(\begin{smallmatrix} 0 \\ 4 \\ -3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R}$, then $(x,y,z)^T \in T$ if and only if the following linear system is solvable

$$\left(\begin{array}{c|c} 1 & x \\ 2 & y-4 \\ 3 & z+3 \end{array}\right) \longrightarrow \left(\begin{array}{c|c} 1 & x \\ 0 & y-4-2x \\ 0 & z+3-3x \end{array}\right) \qquad T = \left\{(x,y,z)^T \in \mathbb{R}^3 \mid \left\{\begin{array}{c} 2x-y=-4 \\ 3x-z=3 \end{array}\right.\right\}$$

**Line through two points & plane through three points**

A line is uniquely determined by 2 (distinct) points $P, Q$ on it. Indeed, $Q-P$ determines the direction of the line and the point $P$ (or $Q$) gives us the application point: parametric equations of the line passing through $P$ and $Q$ are: $P+(Q-P)t$, $t \in \mathbb{R}$.

**Example.** The line $l \subset \mathbb{R}^3$ passing through $P = \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)$ and $Q = \left(\begin{smallmatrix} 4 \\ 5 \\ 6 \end{smallmatrix}\right)$ is

$$l = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}\mathbb{R} \qquad \longrightarrow \qquad l = \left\{(x,y,z)^T \in \mathbb{R}^3 \mid \left\{\begin{array}{c} x-y=-1 \\ x-z=-2 \end{array}\right.\right\}$$

A plane is uniquely determined by 3 *non-collinear* points (i.e. not on the same line) $P, Q, R$ on it. Indeed, $Q-P$ and $R-P$ determine the direction of the plane and the point $P$ (or $Q$ or $R$) gives us the application point: parametric equations of the plane passing through $P$, $Q$ and $R$ are: $P+(Q-P)t_1+(R-P)t_2$, $t_1, t_2 \in \mathbb{R}$.

**Example.** The line $\pi \subset \mathbb{R}^3$ passing through $P = \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right)$, $Q = \left(\begin{smallmatrix} 1 \\ -2 \\ 1 \end{smallmatrix}\right)$ and $R = \left(\begin{smallmatrix} 2 \\ -1 \\ 3 \end{smallmatrix}\right)$ is

$$\pi = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -3 \\ 1 \end{pmatrix}\mathbb{R} + \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}\mathbb{R} \qquad \longrightarrow \qquad \pi = \{(x,y,z)^T \in \mathbb{R}^3 \mid 7x-y-3z=6\}$$

**Example.** The points $P = \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)$, $Q = \left(\begin{smallmatrix} 4 \\ 5 \\ 6 \end{smallmatrix}\right)$ and $R = \left(\begin{smallmatrix} 7 \\ 8 \\ 9 \end{smallmatrix}\right)$ are collinear, indeed by considering $P+(Q-P)t_1+(R-P)t_2$, $t_1, t_2 \in \mathbb{R}$ we get

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}\mathbb{R} + \begin{pmatrix} 6 \\ 6 \\ 6 \end{pmatrix}\mathbb{R} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\mathbb{R}$$

## 6.3 Reciprocal position of affine subspaces

To simplify the notation, from now on we denote by $(x_P, y_P, z_P)^T$ the coordinates of the point $P \in \mathbb{R}^3$.

**Point - point**

Let $P_1 = \mathrm{Sol}(A_1|b_1)$ and $P_2 = \mathrm{Sol}(A_2|b_2)$ be two points, then they are either equal (the linear systems $A_1x = b_1$ and $A_2x = b_2$ are equivalent), or they are different.

**Point - line**

Let $P = (x_P, y_P, z_P)^T$ be a point and $l = \mathrm{Sol}(A|b)$ be a line, then either $P$ belongs to $l$ (i.e. $P$ is a solution of $Ax = b$) or not.

**Point - plane**

Let $P = (x_P, y_P, z_P)^T$ be a point and $\pi = \{(x, y, z)^T \in \mathbb{R}^3 \mid \alpha x + \beta y + \gamma z = \delta\}$ be a plane. Then either $P$ belongs to $\pi$ (i.e. $\alpha x_P + \beta y_P + \gamma z_P = \delta$) or not.

**Line - line**

Let $l_1 = \mathrm{Sol}(A_1|b_1)$ and $l_2 = \mathrm{Sol}(A_2|b_2)$ be two lines. The first thing to check is whether they meet or not, so we need to understand if the linear system

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

has solutions or not. The matrix $A := \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \in \mathcal{M}_{\mathbb{R}}(4,3)$ has rank $\mathrm{rk}(A) = 2$ or $3$, while the augmented matrix $(A|b) = \left(\begin{array}{c|c} A_1 & b_1 \\ A_2 & b_2 \end{array}\right) \in \mathcal{M}_{\mathbb{R}}(4,4)$ has either $\mathrm{rk}(A|b) = \mathrm{rk}(A)$ or $\mathrm{rk}(A|b) = \mathrm{rk}(A) + 1$; there are 4 possibilities for the reciprocal position of $l_1$ and $l_2$.

Case 1: $\mathrm{rk}(A) = \mathrm{rk}(A|b) = 2$: this means that the linear systems $A_1x = b_1$ and $A_2x = b_2$ are equivalent, so $l_1 = l_2$: they are *coincident lines.*

Case 2: $\mathrm{rk}(A) = 2$ and $\mathrm{rk}(A|b) = 3$: by the Rouché-Capelli Theorem (1.11) the two lines do not have common points, but same direction ($\dim \mathrm{Sol}(A|0) = 1 = \dim \mathrm{Sol}(A_1|0) = \dim \mathrm{Sol}(A_2|0)$ implies $\mathrm{Sol}(A_1|0) = \mathrm{Sol}(A_2|0)$) so $l_1$ and $l_2$ are *parallel lines.*

More generally we have:

**Definition 6.3.** Let $S_1 = \mathrm{Sol}(A_1|b_1)$ and $S_2 = \mathrm{Sol}(A_2|b_2)$ be two affine subspace, with corresponding directions $S_{1,0} = \mathrm{Sol}(A_1|0)$ and $S_{2,0} = \mathrm{Sol}(A_2|0)$.

The affine subspaces $S_1$ and $S_2$ are *parallel* if: i) $S_1$ and $S_2$ have no common points, and ii) $S_{1,0} \subset S_{2,0}$ or $S_{2,0} \subset S_{1,0}$.

**Example.** i) The lines $\left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\mathbb{R}$ and $\left(\begin{smallmatrix} 4 \\ -1 \\ 7 \end{smallmatrix}\right) + \left(\begin{smallmatrix} -2 \\ -2 \\ -2 \end{smallmatrix}\right)\mathbb{R}$ are parallel.
ii) The line $\left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\mathbb{R}$ and the plane $\left(\begin{smallmatrix} 4 \\ -1 \\ 7 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 2 \\ 1 \\ 0 \end{smallmatrix}\right)\mathbb{R} + \left(\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right)\mathbb{R}$ are parallel.

Case 3: $\mathrm{rk}(A) = 3 = \mathrm{rk}(A|b)$: by the Rouché-Capelli Theorem (1.11) the two lines have a single common point: $l_1$ and $l_2$ are *incident lines*.

Case 4: $\mathrm{rk}(A) = 3$ and $\mathrm{rk}(A|b) = 4$: by the Rouché-Capelli Theorem (1.11) the two lines do not have common points and different directions $\mathrm{Sol}(A_1|0) \neq \mathrm{Sol}(A_2|0)$: $l_1$ and $l_2$ are *skew lines*.

**Example.** • $\begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $\begin{cases} -2x + y + z = 3 \\ y - z = -1 \end{cases}$ are coincident lines, indeed

$$\mathrm{rk}\left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 1 & 0 & -1 & -2 \\ -2 & 1 & 1 & 3 \\ 0 & 1 & -1 & -1 \end{array}\right) = \mathrm{rk}\left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & -1 & 1 & 1 \\ 0 & 1 & -1 & -1 \end{array}\right) = 2$$

• $\begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $\begin{cases} -2x + y + z = 4 \\ y - z = -2 \end{cases}$ are parallel lines, indeed

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 1 & 0 & -1 & -2 \\ -2 & 1 & 1 & 4 \\ 0 & 1 & -1 & -2 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & -1 & 1 & 2 \\ 0 & 1 & -1 & -2 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

• $l_1 = \begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $l_2 = \left(\begin{smallmatrix} 0 \\ -1 \\ -2 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R}$ are incident lines, indeed by plugging in the equations of $l_2$ into those of $l_1$ we get

$$\begin{cases} (0 + t) - (-1 + 2t) = -1 \\ (0 + t) - (-2 + 3t) = -2 \end{cases} \leftrightarrow \begin{cases} -t = -2 \\ -2t = -4 \end{cases} \leftrightarrow t = 2$$

a single solution. This means that the lines $l_1$ and $l_2$ have a single common point $(2, 3, 4)^T$ (i.e. they are incident lines).

• $l_1 = \begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $l_2 = \left(\begin{smallmatrix} 0 \\ 4 \\ -3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R}$ are skew lines, indeed by plugging in the equations of $l_2$ into those of $l_1$ we get

$$\begin{cases} (0 + t) - (4 + 2t) = -1 \\ (0 + t) - (-3 + 3t) = -2 \end{cases} \leftrightarrow \begin{cases} -t = -5 \\ -2t = 1 \end{cases}$$

which has no solutions, i.e. $l_1$ and $l_2$ have no common points. Moreover, their directions are different, indeed $l_{1,0} = \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\mathbb{R}$ does not contain $\left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)$.

*Remark.* Having parametric equations of the lines $l_1 = P_1 + v_1\mathbb{R}$ and $l_2 = P_2 + v_2\mathbb{R}$, we can read the four cases as follows:
Case 1, coincident lines: $\mathrm{rk}(v_1|v_2) = 1 = \mathrm{rk}(v_1|v_2|P_2 - P_1)$.
Case 2, parallel lines: $\mathrm{rk}(v_1|v_2)$, $\mathrm{rk}(v_1|v_2|P_2 - P_1) = 2$.
Case 3, incident lines: $\mathrm{rk}(v_1|v_2) = 2 = \mathrm{rk}(v_1|v_2|P_2 - P_1)$.
Case 4, skew lines: $\mathrm{rk}(v_1|v_2) = 2$ $\mathrm{rk}(v_1|v_2|P_2 - P_1) = 3$.

Indeed, $\text{rk}(v_1|v_2) = 1$ means that the lines have the same direction, while $\text{rk}(v_1|v_2) = 2$ means that the lines have different directions.

And if, by adding the column $P_2 - P_1$, the rank increases, it means that the vector $P_2 - P_1$ is independent from $\{v_1, v_2\}$, so that the translation of $v_1\mathbb{R}$ by $P_1$ and the translation of $v_2\mathbb{R}$ by $P_2$ separate the lines.

**Example.** • The lines $\left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right) + \left(\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right)\mathbb{R}$ and $\left(\begin{smallmatrix}4\\-1\\7\end{smallmatrix}\right) + \left(\begin{smallmatrix}-2\\-2\\-2\end{smallmatrix}\right)\mathbb{R}$ have the same direction, so they are either coincident or parallel. They are parellel since $\left(\begin{smallmatrix}4\\-1\\7\end{smallmatrix}\right) - \left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right) = \left(\begin{smallmatrix}3\\-3\\4\end{smallmatrix}\right)$ is not a multiple of $\left(\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right)$.

• The lines $\left(\begin{smallmatrix}1\\-1\\2\end{smallmatrix}\right) + \left(\begin{smallmatrix}1\\2\\1\end{smallmatrix}\right)\mathbb{R}$ and $\left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right) + \left(\begin{smallmatrix}2\\1\\1\end{smallmatrix}\right)\mathbb{R}$ have different directions, so they are either incident or skew. Since $\left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right) - \left(\begin{smallmatrix}1\\-1\\2\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\3\\1\end{smallmatrix}\right) \in \text{Span}\left(\left(\begin{smallmatrix}1\\2\\1\end{smallmatrix}\right), \left(\begin{smallmatrix}2\\1\\1\end{smallmatrix}\right)\right)$, the lines are incident.

**Line - plane**

Let $l = \text{Sol}(A_1|b_1)$ be a line and $\pi = \text{Sol}(A_2|b_2)$ be a plane.
The matrix $A := \left(\begin{smallmatrix}A_1\\A_2\end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{R}}(3,3)$ has rank $\text{rk}(A) = 2$ or 3, while the augmented matrix $(A|b) = \left(\begin{smallmatrix}A_1 & b_1\\A_2 & b_2\end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{R}}(3,4)$ has $\text{rk}(A) \le \text{rk}(A|b) \le 3$, so there are 3 cases.

Case 1: $\text{rk}(A) = \text{rk}(A|b) = 2$: this means that the line and the plane have infinitely many common points: the line is contained in $\pi$: $l \subset \pi$.

To be more precise this means that the cartesian equation defining the plane is a linear combination of the cartesian equation defining the line. In formula, let $\begin{cases}\alpha_1 x + \beta_1 y + \gamma_1 z = \delta_1\\ \alpha_2 x + \beta_2 y + \gamma_2 z = \delta_2\end{cases}$ be cartesian equations of the line, then there exists $\lambda, \mu \in \mathbb{R}$ (not both zero) such that $\pi$ is given by the equation:

$$\lambda(\alpha_1 x + \beta_1 y + \gamma_1 z - \delta_1) + \mu(\alpha_2 x + \beta_2 y + \gamma_2 z - \delta_2) = 0 \tag{6.2}$$

**Definition 6.4.** Letting $\lambda, \mu$ vary in $\mathbb{R}$ (not both zero) the equation (6.2) describes all planes containing the line $l = \begin{cases}\alpha_1 x + \beta_1 y + \gamma_1 z = \delta_1\\ \alpha_2 x + \beta_2 y + \gamma_2 z = \delta_2\end{cases}$. This is called the *sheaf of planes through $l$*.

Case 2: $\text{rk}(A) = 2$ and $\text{rk}(A|b) = 3$: this means that the line and the plane have no common points, and that the direction of the line $\text{Sol}(A_1|0)$ is contained in the direction of the plane $\text{Sol}(A_2|0)$, so the line and the plane are *parallel*.

Case 3: $\text{rk}(A) = 3 = \text{rk}(A|b)$: the line and the plane have a single common point, so the line and the plane are *incident*.

*Remark.* Using the same reasoning above, having parametric equations of the line $l = P_1 + v_1\mathbb{R}$ and of the plane $\pi = P_2 + v_2\mathbb{R} + v_3\mathbb{R}$, we can read the three cases as follows:
Case 1, $l \subset \pi$: $\text{rk}(v_1|v_2|v_3) = 2 = \text{rk}(v_1|v_2|v_3|P_2 - P_1)$.
Case 2, line and plane are parallel: $\text{rk}(v_1|v_2|v_3) = 2$ and $\text{rk}(v_1|v_2|v_3|P_2 - P_1) = 3$.
Case 3, line and plane are incident: $\text{rk}(v_1|v_2|v_3) = 3$.

**Example.** • The line $l = \begin{cases} x-y=-1 \\ x-z=-2 \end{cases}$ is contained in the plane $\pi = \{x+y-2z=-3\}$, indeed

$$\operatorname{rk}\left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 1 & 0 & -1 & -2 \\ 1 & 1 & -2 & -3 \end{array}\right) = \operatorname{rk}\left(\begin{array}{ccc|c} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 2 & -2 & -2 \end{array}\right) = 2$$

• The line $l = \begin{cases} x-y=-1 \\ x-z=-2 \end{cases}$ and the plane $\pi = \left(\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R} + \left(\begin{smallmatrix} 1 \\ 1 \\ -3 \end{smallmatrix}\right)\mathbb{R}$ are incident, indeed by plugging in the equations of $\pi$ into those of $l$ we get a single solution (for $s, t$)

$$\begin{cases} (1+s+t)-(0+2s+t)=-1 \\ (1+s+t)-(1+3s-3t)=-2 \end{cases} \leftrightarrow \begin{cases} s=2 \\ -2s+4t=-2 \end{cases} \leftrightarrow \begin{cases} s=2 \\ t=\frac{1}{2} \end{cases}$$

• The line $= \left(\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R}$ and the plane $\pi = x+y-z=10$ parallel, indeed by plugging in the equations of $l$ into that of $\pi$ we get an impossible equation: $10 = (1+t)+(0+2t)-(1+3t) = 2$.

### Plane - plane

Let $\pi_1 = \operatorname{Sol}(A_1|b_1)$ and $\pi_2 = \operatorname{Sol}(A_2|b_2)$ be planes.
The matrix $A := \left(\begin{smallmatrix} A_1 \\ A_2 \end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{R}}(2,3)$ has rank $\operatorname{rk}(A) = 1$ or $2$, while the augmented matrix $(A|b) = \left(\begin{smallmatrix} A_1 & | & b_1 \\ A_2 & | & b_2 \end{smallmatrix}\right) \in \mathcal{M}_{\mathbb{R}}(2,4)$ has $\operatorname{rk}(A) \leq \operatorname{rk}(A|b) \leq 2$, so there are 3 cases.

Case 1: $\operatorname{rk}(A) = \operatorname{rk}(A|b) = 1$: this means that the linear systems $A_1x = b_1$ and $A_2x = b_2$ are equivalent, so $\pi_1 = \pi_2$: they are *coincident planes.*

Case 2: $\operatorname{rk}(A) = 1$ and $\operatorname{rk}(A|b) = 2$: by the Rouché-Capelli Theorem (1.11) the planes do not have common points, but they have same directions ($\dim \operatorname{Sol}(A|0) = 2 = \dim \operatorname{Sol}(A_1|0) = \dim \operatorname{Sol}(A_2|0)$ implies $\operatorname{Sol}(A_1|0) = \operatorname{Sol}(A_2|0)$) so $\pi_1$ and $\pi_2$ are *parallel planes.*

To be more precise, both linear systems defining the planes consists of a single equation, and the fact that $\operatorname{rk}(A) = 1$, means that the equations are of the form:

$$\pi_1 : \alpha x + \beta y + \gamma z = \delta_1 \qquad \pi_2 : \alpha x + \beta y + \gamma z = \delta_2$$

So if $\delta_1 = \delta_2$ the planes coincides, otherwise they are parallel.

**Definition 6.5.** Let $\pi$ be the plane of equation

$$\alpha x + \beta y + \gamma z = \delta.$$

Letting $\delta$ vary in $\mathbb{R}$ we describe all planes parallel to $\pi$. This is a *parallel sheaf of planes.*

Case 3: $\operatorname{rk}(A) = 2 = \operatorname{rk}(A|b)$: by the Rouché-Capelli Theorem (1.11) the have a common line: $\pi_1$ and $\pi_2$ are *incident planes.*

*Remark.* As above, having parametric equations of the planes $\pi_1 = P_1 + v_1\mathbb{R} + v_2\mathbb{R}$ and of the plane $\pi_2 = P_2 + w_1\mathbb{R} + w_2\mathbb{R}$, we can read the three cases as follows:
Case 1, coincident planes: $\operatorname{rk}(v_1|v_2|w_1|w_2) = 2 = \operatorname{rk}(v_1|v_2|w_1|w_2|P_2 - P_1)$.
Case 2, parallel planes: $\operatorname{rk}(v_1|v_2|w_1|w_2) = 2$ and $\operatorname{rk}(v_1|v_2|w_1|w_2|P_2 - P_1) = 3$.
Case 3, incident planes: $\operatorname{rk}(v_1|v_2|w_1|w_2) = 3$.

### 6.3.1 Orthogonality

**Definition 6.6.** Two affine subspaces $S_1 = \mathrm{Sol}(A_1|b_1)$ and $S_2 = \mathrm{Sol}(A_2|b_2)$ are *orthogonal* if the corresponding directions $S_{1,0} = \mathrm{Sol}(A_1|0)$ and $S_{2,0} = \mathrm{Sol}(A_2|0)$ are orthogonal vector subspaces of the euclidean space $\mathbb{R}^3$ (see Definition 5.5).

**Example.** • The lines $l_1 = \left(\begin{smallmatrix}10\\11\\12\end{smallmatrix}\right) + \left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right)\mathbb{R}$ and $l_2 = \left(\begin{smallmatrix}-7\\9\\-11\end{smallmatrix}\right) + \left(\begin{smallmatrix}1\\1\\-1\end{smallmatrix}\right)\mathbb{R}$ are orthogonal, indeed $\langle (1,2,3)^T, (1,1,-1)^T \rangle = 0$.
• Let $l$ be the line of equation $l = \begin{cases} x + y = 10 \\ 2y + 2z = -17 \end{cases}$, then any plane of parametric equation $P + \left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right)\mathbb{R} + \left(\begin{smallmatrix}0\\2\\2\end{smallmatrix}\right)\mathbb{R}$ $(P \in \mathbb{R}^3)$ is orthogonal to $l$.
• Let $\pi$ be the plane of equation $\pi = x - 2y + 5z = 20$, then any line of parametric equation $P + \left(\begin{smallmatrix}1\\-2\\5\end{smallmatrix}\right)\mathbb{R}$ $(P \in \mathbb{R}^3)$ is orthogonal to $\pi$.

*Remark.* Let $\pi := \{\alpha x + \beta y + \gamma z = \delta\} \subset \mathbb{R}^3$ be a plane, the vector $(\alpha, \beta, \gamma)^T$ is orthogonal to $\pi$ and it is a *normal vector* of the plane. It is unique, up to scaling.

## 6.4 Distances

**Definition 6.7.** Let $P = (x_P, y_P, z_P)^T, Q = (x_Q, y_Q, z_Q)^T \in \mathbb{R}^3$ be two points. Their *distance* is

$$d(P,Q) = ||P - Q|| = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2 + (z_P - z_Q)^2} \tag{6.3}$$

Let $S_1, S_2 \subset \mathbb{R}^3$ be affine subspaces, then $d(S_1, S_2) := \min\{d(P,Q) \mid P \in S_1, Q \in S_2\}$.

For example, the distance between a point $P$ and a plane $\pi$ is $d(P,\pi) := \min\{d(P,Q) \mid Q \in \pi\} = d(P, \hat{P})$, where $\hat{P}$ is the orthogonal projection of $P$ onto $\pi$, i.e. it is the intersection point between $\pi$ and the line $l$ orthogonal to $\pi$ and passing through $P$.

#### Point - plane

Let $P = (x_P, y_P, z_P)^T$ be a point and $\pi = \{(x,y,z)^T \in \mathbb{R}^3 \mid \alpha x + \beta y + \gamma z = \delta\}$ be a plane. We want to compute $d(P,\pi) = d(P,\hat{P})$, where $\hat{P}$ is as above.

Being orthogonal to $\pi$ and passing through $P$, the line $l$ has parametric equation $l = P + \left(\begin{smallmatrix}\alpha\\\beta\\\gamma\end{smallmatrix}\right)\mathbb{R}$, so $\hat{P} = P + \hat{t}\left(\begin{smallmatrix}\alpha\\\beta\\\gamma\end{smallmatrix}\right)$ for a $\hat{t} \in \mathbb{R}$, which is determined by imposing $\hat{P} \in \pi$:

$$\delta = \alpha(x_P + \hat{t}\alpha) + \beta(y_P + \hat{t}\beta) + \gamma(z_P + \hat{t}\gamma) = (\alpha x_P + \beta y_P + \gamma z_P) + \hat{t}(\alpha^2 + \beta^2 + \gamma^2)$$

Thus, $\hat{t} = \dfrac{\delta - (\alpha x_P + \beta y_P + \gamma z_P)}{(\alpha^2 + \beta^2 + \gamma^2)}$ and $d(P,\hat{P}) = ||P + \hat{t}(\alpha,\beta,\gamma)^T - P|| = |\hat{t}| \cdot ||(\alpha,\beta,\gamma)^T||$:

$$d(P,\pi) = \frac{|\alpha x_P + \beta y_P + \gamma z_P - \delta|}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}}$$

**Point - line**

Let $P = (x_P, y_P, z_P)^T$ be a point and $l \subset \mathbb{R}^3$ be a line. The distance between them is $d(P, l) := \min\{d(P, Q) \mid Q \in l\} = d(P, \hat{Q})$, where $\hat{Q}$ is the orthogonal projection of $P$ onto $l$, and one can determine the distance $d(P, \hat{Q})$ arguing as above, but we prefer to use a different approach to get a simple formula.

Let $l = Q + v\mathbb{R}$ be a parametric equation of $l$ and define $w := Q - P$, so that $P, Q, \hat{Q}$ are the vertices of a right triangle (with the right angle in $\hat{Q}$) and hypotenuse $v$. Let $\theta$ be the angle between $v$ and $w$ (the one in $Q$)[2], then $d(P, \hat{Q}) = ||w|| \cdot \sin\theta = \frac{||w|| \cdot ||v|| \cdot \sin\theta}{||v||}$:

$$d(P, l) = \frac{||v \times (Q - P)||}{||v||}$$

**Example.** Let $P = (1, 2, 3)^T \in \mathbb{R}^3$, let $\pi \subset \mathbb{R}^3$ be the plane of equation $x - 3y + z = 10$, and let $l \subset \mathbb{R}^3$ be the line of equations $l = \begin{cases} x - 3y + z = 10 \\ x + y = 10 \end{cases}$. The distance $d(P, \pi)$ is

$$d(P, \pi) = \frac{|1 \cdot 1 + (-3) \cdot 2 + 1 \cdot 3 - 10|}{\sqrt{1^2 + (-3)^2 + 1^2}} = \frac{12}{\sqrt{11}}$$

while the distance $d(P, l)$ is

$$l = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \mathbb{R} \quad \Longrightarrow \quad d(P, l) = \frac{||(1, 1, 2)^T \times (1, 3, 2)^T||}{||(1, 1, 2)^T||} = \frac{\sqrt{20}}{\sqrt{6}}.$$

**Line - line**

Let $l_1, l_2 \subset \mathbb{R}^3$ be lines. If they meet ($l_1 \cap l_2 \neq \emptyset$), then $d(l_1, l_2) = 0$. If they do not meet, they are either parallel or skew.

Assume that $l_1$ and $l_2$ are parallel: $l_1 = P_1 + v\mathbb{R}$, $l_2 = P_2 + v\mathbb{R}$. Then $d(l_1, l_2) = d(P_1, l_2)$:

$$d(l_1, l_2) = \frac{||(P_2 - P_1) \times v||}{||v||}$$

Assume that $l_1$ and $l_2$ are skew: $l_1 = P_1 + v_1\mathbb{R}$, $l_2 = P_2 + v_2\mathbb{R}$. Let $\pi$ be the unique plane parallel to $l_1$ and containing $l_2$: $\pi = P_2 + v_2\mathbb{R} + v_1\mathbb{R}$, so that $d(l_1, l_2) = d(P_1, \pi)$.

To apply the previous formula, we need a cartesian equation of $\pi$. A normal vector is $v_1 \times v_2 = \left(\begin{smallmatrix} \alpha \\ \beta \\ \gamma \end{smallmatrix}\right)$, so that $\pi = \alpha x + \beta y + \gamma z = \delta$, where $\delta$ is determined by the fact that $P_2 = (x_{P_2}, y_{P_2}, z_{P_2})^T \in \pi$: $\delta = \alpha x_{P_2} + \beta y_{P_2} + \gamma z_{P_2} = \langle v_1 \times v_2, P_2 \rangle$, so

$$d(P_1, \pi) = \frac{|\alpha P_{1,x} + \beta P_{1,y} + \gamma P_{1,z} - \delta|}{\sqrt{(\alpha^2 + \beta^2 + \gamma^2)}} = \frac{|\langle v_1 \times v_2, P_1 \rangle - \langle v_1 \times v_2, P_2 \rangle|}{||v_1 \times v_2||}$$

We get

$$d(l_1, l_2) = \frac{|\langle v_1 \times v_2, P_1 - P_2 \rangle|}{||v_1 \times v_2||}$$

[2] As remarked on page 67, $\sin\theta \geq 0$.

**Example.** The lines $l_1 = \begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $l_2 = \begin{cases} -2x + y + z = 4 \\ y - z = -2 \end{cases}$ are parallel lines: $l_1 = \left(\begin{smallmatrix} 0 \\ 1 \\ 2 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\mathbb{R}$, $l_2 = \left(\begin{smallmatrix} -1 \\ 0 \\ 2 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\mathbb{R}$, and their distance is

$$d(l_1, l_2) = \frac{||(1,1,0)^T \times (1,1,1)^T||}{||(1,1,1)^T||} = \frac{\sqrt{2}}{\sqrt{3}}$$

The lines $l_1 = \begin{cases} x - y = -1 \\ x - z = -2 \end{cases}$ and $l_2 = \left(\begin{smallmatrix} 0 \\ 4 \\ -3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)\mathbb{R}$ are skew lines, and their distance is

$$d(l_1, l_2) = \frac{|\langle (1,1,1)^T \times (1,2,3)^T, (0,-3,5)^T \rangle|}{||(1,1,1)^T \times (1,2,3)^T||} = \frac{|\langle (1,-2,1)^T, (0,-3,5)^T \rangle|}{||(1,-2,1)^T||} = \frac{11}{\sqrt{6}}$$

**Line - plane**

Let $l \subset \mathbb{R}^3$ be a line, let $\pi \subset \mathbb{R}^3$ be a plane, and let $P_1 \in l$, then:

$$d(l, \pi) = \begin{cases} 0 & \text{if } l \cap \pi \neq \emptyset \\ d(P_1, \pi) & \text{if } l \text{ and } \pi \text{ are parallel} \end{cases}$$

**Plane - plane**

Let $\pi_1, \pi_2 \subset \mathbb{R}^3$ be planes, and let $P_1 \in \pi_1$, then:

$$d(\pi_1, \pi_2) = \begin{cases} 0 & \text{if } \pi_1 \cap \pi_2 \neq \emptyset \\ d(P_1, \pi_2) & \text{if } \pi_1 \text{ and } \pi_2 \text{ are parallel} \end{cases}$$