Linear algebra

S. Richard

Fall Semester 2014 and Spring Semester 2015

Contents

Introduction

These lecture notes correspond to a first course in linear algebra, which does not rely on any prerequisite. All necessary notions are introduced from scratch, and the proofs of most of the statements are provided. Examples are provided in the text while exercises are gathered at the end of each chapter. The main source of inspiration has been the book of S. Lang: *Introduction to linear algebra*[1].

0.1 Motivation

Several problems lead naturally to the basic concepts of linear algebra. We list some examples which are at the root of this course or which provide some motivation for this course.

Solving linear equations Consider the following linear system of equations

$$\begin{cases} 2x + y = 7 \\ -x + 2y = 4 \end{cases} .$$

Its only solution is $\begin{cases} x = 2 \\ y = 3 \end{cases}$. However, if one considers the system

$$\begin{cases} 2x + y = 7 \\ 4x + 2y = 14 \end{cases}$$

then one can find several solutions, as for example $\begin{cases} x = 0 \\ y = 7 \end{cases}$ or $\begin{cases} x = 2 \\ y = 3 \end{cases}$. It is then natural to wonder what are all solutions of this system ? How can one describe this set of solutions and how can one understand it ?

Solutions of linear differential equations Consider a real function f defined on $\mathbb{R}$, *i.e.* $f : \mathbb{R} \ni t \mapsto f(t) \in \mathbb{R}$ satisfying the relation

$$f''(t) = -m^2 f(t)$$

[1] Serge Lang, *Introduction to linear algebra*, second edition, Undergraduate texts in mathematics, Springer.

for some constant $m \in \mathbb{R}$. Can one find all solutions f for this equation ? For example, $f(t) = \cos(mt)$, $f(t) = 3\sin(mt)$ or $f(t) = 2\cos(mt + 3)$ are solutions of this equation, but how can one describe all of them ?

Describing linear transformations Consider the following linear transformation:

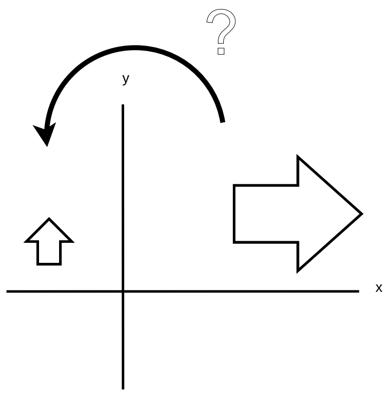


Figure 1: Linear transformation

How can one describe this transformation efficiently ?

Change of bases Consider the same object described in two different reference systems:

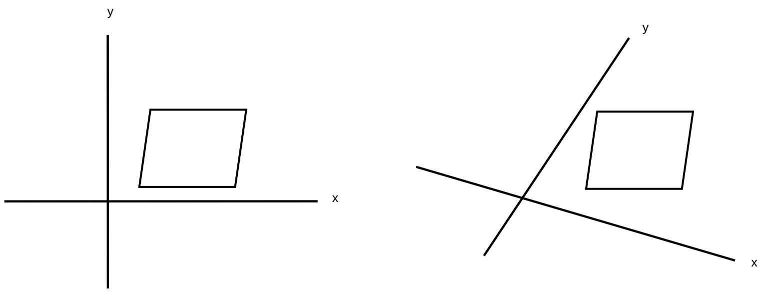


Figure 2: Change of reference system

How can one compare the information on the same object in these two systems ?

Reading a scientific paper How can one understand the following paper: "The \$25'000'000'000 eigenvector: the linear algebra behind Google"[2].

Other motivations Linear algebra is also at the root of quantum mechanics, dynamical systems, linear response theory, linear perturbations theory, ...

[2]Kurt Bryan, Tanya Leise, *The \$25'000'000'000 eigenvector: the linear algebra behind Google*, SIAM REVIEW, Vol. 48, No. 3, pp. 569–581.

Chapter 1

Geometric setting

In this Chapter we recall some basic notions on points or vectors in $\mathbb{R}^n$. The norm of a vector and the scalar product between two vectors are also introduced.

1.1 The Euclidean space $\mathbb{R}^n$

We set $\mathbb{N} := \{1, 2, 3, \dots\}$ for the set of *natural numbers*, also called *positive integers*, and let $\mathbb{R}$ be the set of all real numbers.

Definition 1.1.1. *One sets*

$$\mathbb{R}^n = \big\{(a_1, a_2, \dots, a_n) \mid a_j \in \mathbb{R} \text{ for all } j \in \{1, 2, \dots, n\}\big\}^1.$$

Alternatively, an element of $\mathbb{R}^n$*, also called* a n-tuple *or* a vector*, is a collection of* n *numbers* $(a_1, a_2, \dots, a_n)$ *with* $a_j \in \mathbb{R}$ *for any* $j \in \{1, 2, \dots, n\}$*. The number* n *is called* the dimension of $\mathbb{R}^n$*.*

In the sequel, we shall often write $A \in \mathbb{R}^n$ for the vector $A = (a_1, a_2, \dots, a_n)$. With this notation, the values $a_1, a_2, \dots, a_n$ are called *the components* or *the coordinates* of A. For example, a_1 is the first component of A, or the first coordinate of A. Be aware that $(1, 3)$ and $(3, 1)$ are two different elements of $\mathbb{R}^2$. Note that one often writes (x, y) for elements of $\mathbb{R}^2$ and (x, y, z) for elements of $\mathbb{R}^3$, see Figure 1.1. However this notation is not really convenient in higher dimensions.

The set $\mathbb{R}^n$ can be endowed with two operations, *the addition* and *the multiplication by a scalar.*

Definition 1.1.2. *For any* $A, B \in \mathbb{R}^n$ *with* $A = (a_1, a_2, \dots, a_n)$ *and* $B = (b_1, b_2, \dots, b_n)$ *and for any* $\lambda \in \mathbb{R}$ *one defines the addition of* A *and* B *by*

$$A + B := (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) \in \mathbb{R}^n$$

and the multiplication of A *by the scalar* λ *by*

$$\lambda A := (\lambda a_1, \lambda a_2, \dots, \lambda a_n) \in \mathbb{R}^n.$$

[1] The vertical line | has to be read "such that".

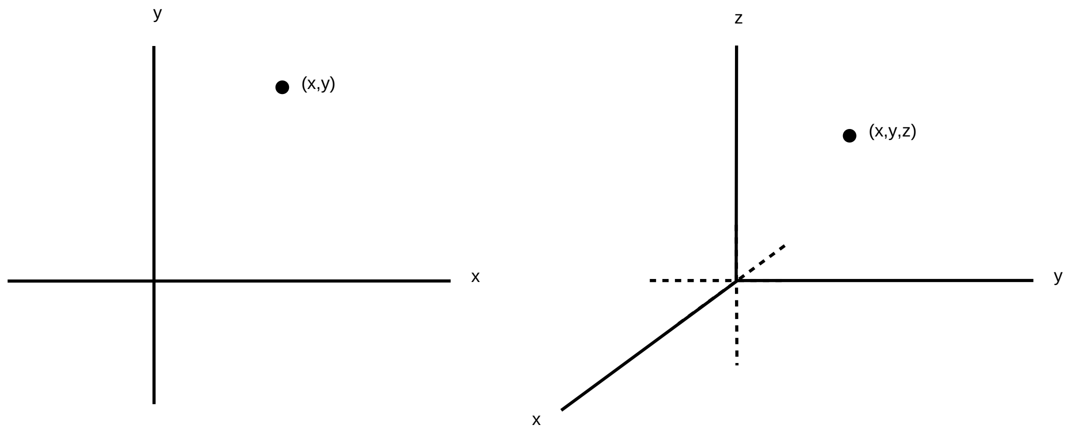


Figure 1.1: Elements of $\mathbb{R}^2$ and $\mathbb{R}^3$

Examples 1.1.3. *(i)* $(1,3)+(2,4)=(3,7)\in\mathbb{R}^2$,

(ii) $(1,2,3,4,5)+(5,4,3,2,1)=(6,6,6,6,6)\in\mathbb{R}^5$,

(iii) $3(1,2)=(3,6)\in\mathbb{R}^2$,

(iv) $\pi(0,0,1)=(0,0,\pi)\in\mathbb{R}^3$.

One usually sets
$$\mathbf{0}=(0,0,\dots,0)\in\mathbb{R}^n$$
and this element satisfies $A+\mathbf{0}=\mathbf{0}+A=A$ for any $A\in\mathbb{R}^n$. If $A=(a_1,a_2,\dots,a_n)$ one also writes $-A$ for the element $-1A=(-a_1,-a_2,\dots,-a_n)$. Then, by an abuse of notation, one writes $A-B$ for $A+(-B)$ if $A,B\in\mathbb{R}^n$, and obviously one has $A-A=\mathbf{0}$. Note that $A+B$ is defined if and only if A and B belong to $\mathbb{R}^n$, but has no meaning if $A\in\mathbb{R}^n$ and $B\in\mathbb{R}^m$ with $n\neq m$.

Properties 1.1.4. *If $A,B,C\in\mathbb{R}^n$ and $\lambda,\mu\in\mathbb{R}$ then one has*

(i) $A+B=B+A$, (commutativity)

(ii) $(A+B)+C=A+(B+C)$, (associativity)

(iii) $\lambda(A+B)=\lambda A+\lambda B$, (distributivity)

(iv) $(\lambda+\mu)A=\lambda A+\mu A$,

(v) $(\lambda\mu)A=\lambda(\mu A)$.

These properties will be proved in Exercise 1.3.

1.2 Located vectors in $\mathbb{R}^n$

A geometric picture can often aid our intuition (but can also be misleading). For example, one often identifies $\mathbb{R}$ with a line, $\mathbb{R}^2$ with a plane and $\mathbb{R}^3$ with the usual 3 dimensional space. In this setting, an element $A \in \mathbb{R}^n$ is often called *a point* in $\mathbb{R}^n$. However, one can also think about the elements of $\mathbb{R}^n$ as *arrows*. In this setting, the element $(3,5) \in \mathbb{R}^2$ can be thought as an arrow starting at the point $(0,0)$ of the usual plane with two axes and ending at the point $(3,5)$ of this plane, see Figure 1.2. With

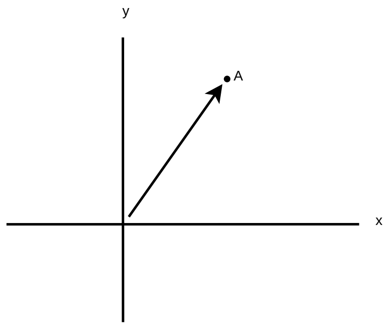


Figure 1.2: A point seen as an arrow

this interpretation in mind, the addition of two elements of $\mathbb{R}^n$ corresponds the addition of two arrows, and the multiplication by a scalar corresponds to the rescaling of an arrow, see Figure 1.3. Note that in the sequel both interpretations (points and arrows) will appear, but this should not lead to any confusion.

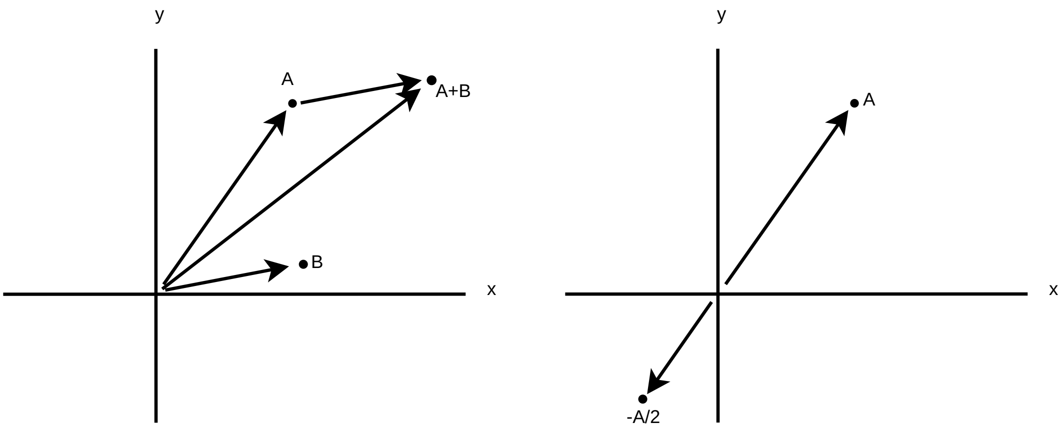


Figure 1.3: Addition of arrows and multiplication by $\lambda = -1/2$

In relation with this geometric interpretation, it is sometimes convenient to have the following notion at hand.

Definition 1.2.1. *For any $A, B \in \mathbb{R}^n$ we set $\overrightarrow{AB}$ for the arrow starting at A and ending at B, and call it the* located vector $\overrightarrow{AB}$.

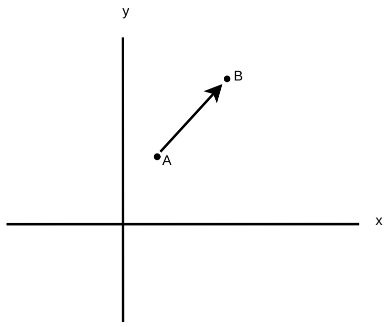


Figure 1.4: The located vector $\overrightarrow{AB}$

With this definition and for any $A \in \mathbb{R}^n$ the located vector $\overrightarrow{\mathbf{0}A}$ corresponds to the arrow mentioned in the previous geometric interpretation. For that reason, the located vector $\overrightarrow{\mathbf{0}A}$ is simply called *a vector* and is often identified with the element A of $\mathbb{R}^n$. Let us now introduce various relations between located vectors:

Definition 1.2.2. *For $A, B, C, D \in \mathbb{R}^n$, the located vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are* equivalent *if $B - A = D - C$. These located vectors are* parallel *if there exists $\lambda \in \mathbb{R}^* \equiv \mathbb{R} \setminus \{0\}$ such that $B - A = \lambda(D - C)$. In particular, they have the* same direction *if $\lambda > 0$ or have* opposite direction *if $\lambda < 0$.*

In Figure 1.5 equivalent located vectors and parallel located vectors are represented. Note that the located vector $\overrightarrow{AB}$ is always equivalent to the located vector $\overrightarrow{\mathbf{0}(B-A)}$

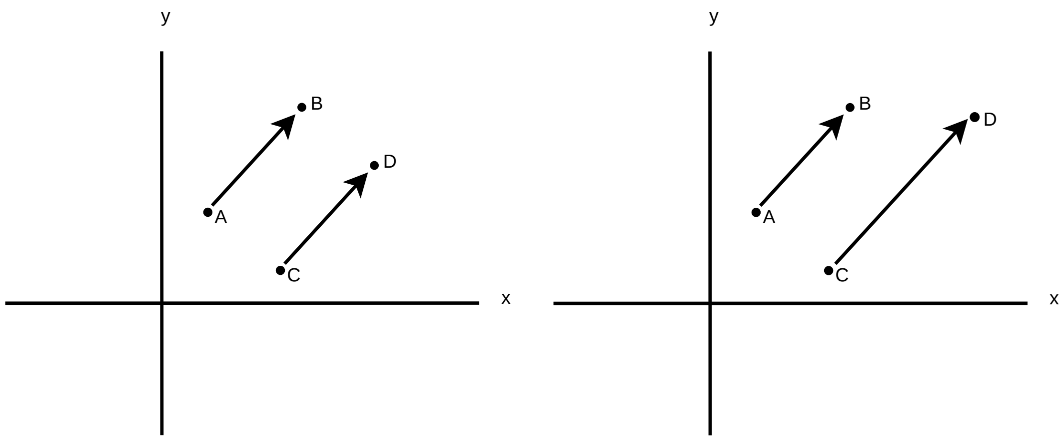


Figure 1.5: Equivalent and parallel located vectors

which is located at *the origin* $\mathbf{0}$, see Figure 1.6. This fact follows from the equality

$$(B - A) - \mathbf{0} = (B - A) = B - A.$$

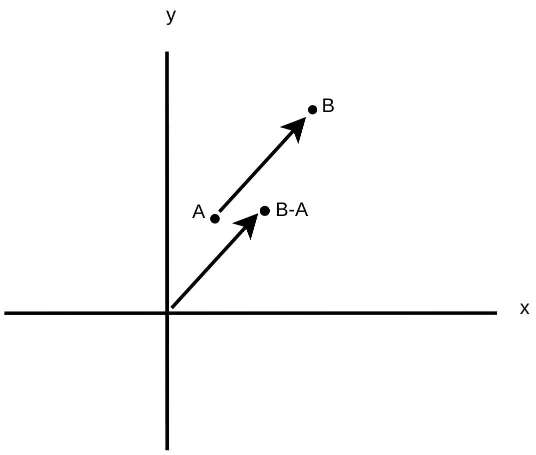


Figure 1.6: Located vector $\overrightarrow{AB}$ equivalent to the located vector $\overrightarrow{\mathbf{0}(B-A)}$

Question: What could be the meaning for two located vectors to be *perpendicular*? Even if one has an intuition in $\mathbb{R}^2$ or $\mathbb{R}^3$, one needs a precise definition for located vectors in $\mathbb{R}^n$.

1.3 Scalar product in $\mathbb{R}^n$

Definition 1.3.1. *For any $A, B \in \mathbb{R}^n$ with $A = (a_1, a_2, \dots, a_n)$ and $B = (b_1, b_2, \dots, b_n)$ one sets*

$$A \cdot B := a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{j=1}^{n} a_j b_j$$

and calls this number the scalar product *between A and B.*

For example, if $A = (1, 2)$ and $B = (3, 4)$, then $A \cdot B = 1{\cdot}3 + 2{\cdot}4 = 3 + 8 = 11$, but if $A = (1, 3)$ and $B = (6, -2)$, then $A \cdot B = 6 - 6 = 0$. Be aware that the previous notation is slightly misleading since the dot $\cdot$ between A and B corresponds to the scalar product while the dot between numbers just corresponds to the usual multiplication of numbers.

Properties 1.3.2. *For any $A, B, C \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ one has*

(i) $A \cdot B = B \cdot A$,

(ii) $A \cdot (B + C) = A \cdot B + A \cdot C$,

(iii) $(\lambda A) \cdot B = A \cdot (\lambda B) = \lambda(A \cdot B)$,

(iv) $A \cdot A \geq 0$, *and* $A \cdot A = 0$ *if and only if* $A = \mathbf{0}$.

These properties will be proved in Exercise 1.6.

Definition 1.3.3. *Two vectors* $A, B \in \mathbb{R}^n$ *are* perpendicular *or* orthogonal *if* $A \cdot B = 0$, *in which case one writes* $A \perp B$. *If* $A, B, C, D \in \mathbb{R}^n$, *the located vectors* $\overrightarrow{AB}$ *and* $\overrightarrow{CD}$ *are* perpendicular *or* orthogonal *if they are equivalent to two perpendicular vectors, in which case one writes* $\overrightarrow{AB} \perp \overrightarrow{CD}$.

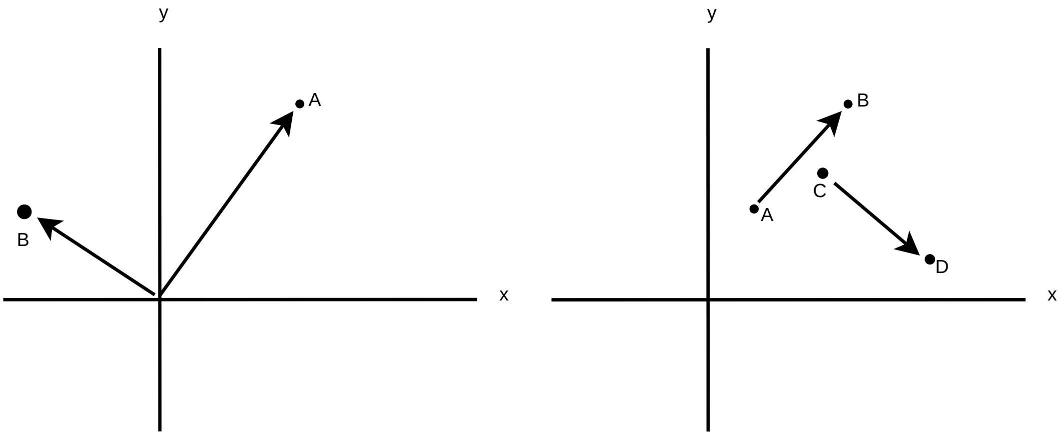


Figure 1.7: Perpendicular vectors and perpendicular located vectors

Remark first that if $A, B \in \mathbb{R}^n$ are perpendicular, then A is also perpendicular to λB for any $\lambda \in \mathbb{R}$. Indeed, from the above properties, it follows that if $A \cdot B = 0$ then $A \cdot (\lambda B) = \lambda(A \cdot B) = 0$. Now, observe also that in the setting of the previous definition, and since $\overrightarrow{AB}$ is equivalent to the vector $\overrightarrow{\mathbf{0}(B-A)}$ and since $\overrightarrow{CD}$ is equivalent to the vector $\overrightarrow{\mathbf{0}(D-C)}$, one has $\overrightarrow{AB} \perp \overrightarrow{CD}$ if and only $\overrightarrow{\mathbf{0}(B-A)}$ is perpendicular to $\overrightarrow{\mathbf{0}(D-C)}$, *i.e.* if and only if

$$(B - A) \cdot (D - C) = 0. \tag{1.3.1}$$

Example 1.3.4. *In* $\mathbb{R}^n$ *let us set* $E_1 = (1, 0, \dots, 0)$, $E_2 = (0, 1, 0, \dots, 0)$, $\dots$, $E_n = (0, \dots, 0, 1)$ *the* n *different vectors obtained by assigning a* 1 *at the coordinate* j *of* E_j *and* 0 *for all its other coordinates. Then, one easily checks that*

$$E_j \cdot E_k = 0 \textit{ whenever } j \neq k \quad \textit{and} \quad E_j \cdot E_j = 1 \textit{ for any } j \in \{1, 2, \dots, n\}.$$

These n *vectors are said to be* mutually orthogonal.

1.4 Euclidean norm in $\mathbb{R}^n$

Recall that for any $A \in \mathbb{R}^n$ one has $A^2 := A \cdot A \geq 0$.

Definition 1.4.1. *The* Euclidean norm *or simply* norm *of a vector* $A \in \mathbb{R}^n$ *is defined by* $\|A\| := \sqrt{A^2}$. *The positive number* $\|A\|$ *is also referred to as the* magnitude *of* A. *A vector of norm* 1 *is called a* unit vector.

Example 1.4.2. *If* $A = (-1, 2, 3) \in \mathbb{R}^3$*, then* $A \cdot A = (-1)^2 + 2^2 + 3^2 = 14$ *and therefore* $\|A\| = \sqrt{14}$*.*

Remark 1.4.3. *If* $n = 2$ *and in the geometric interpretation mentioned in Section 1.2, one observes that the norm* $\|A\|$ *of an element* $A \in \mathbb{R}^2$ *is compatible with Pythagoras theorem.*

Properties 1.4.4. *For any* $A \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$ *one has*

(i) $\|A\| = 0$ *if and only if* $A = \mathbf{0}$*,*

(ii) $\|\lambda A\| = |\lambda| \, \|A\|$*,*

(iii) $\| - A\| = \|A\|$*.*

Note that the third point is a special case of the second point. The proof of these properties will be provided in Exercise 1.8.

Definition 1.4.5. *For any* $A, B \in \mathbb{R}^n$*, the* distance *between* A *and* B*, denoted by* $d(A, B)$*, is defined by* $d(A, B) := \|B - A\|$*.*

Properties 1.4.6. *For any* $A, B, C \in \mathbb{R}^n$ *one has*

(i) $d(A, B) = d(B, A)$*,*

(ii) $d(A, B) = 0$ *if and only if* $A = B$*,*

(iii) $d(A - C, B - C) = d(A, B)$*, and in particular* $d(A, B) = d(\mathbf{0}, B - A)$*.*

The proofs of these properties are left as a free exercise. Now, keeping in mind the geometric interpretation provided in Section 1.2, it is natural to set

$$\|\overrightarrow{AB}\| := d(A, B)$$

and to call this number *the length* of the located vector $\overrightarrow{AB}$. Indeed, it follows from this definition and from Property 1.4.6.(iii) that

$$\|\overrightarrow{AB}\| = d(A, B) = d(\mathbf{0}, B - A) = \|\overrightarrow{\mathbf{0}(B - A)}\| = \|B - A\|.$$

Thus, the length of the located vector $\overrightarrow{AB}$ corresponds to the norm of the vector $(B - A) \in \mathbb{R}^n$. One also observes that any two located vectors which are equivalent have the same length.

Question: If $r > 0$ and $A \in \mathbb{R}^n$, what is

$$\big\{B \in \mathbb{R}^n \mid d(A,B) < r\big\} \quad ?$$

Can one draw a picture of this set for $n = 1$, $n = 2$ or $n = 3$?

Definition 1.4.7. *For $r > 0$ and $A \in \mathbb{R}^n$, one defines*

$$\mathscr{B}(A,r) := \big\{B \in \mathbb{R}^n \mid d(A,B) < r\big\}$$

and call $\mathscr{B}(A,r)$ the (open) ball centered at A and of radius r.

For example, if $n = 2$ then $\mathscr{B}(\mathbf{0},1)$ corresponds to the (open) unit disc in the plane, *i.e.* to the set of points on $\mathbb{R}^2$ which are at a distance strictly less than 1 from the origin $(0,0)$. If $n = 3$ then $\mathscr{B}(\mathbf{0},1)$ corresponds to the (open) unit ball in the usual 3 dimensional space.

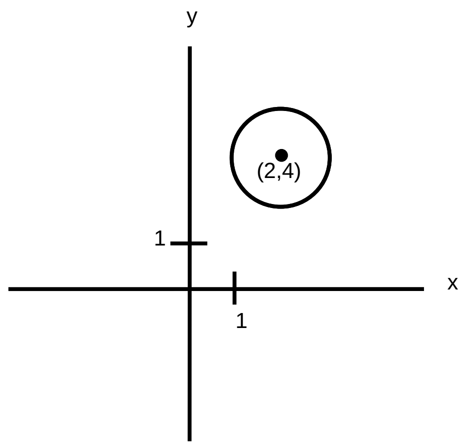


Figure 1.8: The open ball $\mathscr{B}\big((2,4),1\big)$ in $\mathbb{R}^2$

Let us now get a better intuition for the notion of orthogonal vectors. First of all, consider the following property:

Lemma 1.4.8. *For any $A, B \in \mathbb{R}^n$ one has*

$$\|B + A\| = \|B - A\| \Leftrightarrow A \cdot B = 0.$$

Proof. One has

$$\begin{aligned} \|B + A\| = \|B - A\| &\Leftrightarrow \|B + A\|^2 = \|B - A\|^2 \\ &\Leftrightarrow (B + A) \cdot (B + A) = (B - A) \cdot (B - A) \\ &\Leftrightarrow B^2 + 2A \cdot B + A^2 = B^2 - 2A \cdot B + A^2 \\ &\Leftrightarrow 4A \cdot B = 0, \end{aligned}$$

which justifies the statement. □

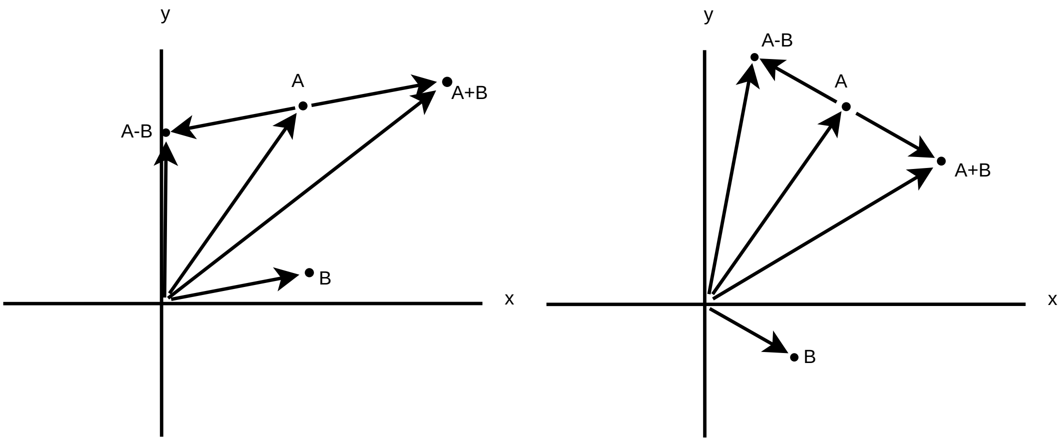


Figure 1.9: The vectors $A + B$ and $A - B$

By considering the geometric setting introduced in Section 1.2, one observes that the condition $\|B + A\| = \|B - A\|$ corresponds to our intuition for the two vectors A and B being perpendicular, see Figure 1.9. More generally, one can prove the general Pythagoras theorem:

Theorem 1.4.9. *Two vectors $A, B \in \mathbb{R}^n$ are mutually orthogonal if and only if the equality $\|A + B\|^2 = \|A\|^2 + \|B\|^2$ holds.*

The proof of this Theorem is provided in Exercise 1.9.

Question: Let $A, B \in \mathbb{R}^n$ with $B \neq \mathbf{0}$. Let P denote the point on the line passing through $\mathbf{0}$ and B, and such that the located vector $\overrightarrow{PA}$ is perpendicular to the located vector $\overrightarrow{\mathbf{0}B}$, see Figure 1.10. Clearly, $P = cB$ for some $c \in \mathbb{R}$, but how can one compute c ?

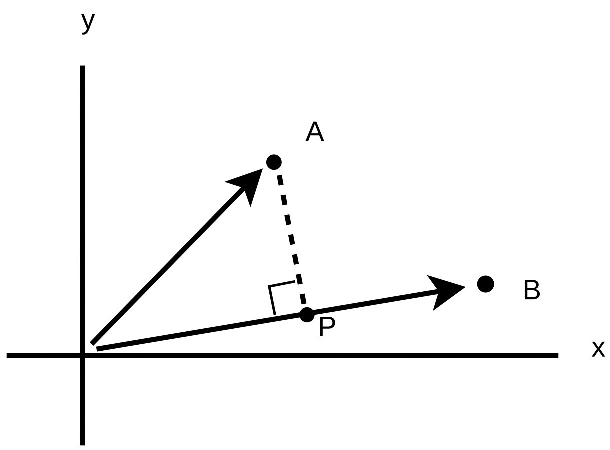


Figure 1.10:

For the answer, it is sufficient to consider the following equivalences:

$$\begin{aligned}\overrightarrow{PA} \perp \overrightarrow{\mathbf{0}B} &\Leftrightarrow \overrightarrow{\mathbf{0}(A-P)} \perp \overrightarrow{\mathbf{0}B} \\ &\Leftrightarrow (A-P)\cdot B = 0 \\ &\Leftrightarrow (A-cB)\cdot B = 0 \\ &\Leftrightarrow A\cdot B = cB^2 \\ &\Leftrightarrow c = \frac{A\cdot B}{\|B\|^2}.\end{aligned}$$

Definition 1.4.10. *Let $A, B \in \mathbb{R}^n$ with $B \neq \mathbf{0}$. Then* the component of A along B *is by definition the number $c := \frac{A\cdot B}{\|B\|^2}$. In this case cB is called* the orthogonal projection of A on B.

Let us recall from plane geometry that if one considers the right (or right-angled) triangle with vertices the points $\mathbf{0}$, A and cB with $A \neq \mathbf{0}$, $B \neq \mathbf{0}$ and with $c > 0$, then the angle θ at the vertex $\mathbf{0}$ satisfies

$$\cos(\theta) = \frac{\|cB\|}{\|A\|} = \frac{c\|B\|}{\|A\|} = \frac{(A\cdot B)\|B\|}{\|B\|^2\,\|A\|} = \frac{A\cdot B}{\|A\|\,\|B\|}.$$

Note that the same argument also holds for $c < 0$, and thus one has for any such triangle

$$\cos(\theta) = \frac{A\cdot B}{\|A\|\,\|B\|}.$$

From the above considerations and since $|\cos(\theta)| \leq 1$, one infers the following result:

Lemma 1.4.11. *For any $A, B \in \mathbb{R}^n$ one has*

$$|A\cdot B| \leq \|A\|\,\|B\|. \tag{1.4.1}$$

Let us also deduce a very useful inequality called *triangle inequality*:

Lemma 1.4.12. *For any $A, B \in \mathbb{R}^n$ one has*

$$\|A+B\| \leq \|A\| + \|B\|.$$

Proof. By taking into account the inequality (1.4.1) one obtains that

$$\begin{aligned}\|A+B\|^2 &= A^2 + B^2 + 2A\cdot B \\ &\leq A^2 + B^2 + 2|A\cdot B| \\ &\leq A^2 + B^2 + 2\|A\|\,\|B\| \\ &= \big(\|A\| + \|B\|\big)^2.\end{aligned}$$

The expected result is then obtained by taking the square root on both sides of the inequality. □

1.5 Parametric representation of a line

Let us consider $P, N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$, and let $t \in \mathbb{R}$.

Question: What does $\{P + tN \mid t \in \mathbb{R}\}$ represent? Can one draw a picture of this set?

Definition 1.5.1. *For any $P, N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$ one defines*

$$L_{P,N} := \big\{P + tN \mid t \in \mathbb{R}\big\}$$

and call this set the line passing through P and having the direction N. *More precisely $L_{P,N}$ is called* the parametric representation *of this line, see Figure 1.11.*

Remark 1.5.2. *(i) If N is replaced by λN for any $\lambda \in \mathbb{R}^*$, then $L_{P,\lambda N}$ describes the same line. In addition, any element of $L_{P,N}$ can be used instead of P and the resulting line will be the same.*

(ii) If $P, Q \in \mathbb{R}^n$, then the line passing through the two points P and Q is given by $L_{P,Q-P}$. Indeed one checks that $L_{P,Q-P} = \{P + t(Q - P) \mid t \in \mathbb{R}\}$, and that this line passes through P at $t = 0$ and passes through Q at $t = 1$.

(iii) For $P, Q \in \mathbb{R}^n$, the set $\{P + t(Q - P) \mid t \in [0, 1]\}$ describes the line segment starting at P and ending at Q.

Remark 1.5.3. *If $n = 2$ a line is often describes by $\{(x, y) \in \mathbb{R}^2 \mid ax + by = c\}$ for some $a, b, c \in \mathbb{R}$. Thus, in dimension 2 a line can be described by this formulation or with $L_{P,N}$ for some $P, N \in \mathbb{R}^2$. Clearly, some relations between a, b, c and P, N can be established. However, note that the above simple description does not exist for $n > 2$ while the definition $L_{P,N}$ holds in arbitrary dimension.*

1.6 Planes and hyperplanes

Let us first recall that two located vectors are orthogonal if they are equivalent to two perpendicular vectors.

Question: Let $P, N \in \mathbb{R}^3$ with $N \neq \mathbf{0}$. How can one describe the plane passing through P and perpendicular to the direction defined by the vector N ?

For the answer, consider a point X belonging to this plane. By definition, the located vector $\overrightarrow{PX}$ is orthogonal to the located vector $\overrightarrow{\mathbf{0}N}$, or equivalently the located vector $\overrightarrow{\mathbf{0}(X-P)}$ is orthogonal to the located vector $\overrightarrow{\mathbf{0}N}$. Now this condition reads $(X-P)\perp N$, which is equivalent to $(X - P) \cdot N = 0$, or by a simple computation to the condition $X \cdot N = P \cdot N$. In summary, the plane passing through P and perpendicular to the direction defined by the vector N is given by

$$\big\{X \in \mathbb{R}^3 \mid X \cdot N = P \cdot N\big\}.$$

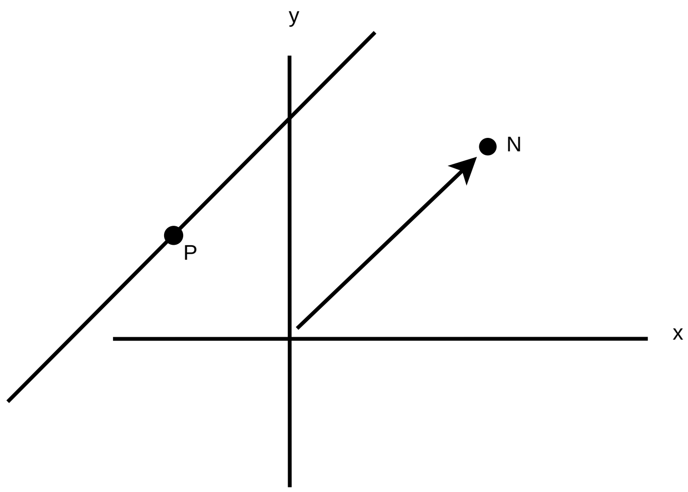


Figure 1.11: Parametric representation of a line

In this case, one also says that the plan is *normal* to the vector N.

Example 1.6.1. *If* $P = (2, 1, -1)$, $N = (-1, 1, 3)$ *and* $X = (x, y, z)$, *then*

$$X \cdot N = P \cdot N \Leftrightarrow (x, y, z) \cdot (-1, 1, 3) = -2 + 1 - 3 \Leftrightarrow -x + y + 3z = -4.$$

Therefore, the plane passing through $(2, 1, -1)$ *and normal to the vector* $(-1, 1, 3)$ *is given by*

$$\big\{(x, y, z) \in \mathbb{R}^3 \mid -x + y + 3z = -4\big\}.$$

Let us now work in arbitrary dimension.

Definition 1.6.2. *For any* $P, N \in \mathbb{R}^n$ *with* $N \neq \mathbf{0}$, *the set*

$$H_{P,N} := \big\{X \in \mathbb{R}^n \mid X \cdot N = P \cdot N\big\}$$

is called the hyperplane *passing through* P *and normal to* N.

Note that if $P = (p_1, p_2, \ldots, p_n)$ and if $N = (n_1, n_2, \ldots, n_n)$, then

$$H_{P,N} = \Big\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid n_1x_1 + n_2x_2 + \cdots + n_nx_n = \sum_{j=1}^{n} p_jn_j\Big\}.$$

Remark 1.6.3. *In the special case* $P \cdot N = 0$, *one observes that the element* $\mathbf{0}$ *belongs to* $H_{P,N}$. *Later on, we shall see that in this case* $H_{P,N}$ *is a* vector space, *see Chapter 3*

Properties 1.6.4. *For any* $P, N \in \mathbb{R}^n$ *with* $N \neq \mathbf{0}$, *and for any* $\lambda \in \mathbb{R}^*$ *one has*

(i) $H_{P,N} = H_{P,\lambda N}$,

(ii) If $P' \in H_{P,N}$, *then* $H_{P',N} = H_{P,N}$.

The proof of these properties will be provided in Exercise 1.16. It is now natural to define various notions related to hyperplanes. The following definitions correspond to the intuition we can have in $\mathbb{R}^2$ or in $\mathbb{R}^3$.

Definition 1.6.5. *Let $P, P', N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$ and with $P' \notin H_{P,N}$. Then the two hyperplanes $H_{P,N}$ and $H_{P',N}$ are* parallel.

Lemma 1.6.6. *Two parallel hyperplanes have an empty intersection.*

Proof. Let $H_{P,N}$ and $H_{P',N}$ be two parallel hyperplanes, and let us assume that there exists $X \in \mathbb{R}^n$ which belongs to both hyperplanes. This assumption means that $X \in H_{P,N}$ and $X \in H_{P',N}$, or equivalently $X \cdot N = P \cdot N$ and $X \cdot N = P' \cdot N$. As a consequence, it follows from these equalities that $P \cdot N = P' \cdot N$.

On the other hand, since the two planes are parallel, the assumption on P' is $P' \notin H_{P,N}$, which means that $P' \cdot N \neq P \cdot N$. Thus one has obtained a contradiction since $P \cdot N = P' \cdot N$ together with $P' \cdot N \neq P \cdot N$ is impossible. As a conclusion, there does not exist any X in the intersection of the two hyperplanes, or equivalently this intersection is empty. □

Example 1.6.7. *For $n = 2$, $P = (0, 0)$, $P' = (0, 1)$ and $N = (1, 1)$, one checks that $P' \cdot N = 1 \neq 0 = P \cdot N$, and thus $P' \notin H_{P,N}$. In addition, if $X = (x, y)$ one easily observes that $X \in H_{P,N}$ if and only $y = -x$ while $X \in H_{P',N}$ if and only if $y = -x + 1$.*

Definition 1.6.8. *Let $P, P', N, N' \in \mathbb{R}^n$ with $N \neq \mathbf{0}$ and $N' \neq \mathbf{0}$. One defines* the angle θ between the hyperplanes $H_{P,N}$ and $H_{P',N'}$ *as the angle between their normal vectors, or more precisely*

$$\cos(\theta) := \frac{N \cdot N'}{\|N\| \, \|N'\|}.$$

From this definition, one observes that the angle between two parallel hyperplanes is equal to 0.

Observation 1.6.9. *Let $P, N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$.*

(i) Since $H_{P,N} = H_{P,\lambda N}$ for any $\lambda \in \mathbb{R}^*$, one has $H_{P,N} = H_{P,\hat{N}}$ with $\hat{N} := \frac{N}{\|N\|}$. Note that $\hat{N}$ is a unit vector (see Definition 1.4.1).

(ii) The hyperplane $H_{P,N}$ divides $\mathbb{R}^n$ into two distinct regions. Indeed, for any $X \in \mathbb{R}^n$ one has either $X \cdot N > P \cdot N$, or $X \cdot N = P \cdot N$ or $X \cdot N < P \cdot N$. In the second case, X belongs to $H_{P,N}$ by definition of this hyperplane. Thus, one is left with the other two regions $\{X \in \mathbb{R}^n \mid X \cdot N > P \cdot N\}$ or $\{X \in \mathbb{R}^n \mid X \cdot N < P \cdot N\}$ and these two regions have an empty intersection.

Question: What is the distance between a point X and a hyperplane $H_{P,N}$?

The natural definition for such a notion can be understood as follows: Consider any point $Y \in H_{P,N}$ and recall that the distance $d(X,Y)$ between X and Y has been defined in Definition 1.4.5. Then, the distance $d(X,H_{P,N})$ between X and the hyperplane $H_{P,N}$ should be the minimal distance between X and any point $Y \in H_{P,N}$, namely

$$d(X,H_{P,N}) := \inf_{Y\in H_{P,N}} d(X,Y),$$

where the notation *inf* has to be read "infimum". In the next Lemma, we give an explicit formula for this distance.

Lemma 1.6.10. *For any $P,N,X \in \mathbb{R}^n$ with $N \neq \mathbf{0}$ one has*

$$d(X,H_{P,N}) = \frac{|(X-P)\cdot N|}{\|N\|}\ .$$

Proof. First of all, observe that if $X \notin H_{P,N}$, there exists $\lambda \in \mathbb{R}^*$ such that $X\cdot N - \lambda = P\cdot N$. In fact, one simply has $\lambda = X\cdot N - P\cdot N = (X-P)\cdot N$. In addition, observe that

$$X\cdot N-\lambda = P\cdot N \Leftrightarrow X\cdot N - \lambda\frac{N\cdot N}{\|N\|^2} = P\cdot N \Leftrightarrow \Big(X-\frac{\lambda}{\|N\|}\frac{N}{\|N\|}\Big)\cdot N = P\cdot N$$

which means that $X-\frac{\lambda}{\|N\|}\frac{N}{\|N\|}$ belongs to $H_{P,N}$ if $\lambda = (X-P)\cdot N$.

From this observation, one infers that $X' := X - \frac{(X-P)\cdot N}{\|N\|}\frac{N}{\|N\|} \in H_{P,N}$ and that

$$d(X,X') = \|X'-X\| = \Big\| -\frac{(X-P)\cdot N}{\|N\|}\frac{N}{\|N\|}\Big\| = \frac{|(X-P)\cdot N|}{\|N\|}.$$

As a consequence, one has $d(X,H_{P,N}) \leq \frac{|(X-P)\cdot N|}{\|N\|}$.

In order to show that this distance is the shortest one, consider any $Y \in H_{P,N}$ and use the general Pythagoras theorem for the right triangle of vertices X, Y and X'. Indeed, since $\overrightarrow{X'Y} \perp \overrightarrow{X'X}$ (because $Y \in H_{X',N}$ and $X - X' = \frac{(X-P)\cdot N}{\|N\|^2}N$) one gets:

$$d(X,Y)^2 = \|Y-X\|^2 = \|Y-X'\|^2 + \|X'-X\|^2 \geq \|X'-X\|^2 = d(X,X')^2$$

from which one infers that $d(X,Y) \geq d(X,X')$. □

Question: What are the intersections of hyperplanes? More precisely, can we find $X \in \mathbb{R}^n$ such that

$$X \in H_{P_1,N_1} \cap H_{P_2,N_2} \cap \cdots \cap H_{P_m,N_m}\ ? \tag{1.6.1}$$

Obviously, if some hyperplanes are parallel, there does not exist any X satisfying this condition. Even if the hyperplanes are not parallel, is it possible that the intersection is empty? Before answering these questions, recall once more that

$$X \in H_{P,N} \Leftrightarrow n_1x_1 + n_2x_2 + \cdots + n_nx_n = \sum_{j=1}^n p_jn_j.$$

and therefore equation (1.6.1) corresponds to a system of linear equations, as we shall see in the sequel.

1.7 Exercises

Exercise 1.1. *Compute $A+B$, $A-B$, $3A$ and $-2B$ in each of the following cases, and illustrate your result with the geometric interpretation.*

1. $A=(2,-1),\ B=(-1,1)$
2. $A=(2,-1,5),\ B=(-1,1,1)$
3. $A=(\pi,3,-1),\ B=(2\pi,-3,7)$

Exercise 1.2. *Let $A=(1,2)$ and $B=(3,1)$. Compute $A+2B$, $A-3B$ and $A+\frac{1}{2}B$ and provide the geometric interpretation.*

Exercise 1.3. *Write the proofs for Properties 1.1.4.*

Exercise 1.4. *In the following cases, determine which located vectors $\overrightarrow{PQ}$ and $\overrightarrow{AB}$ are equivalent.*

1. $P=(1,-1),\ Q=(4,3),\ A=(-1,5),\ B=(5,2)$
2. $P=(1,4),\ Q=(-3,5),\ A=(5,7),\ B=(1,8)$
3. $P=(1,-1,5),\ Q=(-2,3,-4),\ A=(3,1,1),\ B=(0,5,10)$
4. $P=(2,3,-4),\ Q=(-1,3,5),\ A=(-2,3,-1),\ B=(-5,3,8)$

Similarly, determine if the located vectors $\overrightarrow{PQ}$ and $\overrightarrow{AB}$ are parallel.

1. $P=(1,-1),\ Q=(4,3),\ A=(-1,5),\ B=(7,1)$
2. $P=(1,4),\ Q=(-3,5),\ A=(5,7),\ B=(9,6)$
3. $P=(1,-1,5),\ Q=(-2,3,-4),\ A=(3,1,1),\ B=(-3,9,-17)$
4. $P=(2,3,-4),\ Q=(-1,3,5),\ A=(-2,3,-1),\ B=(-11,3,-28)$

Exercise 1.5. *Compute $A\cdot A$ and $A\cdot B$ for the following vectors.*

1. $A=(2,-1),\ B=(-1,1)$
2. $A=(2,-1,5),\ B=(-1,1,1)$
3. $A=(\pi,3,-1),\ B=(2\pi,-3,7)$
4. $A=(1,-1,1),\ B=(2,3,1)$

Which pairs of vectors are perpendicular?

Exercise 1.6. *Write the proofs for Properties 1.3.2.*

Exercise 1.7. *By using the properties of the previous exercise, show the following equalities (we use the notation A^2 for $A \cdot A$).*

1. $(A+B)^2 = A^2 + 2A \cdot B + B^2$
2. $(A-B)^2 = A^2 - 2A \cdot B + B^2$

Exercise 1.8. *Write the proofs for Properties 1.4.4.*

Exercise 1.9. *Write a proof for Theorem 1.4.9.*

Exercise 1.10. *Let us consider the pair (A, B) of elements of $\mathbb{R}^n$.*

1. $A = (2, -1), \ \ B = (-1, 1)$
2. $A = (-1, 3), \ \ B = (0, 4)$
3. $A = (2, -1, 5), \ \ B = (-1, 1, 1)$

For each pair, compute the norm of A, the norm of B, and the orthogonal projection of A along B.

Exercise 1.11. *Find the cosine between the following vectors A and B :*

1. $A = (1, 2), \ \ B = (5, 3)$
2. $A = (1, -2, 3), \ \ B = (-3, 1, 5)$

Exercise 1.12. *Determine the cosine of the angles of the triangle whose vertices are $A = (2, -1, 1)$, $B = (1, -3, -5)$ and $C = (3, -4, -4)$.*

Exercise 1.13. *Let $A_1, \dots, A_r$ be non-zero vectors of $\mathbb{R}^n$ which are all mutually perpendicular, or in other words $A_j \cdot A_k = 0$ if $j \neq k$. Let $c_1, \dots, c_r$ be real numbers such that*

$$c_1 A_1 + c_2 A_2 + \cdots + c_r A_r = \mathbf{0}.$$

Show that $c_j = 0$ for all $j \in \{1, 2, \dots, r\}$.

Exercise 1.14. *Find a parametric representation of the line passing through A and B for*

1. $A = (1, 3, -1), \ \ B = (-4, 1, 2)$
2. $A = (-1, 5, 3), \ \ B = (-2, 4, 7)$

Exercise 1.15. *If P and Q are arbitrary points in $\mathbb{R}^n$, determine the general formula for the midpoint of the line segment between P and Q.*

Exercise 1.16. *Write the proofs for Properties 1.6.4.*

Exercise 1.17. *Determine the cosine of the angle between the two planes defined by*

$$\{(x,y,z) \in \mathbb{R}^3 \mid 2x - y + z = 0\} \quad \textit{and} \quad \{(x,y,z) \in \mathbb{R}^3 \mid x + 2y - z = 1\}.$$

Same question for the planes defined by

$$\{(x,y,z) \in \mathbb{R}^3 \mid x = 1\} \quad \textit{and} \quad \{(x,y,z) \in \mathbb{R}^3 \mid 3x + 2y - 7z = 1\}.$$

Exercise 1.18. *Find the equation of the plane in* $\mathbb{R}^3$ *passing through the three points* $P_1 = (1,2,-1)$, $P_2 = (-1,1,4)$ *and* $P_3 = (1,3,-2)$.

Exercise 1.19. *Let* $P = (-1,1,7)$, $Q = (1,3,5)$ *and* $N = (-1,1,-1)$. *Determine the distance between the point* Q *and the plane* $H_{P,N}$.

Exercise 1.20. *Let* $P = (1,1,1)$, $Q = (1,-1,2)$ *and* $N = (1,2,3)$. *Find the intersection of the line passing through* Q *and having the direction* N *with the plane* $H_{P,N}$.

Exercise 1.21. *Determine the equation of the hyperplane in* $\mathbb{R}^4$ *passing through the point* $(1,1,1,1)$ *and which is parallel to the hyperplane defined by*

$$\{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 \mid 1\,x_1 + 2\,x_2 + 3\,x_3 + 4\,x_4 = 5\}.$$

Similarly, for any $n > 1$ *determine the equation of the hyperplane in* $\mathbb{R}^n$ *passing through the point* $(1,1,\ldots,1)$ *and which is parallel to the hyperplane defined by*

$$\Big\{(x_1,x_2,\ldots,x_n) \in \mathbb{R}^n \mid \sum_{j=1}^{n} j\,x_j = n+1\Big\}.$$

Does something special happen for $n = 2$*?*

Chapter 2

Matrices and linear equations

In this chapter we introduce the notion of matrices and provide an algorithm for solving linear equations.

2.1 Matrices

In this section we introduce the matrices and some of their properties.

Definition 2.1.1. *For any $m, n \in \mathbb{N}$ we set*

$$\mathcal{A} \equiv (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

with $a_{ij} \in \mathbb{R}$ and call it a $m \times n$ matrix. *m corresponds to the number of rows while n corresponds to the number of columns. The number a_{ij} is called* the ij-entry *or* the ij-component *of the matrix $\mathcal{A}$. The set of all $m \times n$ matrices is denoted by $M_{mn}(\mathbb{R})$*[1].

Remark 2.1.2. *(i) $M_{11}(\mathbb{R})$ is identified with $\mathbb{R}$,*

(ii) $(a_1\ a_2\ \dots\ a_n) \equiv (a_{11}\ a_{12}\ \dots\ a_{1n}) \in M_{1n}(\mathbb{R})$ *while* $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \equiv \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} \in M_{m1}(\mathbb{R})$.
Elements of $M_{1n}(\mathbb{R})$ are called row vectors *while elements of $M_{m1}(\mathbb{R})$ are called* column vectors.

(iii) If $m = n$ one speaks about square matrices and sets $M_n(\mathbb{R})$ for $M_{nn}(\mathbb{R})$.

(iv) The matrix $\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$ *is called the 0-matrix, simply denoted by $\mathcal{O}$.*

[1]The symbol $\mathbb{R}$ is written because each entry a_{ij} belongs to $\mathbb{R}$. Note that one can consider more general matrices, as we shall see later on with complex numbers.

In the sequel, we shall tacitly use the following notation:

$$\mathcal{A} \equiv (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \qquad \mathcal{B} \equiv (b_{ij}) = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix},$$

and

$$\mathcal{C} \equiv (c_{ij}) = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix}.$$

The set $M_{mn}(\mathbb{R})$ can be endowed with two operations, namely:

Definition 2.1.3. *For any $\mathcal{A}, \mathcal{B} \in M_{mn}(\mathbb{R})$ and for any $\lambda \in \mathbb{R}$ we define the addition of $\mathcal{A}$ and $\mathcal{B}$ by*

$$\mathcal{A} + \mathcal{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix}$$

and the multiplication of $\mathcal{A}$ by the scalar λ:

$$\lambda \mathcal{A} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{pmatrix}.$$

Remark 2.1.4. *(i) Only matrices of the same size can be added, namely $\mathcal{A} + \mathcal{B}$ is well defined if and only if $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B} \in M_{mn}(\mathbb{R})$.*

(ii) The above rules can be rewritten with the more convenient notations

$$(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}) \quad \textit{and} \qquad \lambda(a_{ij}) = (\lambda a_{ij}) .$$

It is now easily observed that $\mathcal{A} + \mathcal{O} = \mathcal{O} + \mathcal{A} = \mathcal{A}$. In addition, one has $-\mathcal{A} = -1\mathcal{A} = (-a_{ij})$ and $\mathcal{A} - \mathcal{A} = \mathcal{A} + (-\mathcal{A}) = \mathcal{O}$. Some other properties are stated below, and their proofs are left as a free exercise.

Properties 2.1.5. *If $\mathcal{A}, \mathcal{B}, \mathcal{C} \in M_{mn}(R)$ and $\lambda, \mu \in \mathbb{R}$ then one has*

(i) $\mathcal{A} + \mathcal{B} = \mathcal{B} + \mathcal{A}$, (commutativity)

(ii) $(\mathcal{A} + \mathcal{B}) + \mathcal{C} = \mathcal{A} + (\mathcal{B} + \mathcal{C})$, (associativity)

(iii) $\lambda(\mathcal{A}+\mathcal{B}) = \lambda\mathcal{A}+\lambda\mathcal{B}$, (distributivity)

(iv) $(\lambda+\mu)\mathcal{A} = \lambda\mathcal{A}+\mu\mathcal{A}$,

(v) $(\lambda\mu)\mathcal{A} = \lambda(\mu\mathcal{A})$.

Note that the above properties are very similar to the one already mentioned for vectors in Properties 1.1.4. These similarities will be emphasized in the following chapter.

Let us add one more operation on matrices, namely *the transpose* of a matrix.

Definition 2.1.6. *For any $\mathcal{A} = (a_{ij}) \in M_{mn}(\mathbb{R})$, one defines ${}^t\!\mathcal{A} \equiv ({}^t a_{ij}) \in M_{nm}(\mathbb{R})$ the transpose of $\mathcal{A}$ by the relation*

$$ {}^t a_{ij} := a_{ji} \ . $$

In other words, taking the transpose of a matrix consists in changing rows into columns and vice versa.

We also define a product for matrices:

Definition 2.1.7. *For $\mathcal{A} \in M_{mn}(\mathbb{R})$ and for $\mathcal{B} \in M_{np}(\mathbb{R})$ one defines the product of $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{C} := \mathcal{A}\mathcal{B} \in M_{mp}(\mathbb{R})$ with*

$$ c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} \ . $$

Examples 2.1.8. *1.*

$$ \underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}}_{\in M_{32}(\mathbb{R})} \underbrace{\begin{pmatrix} 7 \\ 8 \end{pmatrix}}_{\in M_{21}(\mathbb{R})} = \begin{pmatrix} 1\cdot 7 + 2\cdot 8 \\ 3\cdot 7 + 4\cdot 8 \\ 5\cdot 7 + 6\cdot 8 \end{pmatrix} = \underbrace{\begin{pmatrix} 23 \\ 53 \\ 83 \end{pmatrix}}_{\in M_{31}(\mathbb{R})} $$

2.

$$ \underbrace{\begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}}_{\in M_{1n}(\mathbb{R})} \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}}_{\in M_{n1}(\mathbb{R})} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \ \in M_{11}(\mathbb{R}) $$

3.

$$ \underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}}_{\in M_{nn}(\mathbb{R})} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\in M_{n1}(\mathbb{R})} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\in M_{n1}(\mathbb{R})} $$

with $y_i = \sum_{j=1}^{n} a_{ij} x_j$.

Remark 2.1.9. *(i) If $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B} \in M_{pq}(\mathbb{R})$, then the product $\mathcal{AB}$ can be defined if and only if $n = p$, in which case $\mathcal{AB} \in M_{mq}(\mathbb{R})$.*

(ii) If $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$, then $\mathcal{AB}$ and $\mathcal{BA}$ can be defined and belong to $M_n(\mathbb{R})$. However, in general it is not true that $\mathcal{AB} = \mathcal{BA}$, most of the time $\mathcal{AB} \neq \mathcal{BA}$.

Let us now state some important properties of this newly defined product.

Properties 2.1.10. *(i) For any $\mathcal{A} \in M_{mn}(\mathbb{R})$, $\mathcal{B}, \mathcal{C} \in M_{np}(\mathbb{R})$ and $\lambda \in \mathbb{R}$ one has*

(a) $\mathcal{A}(\mathcal{B} + \mathcal{C}) = \mathcal{AB} + \mathcal{AC}$,

(b) $(\lambda\mathcal{A})\mathcal{B} = \lambda(\mathcal{AB}) = \mathcal{A}(\lambda\mathcal{B})$.

(ii) If $\mathcal{A} \in M_{mn}(\mathbb{R})$, $\mathcal{B} \in M_{np}(\mathbb{R})$ and $\mathcal{C} \in M_{pq}(\mathbb{R})$ one has

$$(\mathcal{AB})\mathcal{C} = \mathcal{A}(\mathcal{BC}).$$

(iii) If $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B} \in M_{np}(\mathbb{R})$ one also has

$${}^t(\mathcal{AB}) = {}^t\mathcal{B}\ {}^t\mathcal{A}.$$

These properties will be proved in Exercise 2.2. Recall now that for the addition, the matrix $\mathcal{O}$ has the property $\mathcal{A} + \mathcal{O} = \mathcal{A} = \mathcal{O} + \mathcal{A}$. We shall now introduce the square matrix $\mathbf{1}_n$ which share a similar property but with respect to the multiplication. Indeed, let us set

$$\mathbf{1}_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

or equivalently $\mathbf{1}_n \in M_n(\mathbb{R})$ is the matrix having 1 on its diagonal, and 0 everywhere else. Then one can show that for any $\mathcal{A} \in M_{mn}(\mathbb{R})$ one has $\mathcal{A}\mathbf{1}_n = \mathcal{A}$ and $\mathbf{1}_m\ \mathcal{A} = \mathcal{A}$, see Exercise 2.3.

For the set of square matrices, we can define the notion of an inverse and state several of their properties.

Definition 2.1.11. *Let $\mathcal{A} \in M_n(\mathbb{R})$. The matrix $\mathcal{B} \in M_n(\mathbb{R})$ is* an inverse for $\mathcal{A}$ *if $\mathcal{AB} = \mathbf{1}_n$ and $\mathcal{BA} = \mathbf{1}_n$.*

Lemma 2.1.12. *The inverse of a matrix, if it exists, is unique*

Proof. Assume that $\mathcal{B}_1, \mathcal{B}_2 \in M_n(\mathbb{R})$ are inverses for $\mathcal{A}$, *i.e.* $\mathcal{AB}_1 = \mathbf{1}_n = \mathcal{B}_1\mathcal{A}$ and $\mathcal{AB}_2 = \mathbf{1}_n = \mathcal{B}_2\mathcal{A}$, then one has

$$\mathcal{B}_1 = \mathcal{B}_1\mathbf{1}_n = \mathcal{B}_1(\mathcal{AB}_2) = (\mathcal{B}_1\mathcal{A})\mathcal{B}_2 = \mathbf{1}_n\mathcal{B}_2 = \mathcal{B}_2$$

which shows that $\mathcal{B}_1$ and $\mathcal{B}_2$ are equal. □

Since the inverse of a matrix $\mathcal{A}$, if it exists, is unique, we can speak about the inverse of $\mathcal{A}$ and denote it by $\mathcal{A}^{-1}$. In such a situation, $\mathcal{A}$ is called *an invertible matrix*.

Remark 2.1.13. *We shall see later on that the property $\mathcal{AB} = \mathbf{1}_n$ automatically implies the property $\mathcal{BA} = \mathbf{1}_n$. Thus, it follows either from $\mathcal{AB} = \mathbf{1}_n$ or from $\mathcal{BA} = \mathbf{1}_n$ that $\mathcal{B}$ is the inverse of $\mathcal{A}$,* i.e. $\mathcal{B} = \mathcal{A}^{-1}$.

Properties 2.1.14. *Let $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ both having an inverse, and let $\lambda \in \mathbb{R}^*$. Then*

(i) $\big(\mathcal{A}^{-1}\big)^{-1} = \mathcal{A}$,

(ii) ${}^t\big(\mathcal{A}^{-1}\big) = \big({}^t\!\mathcal{A}\big)^{-1}$,

(iii) $(\lambda\mathcal{A})^{-1} = \lambda^{-1}\mathcal{A}^{-1}$,

(iv) $(\mathcal{AB})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}$.

Proof. (i) Since $\big(\mathcal{A}^{-1}\big)\mathcal{A} = \mathbf{1}_n = \mathcal{A}\big(\mathcal{A}^{-1}\big)$, it follows that $\mathcal{A}$ is the inverse of $\mathcal{A}^{-1}$, *i.e.* $\big(\mathcal{A}^{-1}\big)^{-1} = \mathcal{A}$.

(ii) Since ${}^t\big(\mathcal{A}^{-1}\big){}^t\!\mathcal{A} = {}^t\big(\mathcal{A}\mathcal{A}^{-1}\big) = {}^t\mathbf{1}_n = \mathbf{1}_n$ and since ${}^t\!\mathcal{A}\,{}^t\big(\mathcal{A}^{-1}\big) = {}^t\big(\mathcal{A}^{-1}\mathcal{A}\big) = {}^t\mathbf{1}_n = \mathbf{1}_n$, it follows that ${}^t\big(\mathcal{A}^{-1}\big)$ is the inverse of ${}^t\!\mathcal{A}$, or in other words $\big({}^t\!\mathcal{A}\big)^{-1} = {}^t\big(\mathcal{A}^{-1}\big)$.

(iii) One has $(\lambda\mathcal{A})(\lambda^{-1})\mathcal{A}^{-1} = \lambda\lambda^{-1}\mathcal{A}\mathcal{A}^{-1} = \mathbf{1}_n = (\lambda^{-1}\mathcal{A}^{-1})(\lambda\mathcal{A})$, which means that $\lambda^{-1}\mathcal{A}^{-1}$ is the inverse for $\lambda\mathcal{A}$.

(iv) One observes that $\big(\mathcal{B}^{-1}\mathcal{A}^{-1}\big)(\mathcal{AB}) = \mathbf{1}_n = (\mathcal{AB})\big(\mathcal{B}^{-1}\mathcal{A}^{-1}\big)$, which shows that $(\mathcal{AB})^{-1}$ is given by $\mathcal{B}^{-1}\mathcal{A}^{-1}$. □

Note that thanks to Remark 2.1.13 one could have simplified the above proof by checking only one condition for each inverse. Let us still introduce some special classes of matrices and the notion of similarity, which are going to play an important role in the sequel.

Definition 2.1.15. *(i) If $\mathcal{A} \equiv (a_{ij}) \in M_n(\mathbb{R})$ with $a_{ij} = 0$ whenever $i \neq j$, then $\mathcal{A}$ is called* a diagonal matrix*,*

(ii) If $\mathcal{A} \equiv (a_{ij}) \in M_n(\mathbb{R})$ with $a_{ij} = 0$ whenever $i > j$, then $\mathcal{A}$ is called an upper triangular matrix*,*

(iii) If $\mathcal{A} \in M_n(\mathbb{R})$ and if there exists $m \in \mathbb{N}$ such that $\mathcal{A}^m := \underbrace{\mathcal{A}\mathcal{A}\ldots\mathcal{A}}_{m \text{ times}} = \mathcal{O}$, then $\mathcal{A}$ is called a nilpotent matrix.

Definition 2.1.16. *For $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ one says that $\mathcal{A}$ and $\mathcal{B}$ are similar if there exists an invertible matrix $\mathcal{U} \in M_n(\mathbb{R})$ such that*

$$\mathcal{B} = \mathcal{U}\mathcal{A}\mathcal{U}^{-1} .$$

Lemma 2.1.17. *Let $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ be two similar matrices. Then* *(i)* $\mathcal{A}$ *is invertible if and only if* $\mathcal{B}$ *is invertible,*

(ii) ${}^t\mathcal{A}$ *and* ${}^t\mathcal{B}$ *are similar,*

(iii) $\mathcal{A}$ *is nilpotent if and only if* $\mathcal{B}$ *is nilpotent.*

Proof. Let us assume that $\mathcal{B} = \mathcal{U}\mathcal{A}\mathcal{U}^{-1}$ for some invertible matrix $\mathcal{U} \in M_n(\mathbb{R})$.

(i) Assume that $\mathcal{A}$ is invertible, and observe that

$$\mathcal{B}(\mathcal{U}\mathcal{A}^{-1}\mathcal{U}^{-1}) = \mathcal{U}\mathcal{A}\mathcal{U}^{-1}\mathcal{U}\mathcal{A}^{-1}\mathcal{U}^{-1} = \mathbf{1}_n$$

which means that $\mathcal{U}\mathcal{A}^{-1}\mathcal{U}^{-1}$ is the inverse of $\mathcal{B}$. As a consequence, $\mathcal{B}$ is invertible. Similarly, if one assumes that $\mathcal{B}$ is invertible, then $\mathcal{U}^{-1}\mathcal{B}^{-1}\mathcal{U}$ is an inverse for $\mathcal{A}$, as it can easily be checked. One then deduces that $\mathcal{A}$ is invertible.

(ii) One observes that

$${}^t\mathcal{B} = {}^t(\mathcal{U}\mathcal{A}\mathcal{U}^{-1}) = {}^t\big(\mathcal{U}^{-1}\big){}^t\mathcal{A}{}^t\mathcal{U} = ({}^t\mathcal{U})^{-1}\,{}^t\mathcal{A}{}^t\big(({}^t\mathcal{U})^{-1}\big)^{-1} = \mathcal{V}\,{}^t\mathcal{A}\mathcal{V}^{-1}$$

with $\mathcal{V} := ({}^t\mathcal{U})^{-1}$ which is invertible. Thus ${}^t\mathcal{A}$ and ${}^t\mathcal{B}$ are similar.

(iii) If $\mathcal{A}^m = \mathcal{O}$, then

$$\mathcal{B}^m = \big(\mathcal{U}\mathcal{A}\mathcal{U}^{-1}\big)^m = \underbrace{(\mathcal{U}\mathcal{A}\mathcal{U}^{-1})(\mathcal{U}\mathcal{A}\mathcal{U}^{-1})\dots(\mathcal{U}\mathcal{A}\mathcal{U}^{-1})}_{m \text{ times}} = \mathcal{U}\mathcal{A}^m\mathcal{U}^{-1} = \mathcal{O}.$$

Similarly, if $\mathcal{B}^m = \mathcal{O}$, then $\mathcal{A}^m = \mathcal{U}^{-1}\mathcal{B}^m\mathcal{U} = \mathcal{O}$, which proves the statement. ☐

2.2 Matrices and elements of $\mathbb{R}^n$

In this section we show how a matrix can be applied to an element of $\mathbb{R}^n$. In fact, such an action has implicitly been mentioned in Example 2.1.8, but we shall now develop this point of view.

First of all, we shall now modify the convention used in the previous chapter. Indeed, for convenience we have written $A = (a_1, a_2, \dots, a_n)$ for any element of $\mathbb{R}^n$. However, from now on we shall write $A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ for elements of $\mathbb{R}^n$. However, the following alternative notation will also be used: $A = {}^t(a_1\ a_2\ \dots\ a_n)$, or equivalently ${}^tA = (a_1\ a_2\ \dots\ a_n)$. Note that is coherent with the notion of transpose of a matrix, since column vector are identified with elements of $M_{n1}(\mathbb{R})$ while row vectors are identified with elements of $M_{1n}(\mathbb{R})$, see Remark 2.1.2.

The main interest in this notation is that a $m \times n$ matrix can now easily be applied to a column vector, and the resulting object is again a column vector. For example

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}}_{\in M_{33}(\mathbb{R})} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\in \mathbb{R}^3} = \underbrace{\begin{pmatrix} 1x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \\ 7x_1 + 8x_2 + 9x_3 \end{pmatrix}}_{\in \mathbb{R}^3}. \tag{2.2.1}$$

More generally, one has

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}}_{\in M_{mn}(\mathbb{R})} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\in \mathbb{R}^n} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{\in \mathbb{R}^m}$$

or in other words by applying a $m \times n$ matrix to an element of $\mathbb{R}^n$ one obtains an element of $\mathbb{R}^m$.

Let us also observe that with the above convention for elements of $\mathbb{R}^n$ in mind, the scalar product $A \cdot B$ of $A, B \in \mathbb{R}^n$ introduced in Definition 1.3.1 can be seen as a product of matrices. Indeed the following equalities hold:

$$A \cdot B = \sum_{j=1}^{n} a_j b_j = (a_1 \ a_2 \ \dots \ a_n) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = {}^t\!A\, B \tag{2.2.2}$$

where the left hand side corresponds to the scalar product of two elements of $\mathbb{R}^n$ while the right hand side corresponds to a product of a matrix in $M_{1n}(\mathbb{R})$ with a matrix in $M_{n1}(\mathbb{R})$.

We can also see that the product of two matrices can be rewritten with an alternative notation. Indeed, for any $\mathcal{A} \in M_{mn}(\mathbb{R})$ let us set $\mathcal{A}^j \in M_{m1}(\mathbb{R})$ for the j^{th} column of $\mathcal{A}$, and $\mathcal{A}_i \in M_{1n}(\mathbb{R})$ for the i^{th} row of $\mathcal{A}$. More explicitly one sets

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}^1 & \mathcal{A}^2 & \dots & \mathcal{A}^n \end{pmatrix} \qquad \text{and} \qquad \mathcal{A} = \begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \vdots \\ \mathcal{A}_m \end{pmatrix}. \tag{2.2.3}$$

With this notation, for any $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B} \in M_{np}(\mathbb{R})$, the matrix $\mathcal{C} := \mathcal{A}\mathcal{B}$ is given by

$$c_{ik} = \mathcal{A}_i \mathcal{B}^k \tag{2.2.4}$$

where the right hand side corresponds to the product of $\mathcal{A}_i \in M_{1n}(\mathbb{R})$ with $\mathcal{B}^k \in M_{n1}(\mathbb{R})$. In other words one can still write

$$c_{ik} = (\text{row } i \text{ of } \mathcal{A}) \begin{pmatrix} \text{col.} \\ k \\ \text{of} \\ \mathcal{B} \end{pmatrix}.$$

2.3 Homogeneous linear equations

In this section, we consider linear systems of equations when the number of unknowns is strictly bigger than the number of equations.

Example 2.3.1. *Let us consider the equation*

$$2x + y - 4z = 0$$

and look for a non-trivial *solution, i.e. a solution* $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ *with* $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. *By writing* $x = \frac{-y+4z}{2}$ *and by choosing* $\begin{pmatrix} y \\ z \end{pmatrix}$ *with* $\begin{pmatrix} y \\ z \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, *one gets for example* $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1 \\ 1 \end{pmatrix}$. *Note that an infinite number of other solutions exist.*

Example 2.3.2. *Let us consider the linear system of equations*

$$\begin{cases} 2x_1 + 3x_2 - x_3 = 0 \\ x_1 + x_2 + x_3 = 0 \end{cases}$$

and look for a non-trivial solution. By multiplying the second equation by 2 *and by subtracting it to the first equation one obtains*

$$\begin{cases} 2x_1 + 3x_2 - x_3 - 2(x_1 + x_2 + x_3) = 0 \\ x_1 + x_2 + x_3 = 0 \end{cases} \Leftrightarrow \begin{cases} x_2 - 3x_3 = 0 \\ x_1 + x_2 + x_3 = 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} x_2 = 3x_3 \\ x_1 + x_2 + x_3 = 0 \end{cases}.$$

Thus, a solution is for example $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \\ 1 \end{pmatrix}$, *but again this is one solution amongst many others.*

More generally, if one starts with a system of m equations for n unknowns $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ with $n > m$, one can eliminate one unknown (say x_1) and obtains $m - 1$ equations for $n-1$ unknowns. By doing this process again, one can then eliminate one more unknown (say x_2) and obtains $m - 2$ equations for $n - 2$ unknowns. Obviously, this can be done again and again...

Question: Can we always find a non-trivial solution in such a situation ? The answer is yes, as we shall see now.

Let us consider the following system of m equations for n unknowns:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{2.3.1}$$

with $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$. By using the notation introduce before, this system can be rewritten as

$$\mathcal{A}X = B$$

with $\mathcal{A} = (a_{ij}) \in M_{mn}(\mathbb{R})$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and $B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$.

Definition 2.3.3. *For any* $\mathcal{A} = (a_{ij}) \in M_{mn}(\mathbb{R})$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ *and* $B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$, *the system*

$$\mathcal{A}X = 0$$

is called the homogeneous linear system associated with the linear system $\mathcal{A}X = B$.

One easily observes that the solution $X = \mathbf{0} \in \mathbb{R}^n$ is always a solution of the homogeneous system.

Theorem 2.3.4. *Let*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{2.3.2}$$

be a homogeneous linear system of m *equations with for* n *unknowns, with* $n > m$. *Then the system has a non-trivial solution (and maybe several).*

Remark 2.3.5. *As already mentioned,* (2.3.2) *is equivalent to* $\mathcal{A}X = \mathbf{0}$, *with* $\mathcal{A} \in M_{mn}(\mathbb{R})$ *and* $X \in \mathbb{R}^n$. *Then, by using the notation introduced in* (2.2.3), *this system is still equivalent to*

$$\begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_m \end{pmatrix} X = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.3.3}$$

where $\mathcal{A}_i \in M_{1n}$ *for* $i \in \{1, \dots m\}$. *Thus,* (2.3.3) *can still be rewritten as the* m *equations* $\mathcal{A}_i X = 0$ *for* $i \in \{1, \dots m\}$, *with the notation analogous to the one already used in* (2.2.4). *In other words,* (2.3.2) *is equivalent to*

$$^t\mathcal{A}_i \cdot X = 0 \qquad \textit{for } i \in \{1, \dots, m\},$$

meaning that X *is orthogonal to all vectors* $^t\mathcal{A}_i \in \mathbb{R}^n$.

Proof of Theorem 2.3.4. The proof consists in an induction procedure.

1) If $m = 1$, then the system reduces to the equation $a_{11}x_1 + \cdots + a_{1n}x_n = 0$. If $a_{11} = \cdots = a_{1n} = 0$, then any $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ is a non-trivial solution. If $a_{11} \neq 0$, then

$$x_1 = \frac{-a_{12}x_2 - \cdots - a_{1n}x_n}{a_{11}}, \tag{2.3.4}$$

and we can choose any $\begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ and then determine x_1 by (2.3.4). The final solution is non-trivial. Note that the choice $a_{11} \neq 0$ is arbitrary, and any other choice would lead to a non-trivial solution.

2) Assume that the statement is true for some $m-1$ equations and $n-1$ unknowns, and let us prove that it is still true for m equations and n unknowns. Again, if all $a_{ij} = 0$, then any $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ is a non-trivial solution. If $a_{11} \neq 0$, let us consider the system

$$\begin{cases} {}^t\mathcal{A}_2 \cdot X - \frac{a_{21}}{a_{11}} {}^t\mathcal{A}_1 \cdot X = 0 \\ \quad\vdots \\ {}^t\mathcal{A}_m \cdot X - \frac{a_{m1}}{a_{11}} {}^t\mathcal{A}_1 \cdot X = 0 \end{cases}$$

with the notations recalled in Remark 2.3.5. Since the coefficients multiplying x_1 are all 0, this system is a system of $m-1$ equations for $n-1$ unknowns. By assumption, such a system has a non-trivial solution which we denote by $\begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix}$. Then, by solving ${}^t\mathcal{A}_1 \cdot X = 0$, one obtains that x_1 is given by (2.3.4) and thus there exists $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ which is a solution of the system.

3) Since m, n were arbitrary with the only condition $n > m$, one has exhibited a non-trivial solution for the original system. □

2.4 Row operations and Gauss elimination

Recall that a system of m equations for n unknowns as written in (2.3.1) is equivalent to the equation

$$\mathcal{A}X = B \tag{2.4.1}$$

with $\mathcal{A} \in M_{mn}(\mathbb{R})$, $B \in \mathbb{R}^m$ and for the unknown $X \in \mathbb{R}^n$.

Question: Given $\mathcal{A}$ and B, can one always find a solution X for the equation (2.4.1) ?

In some special cases, as seen in the previous chapter with $B = \mathbf{0}$ and $n > m$, the answer is yes. We present a here a second special case.

Lemma 2.4.1. *Assume that $m = n$ and that $\mathcal{A} \in M_n(\mathbb{R})$ is invertible. Then the system (2.4.1) admits a unique solution given by $X := \mathcal{A}^{-1}B$.*

Proof. One directly checks that if $X = \mathcal{A}^{-1}B$, then $\mathcal{A}(\mathcal{A}^{-1})B = B$, as expected. On the other hand, if there would exist $X' \in \mathbb{R}^n$ with $X' \neq X$ and satisfying $\mathcal{A}X' = B$, then by applying $\mathcal{A}^{-1}$ on the left of both sides of this equality one gets

$$\mathcal{A}^{-1}(\mathcal{A}X') = \mathcal{A}^{-1}B \Leftrightarrow X' = \mathcal{A}^{-1}B = X$$

which is a contradiction. Thus the solution to (2.4.1) is unique in this case. □

Note that finding $\mathcal{A}^{-1}$ might be complicated, and how can we deal with the general case $m \neq n$? In order to get an efficient way for dealing with linear systems, let us start by recalling a convenient way for solving linear systems. Let us consider the system

$$\begin{cases} 2x + y + 4z + w = -2 \\ -3x + 2y - 3z + w = 1 \\ x + y + z = -1 \end{cases} \tag{2.4.2}$$

and look for a solution to it. By some simple manipulations one gets

$$\begin{cases} 2x + y + 4z + w = -2 \\ -3x + 2y - 3z + w = 1 \\ x + y + z = -1 \end{cases} \overset{r_1-2r_3}{\Longleftrightarrow} \begin{cases} 0x - y + 2z + w = 0 \\ -3x + 2y - 3z + w = 1 \\ x + y + z = -1 \end{cases}$$
$$\overset{r_2+3r_3}{\Longleftrightarrow} \begin{cases} -y + 2z + w = 0 \\ 0x + 5y + 0z + w = -2 \\ x + y + z = -1 \end{cases} \overset{r_1\leftrightarrow r_3}{\Longleftrightarrow} \begin{cases} x + y + z = -1 \\ 5y + w = -2 \\ -y + 2z + w = 0 \end{cases}$$
$$\overset{r_2+5r_3}{\Longleftrightarrow} \begin{cases} x + y + z = -1 \\ 0y + 10z + 6w = -2 \\ -y + 2z + w = 0 \end{cases} \overset{-r_3\leftrightarrow r_2}{\Longleftrightarrow} \begin{cases} x + y + z = -1 \\ y - 2z - w = 0 \\ 10z + 6w = -2 \end{cases}.$$

A special solution for this system is obtained for example by fixing $z = -2$, and then by deducing successively that $w = 3$, $y = -1$ and $x = 2$. In other words a solution to this system is given by $\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -2 \\ 3 \end{pmatrix}$.

Let us now rewrite these manipulations in an equivalent way. *A priori*, it will look longer, but with some practice, the size of the computations will become much shorter. For that purpose, consider *the augmented matrix*

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

obtained by collecting in the same matrix the coefficients of the linear system together with the coefficients on the right hand side of the equality. Then, one can perform the following elementary operations

$$\begin{pmatrix} 2 & 1 & 4 & 1 & -2 \\ -3 & 2 & -3 & 1 & 1 \\ 1 & 1 & 1 & 0 & -1 \end{pmatrix} \overset{r_1-2r_3}{\sim} \begin{pmatrix} 0 & -1 & 2 & 1 & 0 \\ -3 & 2 & -3 & 1 & 1 \\ 1 & 1 & 1 & 0 & -1 \end{pmatrix}$$
$$\overset{r_2+3r_3}{\sim} \begin{pmatrix} 0 & -1 & 2 & 1 & 0 \\ 0 & 5 & 0 & 1 & -2 \\ 1 & 1 & 1 & 0 & -1 \end{pmatrix} \overset{r_1\leftrightarrow r_3}{\sim} \begin{pmatrix} 1 & 1 & 1 & 0 & -1 \\ 0 & 5 & 0 & 1 & -2 \\ 0 & -1 & 2 & 1 & 0 \end{pmatrix}$$
$$\overset{r_2+5r_3}{\sim} \begin{pmatrix} 1 & 1 & 1 & 0 & -1 \\ 0 & 0 & 10 & 6 & -2 \\ 0 & -1 & 2 & 1 & 0 \end{pmatrix} \overset{-r_3\leftrightarrow r_2}{\sim} \begin{pmatrix} 1 & 1 & 1 & 0 & -1 \\ 0 & 1 & -2 & -1 & 0 \\ 0 & 0 & 10 & 6 & -2 \end{pmatrix}$$

from which one deduces the new system of equations

$$\begin{cases} x + y + z = -1 \\ y - 2z - w = 0 \\ 10z + 6w = -2 \end{cases}.$$

Note that this system is the one we had already obtained at the end of the previous computation. For completeness, let us write all its solutions, namely the system is equivalent to

$$\begin{cases} x = \frac{4w-2}{5} \\ y = \frac{-w-2}{5} \\ z = \frac{-3w-1}{5} \\ w \text{ arbitrary element of } \mathbb{R} \end{cases}.$$

Based on this example, let us formalize the procedure.

Definition 2.4.2. *An* elementary row operation *on a matrix consists in one of the following operations:*

(i) multiplying one row by a non-zero number,

(ii) adding (or subtracting) one row to another row,

(iii) interchanging two rows.

Definition 2.4.3. *Two matrices are* row equivalent *if one of them can be obtained from the other by performing a succession of row elementary operations. One writes $\mathcal{A} \sim \mathcal{B}$ if $\mathcal{A}$ and $\mathcal{B}$ are row equivalent.*

Proposition 2.4.4. *Let $\mathcal{A}, \mathcal{A}' \in M_{mn}(\mathbb{R})$ and let $B, B' \in \mathbb{R}^m$. If the augmented matrix $(\mathcal{A}, B)$ and $(\mathcal{A}', B')$, both belonging to $M_{m(n+1)}(\mathbb{R})$, are row equivalent then any solution $X \in \mathbb{R}^n$ of the system $\mathcal{A}X = B$ is a solution of the system $\mathcal{A}'X = B'$, and vice versa.*

The proof of this statement consists simply in checking that the systems of linear equations are equivalent at each step of the procedure. This can be inferred from the example shown above, and can be proved without any difficulty.

Definition 2.4.5. *A matrix is in* row echelon form *if it satisfies the following property: Whenever two successive rows do not consist entirely of* 0*, then the second row starts with a non-zero entry at least one step further than the first row. All the rows consisting entirely of* 0 *are at the bottom of the matrix.*

Examples 2.4.6. *The following matrices are in row echelon form:*

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

Theorem 2.4.7. *Every matrix is row equivalent to a matrix in row echelon form.*

Again, the proof is a simple abstraction of what has been performed on the above example. Note that checking this kind of properties is a good exercise for computer sciences. Indeed, the necessary iterative procedure can be easily implemented by some bootstrap arguments.

Definition 2.4.8. *The first non-zero coefficients occurring on the left of each row on a matrix in row echelon form are called* the leading coefficients.

Corollary 2.4.9. *Every matrix is row equivalent to a matrix in row echelon form and with all leading coefficients equal to* 1*.*

Proof. Use the previous theorem and divide each non-zero row by its leading coefficient. □

Corollary 2.4.10. *Each matrix is row equivalent to a matrix in row echelon form, with leading coefficients equal to* 1*, and with* 0*'s above each leading coefficient.*

We shall say that such matrices are *in the standard form*. Examples of such matrices are

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

Proof of Corollary 2.4.10. Starting from a matrix in row echelon form with all leading coefficients equal to 1, subtract sufficiently many times each row to the rows above it. Do this procedure iteratively, starting with the second row and going downward. □

Example 2.4.11. *Let us finally use this method on an example. In order to solve the linear system*

$$\begin{cases} 2x + y + 4z + w = 0 \\ -3x + 2y - 3z + w = 0 \\ x + y + z = 0 \end{cases},$$

we consider the augmented matrix and some elementary row operations:

$$\begin{pmatrix}2&1&4&1&0\\-3&2&-3&1&0\\1&1&1&0&0\end{pmatrix} \sim \begin{pmatrix}1&1&1&0&0\\0&-1&2&1&0\\0&5&0&1&0\end{pmatrix} \sim \begin{pmatrix}1&1&1&0&0\\0&1&-2&-1&0\\0&0&10&6&0\end{pmatrix} \sim \begin{pmatrix}1&0&0&-4/5&0\\0&1&0&1/5&0\\0&0&1&3/5&0\end{pmatrix}.$$

Then, we immediately infer from the last matrix the general solution

$$\begin{cases} x = 4/5\ w \\ y = -1/5\ w \\ z = -3/5\ w \\ w\ \textit{arbitrary} \end{cases}.$$

Note that adding the last column in the augmented matrix was not useful in this special case.

Note that it is only when the augmented matrix is in the standard form that the solutions of the linear system can be written down very easily. This method for solving linear system of equations is often call *Gauss elimination* or *Gauss-Jordan elimination*[2]. However, Chinese people were apparently using this method already 2000 years ago, see Chapter 1.2 of Bretscher's book[3] for details...

2.5 Elementary matrices

In this section we construct some very simple matrices and show how they can be used in conjunction with Gauss elimination. For $r, s \in \{1, \dots, m\}$ let $I_{rs} \in M_m(\mathbb{R})$ be the matrix whose rs-component is 1 and all the other ones are equal to 0. More precisely one has

$$(I_{rs})_{ij} = 1 \text{ if } i = r \text{ and } j = s, \quad (I_{rs})_{ij} = 0 \text{ otherwise.}$$

These matrices satisfy the relation

$$I_{rs} I_{r's'} = \begin{cases} I_{rs'} & \text{if } s = r' \\ \mathcal{O} & \text{if } s \neq r' \end{cases}.$$

See Exercise 2.21 for the proof of this statement.

Definition 2.5.1. *The following matrices are called* elementary matrices*:*

(i) $\mathbf{1}_m - I_{rr} + cI_{rr}$*, for* $c \neq 0$*,*

(ii) $(\mathbf{1}_m + I_{rs} + I_{sr} - I_{rr} - I_{ss})$*, for* $r \neq s$*,*

(iii) $(\mathbf{1}_m + cI_{rs})$*, for* $r \neq s$ *and* $c \neq 0$*.*

Lemma 2.5.2. *(i) Each elementary matrix is invertible, and its inverse is again an elementary matrix.*

(ii) If $\mathcal{A} \in M_{mn}(\mathbb{R})$*, all elementary row operations on* $\mathcal{A}$ *can be obtained by applying successively elementary matrices on the left of* $\mathcal{A}$*.*

The proof of these statements are provided in Exercises 2.14 and 2.21. Note that the second statements means that if $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B}_1, \dots, \mathcal{B}_p$ are elementary matrices, then $\mathcal{B}_p\mathcal{B}_{p-1} \dots \mathcal{B}_1\mathcal{A}$ is row equivalent to $\mathcal{A}$.

Observation 2.5.3. *Assume that* $\mathcal{B} \in M_m(\mathbb{R})$ *is a square matrix with its last row entirely filled with* 0*, then* $\mathcal{B}$ *is not invertible. Indeed, with the notation introduced in* (2.2.3)*, the assumption means that* $\mathcal{B}_m = {}^t\mathbf{0}$*. Now, by absurd let us assume that* $\mathcal{A} \in M_m(\mathbb{R})$ *is an inverse for* $\mathcal{B}$*, or equivalently that* $\mathcal{B}\mathcal{A} = \mathbf{1}_m$*. Then, since equation*

[2] Johann Carl Friedrich Gauss: 30 April 1777 – 23 February 1855; Wilhelm Jordan: 1 March 1842 – 17 April 1899.

[3] O. Bretscher, *Linear Algebra with Applications*, International Edition, Prentice Hall, 2008.

(2.2.4) *would hold for this product, one would have* $(\mathbf{1}_m)_{ik} = \mathcal{B}_i A^k$*, and in particular for* $i = k = m$ *one would have* $1 = \mathcal{B}_m \mathcal{A}^m$*, which is impossible since* $\mathcal{B}_m$ *is made only of* 0*. Thus, one concludes that there does not exist any inverse for* $\mathcal{B}$*, or equivalently that* $\mathcal{B}$ *is not invertible.*

In the next statement, we provide information about invertibility of square matrices.

Theorem 2.5.4. *(i) Let* $\mathcal{A}, \mathcal{A}' \in M_m(\mathbb{R})$ *be row equivalent. Then* $\mathcal{A}$ *is invertible if and only if* $\mathcal{A}'$ *is invertible,*

(ii) Let $\mathcal{A} \in M_m(\mathbb{R})$ *be upper triangular with non-zero diagonal elements. Then* $\mathcal{A}$ *is invertible,*

(iii) Any $\mathcal{A} \in M_m(\mathbb{R})$ *is invertible if and only if* $\mathcal{A}$ *is row equivalent to* $\mathbf{1}_m$*.*

Proof. (i) This part of the proof is provided in Exercise 2.22.

(ii) Observe first that an upper triangular matrix is already in row echelon form. Then by dividing each row by its leading term one obtains that $\mathcal{A}$ is row equivalent to a matrix in row echelon form and with 1 on its diagonal. Then, by subtracting each row coherently, starting with the second row and going downward, one obtains that $\mathcal{A}$ is row equivalent to $\mathbf{1}_m$. Since $\mathbf{1}_m$ is invertible, it follows from the point (i) that $\mathcal{A}$ is invertible as well.

(iii) $\Longleftarrow$: If $\mathcal{A}$ is row equivalent to $\mathbf{1}_m$ it follows from (i) that $\mathcal{A}$ is invertible. $\Longrightarrow$: By Corollary 2.4.10 we know that $\mathcal{A}$ is row equivalent to a $m \times m$ matrix $\mathcal{B}$ in the standard form. Since $\mathcal{B}$ is a square matrix, it follows that either $\mathcal{B}$ is equal to $\mathbf{1}_m$ or $\mathcal{B}$ has at least its last row filled only with 0. Note that in the former case $\mathcal{B}$ is invertible while in the second case $\mathcal{B}$ is not invertible, see Observation 2.5.3. However, since $\mathcal{A}$ is invertible and row equivalent to $\mathcal{B}$, it follows from (i) that $\mathcal{B}$ is invertible as well, and therefore $\mathcal{B}$ has to be the identity matrix. □

Corollary 2.5.5. *Any invertible* $m \times m$ *matrix can be expressed as a product of elementary matrices.*

Proof. This statement directly follows from the point (iii) of the previous theorem. Indeed, if $\mathcal{B}_p \mathcal{B}_{p-1} \dots \mathcal{B}_1 \mathcal{A} = \mathbf{1}_m$ with each $\mathcal{B}_j$ an elementary matrix, then

$$\mathcal{A} = \mathcal{B}_1^{-1} \mathcal{B}_2^{-1} \dots \mathcal{B}_p^{-1} \mathbf{1}_m = \mathcal{B}_1^{-1} \mathcal{B}_2^{-1} \dots \mathcal{B}_p^{-1},$$

which proves the statement. □

Remark 2.5.6. *It will be useful to observe that if* $\mathcal{B}_p \mathcal{B}_{p-1} \dots \mathcal{B}_1 \mathcal{A} = \mathbf{1}_m$ *for some elementary matrices* $\mathcal{B}_j$ *then*

$$\mathcal{A}^{-1} = \mathcal{B}_p \mathcal{B}_{p-1} \dots \mathcal{B}_1.$$

This observation directly leads to a convenient method for finding the inverse of a matrix $\mathcal{A}$*. Indeed, if* $\mathcal{A} \in M_n(\mathbb{R})$*, by considering the augmented matrix* $(\mathcal{A}, \mathbf{1}_n)$ *with* n *rows but* $2n$ *columns, and by performing elementary row operations such that* $\mathcal{A}$ *is transformed into the matrix* $\mathbf{1}_n$*, then the second part of the matrix will be equal to* $\mathcal{A}^{-1}$*. In other words, one obtains that* $(\mathcal{A}, \mathbf{1}_n)$ *is row equivalent to* $(\mathbf{1}_n, \mathcal{A}^{-1})$*.*

2.6 Exercises

Exercise 2.1. *Let us consider*

$$\mathcal{A} = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{pmatrix} \qquad \text{and} \qquad \mathcal{B} = \begin{pmatrix} -1 & 5 & -2 \\ 1 & 1 & -1 \end{pmatrix}.$$

Compute $\mathcal{A}+\mathcal{B}$, $\mathcal{A}-2\mathcal{B}$, *and* ${}^t\mathcal{A}$.

Exercise 2.2. *Write the proofs for Properties 2.1.10.*

Exercise 2.3. *Let* $\mathcal{A} \in M_{mn}(\mathbb{R})$. *Show that* $\mathbf{1}_m\,\mathcal{A} = \mathcal{A} = \mathcal{A}\mathbf{1}_n$.

Exercise 2.4. *One says that a matrix* $\mathcal{A} \in M_n(\mathbb{R})$ *is* symmetric *if* ${}^t\mathcal{A} = \mathcal{A}$ *and is* skew-symmetric *if* ${}^t\mathcal{A} = -\mathcal{A}$. *Show that for an arbitrary matrix* $\mathcal{A} \in M_n(\mathbb{R})$, *the matrix* $\mathcal{A} + {}^t\mathcal{A}$ *is symmetric while the matrix* $A - {}^t\mathcal{A}$ *is skew-symmetric.*

Exercise 2.5. *Let* $\mathcal{A} \in M_n(\mathbb{R})$.

1. *If* $\mathcal{A}^2 = \mathcal{O}$, *show that* $\mathbf{1}_n - \mathcal{A}$ *is invertible.*
2. *More generally, if* $\mathcal{A}$ *is nilpotent, show that* $\mathbf{1}_n - \mathcal{A}$ *is invertible.*
3. *Suppose that* $\mathcal{A}^2 + 2\mathcal{A} + \mathbf{1}_n = \mathcal{O}$. *Show that* $\mathcal{A}$ *is invertible.*

Exercise 2.6. *If* $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ *are two upper triangular matrices, show that the product* $\mathcal{A}\mathcal{B}$ *is also an upper triangular matrix.*

Exercise 2.7.

1. *Find some* $\mathcal{A} \in M_2(\mathbb{R})$ *such that* $\mathcal{A}^2 = -\mathbf{1}_2$.
2. *Determine all* $\mathcal{A} \in M_2(\mathbb{R})$ *such that* $\mathcal{A}^2 = \mathcal{O}$.

Exercise 2.8. *Let* a, b *be real numbers and let*

$$\mathcal{A} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathcal{B} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

What is $\mathcal{A}\mathcal{B}$ *? Compute* $\mathcal{A}^2$ *and* $\mathcal{A}^3$. *What is* $\mathcal{A}^m$ *for an arbitrary integer* m, *and how to prove it ?*

Exercise 2.9. *One says that a matrix* $\mathcal{A} \in M_n(\mathbb{R})$ *is orthogonal if* ${}^t\mathcal{A} = \mathcal{A}^{-1}$, *or equivalently if* ${}^t\mathcal{A}\mathcal{A} = \mathbf{1}_n$. *Show that if* $\mathcal{A} \in M_n(\mathbb{R})$ *is an orthogonal matrix, then*

1. $\|\mathcal{A}X\| = \|X\|$ *for any* $X \in \mathbb{R}^n$,
2. $(\mathcal{A}X) \cdot (\mathcal{A}Y) = X \cdot Y$ *for any* $X, Y \in \mathbb{R}^n$.

In other words, orthogonal matrices preserve lengths and angles between vectors of $\mathbb{R}^n$.

Exercise 2.10. *A special type of 2×2 matrices represents rotations in the plane. For arbitrary $\theta \in \mathbb{R}$, consider the matrix*

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

1. *Show that for arbitrary θ_1, θ_2 one has $R(\theta_1)\,R(\theta_2) = R(\theta_2)\,R(\theta_1)$,*
2. *Show that for arbitrary θ_1, θ_2 one has $R(\theta_1)\,R(\theta_2) = R(\theta_1 + \theta_2)$,*
3. *Show that for any θ, the matrix $R(\theta)$ has an inverse and write down this inverse.*

Exercise 2.11. *For any $\theta \in \mathbb{R}$, recall that the matrix*

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

represents a rotation by θ in $\mathbb{R}^2$.

1. *For ${}^tX = (1, 2)$, what are its coordinates after a rotation of $\pi/4$?*
2. *For ${}^tY = (-1, 3)$, what are its coordinates after a rotation of $\pi/2$?*

Draw a picture of your results.

Exercise 2.12. *Let*

$$\mathcal{A} = \begin{pmatrix} 2 & 3 & -1 & 1 \\ 1 & 4 & 2 & -2 \\ -1 & 1 & 3 & -5 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

and let $\mathcal{U}$ be one of the matrices shown below. Compute $\mathcal{U}\mathcal{A}$.

$$a)\ \mathcal{U} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad b)\ \mathcal{U} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad c)\ \mathcal{U} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$d)\ \mathcal{U} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad e)\ \mathcal{U} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad f)\ \mathcal{U} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Exercise 2.13. *Do the same exercise with the following matrices $\mathcal{U}$ and $\mathcal{A}$ as above:*

$$a)\ \mathcal{U} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad b)\ \mathcal{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad c)\ \mathcal{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{pmatrix} \quad d)\ \mathcal{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Exercise 2.14. *Let $\mathcal{A} \in M_{mn}(\mathbb{R})$. For $r \in \{1, \dots, m\}$ and $s \in \{1, \dots, m\}$, let $I_{rs} \in M_m(\mathbb{R})$ be the matrix whose rs-component is 1 and all the other ones are equal to 0. Answer the following questions with words :*

1. *What is $I_{rs}\mathcal{A}$?*
2. *For $r \neq s$, what is $(I_{rs} + I_{sr})\mathcal{A}$?*
3. *For $r \neq s$, what is $(\mathbf{1}_m + I_{rs} + I_{sr} - I_{rr} - I_{ss})\mathcal{A}$?*
4. *For $r \neq s$, what is $(\mathbf{1}_m + cI_{rs})\mathcal{A}$, for some $c \in \mathbb{R}$?*

Exercise 2.15. *Find a non-trivial solution for each of the following systems of equations.*

$$a) \qquad \begin{aligned} 2x - 3y + 4z &= 0 \\ 3x + y + z &= 0 \end{aligned}$$

$$b) \qquad \begin{aligned} 2x + y + 4z + w &= 0 \\ -3x + 2y - 3z + w &= 0 \\ x + y + z &= 0 \end{aligned}$$

$$c) \qquad \begin{aligned} -2x + 3y + z + 4w &= 0 \\ x + y + 2z + 3w &= 0 \\ 2x + y + z - 2w &= 0 \end{aligned}$$

Exercise 2.16. *Let $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $B \in \mathbb{R}^m$.*

1. *Assume that $X \in \mathbb{R}^n$ is a solution of $\mathcal{A}X = \mathbf{0}$. Show that for any $c \in \mathbb{R}$, the vector cX is also a solution of this equation.*
2. *Assume that $X, X' \in \mathbb{R}^n$ are solutions of the equations $\mathcal{A}X = \mathbf{0}$ and $\mathcal{A}X' = \mathbf{0}$. Show that $X + X'$ is also a solution of this equation.*
3. *Assume that $Y \in \mathbb{R}^n$ is a solution of the equation $\mathcal{A}Y = B$, and assume that $X \in \mathbb{R}^n$ is a solution of the homogeneous equation $\mathcal{A}X = \mathbf{0}$. Show that $Y + X$ is still a solution of the original equation.*

Exercise 2.17. *In each of the following cases find a row equivalent matrix in the standard form.*

$$a) \begin{pmatrix} 6 & 3 & -4 \\ -4 & 1 & -6 \\ 1 & 2 & -5 \end{pmatrix} \quad b) \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \quad c) \begin{pmatrix} 0 & 1 & 3 & -2 \\ 2 & 1 & -4 & 3 \\ 2 & 3 & 2 & -1 \end{pmatrix} \quad d) \begin{pmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & -2 & 3 \\ 3 & 6 & 2 & -6 & 5 \end{pmatrix}$$

Exercise 2.18. *Find all vectors in* $\mathbb{R}^4$ *which are perpendicular to the vectors*

$$^t(1,1,1,1), \quad ^t(1,2,3,4), \quad ^t(1,9,9,7)$$

Exercise 2.19. *By using Gauss elimination, find all solution for the following systems:*

$$a) \qquad \begin{aligned} x+y-2z &= 5 \\ 2x+3y+4z &= 2 \end{aligned}$$

$$b) \qquad \begin{aligned} x_3+x_4 &= 0 \\ x_2+x_3 &= 0 \\ x_1+x_2 &= 0 \\ x_1+x_4 &= 0 \end{aligned}$$

$$c) \qquad \begin{aligned} x_1+2x_2+2x_4+3x_5 &= 0 \\ x_3+3x_4+2x_5 &= 0 \\ x_3+4x_4-x_5 &= 0 \\ x_5 &= 0 \end{aligned}$$

Exercise 2.20. *Find a polynomial of degree* 3 *whose graph goes through the points* $(0,-1)$, $(1,-1)$, $(-1,-5)$ *and* $(2,1)$.

Exercise 2.21. *For* $r \in \{1,\dots,m\}$ *and* $s \in \{1,\dots,m\}$, *let* $I_{rs} \in M_m(\mathbb{R})$ *be the matrix whose* rs*-component is* 1 *and all the other ones are equal to* 0. *First show that if* $r, s, r', s' \in \{1,\dots,m\}$ *then*

$$I_{rs} I_{r's'} = \begin{cases} I_{rs'} & \text{if } s = r' \\ \mathcal{O} & \text{if } s \neq r' \end{cases}$$

Then, for $c \neq 0$, *consider the following 3 types of matrices :*

1. $\mathbf{1}_m - I_{rr} + cI_{rr}$, *the matrix obtained from the identity matrix by multiplying the* r*-th diagonal component by* c,
2. *For* $r \neq s$, $(\mathbf{1}_m + I_{rs} + I_{sr} - I_{rr} - I_{ss})$, *the matrix obtained from the identity matrix by interchanging the* r*-th row with the* s*-th row,*
3. *For* $r \neq s$, $(\mathbf{1}_m + cI_{rs})$, *the matrix having the* rs*-th component equal to* c, *all other components* 0 *except the diagonal components which are equal to* 1.

Show that these matrices are invertible and exhibit their inverse. If $\mathcal{A} \in M_{mn}(\mathbb{R})$, *show that multiplying the matrix* $\mathcal{A}$ *on the left by one of these matrices corresponds to one of the elementary row operations. For that reason, these matrices are called* **elementary matrices**.

Exercise 2.22. *Let $\mathcal{A}, \mathcal{A}' \in M_n(\mathbb{R})$ be row equivalent. With the help of the previous exercise, prove the following statements : $\mathcal{A}$ is invertible if and only if $\mathcal{A}'$ is invertible.*

Exercise 2.23. *By using elementary row operations, find the inverse for the following matrices :*

$$a) \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 2 & 7 \end{pmatrix} \quad b) \begin{pmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{pmatrix} \quad c) \begin{pmatrix} 2 & 4 & 3 \\ -1 & 3 & 0 \\ 0 & 2 & 1 \end{pmatrix}$$

Exercise 2.24. *Consider the equation*

$$\begin{aligned} x + 2y + 3z &= 4 \\ x + ky + 4z &= 6 \\ x + 2y + (k+2)z &= 6 \end{aligned}$$

where k is an arbitrary constant.

1. *For which values of k does this system have a unique solution ?*
2. *For which values of k does this system have no solution ?*
3. *For which values of k does this system have infinitely many solutions ?*

Exercise 2.25. *A conic is a curve in $\mathbb{R}^2$ that can be described by an equation of the form*

$$f(x, y) = c_1 + c_2x + c_3y + c_4x^2 + c_5xy + c_6y^2 = 0,$$

where at least one of the coefficients c_i is non-zero. Find the conic passing through the following points.

i) $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$.

ii) $(0, 0)$, $(1, 0)$, $(2, 0)$, $(3, 0)$, $(1, 1)$.

Exercise 2.26. *Let $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $X = {}^t(x_1, \ldots, x_n) \in \mathbb{R}^n$. The columns of $\mathcal{A}$ are denoted by $\mathcal{A}^1, \ldots, \mathcal{A}^n$, while the rows of $\mathcal{A}$ are denoted by $\mathcal{A}_1, \ldots, \mathcal{A}_m$. Show that the following three statements are equivalent :*

1. $\mathcal{A}X = \mathbf{0}$,
2. *the vector X is perpendicular to the vector ${}^t\mathcal{A}_j$, for each $j \in \{1, \ldots, m\}$,*
3. *the following linear relation holds :*

$$x_1\mathcal{A}^1 + x_2\mathcal{A}^2 + \cdots + x_n\mathcal{A}^n = \mathbf{0}.$$

Exercise 2.27. *By using elementary row operations, find the inverse for the following matrices :*

$$a) \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 8 & 2 \end{pmatrix} \qquad b) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

Exercise 2.28. *For which values of the parameter k is the following matrix invertible:*

$$\begin{pmatrix} 4 & 3-k \\ 1-k & 2 \end{pmatrix}$$

Exercise 2.29. *To gauge the complexity of a computational task, one can count the number of elementary operations (additions, subtractions, multiplications and divisions) required. For a rough count, one can consider multiplication and divisions only, referring to those jointly as* multiplicative operations. *Start by considering a 2 by 2 invertible matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and check that 8 multiplicative operations are necessary for inverting this matrix by using the Gauss elimination technique.*

(i) How many multiplicative operations are necessary for inverting a 3×3 matrix by the same technique ?

(ii) What about a $n \times n$ matrix ?

(iii) If a very slow computer needs 1 second to invert a 3×3 matrix, how long will it take to invert a 12×12 matrix ?

Exercise 2.30. *Write if the following statements are "true" or "false". Justify briefly your answer, or give a counterexample.*

1. *If $\mathcal{A}$ and $\mathcal{B}$ are symmetric matrices, then $\mathcal{A} + \mathcal{B}$ is symmetric,*
2. *If $\mathcal{A}$ is symmetric and $\mathcal{A} \neq \mathcal{O}$, then $\mathcal{A}$ is invertible,*
3. *If $\mathcal{AB} = \mathcal{O}$, then either $\mathcal{A}$ or $\mathcal{B}$ is the matrix $\mathcal{O}$,*
4. *If $\mathcal{A}^2 = \mathbf{1}$, then $\mathcal{A}$ is invertible,*
5. *If $\mathcal{A}, \mathcal{B}$ are invertible matrices, then $\mathcal{BA}$ is an invertible matrix,*
6. *If $\mathcal{A} \in M_n(\mathbb{R})$, $B \in \mathbb{R}^n$ with $B \neq \mathbf{0}$, and if X and X' satisfy $\mathcal{A}X = B$ and $\mathcal{A}X' = B$, then $(X + X')$ satisfies the same equation,*
7. *If $\mathcal{A}$ is diagonal and if $\mathcal{B}$ is an arbitrary matrix, then the product $\mathcal{AB}$ is diagonal,*
8. *There exists an invertible matrix $\mathcal{A}$ such that $\mathcal{A}^{-1} = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 0 \end{smallmatrix}\right)$,*
9. *Every matrices can be expressed as the product of elementary matrices,*
10. *$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is an orthogonal matrix.*

Chapter 3

Vector spaces

In this Chapter, we provide an abstract framework which encompasses what we have seen on $\mathbb{R}^n$ and for $M_{mn}(\mathbb{R})$.

3.1 Abstract definition

Before introducing the abstract notion of a vector space, let us make the following observation. For any $X, Y, Z \in \mathbb{R}^n$, for any $\mathcal{A}, \mathcal{B}, \mathcal{C} \in M_{mn}(\mathbb{R})$ and for any $\lambda, \mu \in \mathbb{R}$ one has

(i) $(X+Y)+Z = X+(Y+Z)$,

(ii) $X+Y = Y+X$,

(iii) $X+\mathbf{0} = \mathbf{0}+X = X$,

(iv) $X-X = \mathbf{0}$,

(v) $1X = X$,

(vi) $\lambda(X+Y) = \lambda X + \lambda Y$,

(vii) $(\lambda+\mu)X = \lambda X + \mu X$,

(viii) $(\lambda\mu)X = \lambda(\mu X)$.

(i) $(\mathcal{A}+\mathcal{B})+\mathcal{C} = \mathcal{A}+(\mathcal{B}+\mathcal{C})$,

(ii) $\mathcal{A}+\mathcal{B} = \mathcal{B}+\mathcal{A}$,

(iii) $\mathcal{A}+\mathcal{O} = \mathcal{O}+\mathcal{A} = \mathcal{A}$,

(iv) $\mathcal{A}-\mathcal{A} = \mathcal{O}$,

(v) $1\mathcal{A} = \mathcal{A}$,

(vi) $\lambda(\mathcal{A}+\mathcal{B}) = \lambda\mathcal{A} + \lambda\mathcal{B}$,

(vii) $(\lambda+\mu)\mathcal{A} = \lambda\mathcal{A} + \mu\mathcal{A}$,

(viii) $(\lambda\mu)\mathcal{A} = \lambda(\mu\mathcal{A})$.

Note that these properties are borrowed from Chapter 1 and 2 respectively. Another example which would satisfy the same properties is provided by the set of real functions defined on $\mathbb{R}$, together with the addition of such functions and with the multiplication by a scalar. In this case, the element $\mathbf{0}$ (or $\mathcal{O}$) is simply the function which is equal to 0 at any point of $\mathbb{R}$.

Question: Can one give an abstract definition for these rules ?

In the first definition, we give a more general framework in which λ and μ live. You can always think about $\mathbb{R}$ as the main example for the following definition.

Definition 3.1.1. A field[1] $(\mathbb{F}, +, \cdot)$ *is a set* $\mathbb{F}$ *endowed with two operations* $+$ *and* $\cdot$ *such that for any* $\lambda, \mu, \nu \in \mathbb{F}$ *one has*

(i) $\lambda + \mu \in \mathbb{F}$ *and* $\lambda \cdot \mu \in \mathbb{F}$, *(internal operations)*

(ii) $(\lambda + \mu) + \nu = \lambda + (\mu + \nu)$ *and* $(\lambda \cdot \mu) \cdot \nu = \lambda \cdot (\mu \cdot \nu)$, *(associativity)*

(iii) $\lambda + \mu = \mu + \lambda$ *and* $\lambda \cdot \mu = \mu \cdot \lambda$, *(commutativity)*

(iv) There exist $0, 1 \in \mathbb{F}$ *such that* $\lambda + 0 = \lambda$ *and* $1 \cdot \lambda = \lambda$, *(existence of identity elements)*

(v) There exists $-\lambda \in \mathbb{F}$ *such that* $\lambda + (-\lambda) = 0$, *and if* $\lambda \neq 0$ *there exists* $\lambda^{-1} \in \mathbb{F}$ *such that* $\lambda \cdot \lambda^{-1} = 1$, *(existence of inverse elements)*

(vi) $\lambda \cdot (\mu + \nu) = \lambda \cdot \mu + \lambda \cdot \nu$. *(distributivity)*

Note that for simplicity, one usually writes $\lambda - \mu$ instead of $\lambda + (-\mu)$ and λ/μ instead of $\lambda \cdot \mu^{-1}$.

Example 3.1.2. *Some examples of fields are* $(\mathbb{R}, +, \cdot)$ *the set of real numbers together with the usual addition and multiplication,* $(\mathbb{Q}, +, \cdot)$ *the set of fractional numbers* $\mathbb{Q} = \big\{a/b \mid a, b \in \mathbb{Z} \text{ with } b \neq 0\big\}$ *together with the usual addition and multiplication,* $(\mathbb{C}, +, \cdot)$ *the set of complex numbers together with its addition and multiplication (as we shall see at the end of this course). Note that for simplicity, one usually writes* $\mathbb{R}, \mathbb{Q}, \mathbb{C}$*, the other two operations being implicit.*

Let us provide a more general framework for the elements of X, Y, Z or $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and for the properties stated in the table above. However, you can always think about $\mathbb{R}^n$ or $M_{mn}(\mathbb{R})$ as the main examples for the following definition. Note that two slightly different fonts are used for the two different multiplications and for the two different additions.

Definition 3.1.3. A vector space over a field $(\mathbb{F}, +, \cdot)$ *consists in a set* V *endowed with two operations* $\boldsymbol{+} : V \times V \to V$ *and* $\boldsymbol{\cdot} : \mathbb{F} \times V \to V$ *such that if* $X, Y, Z \in V$ *and* $\lambda, \mu \in \mathbb{F}$ *the following properties are satisfied:*

(i) $(X \boldsymbol{+} Y) \boldsymbol{+} Z = X \boldsymbol{+} (Y \boldsymbol{+} Z)$,

(ii) $X \boldsymbol{+} Y = Y \boldsymbol{+} X$,

(iii) There exists (a unique) $\mathbf{0} \in V$ *such that* $X \boldsymbol{+} \mathbf{0} = \mathbf{0} \boldsymbol{+} X = X$,

[1] Without explaining this notion, but tacitly we shall work only with fields of characteristic 0, like $\mathbb{R}$, $\mathbb{Q}$ or $\mathbb{C}$.

(iv) *For all* $X \in V$ *there exists* $-X \in V$ *such that* $X + (-X) = \mathbf{0}$,

(v) $\lambda \cdot X \in V$ *and* $1 \cdot X = X$,

(vi) $\lambda \cdot (X + Y) = \lambda \cdot X + \lambda \cdot Y$,

(vii) $(\lambda + \mu) \cdot X = \lambda \cdot X + \mu \cdot X$,

(viii) $(\lambda \cdot \mu) \cdot X = \lambda \cdot (\mu \cdot X)$.

Before providing some examples, let us just mention a consequence of the previous conditions, namely $0 \cdot X = \mathbf{0}$ for any $X \in V$. Indeed, for any $X \in V$ one has

$$X = 1 \cdot X = (1 + 0) \cdot X = 1 \cdot X + 0 \cdot X = X + 0 \cdot X,$$

from which one infers that $0 \cdot X = \mathbf{0}$. Let us also note that whenever the field $\mathbb{F}$ consists in $\mathbb{R}$, one simply says *a real vector space* instead of a vector space over the field $\mathbb{R}$.

Examples 3.1.4. *(i)* $\mathbb{F} = \mathbb{R}$ *and* $V = \mathbb{R}^n$ *with the addition and multiplication by a scalar, as introduced in Chapter 1,*

(ii) $\mathbb{F} = \mathbb{R}$ *and* $V = M_{mn}(\mathbb{R})$ *with the addition and the multiplication by a scalar, as introduced in Chapter 2. More generally, for any field* $\mathbb{F}$ *the set* $M_{mn}(\mathbb{F})$*, defined exactly as* $M_{mn}(\mathbb{R})$*, is a vector space over* $\mathbb{F}$*,*

(iii) $\mathbb{F} = \mathbb{R}$ *and* V *is the set of real functions defined on* $\mathbb{R}$*, with the addition of functions and the multiplication by scalar,*

(iv) $\mathbb{F} = \mathbb{R}$ *and* $V = \{$*polynomial functions on* $\mathbb{R}\}$

$$= \Big\{ f : \mathbb{R} \to \mathbb{R} \mid f(x) = \sum_{j=0}^{m} a_j x^j \text{ with } a_j \in \mathbb{R} \Big\},$$

(v) $\mathbb{F} = \mathbb{R}$ *and* $V = \{f : \mathbb{R} \to \mathbb{R}$ *continuous* $\}$,

(vi) $\mathbb{F} = \mathbb{R}$ *and* $V = \{f : \mathbb{R} \to \mathbb{R}$ *differentiable* $\}$.

From now and for simplicity we shall no more use two different notations for the two different multiplications and for the two different additions. This simplification should not lead to any confusion. In addition, we shall simply write $\mathbb{F}$ for the field, instead of $(\mathbb{F}, +, \cdot)$, and the multiplication will be denoted without a dot; the sign $\cdot$ will be kept for the scalar product only.

Definition 3.1.5. *Let* V *be a vector space over* $\mathbb{F}$*, and let* W *be a (non-void) subset of* V*. Then* W *is* a subspace of V *if the following conditions are satisfied:*

(i) If $X, Y \in W$*, then* $X + Y \in W$,

(ii) If $X \in W$ and $\lambda \in \mathbb{F}$, then $\lambda X \in W$.

In other words, a subspace of a vector space V is a subset W of V which is stable for the two operations, *i.e.* the addition and the multiplication by a scalar. The next statement will be very useful when checking that a certain set is a vector space. Its proof will be provided in Exercise 3.5.

Lemma 3.1.6. *Any subspace W of a vector space V over a field $\mathbb{F}$ is itself a vector space over $\mathbb{F}$.*

Examples 3.1.7. *(i) $\{{}^t(x_1, x_2, \dots, x_{n-1}, 0) \mid x_j \in \mathbb{R}$ for any $j \in \{1, \dots, n-1\}\} \subset \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$,*

(ii) If $P, N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$, then $H_{P,N}$ is a subspace of $\mathbb{R}^n$ if and only if $\mathbf{0} \in H_{P,N}$. In particular, $H_{\mathbf{0},N}$ is a subspace of $\mathbb{R}^n$,

(iii) The set of upper triangular $n \times n$ matrices is a subspace of $M_n(\mathbb{R})$,

(iv) The set of $n \times n$ symmetric matrices is a subspace of $M_n(\mathbb{R})$.

The following statement deals with the intersection of subspaces or with the sum of subspaces of a vector space.

Lemma 3.1.8. *Let V be a vector space over a field $\mathbb{F}$, and let W_1, W_2 be two subspaces of V. Then*

(i) $W_1 \cap W_2 = \{X \in W_1$ and $X \in W_2\}$ is a subspace of V,

(ii) $W_1 + W_2 = \{X = X_1 + X_2 \mid X_1 \in W_1$ and $X_2 \in W_2\}$ is a subspace of V.

Proof. The proof consists in checking that the two conditions of Definition 3.1.5 are satisfied.

(i) If $X, Y \in W_j$ for $j \in \{1, 2\}$, then $X + Y \in W_j$ because W_j is a subspace. In particular, this implies that if $X, Y \in W_1 \cap W_2$, then $X + Y \in W_1 \cap W_2$. Similarly, in this case one also has $\lambda X \in W_1 \cap W_2$, since W_1 and W_2 are stable for the multiplication by a scalar.

(ii) If $X = X_1 + X_2$ and $Y = Y_1 + Y_2$ with $X_j, Y_j \in W_j$, then $X + Y = X_1 + X_2 + Y_1 + Y_2 = (X_1 + Y_1) + (X_2 + Y_2)$ with $(X_1 + Y_1) \in W_1$ and $(X_2 + Y_2) \in W_2$, which implies that $X + Y \in W_1 + W_2$. Similarly, in this case one also has $\lambda X = \lambda X_1 + \lambda X_2 \in W_1 + W_2$, since both W_1 and W_2 are stable for the multiplication by a scalar. □

3.2 Linear combinations

Let us start with a definition:

Definition 3.2.1. *Let V be a vector space over a field $\mathbb{F}$, and let $X_1, \dots, X_r \in V$. One sets*

$$\mathrm{Vect}(X_1, \dots, X_r) := \big\{\lambda_1 X_1 + \cdots + \lambda_r X_r \mid \lambda_j \in \mathbb{F} \text{ for } j \in \{1, \dots, r\}\big\},$$

and call this set the subspace of V *generated by* $X_1, \dots, X_r$.

Obviously, the first thing to do is to check that this set is indeed a subspace of V.

Lemma 3.2.2. *In the above setting,* $\mathrm{Vect}(X_1, \dots, X_r)$ *is a subspace of* V.

Proof. The proof consists in checking that both conditions of Definition 3.1.5 are satisfied. First of all, if $X = \lambda_1 X_1 + \cdots + \lambda_r X_r$ and $X' = \lambda_1' X_1 + \cdots + \lambda_r' X_r$, then

$$X + X' = \underbrace{(\lambda_1 + \lambda_1')}_{\in \mathbb{F}} X_1 + \cdots + \underbrace{(\lambda_r + \lambda_r')}_{\in \mathbb{F}} X_r \ \in \mathrm{Vect}(X_1, \dots, X_r).$$

Similarly, if $X = \lambda_1 X_1 + \cdots + \lambda_r X_r$ and $\lambda \in \mathbb{F}$, then

$$\lambda X = \underbrace{(\lambda\lambda_1)}_{\in \mathbb{F}} X_1 + \cdots + \underbrace{(\lambda\lambda_r)}_{\in \mathbb{F}} X_r \ \in \mathrm{Vect}(X_1, \dots, X_r).$$

Since both conditions are checked, it is thus a subspace of V. □

Since $\mathrm{Vect}(X_1, \dots, X_r)$ is a subspace of V, it was legitimate to call it as we did. Note that one also says that $\lambda_1, \dots, \lambda_r$ are *the coefficients* of *the linear combination* $\lambda_1 X_1 + \cdots + \lambda_r X_r$.

Remark 3.2.3. *If* $\mathrm{Vect}(X_1, \dots, X_r) = V$, *then one says that V is generated by the elements $X_1, \dots, X_r$, or that $\{X_1, \dots, X_r\}$ is a* generating family.

The following three examples are related to real vector spaces, as it is the case in most of the examples of these lecture notes.

Examples 3.2.4. *(i) Recall that $E_j = {}^t(0, \dots, 1, \dots, 0)$ with the entry 1 at the position j. Then $\{E_j\}_{j=1}^n \equiv \{E_1, E_2, \dots, E_n\}$ is a generating family for $\mathbb{R}^n$.*

(ii) If $N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$, then $\mathrm{Vect}(N)$ *is the line passing through $\mathbf{0}$ and having the direction N,* i.e. $\mathrm{Vect}(N) = L_{\mathbf{0},N}$, *with the $L_{\mathbf{0},N}$ defined in Definition 1.5.1.*

(iii) If $X, Y \in \mathbb{R}^3$ with $X \neq \mathbf{0}$, $Y \neq \mathbf{0}$, and $Y \neq \lambda X$ for any $\lambda \in \mathbb{R}$, then $\mathrm{Vect}(X, Y)$ *defines a plane in $\mathbb{R}^3$ passing through $\mathbf{0}$. In fact, it corresponds to the plane passing through the three points $\mathbf{0}, X, Y$, as seen in Exercise 1.18.*

Remark 3.2.5. *If $\mathbb{F} = \mathbb{R}$ and if one considers $X_1, \dots, X_r \in \mathbb{R}^n$, then one can set*

$$\mathrm{Box}(X_1, \dots, X_r) := \big\{\lambda_1 X_1 + \cdots + \lambda_r X_r \mid \lambda_j \in [0, 1] \text{ for } j \in \{1, \dots, r\}\big\}.$$

This is a subset of $\mathrm{Vect}(X_1, \dots, X_r)$, *called* the hyperbox *or* generalized box *generated by $X_1, \dots X_r$. Note that* $\mathrm{Box}(X_1, \dots, X_r)$ *is not a subspace. It is also easily observed that* $\mathrm{Box}(X_1)$ *corresponds to the segment between $\mathbf{0}$ and X_1 and that* $\mathrm{Box}(X_1, X_2)$ *corresponds to the parallelogram generated by X_1 and X_2, and with one apex at $\mathbf{0}$.*

3.3 Convex sets

In this section we consider only real vector spaces, *i.e.* the field $\mathbb{F}$ is equal to $\mathbb{R}$ for all vector spaces.

Definition 3.3.1. *Let S be a subset of a real vector space V. Then S is* convex *if for any $X, Y \in S$ and for any $t \in [0,1]$ one has*

$$X + t(Y - X) \equiv (1-t)X + tY \ \in S.$$

Examples 3.3.2. *(i) A ball is convex, but a doughnut is not convex,*

(ii) For any $X_1, \dots, X_r$ in a real vector space V, $\mathrm{Vect}(X_1, \dots, X_r)$ is convex. Indeed, if $X = \lambda_1 X_1 + \dots + \lambda_r X_r$ and $X' = \lambda_1' X_1 + \dots + \lambda_r' X_r$ with $\lambda_j, \lambda_j' \in \mathbb{R}$ then

$$(1-t)X + tX' = \underbrace{\big((1-t)\lambda_1 + t\lambda_1'\big)}_{\in \mathbb{R}} X_1 + \underbrace{\big((1-t)\lambda_r + t\lambda_r'\big)}_{\in \mathbb{R}} X_r \ \in \mathrm{Vect}(X_1, \dots, X_r),$$

(iii) For any $X_1, \dots, X_r$ in a real vector space V, $\mathrm{Box}(X_1, \dots, X_r)$ is convex. Indeed, in the framework of the previous example, observe that if $0 \le \lambda_j \le 1$ and $0 \le \lambda_j' \le 1$ then one has for any $t \in [0,1]$

$$0 \le (1-t)\lambda_j + t\lambda_j' \le (1-t)1 + t1 = 1.$$

As a consequence, one infers that

$$(1-t)X + tX' = \underbrace{\big((1-t)\lambda_1 + t\lambda_1'\big)}_{\in [0,1]} X_1 + \underbrace{\big((1-t)\lambda_r + t\lambda_r'\big)}_{\in [0,1]} X_r \ \in \mathrm{Box}(X_1, \dots, X_r).$$

Definition 3.3.3. *Let V be a real vector space, and let $X_1, \dots, X_r \in V$. We set*

$$\mathrm{CS}(X_1, \dots, X_r) := \big\{\lambda_1 X_1 + \dots + \lambda_r X_r \mid 0 \le \lambda_j \le 1 \text{ for } j \in \{1, \dots, r\} \\ \text{and } \lambda_1 + \lambda_2 + \dots + \lambda_r = 1\big\}$$

and call is the convex set *generated or spanned by $X_1, \dots, X_r$.*

Note that by definition, the following inclusions always hold

$$\mathrm{CS}(X_1, \dots, X_r) \subset \mathrm{Box}(X_1, \dots, X_r) \subset \mathrm{Vect}(X_1, \dots, X_r).$$

Example 3.3.4. *If $V = \mathbb{R}^n$, then $\mathrm{CS}(X_1)$ corresponds just to the point X_1, $\mathrm{CS}(X_1, X_2)$ corresponds to the segment between X_1 and X_2 while $\mathrm{CS}(X_1, X_2, X_3)$ corresponds to the triangle of apexes X_1, X_2 and X_3.*

Obviously, one has to show immediately the following statement:

Lemma 3.3.5. *If V is a real vector space and $X_1, \dots, X_r \in V$, then* $\mathrm{CS}(X_1, \dots, X_r)$ *is convex.*

In fact, $\mathrm{CS}(X_1, \dots, X_r)$ is the smallest convex set containing $X_1, \dots, X_r$.

Proof. Let $X = \lambda_1 X_1 + \dots + \lambda_r X_r$ with $0 \leq \lambda_j \leq 1$ and $\lambda_1 + \lambda_2 + \dots + \lambda_r = 1$, and let $X' = \lambda'_1 X_1 + \dots + \lambda'_r X_r$ with $0 \leq \lambda'_j \leq 1$ and $\lambda'_1 + \lambda'_2 + \dots + \lambda'_r = 1$. Then for any $t \in [0, 1]$ one has

$$(1-t)X + tX' = \sum_{j=1}^{r} \big((1-t)\lambda_j + t\lambda'_j\big) X_j$$

with $0 \leq (1-t)\lambda_j + t\lambda'_j \leq (1-t)1 + t1 = 1$ and

$$\sum_{j=1}^{r} \big((1-t)\lambda_j + t\lambda'_j\big) = (1-t)\sum_{j=1}^{r} \lambda_j + t \sum_{j=1}^{r} \lambda'_j = (1-t)1 + t1 = 1.$$

As a consequence, $(1-t)X + tX' \in \mathrm{CS}(X_1, \dots, X_r)$, which means that $\mathrm{CS}(X_1, \dots, X_r)$ is convex. □

3.4 Linear independence

The following definition will be of importance in the sequel.

Definition 3.4.1. *Let V be a vector space over a field $\mathbb{F}$, and let $X_1, \dots, X_r \in V$. The elements $X_1, \dots, X_r$ are* linearly dependent *if there exist $\lambda_1, \dots, \lambda_r \in \mathbb{F}$ not all equal to 0 such that*

$$\lambda_1 X_1 + \dots + \lambda_r X_r = \mathbf{0}. \tag{3.4.1}$$

The elements $X_1, \dots, X_r$ are said linearly independent *if there do not exist such scalars $\lambda_1, \dots, \lambda_r$.*

Note that alternatively, the vectors $X_1, \dots, X_r$ are linearly independent if whenever (3.4.1) is satisfied, then one must have $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$. In this case, one also says that the family $\{X_1, \dots, X_r\}$ is linearly independent.

Examples 3.4.2. *(i) For $V = \mathbb{R}^n$, the family $\{E_j\}_{j=1}^n$ is linearly independent,*

(ii) For $V = M_n(\mathbb{R})$, the family $\{I_{rs}\}_{r,s=1}^n$ of elementary matrices introduced in Section 2.5 is linearly independent,

(iii) For $V = \mathbb{R}^2$, the elements $\binom{1}{0}$, $\binom{0}{1}$, and $\binom{1}{1}$ are linearly dependent since

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

(iv) *Let f_1, f_2 be two continuous real functions on $\mathbb{R}$. In this case f_1, f_2 are linearly dependent if there exists $\lambda_1, \lambda_2 \in \mathbb{R}$ with $(\lambda_1, \lambda_2) \neq (0,0)$ such that $\lambda_1 f_1 + \lambda_2 f_2 = \mathbf{0}$, or more precisely*

$$\lambda_1 f_1(x) + \lambda_2 f_2(x) = 0 \quad \text{for all } x \in \mathbb{R}.$$

For example, if $f_1(x) = \cos(x)$ and $f_2(x) = \sin(x)$, then f_1 and f_2 are linearly independent even if $0\cos(0) + \lambda \sin(0) = 0$ for arbitrary $\lambda \in \mathbb{R}$.

Definition 3.4.3. *Let V be a vector space over a field $\mathbb{F}$, and let $X_1, \dots, X_r \in V$. If $\mathrm{Vect}(X_1, \dots, X_r) = V$ and if $X_1, \dots, X_r$ are linearly independent, then $\{X_1, \dots, X_r\}$ is called* a basis *for V. Alternatively, one also says that the family $\{X_1, \dots, X_r\}$ constitutes or forms a basis for V.*

Examples 3.4.4. *(i) $\{E_j\}_{j=1}^n$ forms a basis for $\mathbb{R}^n$,*

(ii) $\{I_{rs}\}_{r,s=1}^n$ forms a basis for $M_n(\mathbb{R})$,

(iii) $\{x \mapsto x^n\}_{n=0}^\infty$ forms a basis for the vector space of all polynomials on $\mathbb{R}$.

Let us consider a special case of the previous definition in the case $n = 2$. The content of the following lemma will be useful later on, and its proof will be provided in Exercise 3.9.

Lemma 3.4.5. *Let $\binom{a}{b}, \binom{c}{d} \in \mathbb{R}^2$, with $a, b, c, d \in \mathbb{R}$.*

(i) The two vectors are linearly independent if and only if $ad - bc \neq 0$,

(ii) If the two vectors are linearly independent, they form a basis of $\mathbb{R}^2$.

Given a basis of a vector space V, any point X can be expressed as a linear combinations of elements of this basis. More precisely, one sets:

Definition 3.4.6. *Let $\{X_1, \dots, X_r\}$ be a basis for a vector space V over $\mathbb{F}$. Then, for $X = \lambda_1 X_1 + \dots + \lambda_r X_r$ the coefficients $\{\lambda_1, \dots, \lambda_r\}$ are called* the coordinates of X with respect to the basis $\{X_1, \dots, X_r\}$ of V.

In order to speak about "the" coordinates, the following lemma is necessary.

Lemma 3.4.7. *The coordinates of a vector with respect to a basis are unique.*

Proof. Let $\{X_1, \dots, X_r\}$ be a basis, and assume that

$$X = \lambda_1 X_1 + \dots + \lambda_r X_r = \lambda_1' X_1 + \dots + \lambda_r' X_r.$$

It then follows that

$$X - X = \mathbf{0} = (\lambda_1 - \lambda_1')X_1 + \dots + (\lambda_r - \lambda_r')X_r.$$

By independence of $X_1, \dots, X_r$, it follows that $(\lambda_j - \lambda_j') = 0$ for all $j \in \{1, \dots, r\}$, which means that $\lambda_j = \lambda_j'$. Thus, the coordinates of X with respect to a basis are unique. □

3.5 Dimension

Question: Can one find 3 linearly independent elements in $\mathbb{R}^2$? For instance, if $A = \binom{1}{2}$, $B = \binom{-5}{7}$ and $C = \binom{10}{4}$, are they linearly independent vectors ? The answer is no, and there is no need to do any computation for getting this answer. Indeed, let us consider the more general setting provided by $A = \binom{a_1}{a_2}$, $B = \binom{b_1}{b_2}$ and $C = \binom{c_1}{c_2}$. Then one has

$$\lambda_1 \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \lambda_2 \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \lambda_3 \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \mathbf{0} \Longleftrightarrow \begin{cases} \lambda_1 a_1 + \lambda_2 b_1 + \lambda_3 c_1 = 0 \\ \lambda_1 a_2 + \lambda_2 b_2 + \lambda_3 c_2 = 0 \end{cases}.$$

Note that this corresponds to a system of two equations for the three unknowns λ_1, λ_2 and λ_3. As seen in Theorem 2.3.4, such a homogeneous system of equations has always a non-trivial solution, which means that there exists a non trivial solution for the corresponding equation 3.4.1. As a consequence, three vectors in $\mathbb{R}^2$ can never be independent.

More generally, one has:

Theorem 3.5.1. *Let $\{X_1, \dots, X_r\}$ be a basis of a vector space V over $\mathbb{F}$. Consider also $Y_1, \dots, Y_m \in V$ and assume that $m > r$. Then $Y_1, \dots, Y_m$ are linearly dependent.*

Note that if $m \leq r$, the statement does not imply that $Y_1, \dots, Y_m$ are linearly independent.

Now, in order to give the proof in its full generality, we need to extend the definition of $M_{mn}(\mathbb{R})$ to $M_{mn}(\mathbb{F})$, for an arbitrary field $\mathbb{F}$. In fact, since elements in a field can be added and multiplied, all definitions related to $M_{mn}(\mathbb{R})$ can be translated directly into the same definitions for $M_{mn}(\mathbb{F})$. The only modification is that any entry a_{ij} of a matrix $\mathcal{A}$ belongs to $\mathbb{F}$ instead of $\mathbb{R}$, and the multiplication of a matrix by a scalar $\lambda \in \mathbb{R}$ is now replaced by the multiplication by an element of $\mathbb{F}$. Then, most of the statements of Section 2 are valid (simply by replacing $\mathbb{R}$ by $\mathbb{F}$), and in particular Theorem 2.3.4 can be obtained in this more general context. This theorem is precisely the one required for the proof of the above statement.

Proof. Since $X_1, \dots, X_r$ generate V, there exists $a_{ij} \in \mathbb{F}$ such that

$$Y_j = a_{1j} X_1 + \cdots + a_{rj} X_r \quad \text{for any } j \in \{1, \dots, m\}.$$

Then, let us consider $\lambda_1, \dots, \lambda_m \in \mathbb{F}$ and observe that

$$\begin{aligned} & \lambda_1 Y_1 + \cdots + \lambda_m Y_m = \mathbf{0} \\ \Longleftrightarrow & \lambda_1 (a_{11} X_1 + \cdots + a_{r1} X_r) + \cdots + \lambda_m (a_{1m} X_1 + \cdots + a_{rm} X_r) = \mathbf{0} \\ \Longleftrightarrow & (\lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_m a_{1m}) X_1 + \cdots + (\lambda_1 a_{r1} + \dots \lambda_m a_{rm}) X_r = \mathbf{0} \end{aligned}$$

which is equivalent to the following expression, by linear independence of $X_1, \dots, X_r$:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Finally, since $m > r$, it follows from a simple adaptation of Theorem 2.3.4 to $M_n(\mathbb{F})$ that this system of equation always has a non-trivial solution. However, this means precisely that the elements $Y_1, \dots, Y_m$ are linearly dependent. □

Corollary 3.5.2. *Let V be a vector space and suppose that $\{X_1, \dots, X_n\}$ is a basis for V. Then any other basis for V also contains n elements.*

Proof. Let $\{Y_1, \dots, Y_m\}$ be a second basis for V. If $m > n$ then $Y_1, \dots, Y_m$ can not be linearly independent, by the previous theorem. Similarly, if $m < n$ then $X_1, \dots, X_n$ can not be linearly independent, also by the previous theorem. Since any basis is made of linearly independent vectors, one obtains that the only possibility is $m = n$. □

We now define a notion which has been implicitly used from the beginning for $\mathbb{R}^n$.

Definition 3.5.3. *Let V be a vector space over a field $\mathbb{F}$, and let $\{X_1, \dots, X_n\}$ be a basis for V. Then one says that V is* of dimension n*, since this number is independent of the choice of a particular basis for V. The dimension of the vector space V is denoted by* $\dim(V)$*.*

Remark 3.5.4. *In these lecture notes except in a few exercises, all vector spaces are of finite dimension. This fact is tacitly assumed in many statements later on, but note that vector spaces of infinite dimensions often appear in physics or in mathematics.*

Examples 3.5.5. *(i) $\mathbb{R}^n$ is of dimension n, $\mathbb{F}^n$ is also of dimension n,*

(ii) For any vector space V and any $X \in V$, the dimension of $\mathrm{Vect}(X)$ *is* 1*,*

(iii) Any plane in $\mathbb{R}^3$ (passing through the origin) is of dimension 2, while a line passing through the origin is of dimension 1.

The following result is often useful, when the dimension of the vector space is already known.

Lemma 3.5.6. *Let V be a vector space of dimension n, and let $X_1, \dots, X_n \in V$ be linearly independent. Then $\{X_1, \dots, X_n\}$ is a basis for V.*

Proof. One only has to show that $\mathrm{Vect}(X_1, \dots, X_n) = V$. By contradiction, assume that there exists $Y \in V$ such that $Y \notin \mathrm{Vect}(X_1, \dots, X_n)$. Then, the vectors $X_1, \dots, X_n, Y$ are linearly independent, and $\{X_1, \dots, X_n, Y\}$ would generate a basis for V of dimension $n + 1$, which is impossible by the previous Corollary. □

3.6 The rank of a matrix

Let $\mathcal{A} \in M_{mn}(\mathbb{F})$, and recall from Section 2.2 that the columns of $\mathcal{A}$ have been denoted by $\mathcal{A}^1, \dots, \mathcal{A}^n$. Each column is an element of $\mathbb{F}^m$ (a column vector with m entries in $\mathbb{F}$) and the family $\{\mathcal{A}^1, \dots, \mathcal{A}^n\}$ generates a subspace of $\mathbb{F}^m$, which we have denoted by

$\mathrm{Vect}(\mathcal{A}^1, \dots, \mathcal{A}^n)$. In this case this subspace is called *the column space*. Alternatively, the rows of $\mathcal{A}$ generate the subspace $\mathrm{Vect}({}^t\!\mathcal{A}_1, \dots, {}^t\!\mathcal{A}_m)$ of $\mathbb{F}^n$, which is called *the row space*. The dimension of the first subspace is called *the column rank*, while the dimension of the second subspace is called *the row rank*.

By what we have seen in the previous sections, the column rank corresponds to the maximal number of linearly independent columns, while the row rank corresponds to the maximal number of linearly independent rows. Our aim in this section is to study these numbers.

Lemma 3.6.1. *Elementary row operations do not change the row rank of a matrix.*

Proof. For this proof, it is sufficient to observe that

$$\begin{aligned} &\mathrm{Vect}({}^t\!\mathcal{A}_1, \dots, {}^t\!\mathcal{A}_j, \dots, {}^t\!\mathcal{A}_k, \dots, {}^t\!\mathcal{A}_m) \\ &= \mathrm{Vect}({}^t\!\mathcal{A}_1, \dots, {}^t\!\mathcal{A}_k, \dots, {}^t\!\mathcal{A}_j, \dots, {}^t\!\mathcal{A}_m) \\ &= \mathrm{Vect}({}^t\!\mathcal{A}_1, \dots, c\,{}^t\!\mathcal{A}_j, \dots, {}^t\!\mathcal{A}_k, \dots, {}^t\!\mathcal{A}_m) \\ &= \mathrm{Vect}({}^t\!\mathcal{A}_1, \dots, {}^t\!\mathcal{A}_j + c\,{}^t\!\mathcal{A}_k, \dots, {}^t\!\mathcal{A}_k, \dots, {}^t\!\mathcal{A}_m) \end{aligned}$$

for any $c \neq 0$. Since these subspaces are the same, their dimension coincide. □

Similarly, one can show that elementary row operations do not change the column rank of a matrix. Note that this proof is less easy since elementary row operations have been defined on rows, and not on columns.

Theorem 3.6.2. *For any $\mathcal{A} \in M_{mn}(\mathbb{F})$, the column rank and the row rank of $\mathcal{A}$ are equal.*

Thanks to this statement, it is sufficient to speak about *the rank of a matrix*, denoted by $\mathrm{rank}(\mathcal{A})$, there is no need to specify if it is the column rank or the row rank.

Proof. First of all, recall that the matrix $\mathcal{A}$ is row equivalent to a matrix $\mathcal{B}$ in the standard form, see Corollary 2.4.10. Since elementary row operations do not change the row rank or the column rank, the matrix $\mathcal{B}$ has the same row rank and column rank as the original matrix $\mathcal{A}$. Then, it is easily observed that the number of leading coefficients of $\mathcal{B}$ is equal to the number of linearly independent rows, but also to the number of linearly independent columns of $\mathcal{B}$. Therefore, the number of leading coefficients of $\mathcal{B}$ is equal to the row rank of $\mathcal{B}$ and to the column rank of $\mathcal{B}$. It follows that these two numbers are equal, and that the row rank of $\mathcal{A}$ and the column rank of $\mathcal{A}$ are also equal to this number. □

3.7 Exercises

Exercise 3.1. *Show that the following sets of elements of* $\mathbb{R}^3$ *form subspaces :*

i) $S_1 := \{{}^t(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$,

ii) $S_2 := \{{}^t(x, y, z) \in \mathbb{R}^3 \mid x = y \text{ and } 2y = z\}$,

iii) $S_3 := \{{}^t(x, y, z) \in \mathbb{R}^3 \mid x + y = 3z\}$.

Exercise 3.2. *Let* V *be a subspace of* $\mathbb{R}^n$*, and let* W *be the set of all elements of* $\mathbb{R}^n$ *which are perpendicular to all elements of* V*. Show that* W *itself is a subspace of* $\mathbb{R}^n$*. This subspace is often denoted by* $V^\perp$ *and called the orthogonal complement of* V *in* $\mathbb{R}^n$*.*

Exercise 3.3. *Let* $A_1, \ldots, A_r$ *be generators of a subspace* V *of* $\mathbb{R}^n$*. Let* W *be the set of all elements in* $\mathbb{R}^n$ *which are perpendicular to* $A_1, \ldots, A_r$*. Show that* $W = V^\perp$*.*

Exercise 3.4. *Show that the set of all real polynomials is a subspace of the vector space of all real and continuous functions on* $\mathbb{R}$*. Exhibit a generating family for this subspace.*

Exercise 3.5. *Let* V *be a vector space over a field* $\mathbb{F}$*. Show that any subspace of* V *is itself a vector space.*

Exercise 3.6. *Let* S *be a convex set in a real vector space* V*.*

i) *For* $\lambda \in \mathbb{R}$*, show that* λS *is a convex set in* V*, with* $\lambda S = \{\lambda X \mid X \in S\}$*.*

ii) *For* $Y \in V$*, show that* $S + Y$ *is a convex set in* V*, with* $S + Y = \{X + Y \mid X \in S\}$*.*

Exercise 3.7. *Show that the intersection of two convex sets is still convex.*

Exercise 3.8. *Show that the vectors* $\begin{pmatrix}1\\0\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\\0\end{pmatrix}$ *and* $\begin{pmatrix}1\\1\\1\end{pmatrix}$ *form a basis of* $\mathbb{R}^3$*.*

Exercise 3.9. *Let* $\binom{a}{b}, \binom{c}{d} \in \mathbb{R}^2$*. Show that these two vectors are linearly independent if and only if* $ad - bc \neq 0$*.*

Exercise 3.10. *Express the coordinates of* Y *in the basis generated by* X_1 *and* X_2 *:*

i) $Y = \begin{pmatrix}1\\0\end{pmatrix}$, $X_1 = \begin{pmatrix}1\\1\end{pmatrix}$ *and* $X_2 = \begin{pmatrix}0\\1\end{pmatrix}$,

ii) $Y = \begin{pmatrix}2\\1\end{pmatrix}$, $X_1 = \begin{pmatrix}1\\-1\end{pmatrix}$ *and* $X_2 = \begin{pmatrix}1\\1\end{pmatrix}$.

Exercise 3.11. *Let* $X_1, \ldots, X_r$ *be non-zero elements of* $\mathbb{R}^n$ *and assume that* $X_j \cdot X_k = 0$ *for each* $j \neq k$*. Show that these elements are linearly independent.*

Exercise 3.12. *Determine the dimension of the following subspaces:*

i) $S_1 := \{{}^t(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$, *ii)* $S_2 := \{^t(x, y, z) \in \mathbb{R}^3 \mid x = y \text{ and } 2y = z\}$,

iii) $S_3 := \{^t(x, y, z) \in \mathbb{R}^3 \mid x + y = 3z\}$.

Exercise 3.13. *Determine the rank of the following matrices :*

$$a)\ \begin{pmatrix} 6 & 3 & -4 \\ -4 & 1 & -6 \\ 1 & 2 & -5 \end{pmatrix} \quad b)\ \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \quad c)\ \begin{pmatrix} 0 & 1 & 3 & -2 \\ 2 & 1 & -4 & 3 \\ 2 & 3 & 2 & -1 \end{pmatrix} \quad d)\ \begin{pmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & -2 & 3 \\ 3 & 6 & 2 & -6 & 5 \end{pmatrix}$$

Exercise 3.14. *A* doubly stochastic matrix *is a* $n \times n$ *matrix* $\mathcal{A} = (a_{jk})$ *such that* $a_{jk} \in [0, 1]$ *and such that the sum of the elements of each line is equal to* 1*, as well as the sum of the elements of each column.*

(i) Show that the product of two doubly stochastic matrices is still a doubly stochastic matrix,

(ii) Show that the set of all doubly stochastic matrices is a convex set.

Chapter 4

Linear maps

Before concentrating on linear maps, we provide a more general setting.

4.1 General maps

We start with the general definition of a map between two sets, and introduce some notations.

Definition 4.1.1. *Let S, S' be two sets.* A map T from S to S' *is a rule which associates to each element of S an element of S'. The notation*

$$\mathrm{T} : S \ni X \mapsto \mathrm{T}(X) \in S'$$

will be used for such a map. If $X \in S$, then $\mathrm{T}(X) \in S'$ is called the image of X by T. *The set S is often called* the domain of T *and is also denoted by* Dom(T), *while*

$$\mathrm{T}(S) := \big\{\mathrm{T}(X) \mid X \in S\big\}$$

is often called the range of T *and is also denoted by* Ran(T).

Examples 4.1.2. *(i) The function $f : \mathbb{R} \ni x \mapsto f(x) = x^2 - 3x + 2 \in \mathbb{R}$ is a map from $\mathbb{R}$ to $\mathbb{R}$,*

(ii) Any $\mathcal{A} \in M_{mn}(\mathbb{R})$ defines a map $\mathrm{L}_{\mathcal{A}} : \mathbb{R}^n \to \mathbb{R}^m$ by $\mathrm{L}_{\mathcal{A}}(X) := \mathcal{A}X$ for any $X \in \mathbb{R}^n$. More generally, for any field $\mathbb{F}$ and any $\mathcal{A} \in M_{mn}(\mathbb{F})$, one defines a map $\mathrm{L}_{\mathcal{A}} : \mathbb{F}^n \to \mathbb{F}^m$ by $\mathrm{L}_{\mathcal{A}}(X) := \mathcal{A}X$ for any $X \in \mathbb{F}^n$,

(iii) The rule $\mathrm{F} : \mathbb{R}^3 \ni \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \mathrm{F}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x^2+y \\ x+y+z+3 \end{pmatrix} \in \mathbb{R}^2$ is a map,

(iv) Let $C^1(\mathbb{R}) := \{$continuous functions f on $\mathbb{R} \mid f'$ exists and is continuous $\}$ and let $C(\mathbb{R}) := \{$continuous functions f on $\mathbb{R}\}$. Then the following rule defines a map:

$$\mathrm{D} : C^1(\mathbb{R}) \ni f \mapsto \mathrm{D}f = f' \in C(\mathbb{R})$$

(v) *For any* $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ *one can define a map by*

$$\mathrm{T}_{\mathcal{A},\mathcal{B}} : M_n(\mathbb{R}) \ni X \mapsto \mathrm{T}_{\mathcal{A},\mathcal{B}}(X) = \mathcal{A}X + \mathcal{B} \in M_n(\mathbb{R}),$$

(vi) *The function* $g : \mathbb{R}^* \ni x \mapsto g(x) = \frac{3x-2}{x} \in \mathbb{R}$ *is a map from* $\mathbb{R}^*$ *to* $\mathbb{R}$, *but is not a map from* $\mathbb{R}$ *to* $\mathbb{R}$ *because* $g(0)$ *is not defined,*

(vii) *For any fixed* $Y \in \mathbb{R}^n$, *a map is defined by* $\mathrm{T}_Y : \mathbb{R}^n \ni X \mapsto \mathrm{T}_Y(X) = X + Y \in \mathbb{R}^n$, *and is called* the translation by Y.

Remark 4.1.3. *For a map* $\mathrm{T} : S \to S'$, *the determination of* $\mathrm{Ran}(\mathrm{T})$ *is not always an easy task. For example if one considers* $f : \mathbb{R} \to \mathbb{R}$ *with* $f(x) = x^2 - 3x + 2$, *then one has to look for the minimum of* f, *which is* $-1/4$ *obtained for* $x = 3/2$, *and one can then set* $\mathrm{Ran}(f) = [-1/4, \infty)$. *Similarly, if* $\mathcal{A} \in M_{mn}(\mathbb{R})$, *then what is the range of* $\mathrm{L}_\mathcal{A}$, i.e. *the set of* $Y \in \mathbb{R}^m$ *such that* $Y = \mathcal{A}X$ *for some* $X \in \mathbb{R}^n$ *?*

We end this section with a natural definition.

Definition 4.1.4. *Let* $\mathrm{T} : S \to S'$ *be a map, let* $W \subset S$ *be a subset of* S *and let* Z *be a subset of* S'. *Then the set* $\mathrm{T}(W) := \{\mathrm{T}(X) \mid X \in W\}$ *is called* the image of W by T, *while the set*

$$\mathrm{T}^{-1}(Z) := \big\{X \in S \mid \mathrm{T}(X) \in Z\big\}$$

is called the preimage of Z by T.

4.2 Linear maps

From now on, we shall concentrate on the simplest maps, the linear ones. Note that in order to state the next definition, one has to deal with vector spaces instead of arbitrary sets, and in addition the two vector spaces have to be defined on the same field.

Definition 4.2.1. *Let* V, W *be two vector spaces over the same field* $\mathbb{F}$. *A map* $\mathrm{T} : V \to W$ *is a linear map if the following two conditions are satisfied:*

(i) $\mathrm{T}(X + Y) = \mathrm{T}(X) + \mathrm{T}(Y)$ *for any* $X, Y \in V$,

(ii) $\mathrm{T}(\lambda X) = \lambda \mathrm{T}(X)$ *for any* $X \in V$ *and* $\lambda \in \mathbb{F}$.

Note that the examples (*ii*) and (*iv*) of Examples 4.1.2 were already linear maps. Let us still mention the map $\mathrm{Id} : V \to V$ (also denoted by $\mathbf{1}$) defined by $\mathrm{Id}(X) = X$ for any $X \in V$, which is clearly linear, and the map $\mathcal{O} : V \to W$ defined by $\mathcal{O}(X) = \mathbf{0}$ for any $X \in V$, which is also linear.

Let us now observe that linear maps are rather simple maps.

Lemma 4.2.2. *Let* V, W *be vector spaces over the same field* $\mathbb{F}$, *and let* $\mathrm{T} : V \to W$ *be a linear map. Then,* *(i)* $\mathrm{T}(\mathbf{0}) = \mathbf{0}$,

(ii) $\mathrm{T}(-X) = -\mathrm{T}(X)$ *for any* $X \in V$.

Proof. (i) It is sufficient to observe that

$$\mathrm{T}(\mathbf{0}) = \mathrm{T}(\mathbf{0} + \mathbf{0}) = \mathrm{T}(\mathbf{0}) + \mathrm{T}(\mathbf{0}) = 2\mathrm{T}(\mathbf{0})$$

which implies the result.

(i) Observe that

$$\mathbf{0} = \mathrm{T}(\mathbf{0}) = \mathrm{T}(X - X) = \mathrm{T}(X) + \mathrm{T}(-X)$$

which directly leads to the result. □

Let go us step further in abstraction and consider families of linear maps. For that purpose, let us first define an addition of linear maps, and the multiplication of a linear map by a scalar. Namely, if V, W are vector spaces over the same field $\mathbb{F}$ and if $\mathrm{T}_1, \mathrm{T}_2$ are linear maps from V to W, one sets

$$\big(\mathrm{T}_1 + \mathrm{T}_2\big)(X) = \mathrm{T}_1(X) + \mathrm{T}_2(X) \qquad \text{for any } X \in V. \tag{4.2.1}$$

If $\lambda \in \mathbb{F}$ and if $\mathrm{T} : V \to W$ is linear, one also sets

$$\big(\lambda\mathrm{T}\big)(X) = \lambda\mathrm{T}(X) \qquad \text{for any } X \in V. \tag{4.2.2}$$

It is then easily observed that $\mathrm{T}_1 + \mathrm{T}_2$ is still a linear map, and that $\lambda\mathrm{T}$ is also a linear map. We can then even say more:

Proposition 4.2.3. *Let* V, W *be vector spaces over the same field* $\mathbb{F}$. *Then*

$$\mathcal{L}(V, W) := \big\{\mathrm{T} : V \to W \mid \mathrm{T} \textit{ is linear } \big\},$$

is a vector space over $\mathbb{F}$, *once endowed with the addition defined by* (4.2.1) *and the multiplication by a scalar defined in* (4.2.2).

Before giving the proof, let us observe that if $V = \mathbb{R}^n$ and $W = \mathbb{R}^n$, then $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ corresponds to the set of all $\mathrm{L}_\mathcal{A}$ with $\mathcal{A} \in M_{mn}(\mathbb{R})$. Note that this statement also holds for arbitrary field $\mathbb{F}$, *i.e.*

$$\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m) = \big\{\mathrm{L}_\mathcal{A} \mid \mathcal{A} \in M_{mn}(\mathbb{F})\big\}.$$

Proof. The proof consists in checking all conditions of Definition 3.1.3. For that purpose, consider $\mathrm{T}, \mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3$ be linear maps from V to W, and let $\lambda, \mu \in \mathbb{F}$. Let also X be an arbitrary element of V.

(i) One has

$$\begin{aligned}\big[(\mathrm{T}_1 + \mathrm{T}_2) + \mathrm{T}_3\big](X) &= (\mathrm{T}_1 + \mathrm{T}_2)(X) + \mathrm{T}_3(X) = \mathrm{T}_1(X) + \mathrm{T}_2(X) + \mathrm{T}_3(X) \\ &= \mathrm{T}_1(X) + (\mathrm{T}_2 + \mathrm{T}_3)(X) = \big[\mathrm{T}_1 + (\mathrm{T}_2 + \mathrm{T}_3)\big](X).\end{aligned}$$

Since X is arbitrary, it follows that $(\mathrm{T}_1 + \mathrm{T}_2) + \mathrm{T}_3 = \mathrm{T}_1 + (\mathrm{T}_2 + \mathrm{T}_3)$.

(ii) One has

$$(\mathrm{T}_1 + \mathrm{T}_2)(X) = \mathrm{T}_1(X) + \mathrm{T}_2(X) = \mathrm{T}_2(X) + \mathrm{T}_1(X) = (\mathrm{T}_2 + \mathrm{T}_1)(X)$$

which implies that $\mathrm{T}_1 + \mathrm{T}_2 = \mathrm{T}_2 + \mathrm{T}_1$.

(iii) We already know that $\mathcal{O} : V \to W$ is linear, which means that $\mathcal{O} \in \mathcal{L}(V, W)$. In addition, one clearly has $\mathrm{T} + \mathcal{O} = \mathcal{O} + \mathrm{T} = \mathrm{T}$.

(iv) By setting $[-\mathrm{T}](X) := -\mathrm{T}(X)$, one readily observes that $-\mathrm{T} \in \mathcal{L}(V, W)$ and by using the addition (4.2.1) one infers that $\mathrm{T} + (-\mathrm{T}) = \mathcal{O}$.

(v) Similarly, $\lambda\mathrm{T} \in \mathcal{L}(V, W)$ and $1\mathrm{T} = \mathrm{T}$.

The remaining three properties are easily checked by using the definition 4.2.2 and the basic properties of vector spaces. □

Question : If $\dim(V) = n$ and if $\dim(W) = m$, what is the dimension of $\mathcal{L}(V, W)$?

Let us now consider a linear map $\mathrm{T} : V \to \mathbb{R}^n$ with V a real vector space. Since for each $X \in V$ one has $\mathrm{T}(X) \in \mathbb{R}^n$, one often sets

$$\mathrm{T}(X) = \begin{pmatrix} \mathrm{T}_1(X) \\ \vdots \\ \mathrm{T}_n(X) \end{pmatrix} \tag{4.2.3}$$

with $\mathrm{T}_j(X) := \mathrm{T}(X)_j$ the j^{th} component of T evaluated at X. Thus, T defines a family of maps $\mathrm{T}_j : V \to \mathbb{R}$, and reciprocally, any family $\{\mathrm{T}_j\}_{j=1}^n$ with $\mathrm{T}_j : V \to \mathbb{R}$ defines a map $\mathrm{T} : V \to \mathbb{R}^n$ by (4.2.3). Sometimes, the maps $\mathrm{T}_1, \dots, \mathrm{T}_n$ are called *the components of* T.

Example 4.2.4. *If* $\mathrm{T} : \mathbb{R}^2 \to \mathbb{R}^3$ *is defined by*

$$\mathrm{T}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - y \\ 3x + 4y \\ x - 5y \end{pmatrix},$$

then $\mathrm{T} = {}^t(\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3)$ *with* $\mathrm{T}_1\binom{x}{y} = 2x - y$, $\mathrm{T}_2\binom{x}{y} = 3x + 4y$ *and* $\mathrm{T}_3\binom{x}{y} = x - 5y$.

More generally:

Lemma 4.2.5. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathrm{T} : V \to \mathbb{F}^n$ with $\mathrm{T} = {}^t(\mathrm{T}_1, \dots, \mathrm{T}_n)$ the components of* T. *Then* T *is a linear map if and only if each* T_j *is a linear map.*

Proof. One has $\mathrm{T}(X + Y) = {}^t\big(\mathrm{T}_1(X + Y), \dots, \mathrm{T}_n(X + Y)\big)$ and $\mathrm{T}(X) + \mathrm{T}(Y) = {}^t\big(\mathrm{T}_1(X) + \mathrm{T}_1(Y), \dots, \mathrm{T}_n(X) + \mathrm{T}_n(Y)\big)$. It then follows that

$$\mathrm{T}(X + Y) = \mathrm{T}(X) + \mathrm{T}(Y) \Longleftrightarrow \begin{pmatrix} \mathrm{T}_1(X + Y) \\ \vdots \\ \mathrm{T}_n(X + Y) \end{pmatrix} = \begin{pmatrix} \mathrm{T}_1(X) + \mathrm{T}_1(Y) \\ \vdots \\ \mathrm{T}_n(X) + \mathrm{T}_n(Y) \end{pmatrix},$$

which corresponds to half of the statement. A similar argument holds for the multiplication by a scalar. □

4.3 Kernel and range of a linear map

Let V, W be two vector spaces over the same field $\mathbb{F}$ and let $\mathrm{T} : V \to W$ be a linear map. Recall that

$$\mathrm{Ran(T)} := \{Y \in W \mid Y = \mathrm{T}(X) \text{ for some } X \in V\}$$

and

$$\mathrm{Ker(T)} := \{X \in V \mid \mathrm{T}(X) = \mathbf{0}\}.$$

Lemma 4.3.1. *In the previous setting,* Ker(T) *is a subspace of* V *while* Ran(T) *is a subspace of* W.

Proof. The first part of the statement is proved in Exercise 4.4. For the second part of the statement, consider $Y_1, Y_2 \in \mathrm{Ran(T)}$, *i.e.* there exist $X_1, X_2 \in V$ such that $Y_1 = \mathrm{T}(X_1)$ and $Y_2 = \mathrm{T}(X_2)$. Then one has

$$Y_1 + Y_2 = \mathrm{T}(X_1) + \mathrm{T}(X_2) = \mathrm{T}(X_1 + X_2)$$

with $X_1 + X_2 \in V$. In other words, $Y_1 + Y_2$ belongs to Ran(T). Similarly, for $\lambda \in \mathbb{F}$ and any $Y = \mathrm{T}(X)$ with $X \in V$ one has

$$\lambda Y = \lambda \mathrm{T}(X) = \mathrm{T}(\lambda X)$$

with $\lambda X \in V$. Again, it follows that $\lambda Y \in \mathrm{Ran(T)}$, from which one concludes that Ran(T) is a subspace of W. □

Examples 4.3.2. *(i) Let* $N \in \mathbb{R}^n$ *with* $N \neq \mathbf{0}$*, and let us set* $\mathrm{T}_N : \mathbb{R}^n \to \mathbb{R}$ *by* $\mathrm{T}_N(X) = N \cdot X$*. In this case,* T_N *is a linear map. Indeed, one has*

$$\mathrm{T}_N(X + Y) = N \cdot (X + Y) = N \cdot X + N \cdot Y = \mathrm{T}_N(X) + \mathrm{T}_N(Y),$$

and similarly $\mathrm{T}_N(\lambda X) = N \cdot (\lambda X) = \lambda(N \cdot X) = \lambda \mathrm{T}_N(X)$*. Then one observes that*

$$\mathrm{Ker}(\mathrm{T}_N) = \{X \in \mathbb{R}^n \mid N \cdot X = 0\} = \{X \in \mathbb{R}^n \mid X \cdot N = \mathbf{0} \cdot N\} = H_{N,\mathbf{0}}.$$

On the other hand, $\mathrm{Ran}(\mathrm{T}_N) = \mathbb{R}$*, as it can easily be checked by considering elements* X *of the form* λN*, for any* $\lambda \in \mathbb{R}$*.*

(ii) Let $\mathcal{A} \in M_{mn}(\mathbb{R})$ *and let us set* $\mathrm{L}_\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ *defined by* $\mathrm{L}_\mathcal{A}(X) = \mathcal{A}X$ *for any* $X \in \mathbb{R}^n$*. As already mentioned, this map is linear, and one has* $\mathrm{Ker}(\mathrm{L}_\mathcal{A}) = \{X \in \mathbb{R}^n \mid \mathcal{A}X = \mathbf{0}\}$, i.e. $\mathrm{Ker}(\mathrm{L}_\mathcal{A})$ *are the solutions of the linear system* $\mathcal{A}X = \mathbf{0}$*.*

Remark 4.3.3. *The kernel of a linear map is never empty, indeed it always contains the element* $\mathbf{0}$*.*

Lemma 4.3.4. *Let* $\mathrm{T} : V \to W$ *be a linear map between vector spaces over the same field* $\mathbb{F}$*, and assume that* $\mathrm{Ker}(\mathrm{T}) = \{\mathbf{0}\}$*. If* $\{X_1, \dots, X_r\}$ *are linearly independent elements of* V*, then* $\{\mathrm{T}(X_1), \dots, \mathrm{T}(X_n)\}$ *are linearly independent elements of* W*.*

Proof. Let $\lambda_1, \dots, \lambda_n$ such that

$$\lambda_1 \mathrm{T}(X_1) + \lambda_2 \mathrm{T}(X_2) + \dots + \lambda_n \mathrm{T}(X_n) = \mathbf{0}.$$

By linearity, this is equivalent to $\mathrm{T}\big(\lambda_1 X_1 + \dots + \lambda_n X_n\big) = \mathbf{0}$, but since the kernel of T is reduced to $\mathbf{0}$ it means that $\lambda_1 X_1 + \dots + \lambda_n X_n = \mathbf{0}$. Finally, by the linear independence of $X_1, \dots, X_n$ it follows that $\lambda_j = 0$ for any $j \in \{1, \dots, n\}$. As a consequence, the elements $\mathrm{T}(X_1), \dots, \mathrm{T}(X_n)$ of W are linearly independent. □

Let us now come to an important result of this section. For this, we just recall that for a vector space, its dimension corresponds to the number of elements of any of its bases. It also corresponds to the maximal number of linearly independent elements of this vector space.

Theorem 4.3.5. *Let* $\mathrm{T} : V \to W$ *be a linear map between two vector spaces over the same field* $\mathbb{F}$*, and assume that* V *is of finite dimension. Then*

$$\dim\big(\mathrm{Ker}(\mathrm{T})\big) + \dim\big(\mathrm{Ran}(\mathrm{T})\big) = \dim(V).$$

Proof. Let $\{Y_1, \dots, Y_n\}$ be a basis for $\mathrm{Ran}(\mathrm{T})$, and let $X_1, \dots, X_n \in V$ such that $\mathrm{T}(X_j) = Y_j$ for any $j \in \{1, \dots, n\}$. Let also $\{K_1, \dots, K_m\}$ be a basis for $\mathrm{Ker}(\mathrm{T})$. Note that if one shows that $\{X_1, \dots, X_n, K_1, \dots, K_m\}$ is a basis for V, then the statement is proved (with $\dim(V) = m + n$).

So, let X be an arbitrary element of V. Then there exist $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ such that $\mathrm{T}(X) = \lambda_1 Y_1 + \dots + \lambda_n X_n$, since $\{Y_1, \dots, Y_n\}$ is a basis for $\mathrm{Ran}(\mathrm{T})$. It follows that

$$\begin{aligned}
\mathbf{0} &= \mathrm{T}(X) - \lambda_1 X_1 - \dots - \lambda_n Y_n \\
&= \mathrm{T}(X) - \lambda_1 \mathrm{T}(X_1) - \dots - \lambda_n \mathrm{T}(X_n) \\
&= \mathrm{T}\big(X - \lambda_1 X_1 - \dots - \lambda_n X_n\big),
\end{aligned}$$

which means that $X - \lambda_1 X_1 - \dots - \lambda_n X_n$ belongs to $\mathrm{Ker}(\mathrm{T})$. As a consequence, there exist $\lambda_1', \dots, \lambda_m' \in \mathbb{F}$ such that

$$X - \lambda_1 X_1 - \dots - \lambda_n X_n = \lambda_1' K_1 + \dots + \lambda_m' K_m,$$

since $\{K_1, \dots, K_m\}$ is a basis for $\mathrm{Ker}(\mathrm{T})$. Consequently, one gets

$$X = \lambda_1 X_1 + \dots + \lambda_n X_n + \lambda_1' K_1 + \dots + \lambda_m' K_m,$$

or in other words

$$\mathrm{Vect}\big(X_1, \dots, X_n, K_1, \dots, K_m\big) = V.$$

Let us now show that these vectors are linearly independent. By contraposition, assume that

$$\lambda_1 X_1 + \cdots + \lambda_n X_n + \lambda'_1 K_1 + \cdots + \lambda'_m K_m = \mathbf{0} \tag{4.3.1}$$

for some $\lambda_1, \ldots, \lambda_n, \lambda'_1, \ldots, \lambda'_m$. Then one infers from (4.3.1) that

$$\begin{aligned}\mathbf{0} &= \mathrm{T}(\mathbf{0}) \\ &= \mathrm{T}\big(\lambda_1 X_1 + \cdots + \lambda_n X_n + \lambda'_1 K_1 + \cdots + \lambda'_m K_m\big) \\ &= \mathrm{T}\big(\lambda_1 X_1 + \cdots + \lambda_n X_n\big) + \mathbf{0} \\ &= \lambda_1 \mathrm{T}(X_1) + \cdots + \lambda_n \mathrm{T}(X_n) \\ &= \lambda_1 Y_1 + \cdots + \lambda_n Y_n.\end{aligned}$$

Since $Y_1, \ldots, Y_n$ are linearly independent, one already concludes that $\lambda_j = 0$ for any $j \in \{1, \ldots, n\}$. It then follows from (4.3.1) that $\lambda'_1 K_1 + \cdots + \lambda'_m K_m = \mathbf{0}$, which implies that $\lambda'_i = 0$ for any $i \in \{1, \ldots, m\}$ since the vectors K_i are linearly independent.

In summary, one has shown that V is generated by the family of linearly independent elements $X_1, \ldots, X_n, K_1, \ldots, K_m$ of V. Thus, these elements define a basis, as expected. □

4.4 Rank and linear maps

Let us come back to matrices over $\mathbb{F}$. For any $\mathcal{A} \in M_{mn}(\mathbb{F})$, recall that we denote by $\mathcal{A}^j$ the j^{th} column of $\mathcal{A}$ and by $\mathcal{A}_k$ the k^{th} row of $\mathcal{A}$. We also denote by $\mathrm{L}_\mathcal{A} : \mathbb{F}^n \to \mathbb{F}^m$ the linear map defined by $\mathrm{L}_\mathcal{A}(X) = \mathcal{A}X$. Observe finally that $\{E_j\}_{j=1}^n$ is a basis of $\mathbb{F}^n$ (note that the 1 at the entry j of E_j is the 1 of the field $\mathbb{F}$). Thus, for any $X \in \mathbb{F}^n$ one has

$$X = {}^t(x_1, \ldots, x_n) = x_1 E_1 + x_2 E_2 + \cdots + x_n E_n$$

and in addition

$$\begin{aligned}\mathrm{L}_\mathcal{A}(X) &= \mathcal{A}\big(x_1 E_1 + x_2 E_2 + \cdots + x_n E_n\big) \\ &= x_1 \mathcal{A} E_1 + x_2 \mathcal{A} E_2 + \cdots + x_n \mathcal{A} E_n \\ &= x_1 \mathcal{A}^1 + x_2 \mathcal{A}^2 + \ldots x_n \mathcal{A}^n.\end{aligned}$$

With such equalities, one directly infers the following statement:

Lemma 4.4.1. *The range of* $\mathrm{L}_\mathcal{A}$ *corresponds to the subspace generated by the columns of* $\mathcal{A}$*.*

Proof. It is enough to remember the following equality

$$\mathrm{Ran}(\mathrm{L}_\mathcal{A}) = \big\{\mathrm{L}_\mathcal{A}(X) \mid X \in \mathbb{F}^n\big\}$$

and to take into account the computation performed before the statement. □

Considering the dimensions of these spaces one directly gets:

Corollary 4.4.2. *The dimension of the range of the linear map* $\mathrm{L}_{\mathcal{A}}$ *is equal to the rank of* $\mathcal{A}$*,* i.e.

$$\dim\big(\mathrm{Ran}(\mathrm{L}_{\mathcal{A}})\big) = \mathrm{rank}(\mathcal{A}).$$

Theorem 4.4.3. *Let* $\mathcal{A} \in M_{mn}(\mathbb{F})$ *with* $\mathrm{rank}(\mathcal{A}) = r$*. Then one has* $\dim\big(\mathrm{Ker}(\mathrm{L}_{\mathcal{A}})\big) = n - r$*.*

Proof. Since $\mathrm{L}_{\mathcal{A}} : \mathbb{F}^n \to \mathbb{F}^m$ is a linear map, one has from Theorem 4.3.5

$$\dim\big(\mathrm{Ker}(\mathrm{L}_{\mathcal{A}})\big) + \underbrace{\dim\big(\mathrm{Ran}(\mathrm{L}_{\mathcal{A}})\big)}_{=r} = n,$$

from which the statement follows. □

Example 4.4.4. *What is the dimension of the space of solutions of the system*

$$\begin{cases} 2x_1 - x_2 + x_3 + 2x_4 = 0 \\ x_1 + x_2 - 2x_3 - x_4 = 0 \end{cases} \qquad ?$$

Since this system is equivalent to $\mathrm{L}_{\mathcal{A}}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *with* $\mathcal{A} = \begin{pmatrix} 2 & -1 & 1 & 2 \\ 1 & 1 & -2 & -1 \end{pmatrix}$ *and since* $\mathrm{rank}(\mathcal{A}) = 2$*, one directly infers from the previous result that* $\dim\big(\mathrm{Ker}(\mathrm{L}_{\mathcal{A}})\big) = 4-2 = 2$*. This corresponds to the dimension of the space of solutions of the homogeneous equation.*

One ends up this section with an important result:

Theorem 4.4.5. *Let* $\mathcal{A} \in M_{mn}(\mathbb{F})$ *and* $B \in \mathbb{F}^m$*, and consider the equation* $\mathcal{A}X = B$ *for some* $X \in \mathbb{F}^n$*. If this equation has a solution, then its set of all solutions is of dimension equal to* $\dim\big(\mathrm{Ker}(\mathrm{L}_{\mathcal{A}})\big)$*.*

Proof. Assume that $Y_0 \in \mathbb{F}^n$ satisfies $\mathcal{A}Y_0 = B$. Then if $Y \in \mathbb{F}^n$ satisfies $\mathcal{A}Y = \mathbf{0}$, one infers that $\mathcal{A}(Y_0 + Y) = B$, which means that $Y_0 + Y$ is a solution of the original problem, for any $Y \in \mathrm{Ker}(\mathrm{L}_{\mathcal{A}})$. Now, if one can show that all solutions of $\mathcal{A}X = B$ are of the form $X = Y_0 + Y$ for some $Y \in \mathrm{Ker}(\mathrm{L}_{\mathcal{A}})$, then the statement is proved. For that purpose, it is sufficient to observe that if $X \in \mathbb{F}^n$ satisfies $\mathcal{A}X = B$, then one has $\mathcal{A}(X - Y_0) = B - B = \mathbf{0}$, or in other words $X - Y_0 =: Y$ for some $Y \in \mathrm{Ker}(\mathrm{L}_{\mathcal{A}})$. As a consequence, one infers that $X = Y_0 + Y$ with $Y \in \mathrm{Ker}(\mathrm{L}_{\mathcal{A}})$, as expected. □

4.5 Matrix associated with a linear map

Let us start with a question: If V, W are vector spaces over a field $\mathbb{F}$ and if $\mathrm{T} : V \to W$ is a linear map, how can one associate with this linear map a matrix ?

In fact, this can be done only once a choice of bases for V and W has been done, and the resulting matrix will depend on the choice of bases, as we shall see. So, let us introduce a new notation: a basis for a vector space V over $\mathbb{F}$ will be denoted by $\mathcal{V} := \{V_1, \dots, V_n\}$ with $\{V_1, \dots, V_n\}$ a family of linearly independent elements of V which generate V. In addition, let us denote by $\mathcal{X}$ an arbitrary element of V (which was simply denoted by X up to now). Then, since $\mathcal{V}$ is a basis for V there exists $X := {}^t(x_1, \dots, x_n) \in \mathbb{F}^n$ such that

$$\mathcal{X} = x_1V_1 + x_2V_2 + \cdots + x_nV_n.$$

The vector $X \in \mathbb{F}^n$ is called *the coordinate vector of $\mathcal{X}$ with respect to the basis $\mathcal{V}$ of V*, and we shall use the notation

$$(\mathcal{X})_{\mathcal{V}} = X$$

meaning precisely that the coordinates of $\mathcal{X}$ with respect to the basis $\mathcal{V}$ are X.

Remark 4.5.1. *Clearly, if $V = \mathbb{R}^n$ and if $V_j = E_j$, one just says that X are the coordinates of $\mathcal{X}$ and one uses to identify $\mathcal{X}$ and X. This is what we have done until now since we have only considered the usual basis $\{E_j\}_{j=1}^n$ on $\mathbb{R}^n$. However, if one needs to consider different bases on $\mathbb{R}^n$, the above notations are necessary. Note for example that $\mathcal{X}$ exists without any choice of a particular basis, while X depends on such a choice.*

Now, if $\mathcal{Y}$ is another element of V with $(\mathcal{Y})_{\mathcal{V}} = Y = {}^t(y_1, \dots, y_n)$, let us observe that

$$(\mathcal{X} + \mathcal{Y})_{\mathcal{V}} = X + Y \qquad \text{and} \qquad (\lambda\mathcal{X})_{\mathcal{V}} = \lambda X \tag{4.5.1}$$

for any $\lambda \in \mathbb{F}$. Indeed, this follows from the equalities

$$\begin{aligned}\mathcal{X} + \mathcal{Y} &= x_1V_1 + \cdots + x_nV_n + y_1V_1 + \cdots + y_nV_n \\ &= (x_1 + y_1)V_1 + \cdots + (x_n + y_n)V_n\end{aligned}$$

and

$$\lambda\mathcal{X} = \lambda(x_1V_1 + \cdots + x_nV_n) = (\lambda x_1)V_1 + \cdots + (\lambda x_n)V_n.$$

Thus, choosing a basis $\mathcal{V}$ for V allows one to identity any point of V with an element of $\mathbb{F}^n$ via its coordinate vector. By taking (4.5.1) into account, one also observes that $\mathcal{V}$ allows one to define a linear map $(\cdot)_{\mathcal{V}} : V \to \mathbb{F}^n$.

We also consider a vector space W over $\mathbb{F}$ endowed with a basis $\mathcal{W} := \{W_1, \dots, W_m\}$. In this case, for any $\mathcal{Z} \in \mathcal{W}$ we set $(\mathcal{Z})_{\mathcal{W}} = Z = {}^t(z_1, \dots, z_m) \in \mathbb{F}^m$ for the coordinate vector of $\mathcal{Z}$ with respect to the basis $\mathcal{W}$ of W. Thus, if $\mathrm{T} : V \to W$ is a linear map, there exists $\mathcal{T} := (t_{ij}) \in M_{mn}(\mathbb{F})$, called *the matrix associated with* T *with respect to the basis $\mathcal{V}$ of V and $\mathcal{W}$ of W* defined by

$$\mathrm{T}(V_j) = \sum_{i=1}^m t_{ij}W_i = \sum_{i=1}^m {}^t t_{ji}W_i \tag{4.5.2}$$

for any $j \in \{1, \dots, n\}$. On the other hand, we shall show just below that the following equality also holds

$$\big(\mathrm{T}(\mathcal{X})\big)_{\mathcal{W}} = \mathcal{T}(\mathcal{X})_{\mathcal{V}}. \tag{4.5.3}$$

In other words, the action of T on a basis of V is given in terms of ${}^t\mathcal{T}$ by relation (4.5.2), while the action of T on the coordinate vectors is given in terms of $\mathcal{T}$ by relation (4.5.3). Note that this is related to the more general notion of *covariant* or *contravariant* transformations.

For the proof of (4.5.3) it is enough to observe that one has

$$\begin{aligned}\mathrm{T}(\mathcal{X}) = \mathrm{T}\Big(\sum_{j=1}^n x_j V_j\Big) &= \sum_{j=1}^n x_j \mathrm{T}(V_j)\\ &= \sum_{j=1}^n x_j \sum_{i=1}^m t_{ij} W_i = \sum_{i=1}^m \Big(\sum_{j=1}^n t_{ij}\, x_j\Big) W_i = \sum_{i=1}^m (\mathcal{T}X)_i W_i,\end{aligned}$$

which implies that

$$\big(\mathrm{T}(\mathcal{X})\big)_{\mathcal{W}} = \mathcal{T}X = \mathcal{T}(\mathcal{X})_{\mathcal{V}}. \tag{4.5.4}$$

Example 4.5.2. *If $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ and $V_j = E_j$ while $W_i = E_i$ for any $j \in \{1,\dots,n\}$ and $j \in \{1,\dots,m\}$, and if* T *is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ then one deduces from* (4.5.2) *that*

$$\mathrm{T}(E_j) = \sum_{i=1}^m t_{ij} E_i = \begin{pmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{pmatrix} = \mathcal{T}^j$$

where $\mathcal{T}^j$ corresponds to the j^{th} column of the matrix $\mathcal{T}$. In other words one has

$$\mathcal{T} = \big(\mathrm{T}(E_1)\ \mathrm{T}(E_2)\ \dots\ \mathrm{T}(E_n)\big)\,.$$

Example 4.5.3. *If V is a real vector space with basis $\mathcal{V} = \{V_1, V_2, V_3\}$ and if $\mathrm{T} : V \to V$ is the linear map such that*

$$\mathrm{T}(V_1) = 2V_1 - V_2, \quad \mathrm{T}(V_2) = V_1 + V_2 - 4V_3, \quad \mathrm{T}(V_3) = 5V_1 + 4V_2 + 2V_3,$$

then the matrix associated with T *with respect to the basis $\mathcal{V}$ is given by*

$$\mathcal{T} = \begin{pmatrix} 2 & 1 & 5 \\ -1 & 1 & 4 \\ 0 & -4 & 2 \end{pmatrix}.$$

Let us still consider the notion of a change of basis. Indeed, given the matrix associated to a linear map in a prescribed basis, it is natural to wonder about the matrix associated to the same linear map but with respect to another basis. So, let $\mathcal{V} = \{V_1, \dots, V_n\}$ and $\mathcal{V}' = \{V_1', \dots, V_n'\}$ be two basis of the same vector space V. Let $\mathcal{B} = (b_{ij}) \in M_n(\mathbb{F})$ be the matrix defined by

$$V_j' = \sum_{i=1}^n b_{ij} V_i \equiv \sum_{i=1}^n {}^t b_{ji} V_i.$$

It is easily observed that the matrix $\mathcal{B}$ is invertible. Then, for any $\mathcal{X} \in V$ with $X = (\mathcal{X})_\mathcal{V}$ and $X' = (\mathcal{X})_{\mathcal{V}'}$, one has

$$\sum_{i=1}^n x_i V_i = \mathcal{X} = \sum_{j=1}^n x'_j V'_j = \sum_{j=1}^n x'_j \sum_{i=1}^n b_{ij} V_i = \sum_{i=1}^n \Big(\sum_{j=1}^n b_{ij} x'_j \Big) V_i.$$

Since the vectors $V_1, \dots, V_n$ are linearly independent, this implies that

$$X = \mathcal{B}X' \quad \text{or equivalently} \quad (\mathcal{X})_\mathcal{V} = \mathcal{B}(\mathcal{X})_{\mathcal{V}'}. \tag{4.5.5}$$

Let us now consider a linear map $\mathrm{T} : V \to V$, and let $\mathcal{T}$ be the matrix associated with T with respect to the basis $\mathcal{V}$, and let $\mathcal{T}'$ be the matrix associated to T with respect to the basis $\mathcal{V}'$. The original question corresponds then to the link between $\mathcal{T}$ and $\mathcal{T}'$? In order to answer this question, observe that for any $\mathcal{X} \in V$ one gets by equations (4.5.4) and (4.5.5) that

$$\mathcal{T}\mathcal{B}X' = \mathcal{T}X = \big(\mathrm{T}(\mathcal{X})\big)_\mathcal{V} = \mathcal{B}\big(\mathrm{T}(\mathcal{X})\big)_{\mathcal{V}'} = \mathcal{B}\mathcal{T}'X'.$$

Since X' is arbitrary, one infers that $\mathcal{T}\mathcal{B} = \mathcal{B}\mathcal{T}'$, or equivalently

$$\mathcal{T}' = \mathcal{B}^{-1}\mathcal{T}\mathcal{B}. \tag{4.5.6}$$

One deduces in particular that the matrix $\mathcal{T}$ and $\mathcal{T}'$ are similar, see Definition 2.1.16.

Note that a similar (but slightly more complicated) computation can be realized for a linear map between two vector spaces V and W over the same field $\mathbb{F}$ endowed with two different bases $\mathcal{V}, \mathcal{V}'$ and $\mathcal{W}, \mathcal{W}'$.

4.6 Composition of linear maps

Let us now consider three sets U, V, W and let $\mathrm{F} : U \to V$ and $\mathrm{G} : V \to W$ be maps. Then the map

$$\mathrm{G} \circ \mathrm{F} : U \to W,$$

defined by $\big(\mathrm{G} \circ \mathrm{F}\big)(X) = \mathrm{G}\big(\mathrm{F}(X)\big)$ for any $X \in U$, is called *the composition map of* F *with* G. Notice that if $W \not\subset U$ the composition map $\mathrm{F} \circ \mathrm{G}$ has simply no meaning.

Examples 4.6.1. *(i) Let $U = V = W = \mathbb{R}$ and* F, G *be two real functions defined on $\mathbb{R}$. Then* G ∘ F *just corresponds to the composition of functions.*

(ii) If $U = \mathbb{R}^n$, $V = \mathbb{R}^m$, $W = \mathbb{R}^p$, $\mathcal{A} \in M_{mn}(\mathbb{R})$ and $\mathcal{B} \in M_{pm}(\mathbb{R})$, then for any $X \in \mathbb{R}^n$ one has

$$\big(\mathrm{L}_\mathcal{B} \circ \mathrm{L}_\mathcal{A}\big)(X) = \mathrm{L}_\mathcal{B}\big(\mathrm{L}_\mathcal{A}(X)\big) = \mathcal{B}\mathcal{A}X = (\mathcal{B}\mathcal{A})X = \mathrm{L}_{\mathcal{B}\mathcal{A}}(X). \tag{4.6.1}$$

Let us now observe an important property of the composition of maps, namely the associativity. Indeed, If U, V, W, S are sets and $\mathrm{F} : U \to V$, $\mathrm{G} : V \to W$ and $\mathrm{H} : W \to S$ are maps, one has

$$(\mathrm{H} \circ \mathrm{G}) \circ \mathrm{F} = \mathrm{H} \circ (\mathrm{G} \circ \mathrm{F}).$$

Indeed, for any $X \in U$ one has

$$\big[(\mathrm{H} \circ \mathrm{G}) \circ \mathrm{F}\big](X) = \big(\mathrm{H} \circ \mathrm{G}\big)(\mathrm{F}(X)) = \mathrm{H}\big(\mathrm{G}(\mathrm{F}(X))\big)$$

and

$$\big[\mathrm{H} \circ (\mathrm{G} \circ \mathrm{F})\big](X) = \mathrm{H}\big((\mathrm{G} \circ \mathrm{F})(X)\big) = \mathrm{H}\big(\mathrm{G}(\mathrm{F}(X))\big),$$

and the equality of the two right hand sides implies the statement.

Lemma 4.6.2. *Let U, V, W be vector spaces over a field $\mathbb{F}$, and let $\mathrm{G} : U \to V$, $\mathrm{G}' : U \to V$, $\mathrm{H} : V \to W$ and $\mathrm{H}' : V \to W$ be linear maps. Then*

(i) $\mathrm{H} \circ \mathrm{G} : U \to W$ *is a linear map,*

(ii) $(\mathrm{H} + \mathrm{H}') \circ \mathrm{G} = \mathrm{H} \circ \mathrm{G} + \mathrm{H}' \circ \mathrm{G}$,

(iii) $\mathrm{H} \circ (\mathrm{G} + \mathrm{G}') = \mathrm{H} \circ \mathrm{G} + \mathrm{H} \circ \mathrm{G}'$,

(iv) $(\lambda \mathrm{H}) \circ \mathrm{G} = \mathrm{H} \circ (\lambda \mathrm{G}) = \lambda(\mathrm{H} \circ \mathrm{G})$, *for all* $\lambda \in \mathbb{F}$.

The proof will be provided in Exercise 4.17.

Remark 4.6.3. *If V is a vector space and if $\mathrm{T} : V \to V$ is a linear map, then $\mathrm{T}^n = \underbrace{\mathrm{T} \circ \mathrm{T} \cdots \circ \mathrm{T}}_{n \text{ terms}}$ is a linear map from V to V. By convention, one sets $\mathrm{T}^0 = \mathbf{1}$, and observes that $\mathrm{T}^{r+s} = \mathrm{T}^r \circ \mathrm{T}^s = \mathrm{T}^s \circ \mathrm{T}^r$.*

4.7 Inverse of a linear map

Definition 4.7.1. *For a map $\mathrm{F} : V \to W$ between two sets V and W, one says that* F *has* an inverse *if there exists $\mathrm{G} : W \to V$ such that $\mathrm{G} \circ \mathrm{F} = \mathbf{1}_V$ and $\mathrm{F} \circ \mathrm{G} = \mathbf{1}_W$. In this case, one also says that* F *is* invertible *and write F^{-1} for this inverse.*

Example 4.7.2. *If $\mathcal{A} \in M_n(\mathbb{R})$ is invertible, then the linear map $\mathrm{L}_{\mathcal{A}} : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, with inverse $\mathrm{L}_{\mathcal{A}^{-1}}$. This follows from equation* (4.6.1)*, or more precisely*

$$\mathrm{L}_{\mathcal{A}} \circ \mathrm{L}_{\mathcal{A}^{-1}} = \mathrm{L}_{\mathcal{A}\mathcal{A}^{-1}} = \mathbf{1} = \mathrm{L}_{\mathcal{A}^{-1}\mathcal{A}} = \mathrm{L}_{\mathcal{A}^{-1}} \circ \mathrm{L}_{\mathcal{A}}.$$

Due to the following lemma, there is no ambiguity in speaking about the inverse (and not only about an inverse) of a invertible map.

Lemma 4.7.3. *Let $\mathrm{F} : V \to W$ be an invertible map between two sets V et W. Then this inverse is unique.*

Proof. Let us assume that there exists $\mathrm{G} : W \to V$ and $\mathrm{G}' : W \to V$ such that $\mathrm{G} \circ \mathrm{F} = \mathbf{1}_V$, $\mathrm{F} \circ \mathrm{G} = \mathbf{1}_W$, $\mathrm{G}' \circ \mathrm{F} = \mathbf{1}_V$, and $\mathrm{F} \circ \mathrm{G}' = \mathbf{1}_W$. Then one gets

$$\mathrm{G} = \mathbf{1}_V \circ \mathrm{G} = (\mathrm{G}' \circ \mathrm{F}) \circ \mathrm{G} = \mathrm{G}' \circ (\mathrm{F} \circ \mathrm{G}) = \mathrm{G}' \circ \mathbf{1}_W = \mathrm{G}'$$

from which the result follows. □

Let us now come to two more refined notions related to a maps, linear or not.

Definition 4.7.4. *A map* $\mathrm{F} : V \to W$ *between two sets is* injective *or* one-to-one *if* $\mathrm{F}(X_1) \neq \mathrm{F}(X_2)$ *whenever* $X_1, X_2 \in V$ *with* $X_1 \neq X_2$. *The map* F *is called* surjective *if for any* $Y \in W$ *there exists at least one* $X \in V$ *such that* $\mathrm{F}(X) = Y$. *The map* F *is* bijective *if it is both injective and surjective.*

The following result links the notions of invertibility and of bijectivity.

Theorem 4.7.5. *A map* $\mathrm{F} : V \to W$ *between two sets is invertible if and only if* F *is bijective.*

Proof. (i) Assume first that F is bijective. In particular, since F is surjective, for any $Y \in W$, there exists $X \in V$ such that $\mathrm{F}(X) = Y$. Note that X is unique because F is also injective. Thus if one sets $\mathrm{F}^{-1}(Y) := X$ then one has

$$(\mathrm{F}^{-1} \circ \mathrm{F})(X) = \mathrm{F}^{-1}\big(\mathrm{F}(X)\big) = \mathrm{F}^{-1}(Y) = X$$

which implies that $\mathrm{F}^{-1} \circ \mathrm{F} = \mathbf{1}_V$, and similarly

$$\big(\mathrm{F} \circ \mathrm{F}^{-1}\big)(Y) = \mathrm{F}\big(\mathrm{F}^{-1}(Y)\big) = \mathrm{F}(X) = Y$$

which implies that $\mathrm{F} \circ \mathrm{F}^{-1} = \mathbf{1}_W$. One has thus define an inverse for F.

(ii) Let us now assume that F is invertible, with inverse denoted by F^{-1}. Let first $X_1, X_2 \in V$ with $\mathrm{F}(X_1) = \mathrm{F}(X_2)$. One then deduces that

$$X_1 = \mathbf{1}_V X_1 = \big(\mathrm{F}^{-1} \circ \mathrm{F}\big) X_1 = \mathrm{F}^{-1}\big(\mathrm{F}(X_1)\big) = \mathrm{F}^{-1}\big(\mathrm{F}(X_2)\big) = \big(\mathrm{F}^{-1} \circ \mathrm{F}\big)(X_2) = X_2,$$

and thus F is injective. Secondly, let $Y \in W$, and observe that

$$Y = \mathbf{1}_W Y = \big(\mathrm{F} \circ \mathrm{F}^{-1}\big)(Y) = \mathrm{F}\big(\mathrm{F}^{-1}(Y)\big)$$

which implies that $Y = \mathrm{F}(X)$ for X given by $\mathrm{F}^{-1}(X)$. Thus F is surjective. Since F is both injective and surjective, F is bijective. □

For linear maps the general theory simplifies a lot, as we shall see now.

Theorem 4.7.6. *Let* V, W *be two vector spaces over the same field* $\mathbb{F}$, *and let* $\mathrm{T} : V \to W$ *be an invertible linear map. Then its inverse* $\mathrm{T}^{-1} : W \to V$ *is also a linear map.*

Proof. Let $Y_1, Y_2 \in W$ and set $X_1 := \mathrm{T}^{-1}(Y_1)$ and $X_2 := \mathrm{T}^{-1}(X_2)$. Since $\mathrm{T} \circ \mathrm{T}^{-1} = \mathbf{1}_W$ one has for $j \in \{1, 2\}$

$$Y_j = \big(\mathrm{T} \circ \mathrm{T}^{-1}\big)(Y_j) = \mathrm{T}\big(\mathrm{T}^{-1}(Y_j)\big) = \mathrm{T}(X_j).$$

Then, one infers that

$$\begin{aligned} \mathrm{T}^{-1}\big(Y_1 + Y_2\big) = \mathrm{T}^{-1}\big(\mathrm{T}(X_1) + \mathrm{T}(X_2)\big) &\underbrace{=}_{linearity} \mathrm{T}^{-1}\big(\mathrm{T}(X_1 + X_2)\big) \\ = \mathrm{T}^{-1} \circ \mathrm{T}(X_1 + X_2) = X_1 + X_2 &= \mathrm{T}^{-1}(Y_1) + \mathrm{T}^{-1}(Y_2). \end{aligned} \tag{4.7.1}$$

Similarly one has for any $\lambda \in \mathbb{F}$ and $Y \in W$ (with $Y := \mathrm{T}(X)$)

$$\begin{aligned} \mathrm{T}^{-1}(\lambda Y) = \mathrm{T}^{-1}\big(\lambda \mathrm{T}(X)\big) &\underbrace{=}_{linearity} \mathrm{T}^{-1}\big(\mathrm{T}(\lambda X)\big) \\ = \big(\mathrm{T}^{-1} \circ \mathrm{T}\big)(\lambda X) &= \lambda X = \lambda \mathrm{T}^{-1}(Y). \end{aligned} \tag{4.7.2}$$

It is then sufficient to observe that (4.7.1) and (4.7.2) correspond to the linearity conditions for T^{-1}. □

In the next statement we give an equivalent property for the injectivity of a linear map.

Lemma 4.7.7. *A linear map* $\mathrm{T} : V \to W$ *between two vector spaces over the same field is injective if and only if* $\mathrm{Ker}(\mathrm{T}) = \{\mathbf{0}\}$.

Proof. (i) The first part of the proof is a contraposition argument: instead of proving $A \Rightarrow B$ we show equivalently that $\bar{B} \Rightarrow \bar{A}$. Thus, let us assume first that $\mathrm{Ker}(\mathrm{T}) \neq \{\mathbf{0}\}$, then there exists $X_0 \neq \mathbf{0}$ such that $\mathrm{T}(X_0) = \mathbf{0}$. In addition, for any $X \in V$ one has

$$\mathrm{T}(X + X_0) = \mathrm{T}(X) + \mathrm{T}(X_0) = \mathrm{T}(X) + \mathbf{0} = \mathrm{T}(X).$$

Since $X \neq X + X_0$ but $\mathrm{T}(X) = \mathrm{T}(X + X_0)$, one concludes that T is not injective. By contraposition, one has shown that T injective implies that $\mathrm{Ker}(\mathrm{T}) = \{\mathbf{0}\}$.

(ii) Assume now that $\mathrm{Ker}(\mathrm{T}) = \{\mathbf{0}\}$, and consider $X_1, X_2 \in V$ with $X_1 \neq X_2$. Then one has

$$\mathrm{T}(X_1) - \mathrm{T}(X_2) = \mathrm{T}(X_1 - X_2) \neq \mathbf{0}$$

since $X_1 - X_2 \neq \mathbf{0}$. As a consequence, $\mathrm{T}(X_1) \neq \mathrm{T}(X_2)$. □

Let us provide a final theorem for this section, which is useful in many situations.

Theorem 4.7.8. *Let* $\mathrm{T} : V \to W$ *be a linear map between the vector spaces* V *and* W, *and assume that* $\dim(V) = \dim(W) < \infty$. *Then the following assertions are equivalent:*

(i) $\mathrm{Ker}(\mathrm{T}) = \{\mathbf{0}\}$,

(ii) T *is invertible,*

(iii) T *is surjective.*

Proof. The implication $(ii) \Rightarrow (i)$ and $(ii) \Rightarrow (iii)$ are direct consequences of Theorem 4.7.5 and Lemma 4.7.7.

Assume now (i), and recall from Lemma 4.7.7 that this condition corresponds to T injective. Then from Theorem 4.3.5 and more precisely from the equality

$$\underbrace{\dim\big(\mathrm{Ker}(\mathrm{T})\big)}_{0} + \dim\big(\mathrm{Ran}(\mathrm{T})\big) = \dim(V)$$

one deduces that $\dim\big(\mathrm{Ran}(\mathrm{T})\big) = \dim(V) = \dim(W)$, where the assumption about the dimension has been taken into account. It is enough then to observe that

$$\dim\big(\mathrm{Ran}(\mathrm{T})\big) = \dim(W)$$

means that T is surjective. Since T is also injective, it follows that T is bijective. Since bijectivity corresponds to invertibility by Theorem 4.7.5, one infers that (ii) holds.

Assume now that (iii) holds. By taking again Theorem 4.3.5 into account, one deduces that from the equality

$$\dim\big(\mathrm{Ran}(\mathrm{T})\big) = \dim(W) = \dim(V)$$

that $\dim\big(\mathrm{Ker}(\mathrm{T})\big) = 0$, meaning that T is injective. Again, it implies that T is bijective, and thus invertible, and thus that (ii) holds. ☐

Corollary 4.7.9. *For any $\mathcal{A} \in M_n(\mathbb{F})$, the following statements are equivalent:*

(i) There exists $\mathcal{B} \in M_n(\mathbb{F})$ such that $\mathcal{B}\mathcal{A} = \mathbf{1}_n$,

(ii) There exists $\mathcal{C} \in M_n(\mathbb{F})$ such that $\mathcal{A}\mathcal{C} = \mathbf{1}_n$.

In addition, whenever (i) or (ii) holds, then $\mathcal{B} = \mathcal{C}$, and $\mathcal{A}$ is invertible with $\mathcal{A}^{-1} = \mathcal{B} = \mathcal{C}$.

4.8 Exercises

Exercise 4.1. *Let* $\mathrm{F}:\mathbb{R}^2\to\mathbb{R}^2$ *be the map defined by* $\mathrm{F}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\left(\begin{smallmatrix}2x\\3y\end{smallmatrix}\right)$ *for any* $\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)\in\mathbb{R}^2$. *Describe the image by* F *of the points lying on the unit circle centered at* $\mathbf{0}$, *i.e.* $\left\{\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)\in\mathbb{R}^2 \mid x^2+y^2=1\right\}$.

Exercise 4.2. *Let* $\mathrm{F}:\mathbb{R}^2\to\mathbb{R}^2$ *be the map defined by* $\mathrm{F}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\left(\begin{smallmatrix}xy\\y\end{smallmatrix}\right)$ *for any* $\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)\in\mathbb{R}^2$. *Describe the image by* F *of the line* $\{\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)\in\mathbb{R}^2 \mid x=2\}$.

Exercise 4.3. *Let* V *be a vector space of dimension* n, *and let* $\{X_1,\dots,X_n\}$ *be a basis for* V. *Let* F *be a linear map from* V *into itself. Show that* F *is uniquely defined if one knows* $\mathrm{F}(X_j)$ *for* $j\in\{1,\dots,n\}$. *Is it also true if* F *is an arbitrary map from* V *into itself ?*

Exercise 4.4. *Let* V,W *be vector spaces over the same field, and let* $\mathrm{T}:V\to W$ *be a linear map. Show that the following set is a subspace of* V*:*

$$\{X\in V \mid \mathrm{T}(X)=\mathbf{0}\}.$$

This subspace is called the kernel of T.

Exercise 4.5. *Show that the image of a convex set under a linear map is a convex set.*

Exercise 4.6. *Determine which of the following maps are linear:*

a) $\mathrm{F}:\mathbb{R}^3\to\mathbb{R}^2$ *defined by* $\mathrm{F}\begin{pmatrix}x\\y\\z\end{pmatrix}=\left(\begin{smallmatrix}x\\z\end{smallmatrix}\right)$,

b) $\mathrm{F}:\mathbb{R}^4\to\mathbb{R}^4$ *defined by* $\mathrm{F}(X)=-X$ *for all* $X\in\mathbb{R}^4$,

c) $\mathrm{F}:\mathbb{R}^3\to\mathbb{R}^3$ *defined by* $\mathrm{F}(X)=X+\begin{pmatrix}0\\-1\\0\end{pmatrix}$ *for all* $X\in\mathbb{R}^3$,

d) $\mathrm{F}:\mathbb{R}^2\to\mathbb{R}^2$ *defined by* $\mathrm{F}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\left(\begin{smallmatrix}2x\\y-x\end{smallmatrix}\right)$,

e) $\mathrm{F}:\mathbb{R}^2\to\mathbb{R}^2$ *defined by* $\mathrm{F}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\left(\begin{smallmatrix}y\\x\end{smallmatrix}\right)$,

f) $\mathrm{F}:\mathbb{R}^2\to\mathbb{R}$ *defined by* $\mathrm{F}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=xy$.

Exercise 4.7. *Determine the kernel and the range of the maps defined in the previous exercise.*

Exercise 4.8. *Consider the subset of* $\mathbb{R}^n$ *consisting of all vectors* ${}^t(x_1,\dots,x_n)$ *such that* $x_1+x_2+\cdots+x_n=0$. *Is it a subspace of* $\mathbb{R}^n$ *? If so, what is its dimension ?*

Exercise 4.9. *Let* $\mathrm{P}:M_n(\mathbb{R})\to M_n(\mathbb{R})$ *be the map defined for any* $\mathcal{A}\in M_n(\mathbb{R})$ *by*

$$\mathrm{P}(\mathcal{A})=\frac{1}{2}\big(\mathcal{A}+{}^t\!\mathcal{A}\big).$$

1. *Show that* P *is a linear map.*

2. *Show that the kernel of* P *consists in the vector space of all skew-symmetric matrices.*

3. *Show that the range of* P *consists in the vector space of all symmetric matrices.*

4. *What is the dimension of the vector space of all symmetric matrices, and the dimension of the vector space of all skew-symmetric matrices ?*

Exercise 4.10. *Let $C^\infty(\mathbb{R})$ be the vector space of all real functions on $\mathbb{R}$ which admit derivatives of all orders. Let $\mathrm{D} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the map which associates to any $f \in C^\infty(\mathbb{R})$ its derivative, i.e. $\mathrm{D}f = f'$.*

1. *Is* D *a linear map ?*

2. *What is the kernel of* D *?*

3. *What is the kernel of D^n, for any $n \in \mathbb{N}$, and what is the dimension of this vector space ?*

Exercise 4.11. *Consider the map $\mathrm{F} : \mathbb{R}^3 \to \mathbb{R}^4$ defined by*

$$\mathrm{F}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ x-y \\ x-z \\ x-y-z \end{pmatrix}.$$

1. *Is* F *a linear map ? (Justify your answer)*

2. *Determine the kernel of* F.

3. *Determine the range of* F.

Exercise 4.12. *What is the dimension of the space of solutions of the following systems of linear equations ? In each case, find a basis for the space of solutions.*

$$a)\ \begin{cases} 2x+y-z &= 0 \\ 2x+y+z &= 0 \end{cases} \qquad b)\ \begin{cases} x-y+z &= 0 \end{cases} \qquad c)\ \begin{cases} 4x+7y-\pi z &= 0 \\ 2x-y+z &= 0 \end{cases}$$

and

$$d)\ \begin{cases} x+y+z &= 0 \\ x-y &= 0 \\ y+z &= 0 \end{cases}$$

Exercise 4.13. *Let $\mathcal{A}$ be the matrix given by $\mathcal{A} = \left(\begin{smallmatrix} 0 & 1 & 3 & -2 \\ 2 & 1 & -4 & 3 \\ 2 & 3 & 2 & -1 \end{smallmatrix}\right)$ and consider the linear map $\mathrm{L}_\mathcal{A} : \mathbb{R}^4 \to \mathbb{R}^3$ defined by $\mathrm{L}_\mathcal{A}X = \mathcal{A}X$ for all $X \in \mathbb{R}^4$.*

1. *Determine the rank of $\mathcal{A}$ and the dimension of the range of $\mathrm{L}_\mathcal{A}$.*

2. *Deduce the dimension of the kernel of* $\mathrm{L}_{\mathcal{A}}$*, and exhibit a basis for the kernel of* $\mathrm{L}_{\mathcal{A}}$*.*

3. *Find the set of all solutions of the equation* $\mathcal{A}X = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$*.*

Exercise 4.14. *Let* $\mathrm{F} : \mathbb{R}^3 \to \mathbb{R}^2$ *be the map indicated below. What is the matrix associated with* F *in the canonical bases of* $\mathbb{R}^3$ *and* $\mathbb{R}^2$ *?*

$$a)\quad \mathrm{F}(E_1) = \begin{pmatrix} 1 \\ -3 \end{pmatrix},\ \mathrm{F}(E_2) = \begin{pmatrix} -4 \\ 2 \end{pmatrix},\ \mathrm{F}(E_3) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

and

$$b)\quad \mathrm{F}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 - 2x_2 + x_3 \\ 4x_1 - x_2 + 5x_3 \end{pmatrix}.$$

Exercise 4.15. *Let* $\mathrm{L} : \mathbb{R}^3 \to \mathbb{R}^3$ *be a linear map which associated matrix has the form* $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ *with respect to the canonical basis of* $\mathbb{R}^3$*. What is the matrix associated with* L *in the basis generated by the three vectors* $V_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$*,* $V_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$*,* $V_3 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$

Exercise 4.16. *For any* $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$*, one says that* $\mathcal{A}$ *and* $\mathcal{B}$ commute *if* $\mathcal{AB} = \mathcal{BA}$*.*

a) *Show that the set of all matrices which commute with* $\mathcal{A}$ *is a subspace of* $M_n(\mathbb{R})$*,*

b) *If* $\mathcal{A} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$*, exhibit a basis of the subspace of all matrices which commute with* $\mathcal{A}$*.*

Exercise 4.17. *Let* U, V, W *be vector spaces over a field* $\mathbb{F}$*, and let* $\mathrm{G} : U \to V$*,* $\mathrm{G}' : U \to V$*,* $\mathrm{H} : V \to W$ *and* $\mathrm{H}' : V \to W$ *be linear maps. Show that*

(i) $\mathrm{H} \circ \mathrm{G} : U \to W$ *is a linear map,*

(ii) $(\mathrm{H} + \mathrm{H}') \circ \mathrm{G} = \mathrm{H} \circ \mathrm{G} + \mathrm{H}' \circ \mathrm{G}$*,*

(iii) $\mathrm{H} \circ (\mathrm{G} + \mathrm{G}') = \mathrm{H} \circ \mathrm{G} + \mathrm{H} \circ \mathrm{G}'$*,*

(iv) $(\lambda \mathrm{H}) \circ \mathrm{G} = \mathrm{H} \circ (\lambda \mathrm{G}) = \lambda(\mathrm{H} \circ \mathrm{G})$*,* *for all* $\lambda \in \mathbb{F}$*.*

Exercise 4.18. *Let* V *be a real vector space, and let* $\mathrm{P} : V \to V$ *be a linear map satisfying* $\mathrm{P}^2 = \mathrm{P}$*. Such a linear map is called* a projection*.*

(i) *Show that* $\mathbf{1} - \mathrm{P}$ *is also a projection, and that* $(\mathbf{1} - \mathrm{P})\mathrm{P} = \mathrm{P}(\mathbf{1} - \mathrm{P}) = \mathbf{0}$*,*

(ii) *Show that* $V = \mathrm{Ker}(\mathrm{P}) + \mathrm{Ran}(\mathrm{P})$*,*

(iii) *Show that the intersection of* $\mathrm{Ker}(\mathrm{P})$ *and* $\mathrm{Ran}(\mathrm{P})$ *is* $\{\mathbf{0}\}$*.*

Exercise 4.19. *Let* $\mathrm{L}:\mathbb{R}^2\to\mathbb{R}^2$ *be the linear map defined by* $\mathrm{L}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\left(\begin{smallmatrix}x+y\\x-y\end{smallmatrix}\right)$. *Show that* L *is invertible and find its inverse. Same question with the map* $\mathrm{L}:\mathbb{R}^3\to\mathbb{R}^3$ *defined by* $\mathrm{L}\left(\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right)=\left(\begin{smallmatrix}x-y\\x+z\\x+y+3z\end{smallmatrix}\right)$.

Exercise 4.20. *Let* F,G *be invertible linear maps from a vector space into itself. Show that* $(\mathrm{G}\circ\mathrm{F})^{-1}=\mathrm{F}^{-1}\circ\mathrm{G}^{-1}$.

Exercise 4.21. *Show that the matrix* $\mathcal{B}:\mathbb{R}^n\to\mathbb{R}^n$ *defining a change of basis in* $\mathbb{R}^n$ *is always invertible.*

Exercise 4.22. *Let* V *be the set of all infinite sequences of real numbers* $(x_1,x_2,x_3,\dots)$. *We endow* V *with the pointwise addition and multiplication, i.e.*

$$(x_1,x_2,x_3,\dots)+(x_1',x_2',x_3',\dots)=(x_1+x_1',x_2+x_2',x_3+x_3',\dots)$$

and $\lambda(x_1,x_2,x_3,\dots)=(\lambda x_1,\lambda x_2,\lambda x_3,\dots)$, *which make* V *an infinite dimensional vector space.*

Define the map $\mathrm{F}:V\to V$, *called* shift operator, *by*

$$\mathrm{F}(x_1,x_2,x_3,\dots)=(0,x_1,x_2,x_3,\dots).$$

(i) Is F *a linear map ?*

(ii) Is F *injective, and what is the kernel of* F *?*

(iii) Is F *surjective ?*

(iv) Show that there is a linear map $\mathrm{G}:V\to V$ *such that* $\mathrm{G}\circ\mathrm{F}=\mathbf{1}$.

(v) Does the map G *have the property that* $\mathrm{F}\circ\mathrm{G}=\mathbf{1}$ *?*

(vi) What is different from the finite dimensional case, i.e. when V *is of finite dimension ?*

Exercise 4.23. *Consider the matrices*

$$\mathcal{A}=\begin{pmatrix}2&0\\0&2\end{pmatrix},\quad \mathcal{B}=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad \mathcal{C}=\begin{pmatrix}-1&0\\0&1\end{pmatrix},\quad \mathcal{D}=\begin{pmatrix}0&1\\-1&0\end{pmatrix},$$
$$\mathcal{E}=\begin{pmatrix}1&0.2\\0&1\end{pmatrix},\quad \mathcal{F}=\begin{pmatrix}1&-1\\1&1\end{pmatrix},$$

and show their effect on the letter L *defined by the three points* $\left(\begin{smallmatrix}0\\2\end{smallmatrix}\right),\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right),\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ *of* $\mathbb{R}^2$.

Exercise 4.24. *Let* $N=\left(\begin{smallmatrix}n_1\\n_2\end{smallmatrix}\right)$ *be a vector in* $\mathbb{R}^2$ *with* $\|N\|=1$, *and let* ℓ *be the line in* $\mathbb{R}^2$ *passing trough* $\mathbf{0}\in\mathbb{R}^2$ *and parallel to* N. *Then any vector* $X\in\mathbb{R}^2$ *can be written uniquely as* $X=X_\parallel+X_\perp$, *where* $X_\parallel$ *is a vector parallel to* ℓ *and* $X_\perp$ *is a vector perpendicular to* ℓ. *Show that there exists a projection* $\mathrm{P}\in M_2(\mathbb{R})$ *such that* $X_\parallel=\mathrm{P}X$, *and express* P *in terms of* n_1 *and* n_2.

Exercise 4.25. *1) Do the same exercise in* $\mathbb{R}^3$ *with* N *given by* $\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$.

2) Show that there also exists a projection Q *such that* $X_\perp = \mathrm{Q}\, X$. *If* $H_{\mathbf{0},N}$ *is the plane passing through* $\mathbf{0} \in \mathbb{R}^3$ *and perpendicular to* N, *show that* $X_\perp \in H_{\mathbf{0},N}$.

Exercise 4.26. *In the framework of the previous exercise, a reflection of* X *about* $H_{\mathbf{0},N}$ *is defined by the vector* $X_{\mathrm{ref}} := X_\perp - X_\parallel$. *Show that* $\|X_{\mathrm{ref}}\| = \|X\|$, *and provide the expression for the linear map transforming* X *into* X_{ref}.

Exercise 4.27. *Prove Corollary 4.7.9.*

Exercise 4.28. *Block matrices are matrices which are partitioned into rectangular sub-matrices called* blocks. *For example, let* $\mathcal{A} \in M_{n+m}(\mathbb{R})$ *be the block matrix*

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix}$$

with $\mathcal{A}_{11} \in M_n(\mathbb{R})$, $\mathcal{A}_{22}(\mathbb{R}) \in M_m(\mathbb{R})$, $\mathcal{A}_{12} \in M_{n\times m}(\mathbb{R})$, *and* $\mathcal{A}_{21} \in M_{m\times n}(\mathbb{R})$. *Such matrices can be multiplied as if every blocks where scalars (with the usual multiplication of matrices), as long as the products are well defined. For example, check this statement by computing the product* $\mathcal{AB}$ *in two different ways with the following matrices:* $\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \end{pmatrix}$ *with* $\mathcal{A}_{11} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *and* $\mathcal{A}_{12} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, *and* $\mathcal{B} = \begin{pmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} \\ \mathcal{B}_{21} & \mathcal{B}_{22} \end{pmatrix}$ *with* $\mathcal{B}_{11} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$, $\mathcal{B}_{12} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$, $\mathcal{B}_{21} = \begin{pmatrix} 7 & 8 \end{pmatrix}$, *and* $\mathcal{B}_{22} = \begin{pmatrix} 9 \end{pmatrix}$.

Exercise 4.29. *Let* $\mathcal{A} \in M_{n+m}(\mathbb{R})$ *be the block matrix*

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{O} & \mathcal{A}_{22} \end{pmatrix}$$

with $\mathcal{A}_{11} \in M_n(\mathbb{R})$, $\mathcal{A}_{22}(\mathbb{R}) \in M_m(\mathbb{R})$ *and* $\mathcal{A}_{12} \in M_{n\times m}(\mathbb{R})$.

(i) For which choice of $\mathcal{A}_{11}$, $\mathcal{A}_{12}$ *and* $\mathcal{A}_{22}$ *is* $\mathcal{A}$ *invertible ?*

(ii) If $\mathcal{A}$ *is invertible, what is* $\mathcal{A}^{-1}$, *in terms of* $\mathcal{A}_{11}$, $\mathcal{A}_{12}$ *and* $\mathcal{A}_{22}$ *?*

Chapter 5

Scalar product and orthogonality

5.1 Scalar product

Recall that the notion of a vector space has been introduced as an abstract version of the properties shared both by $\mathbb{R}^n$ and by $M_{mn}(\mathbb{R})$. Similarly, we have introduced the scalar product on $\mathbb{R}^n$ already in Chapter 1, let us now consider an abstract version of it. For simplicity, we introduce it on real vector spaces, but a slightly more general version will be considered once the complex numbers will be at our disposal.

Definition 5.1.1. A scalar product *on a real vector space* V *is a map* $\langle\cdot,\cdot\rangle : V\times V \to \mathbb{R}$ *such that for any* $X, Y, Z \in V$ *and* $\lambda \in \mathbb{R}$ *one has*

(i) $\langle X, Y\rangle = \langle Y, X\rangle$,

(ii) $\langle X+Y, Z\rangle = \langle X, Z\rangle + \langle Y, Z\rangle$,

(iii) $\langle \lambda X, Y\rangle = \lambda\langle X, Y\rangle$,

(iv) $\langle X, X\rangle \geq 0$ *and* $\langle X, X\rangle = 0$ *if and only if* $X = \mathbf{0}$.

Example 5.1.2. *For* $V = \mathbb{R}^n$ *and* $X, Y \in V$ *one sets* $\langle X, Y\rangle := X\cdot Y$ *and one can check that the four conditions above are satisfied.*

Example 5.1.3. *For* $a, b \in \mathbb{R}$ *with* $a < b$ *one considers* $V = C\big([a,b];\mathbb{R}\big)$ *and for any* $f, g \in V$ *one defines*

$$\langle f, g\rangle := \int_a^b f(x)g(x)\mathrm{d}x.$$

It is easily checked that this defines a scalar product on V*, see Exercise 5.5. For information, this scalar product extends to the set of* L^2*-functions (the set of square integrable functions).*

Definition 5.1.4. *If* V *is a real vector space endowed with a scalar product, one says that* $X, Y \in V$ *are* orthogonal *if* $\langle X, Y\rangle = 0$*, and one writes* $X \perp Y$*. If* S *is a subset of* V*, one writes*

$$S^{\perp} := \{Y \in V \mid \langle X, Y\rangle = 0 \text{ for all } X \in S\}$$

and call it the orthogonal subspace of S.

One easily shows that $S^\perp$ is always a subspace of V.

Definition 5.1.5. *For any real vector space V endowed with a scalar product and for any $X \in V$ we set*

$$\|X\| := \sqrt{\langle X, X\rangle}$$

and call it the norm of X *(associated with the scalar product $\langle \cdot, \cdot \rangle$).*

Lemma 5.1.6. *For any real vector space V endowed with a scalar product, for any $X, Y \in V$ and for $\lambda \in \mathbb{R}$ one has*

(i) $\|\lambda X\| = |\lambda|\,\|X\|$,

(ii) $\|X + Y\|^2 = \|X\|^2 + \|Y\|^2$ *if and only if* $X \perp Y$ *(Pythagoras theorem)*

(iii) $\|X + Y\|^2 + \|X - Y\|^2 = 2\|X\|^2 + 2\|Y\|^2$,

(iv) $\|X + Y\| \le \|X\| + \|Y\|$.

The proof will be provided in Exercise 5.1. The following statement is a generalization of a property already derived in the context of $\mathbb{R}^n$.

Lemma 5.1.7. *For any real vector space V endowed with a scalar product and for any $X, Y \in V$ one has*

$$|\langle X, Y\rangle| \le \|X\|\,\|Y\|. \tag{5.1.1}$$

Proof. Let us first consider the trivial case $Y = \mathbf{0}$ for which (5.1.1) is an equality with 0 on both sides.

Now, assume that $Y \neq \mathbf{0}$ and set $c := \frac{\langle X,Y\rangle}{\|Y\|^2}$. Then let us observe that $(X - cY)\perp Y$, since

$$\langle X - cY, Y\rangle = \langle X, Y\rangle - \frac{\langle X, Y\rangle}{\|Y\|^2}\langle Y, Y\rangle = 0.$$

It follows by Pythagoras theorem that

$$\|X\|^2 = \|(X - cY) + cY\|^2 = \|X - cY\|^2 + \|cY\|^2 = \|X - cY\|^2 + c^2\|Y\|^2,$$

which implies that $\|X\|^2 \ge c^2\|Y\|$, or equivalently $\|X\| \ge |c|\,\|Y\|$. Note that this inequality can also be rewritten as $|c| \le \frac{\|X\|}{\|Y\|}$.

By collecting these information one gets

$$|\langle X, Y\rangle| = |c|\,\|Y\|^2 \le \frac{\|X\|}{\|Y\|}\|Y\|^2 = \|X\|\,\|Y\|,$$

which corresponds to the claim. ☐

5.2 Orthogonal bases

Definition 5.2.1. *Let V be a real vector space endowed with a scalar product, and let $\{V_1, \dots, V_n\}$ be a basis for V. The basis is called* orthogonal *if $\langle V_i, V_j\rangle = 0$ whenever $i, j \in \{1, \dots, n\}$ and $i \neq j$. If in addition $\langle V_i, V_i\rangle = 1$ for any $i \in \{1, \dots, n\}$ the basis is called* orthonormal.

Example 5.2.2. *The standard basis $\{E_1, \dots, E_n\}$ of $\mathbb{R}^n$ is an orthonormal basis.*

The following result is of conceptual importance, and rather well-known.

Theorem 5.2.3 (Graham-Schmidt)**.** *Let V be a real vector space of dimension n endowed with a scalar product. Then there exists an orthonormal basis for V.*

The proof consists in the explicit construction of an orthonormal basis.

Proof. Let $\{V_1, \dots, V_n\}$ be an arbitrary basis for V (such a basis exists since otherwise the dimension of V would not be defined), and let us set

$$\begin{aligned} V_1' &:= \frac{1}{\|V_1\|} V_1 \\ V_2' &:= \frac{1}{\|V_2 - \langle V_2, V_1'\rangle V_1'\|}\big(V_2 - \langle V_2, V_1'\rangle V_1'\big) \\ &\vdots \\ V_n' &:= \frac{1}{\big\|V_n - \sum_{i=1}^{n-1}\langle V_n, V_i'\rangle V_i'\big\|}\Big(V_n - \sum_{i=1}^{n-1}\langle V_n, V_i'\rangle V_i'\Big), \end{aligned}$$

where the prefactors are chosen such that $\|V_j'\| = 1$ (note that $V_j - \sum_{i=1}^{j-1}\langle V_j, V_i'\rangle V_i'$ is always different from $\mathbf{0}$ since otherwise V_j would be a linear combination of $V_1, \dots, V_{j-1}$ which is not possible by assumption). Then, it simply remains to observe that $V_j' \perp V_k'$ for any $j \neq k$. As a consequence, the elements V_j' generate an orthonormal basis for V, as expected. □

5.3 Bilinear maps

The notion of bilinear maps will be useful for calculus II.

Definition 5.3.1. *Let V, W, U be vector spaces over the same field $\mathbb{F}$. A map $\mathrm{T} : V \times W \to U$ is* bilinear *if it is linear in each argument, namely for any $X, X_1, X_2 \in V$, any $Y, Y_1, Y_2 \in W$ and $\lambda \in \mathbb{F}$ one has*

(i) $\mathrm{T}(X_1 + X_2, Y) = \mathrm{T}(X_1, Y) + \mathrm{T}(X_2, Y)$,

(ii) $\mathrm{T}(\lambda X, Y) = \lambda \mathrm{T}(X, Y)$,

(iii) $\mathrm{T}(X, Y_1 + Y_2) = \mathrm{T}(X, Y_1) + \mathrm{T}(X, Y_2)$,

(iv) $\mathrm{T}(X, \lambda Y) = \lambda \mathrm{T}(X, Y)$.

Example 5.3.2. *The scalar product* $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is a bilinear map on the Euclidean space* $\mathbb{R}^n$.

Example 5.3.3. *If* $\mathcal{A} \in M_{mn}(\mathbb{F})$ *one can define a bilinear map* $\mathrm{F}_{\mathcal{A}} : \mathbb{F}^m \times \mathbb{F}^n \to \mathbb{F}$ *for any* $X \in \mathbb{F}^m$ *and* $Y \in \mathbb{F}^n$ *by*

$$\mathrm{F}_{\mathcal{A}}(X, Y) = {}^tX\mathcal{A}Y \equiv \underbrace{{}^tX}_{\in M_{1m}(\mathbb{F})} \underbrace{\mathcal{A}}_{\in M_{mn}(\mathbb{F})} \underbrace{Y}_{\in M_{n1}(\mathbb{F})} \in \mathbb{F}. \tag{5.3.1}$$

Note that it is easily checked that $\mathrm{F}_{\mathcal{A}}$ *is indeed a bilinear map. For example, if* $\mathcal{A} = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$, $X = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ *and* $Y = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$, *then*

$$\mathrm{F}_{\mathcal{A}}(X, Y) = (1\ 0) \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (1\ 0) \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2.$$

More generally, observe that if $\mathcal{A} = (a_{ij})$, $X = {}^t(x_1, \dots, x_m)$ *and* $Y = {}^t(y_1, \dots, y_n)$ *then*

$${}^tX\mathcal{A}Y = X \cdot (\mathcal{A}Y) = \sum_{i=1}^m x_i (\mathcal{A}Y)_i = \sum_{i=1}^m x_i \sum_{j=1}^n a_{ij}\, y_j = \sum_{i=1}^m \sum_{j=1}^n a_{ij}\, x_i\, y_j.$$

We shall now see that many bilinear maps are of the form presented in the previous example. For that purpose, recall from Section 4.5 that if $\mathcal{V} = \{V_1, \dots, V_m\}$ is a basis for a vector space V over $\mathbb{F}$ and if $\mathcal{X} \in V$ then the coordinate vector of $\mathcal{X}$ is the element $X = {}^t(x_1, \dots, x_m) \in \mathbb{F}^m$ such that $\mathcal{X} = x_1 V_1 + \dots + x_m V_m$. One has already introduced the notation $(\mathcal{X})_{\mathcal{V}} = X$. Similarly, for a basis $\mathcal{W} = \{W_1, \dots, W_n\}$ of a vector space W over $\mathbb{F}$ and for any $\mathcal{Y} \in W$ one sets $(\mathcal{Y})_{\mathcal{W}} = Y = {}^t(y_1, \dots, y_n) \in \mathbb{F}^n$ for its coordinate vector.

Lemma 5.3.4. *Let* V, W *be vector spaces over a field* $\mathbb{F}$ *and let* $\mathrm{F} : V \times \mathcal{W} \to \mathbb{F}$ *be a bilinear map. If* $\mathcal{V} = \{V_1, \dots, V_m\}$ *is a basis for* V, *and if* $\mathcal{W} = \{W_1, \dots, W_n\}$ *is a basis for* W *then there exists* $\mathcal{A} \in M_{mn}(\mathbb{F})$ *such that*

$$\mathrm{F}(\mathcal{X}, \mathcal{Y}) = {}^tX\mathcal{A}Y$$

for any $\mathcal{X} \in V$, *any* $\mathcal{Y} \in W$ *and with* $X = (\mathcal{X})_{\mathcal{V}}$ *and* $Y = (\mathcal{Y})_{\mathcal{W}}$.

Proof. By taking the bilinearity of F into account, one has

$$\mathrm{F}(\mathcal{X}, \mathcal{Y}) = \mathrm{F}\Big(\sum_{i=1}^m x_i V_i, \sum_{j=1}^n y_j W_j\Big) = \sum_{i=1}^m \sum_{j=1}^n x_i y_j \mathrm{F}(V_i, W_j).$$

Thus, by setting $a_{ij} = \mathrm{F}(V_i, W_j) \in \mathbb{F}$ one deduces that

$$\mathrm{F}(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}\, x_i\, y_j = {}^tX\mathcal{A}Y$$

with $\mathcal{A} = (a_{ij})$. □

Remark 5.3.5. *If V, W, U are vector spaces over the same field $\mathbb{F}$ and if $\mathrm{F}_i : V \times W \to U$ are bilinear maps for $i = 1, 2$, then $\mathrm{F}_1 + \mathrm{F}_2 : V \times W \to U$ is a bilinear map, and $\lambda \mathrm{F}_i$ is also a bilinear map. Thus, the set of bilinear maps from $V \times W$ to U is a vector space.*

Let us end this section with two questions:

Question: Let $V = W = \mathbb{R}^n$ and consider the map F defined by the usual scalar product

$$\mathrm{F}(X, Y) = X \cdot Y \qquad \text{for any } X, Y \in \mathbb{R}^n.$$

In view of Lemma 5.3.4, what is the matrix associated with this bilinear map with respect to the canonical basis of $\mathbb{R}^n$?

Question: How does a bilinear map change when one performs a change of bases for the vector spaces V and W ?

5.4 Exercises

Exercise 5.1. *Let V be a real vector space endowed with a scalar product. Prove the following relations for $X, Y \in V$ and $\lambda \in \mathbb{R}$:*

(i) $\|\lambda X\| = |\lambda|\,\|X\|$,

(ii) $\|X+Y\|^2 = \|X\|^2 + \|Y\|^2$ *if and only if* $X \perp Y$,

(iii) $\|X+Y\|^2 + \|X-Y\|^2 = 2\|X\|^2 + 2\|Y\|^2$,

(iv) $\|X+Y\| \leq \|X\| + \|Y\|$.

Exercise 5.2. *Let $\mathcal{A} = (a_{jk}) \in M_n(\mathbb{R})$ and define $\mathrm{Tr}(\mathcal{A}) = \sum_{j=1}^n a_{jj}$, where $\mathrm{Tr}(\mathcal{A})$ is called* the trace of $\mathcal{A}$. *Show the following properties:*

(i) $\mathrm{Tr} : M_n(\mathbb{R}) \to \mathbb{R}$ *is a linear map,*

(ii) $\mathrm{Tr}(\mathcal{A}\mathcal{B}) = \mathrm{Tr}(\mathcal{B}\mathcal{A})$, *for any* $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$,

(iii) If $\mathcal{C} \in M_n(\mathbb{R})$ *is an invertible matrix, then* $\mathrm{Tr}(\mathcal{C}^{-1}\mathcal{A}\mathcal{C}) = \mathrm{Tr}(\mathcal{A})$,

(iv) If $M_n^s(\mathbb{R})$ *denotes the vector space of all $n \times n$ symmetric matrices, then the map*

$$M_n^s(\mathbb{R}) \times M_n^s(\mathbb{R}) \ni (\mathcal{A}, \mathcal{B}) \mapsto \mathrm{Tr}(\mathcal{A}\mathcal{B}) \in \mathbb{R}$$

defines a scalar product on $M_n^s(\mathbb{R})$. We recall that a matrix $\mathcal{A}$ is symmetric if $\mathcal{A} = {}^t\mathcal{A}$.

Exercise 5.3. *Find an orthonormal basis for the subspace of $\mathbb{R}^4$ defined by the three vectors* $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}$ *and* $\begin{pmatrix} 1 \\ -2 \\ 0 \\ 0 \end{pmatrix}$.

Exercise 5.4. *Find an orthonormal basis for the space of solutions of the following systems:*

$$a)\ \begin{cases} 2x+y-z &= 0 \\ 2x+y+z &= 0 \end{cases} \quad b)\ \begin{cases} x-y+z &= 0 \end{cases} \quad c)\ \begin{cases} 4x+7y-\pi z &= 0 \\ 2x-y+z &= 0 \end{cases}$$

$$d)\ \begin{cases} x+y+z &= 0 \\ x-y &= 0 \\ y+z &= 0 \end{cases}$$

Exercise 5.5. *We consider the real vector space $V := C([0,1])$ made of continuous real functions on $[0,1]$ and endow it with the map*

$$V \times V \ni (f,g) \mapsto \langle f, g \rangle := \int_0^1 f(x)\,g(x)\,\mathrm{d}x \ \in \mathbb{R}.$$

Show that

(i) $\langle \cdot, \cdot \rangle$ *is a scalar product on* V,

(ii) If W *is the subspace of* V *generated by the three functions* $x \mapsto 1$ *(constant function),* $x \mapsto x$ *(identity function), and* $x \mapsto x^2$*, find an orthonormal basis for* W.

Exercise 5.6. *For any symmetric matrix* $\mathcal{A} = (a_{ij}) \in M_n(\mathbb{R})$*, we define the map*

$$\mathrm{F}_{\mathcal{A}} : \mathbb{R}^n \times \mathbb{R}^n \ni (X, Y) \mapsto \mathrm{F}_{\mathcal{A}}(X, Y) := {}^t X \mathcal{A} Y \in \mathbb{R}.$$

(i) Show that $\mathrm{F}_{\mathcal{A}}$ *is a bilinear map,*

(ii) Show that $\mathrm{F}_{\mathcal{A}}(X, Y) = \mathrm{F}_{\mathcal{A}}(Y, X)$ *for any* $X, Y \in \mathbb{R}^n$.

(iii) When does $\mathrm{F}_{\mathcal{A}}$ *define a scalar product ?*

(iv) If $\mathcal{A}$ *is one of the following matrices, does* $\mathrm{F}_{\mathcal{A}}$ *define a scalar product ?*

$$\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \qquad \mathcal{A} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Chapter 6

The determinant

6.1 Multilinear maps

In this first section, we generalize the notions of linear maps and bilinear maps.

Definition 6.1.1. *Let V be a vector space over a field $\mathbb{F}$, and let $n \in \mathbb{N}$. A map*

$$\mathrm{T} : \underbrace{V \times V \times \cdots \times V}_{n \text{ terms}} \to \mathbb{F}$$

*is n-*linear *if it is linear in each argument,* i.e.

$$\begin{aligned}&\mathrm{T}(X_1, X_2, \dots, X_j + X_j', \dots, X_n) \\ &= \mathrm{T}(X_1, X_2, \dots, X_j, \dots, X_n) + \mathrm{T}(X_1, X_2, \dots, X_j', \dots, X_n)\end{aligned}$$

and

$$\mathrm{T}(X_1, X_2, \dots, \lambda X_j, \dots, X_n) = \lambda \mathrm{T}(X_1, X_2, \dots, X_j, \dots, X_n)$$

for any $X_1, \dots, X_j, X_j', \dots, X_n \in V$, $\lambda \in \mathbb{F}$ and $j \in \{1, \dots, n\}$. The set of all n-linear maps is denoted by $\mathrm{Mult}_n(V)$.

Note that if $n = 1$ one speaks about a linear map, while $n = 2$ corresponds to a bilinear map. Without difficulty one can show that the set of $\mathrm{Mult}_n(V)$ is a vector space.

Definition 6.1.2. *An element $\mathrm{T} \in \mathrm{Mult}_n(V)$ is* alternating *if*

$$\mathrm{T}(X_1, \dots, X_i, \dots, X_j, \dots, X_n) = 0$$

whenever $X_i = X_j$ for some $i, j \in \{1, \dots, n\}$ with $i \neq j$.

Example 6.1.3. *If $\mathcal{A} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, then the bilinear map $\mathrm{F}_{\mathcal{A}} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined in* (5.3.1) *is alternating. Indeed, if $X = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ for any $x, y \in \mathbb{R}$, then one has*

$$\mathrm{F}_{\mathcal{A}}(X, X) = (x\ y) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (x\ y) \left(\begin{smallmatrix} y \\ -x \end{smallmatrix}\right) = xy - yx = 0.$$

On the other hand, observe also that if $X = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ *and* $Y = \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right)$*, then*

$$\mathrm{F}_{\mathcal{A}}(X, Y) = (a\ b)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} = ad - bc.$$

Lemma 6.1.4. *Let* V *be a vector space, and let* $\mathrm{T} \in \mathrm{Mult}_n(V)$ *be alternating. If* $X_1, \dots, X_n \in V$ *is a linearly dependent family, then* $\mathrm{T}(X_1, \dots, X_n) = 0$.

Proof. Since the vectors are linearly dependent, it means that one of them, let's say X_1, is a linear combination of the others: $X_1 = \sum_{i=2}^n \lambda_i X_i$ for some scalars λ_i. Then one has

$$\begin{aligned} \mathrm{T}(X_1, X_2, \dots, X_n) &= \mathrm{T}\Big(\sum_{i=2}^n \lambda_i X_i, X_2 \dots, X_n\Big) \\ &= \sum_{i=2}^n \lambda_i \mathrm{T}(X_i, X_2, \dots, X_n) = \sum_{i=2}^n \lambda_i 0 = 0. \end{aligned}$$

□

Note that a simple consequence of this lemma is that if $\dim(V) = m$ and if $\mathrm{T} \in \mathrm{Mult}_n(V)$ for some $n > m$ one must have $\mathrm{T}(X_1, \dots, X_n) = 0$ whenever T is alternating. Indeed, there does not exist a family of n linearly independent vectors in a vector space of dimension $m < n$.

Lemma 6.1.5. *Let* V *be a vector space, and let* $T \in \mathrm{Mult}_n(V)$ *be alternating. For any* $X_1, \dots, X_n \in V$ *one has*

$$\mathrm{T}(X_1, \dots, X_j, \dots, X_k, \dots, X_n) = -\mathrm{T}(X_1, \dots, X_k, \dots, X_j, \dots, X_n),$$

or in other words T *changes its sign when two arguments are exchanged.*

Proof. One has by linearity and since T is alternating:

$$\begin{aligned} 0 =&\mathrm{T}(X_1, \dots, X_j + X_k, \dots, X_j + X_k, \dots, X_n) \\ =&\mathrm{T}(X_1, \dots, X_j, \dots, X_j, \dots, X_n) + \mathrm{T}(X_1, \dots, X_k, \dots, X_k, \dots, X_n) \\ &+ \mathrm{T}(X_1, \dots, X_j, \dots, X_k, \dots, X_n) + \mathrm{T}(X_1, \dots, X_k, \dots, X_j, \dots, X_n) \\ =&0 + 0 + \mathrm{T}(X_1, \dots, X_j, \dots, X_k, \dots, X_n) + \mathrm{T}(X_1, \dots, X_k, \dots, X_j, \dots, X_n) \end{aligned}$$

from which the statement follows directly. □

Let us finally state two useful properties of alternating maps which follow almost directly from the definition.

Lemma 6.1.6. *Let* V *be a vector space over a field* $\mathbb{F}$*, and let* $\mathrm{T} \in \mathrm{Mult}_n(V)$ *be alternating. Then for any* $X_1, \dots, X_n \in V$ *and any* $\lambda, \lambda_i \in \mathbb{F}$ *one has*

(i)

$$\mathrm{T}(X_1, \dots, \lambda X_j, \dots, X_n) = \lambda \mathrm{T}(X_1, \dots, X_j, \dots, X_n),$$

(ii)

$$\mathrm{T}\Big(X_1 + \sum_{i=2}^{n} \lambda_i X_i, X_2, \dots, X_n\Big) = \mathrm{T}(X_1, X_2, \dots, X_n),$$

and such linear combination can be performed at any entry.

Proof. The first statement is nothing but the linearity of T in its j^{th}-argument. The second statement is a consequence of the alternating property of T. □

6.2 The determinant

Let $\mathbb{F}$ be a field, and recall that a map

$$\mathrm{T} : \underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{m \; terms} \to \mathbb{F}$$

is multlinear alternating if T is linear in each of its m arguments and if

$$\mathrm{T}(X_1, \dots, X_i, \dots, X_j, \dots, X_m) = 0$$

whenever $X_i = X_j$ for some $i \neq j$. In the special case $m = n$, a very strong statement holds. For this, recall that the standard basis $\{E_j\}_{j=1}^n$ of $\mathbb{F}^n$ is given by $(E_j)_i = 1$ if $i = j$ and $(E_j)_i = 0$ if $i \neq j$.

Theorem 6.2.1. *For any field* $\mathbb{F}$ *there exists a unique* $\mathrm{T} \in \mathrm{Mult}_n(\mathbb{F}^n)$ *alternating such that*

$$\mathrm{T}(E_1, E_2, \dots, E_n) = 1.$$

In order to prove this statement, we need to introduce one more notation. For any indices $i_1, i_2, \dots, i_n$ with $i_j \in \{1, 2, \dots, n\}$ we define the number $\varepsilon_{i_1 i_2 \dots i_n}$ by

$$\begin{aligned} \varepsilon_{i_1 i_2 \dots i_n} &= 0 && \text{if two of the indices are equal,} \\ \varepsilon_{i_1 i_2 \dots i_n} &= (-1)^m && \text{if } \{i_1, \dots, i_n\} = \{1, \dots, n\} \text{ and if } m \text{ is the number of transpo-} \\ & && \text{sitions (exchanges) needed to reorder } i_1 i_2 \dots i_n \text{ into } 1\,2 \dots n. \end{aligned}$$

Note that the equality $\{i_1, \dots, i_n\} = \{1, \dots, n\}$ has to be understood as an equality between sets, without any consideration about the order. For example, $\{1, 2\} = \{2, 1\}$ because both sets contain the same elements.

Example 6.2.2.

$$\begin{aligned} &\varepsilon_{12} = 1, \varepsilon_{21} = -1, \varepsilon_{11} = 0 = \varepsilon_{22} \\ &\varepsilon_{123} = 1 = \varepsilon_{231} = \varepsilon_{312}, \varepsilon_{132} = -1 = \varepsilon_{321} = \varepsilon_{213}, \varepsilon_{122} = 0 = \varepsilon_{111} = \cdots \end{aligned}$$

Proof of Theorem 6.2.1. Let $X_1, \ldots, X_n \in \mathbb{F}^n$, and let $\lambda_{ij} \in \mathbb{F}$ for $i, j \in \{1, \ldots, n\}$ such that $X_j = \sum_{i=1}^n \lambda_{ji} E_i$. Thus, if T is any multilinear map one has

$$\begin{aligned}\mathrm{T}(X_1, \ldots, X_n) &= \mathrm{T}\Big(\sum_{i_1=1}^n \lambda_{1i_1} E_{i_1}, \ldots, \sum_{i_n=1}^n \lambda_{ni_n} E_{i_n}\Big)\\ &= \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n \lambda_{1i_1} \ldots \lambda_{ni_n} \mathrm{T}(E_{i_1}, \ldots, E_{i_n}). \qquad (6.2.1)\end{aligned}$$

In addition, if T is alternating as well, then $\mathrm{T}(E_{i_1}, \ldots, E_{i_n}) = 0$ unless $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$ and in this case one has

$$\mathrm{T}(E_{i_1}, \ldots, E_{i_n}) = \varepsilon_{i_1 i_2 \ldots i_n} \mathrm{T}(E_1, E_2, \ldots, E_n).$$

Finally, by imposing $\mathrm{T}(E_1, E_2, \ldots, E_n) = 1$ one gets from (6.2.1) that

$$\begin{aligned}\mathrm{T}(X_1, \ldots, X_n) &= \sum_{\{i_1,\ldots,i_n\}=\{1,\ldots,n\}} \lambda_{1i_1} \ldots \lambda_{ni_n} \varepsilon_{i_1 i_2 \ldots i_n}\\ &= \sum_{\{i_1,\ldots,i_n\}=\{1,\ldots,n\}} \varepsilon_{i_1 i_2 \ldots i_n} \lambda_{1i_1} \ldots \lambda_{ni_n}. \qquad (6.2.2)\end{aligned}$$

Note that the summation has to be performed on the set of all permutations of the n numbers $1, 2, \ldots, n$. One concludes by observing that the r.h.s. of (6.2.2) does not depend on T, showing that there exists only one T satisfying the stated conditions. □

Corollary 6.2.3. *There exists a unique map* $\mathrm{Det} : M_n(\mathbb{F}) \to \mathbb{F}$ *which is n-linear alternating as a function of the columns, and which is equal to* 1 *for the identity matrix* $\mathbf{1}_n$. *This map is called* the determinant.

Proof. It is sufficient to identify a matrix $\mathcal{A} \in M_n(\mathbb{F})$ with its n columns $\mathcal{A}^j$, each one belonging to $\mathbb{F}^n$, and to use the previous theorem. □

Note that the following two notations are used for the determinant of a matrix $\mathcal{A}$: either $\mathrm{Det}(\mathcal{A})$ or $|\mathcal{A}|$. In the next statement, we simply adapt the properties proved for n-linear maps to the determinant.

Lemma 6.2.4. *Let* $\mathcal{A} \in M_n(\mathbb{F})$ *with* $\mathcal{A} = (\mathcal{A}^1\ \mathcal{A}^2\ \ldots \mathcal{A}^n)$. *Then*

(i) $\mathrm{Det}(\mathcal{A}) = 0$ *if the n columns of $\mathcal{A}$ are linearly dependent,*

(ii)
$$\mathrm{Det}(\mathcal{A}^1 \ldots \mathcal{A}^j \ldots \mathcal{A}^k \ldots \mathcal{A}^n) = -\mathrm{Det}(\mathcal{A}^1 \ldots \mathcal{A}^k \ldots \mathcal{A}^j \ldots \mathcal{A}^n),$$
or in other words the sign of the determinant changes whenever two columns of the matrix are exchanged,

(iii)
$$\mathrm{Det}(\mathcal{A}^1 \ldots \lambda\mathcal{A}^j \ldots \mathcal{A}^n) = \lambda \mathrm{Det}(\mathcal{A}^1 \ldots \mathcal{A}^j \ldots \mathcal{A}^n),$$

(iv) $\mathrm{Det}(\mathcal{A})$ *is not changed if one adds to a column a linear combination of the other columns.*

Let us also state two formulas for the computation of the determinant (see also Exercise 6.5). For this purpose, we introduce one more notation: For a matrix $\mathcal{A} \in M_n(\mathbb{F})$ and for $i, j \in \{1, \ldots, n\}$ we denote by $\mathcal{A}(i,j) \in M_{n-1}(\mathbb{F})$ the matrix obtained by disregarding the row i and the column j of $\mathcal{A}$. Then the following formulas hold: for any $\mathcal{A} = (a_{ij})$ one has

$$\mathrm{Det}(\mathcal{A}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \,\mathrm{Det}\big(\mathcal{A}(i,j)\big) \qquad \text{for any fixed } i \in \{1, \ldots, n\}, \tag{6.2.3}$$

or

$$\mathrm{Det}(\mathcal{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \,\mathrm{Det}\big(\mathcal{A}(i,j)\big) \qquad \text{for any fixed } j \in \{1, \ldots, n\}. \tag{6.2.4}$$

Note that formula (6.2.3) corresponds to a development of the determinant with respect to the row i of $\mathcal{A}$, while (6.2.4) corresponds to the development of the determinant with respect to the column j of $\mathcal{A}$.

Examples 6.2.5. *(i) If* $\mathcal{A} \in M_1(\mathbb{F})$, i.e. $\mathcal{A} = (a) \in \mathbb{F}$, *then* $\mathrm{Det}(\mathcal{A}) = a$,

(ii) In $\mathcal{A} \in M_2(\mathbb{F})$ *with* $\mathcal{A} = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$, *then*

$$\begin{aligned} \mathrm{Det}(\mathcal{A}) &= (-1)^2 a_{11} \,\mathrm{Det}\big(\mathcal{A}(1,1)\big) + (-1)^3 a_{12} \,\mathrm{Det}\big(\mathcal{A}(1,2)\big) \\ &= a_{11}a_{22} - a_{12}a_{21} \\ &= a_{11}a_{22} - a_{21}a_{12}, \end{aligned}$$

(iii) If $\mathcal{A} \in M_3(\mathbb{F})$ *with* $\mathcal{A} = \left(\begin{smallmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{smallmatrix}\right)$, *then*

$$\begin{aligned} &\mathrm{Det}(\mathcal{A}) \\ =&(-1)^2 a_{11} \,\mathrm{Det}\big(\mathcal{A}(1,1)\big) + (-1)^3 a_{12} \,\mathrm{Det}\big(\mathcal{A}(1,2)\big) + (-1)^4 a_{13} \,\mathrm{Det}\big(\mathcal{A}(1,3)\big) \\ =&a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \\ =&a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}. \end{aligned}$$

Remark that in the above examples, we have performed the development with respect to the first row, but the same result would have been obtained if the development was performed with respect to any other row or column.

In the sequel, we shall obtain various additional properties of the determinant.

Lemma 6.2.6. *Let* $\mathcal{A} \in M_n(\mathbb{F})$, *then* $\mathrm{Det}(\mathcal{A}) = \mathrm{Det}({}^t\mathcal{A})$.

Proof. The proof is performed by induction. Clearly, for $n = 1$ the statement is true since the matrix just corresponds to a single scalar. So we can assume that the statement is true for any matrix in $M_{n-1}(\mathbb{F})$ and prove it for any element of $M_n(\mathbb{F})$. Let $\mathcal{A} = (a_{ij}) \in M_n(\mathbb{F})$ and let us set $\mathcal{B} = (b_{ij})$ with $\mathcal{B} = {}^t\!\mathcal{A}$. Then, by the formula (6.2.3) with $i = 1$ one gets

$$\mathrm{Det}(\mathcal{A}) = a_{11}\,\mathrm{Det}\big(\mathcal{A}(1,1)\big) - a_{12}\,\mathrm{Det}\big(\mathcal{A}(1,2)\big) + \cdots + (-1)^{n+1}a_{1n}\,\mathrm{Det}\big(\mathcal{A}(1,n)\big) \quad (6.2.5)$$

and by formula (6.2.4) with $j = 1$ one gets

$$\mathrm{Det}(\mathcal{B}) = b_{11}\,\mathrm{Det}\big(\mathcal{B}(1,1)\big) - b_{21}\,\mathrm{Det}\big(\mathcal{B}(2,1)\big) + \cdots + (-1)^{n+1}b_{n1}\,\mathrm{Det}\big(\mathcal{B}(n,1)\big). \quad (6.2.6)$$

Now, observe that $a_{1j} = b_{j1}$ because $\mathcal{B}$ is the transpose of $\mathcal{A}$, and similarly $\mathcal{B}(j,1) = {}^t\!\mathcal{A}(1,j) \in M_{n-1}(\mathbb{F})$. Since by assumption one has

$$\mathrm{Det}\big(\mathcal{A}(1,j)\big) = \mathrm{Det}\big({}^t\!\mathcal{A}(1,j)\big) = \mathrm{Det}\big(\mathcal{B}(j,1)\big)$$

one directly infers from (6.2.5) and (6.2.6) that $\mathrm{Det}(\mathcal{A}) = \mathrm{Det}(\mathcal{B})$, which corresponds to the statement. □

Corollary 6.2.7. *All the properties of* $\mathrm{Det}(\mathcal{A})$ *with respect to the columns of* $\mathcal{A}$ *also hold with respect to the rows of* $\mathcal{A}$.

In order to state an important result linking $\mathrm{Det}(\mathcal{A})$ and the invertibility of $\mathcal{A}$, let us recall some results of Chapter 2 but in the general framework of an arbitrary field $\mathbb{F}$.

1) Recall that the elementary matrices have been introduced in Definition 2.5.1 and that their definition holds for any field. One shows in Exercise 6.4 that

(i) $\mathrm{Det}(\mathbf{1}_n - I_{rr} + cI_{rr}) = c$, for $c \in \mathbb{F}$ with $c \neq 0$,

(ii) $\mathrm{Det}(\mathbf{1}_n + I_{rs} + I_{sr} - I_{rr} - I_{ss}) = -1$, for $r \neq s$,

(iii) $\mathrm{Det}(\mathbf{1}_n + cI_{rs}) = 1$, for $r \neq s$ and any $c \in \mathbb{F}$.

In addition, one also observes that for any $\mathcal{A} \in M_n(\mathbb{F})$ and any elementary matrix $\mathcal{B} \in M_n(\mathbb{F})$ one has

$$\mathrm{Det}(\mathcal{B}\mathcal{A}) = \mathrm{Det}(\mathcal{B})\,\mathrm{Det}(\mathcal{A}). \quad (6.2.7)$$

Note that this property can be inferred from the general property of the determinant and from the action of an elementary matrix on $\mathcal{A}$, as seen in Exercise 2.14.

2) For any $\mathcal{A} \in M_n(\mathbb{F})$, let us recall that there exist a family of elementary matrices $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_p \in M_n(\mathbb{F})$ such that $\mathcal{A}' := \mathcal{B}_p\mathcal{B}_{p-1}\dots\mathcal{B}_1\mathcal{A}$ is a matrix in the standard form, see Corollary 2.4.10 and the subsequent definition. In particular, it has been shown in Theorem 2.5.4 that $\mathcal{A}' = \mathbf{1}_n$ if and only if $\mathcal{A}$ is invertible. Equivalently, $\mathcal{A}$ is not invertible if and only if $\mathcal{A}'$ contains some 0 on its diagonal.

3) One easily observes that $\mathrm{Det}(\mathcal{A}') = 1$ if $\mathcal{A}' = \mathbf{1}_n$ and that $\mathrm{Det}(\mathcal{A}') = 0$ if $\mathcal{A}'$ contains some 0 on its diagonal.

Proposition 6.2.8. *For any field $\mathbb{F}$ and any $\mathcal{A} \in M_n(\mathbb{F})$, the following statements are equivalent:*

(i) $\mathrm{Det}(\mathcal{A}) \neq 0$,

(ii) $\mathcal{A}$ *is invertible,*

(iii) The columns $\mathcal{A}^1, \mathcal{A}^2, \dots, \mathcal{A}^n$ *of* $\mathcal{A}$ *are linearly independent,*

(iv) The rows $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ *of* $\mathcal{A}$ *are linearly independent.*

Proof. Since there exist elementary matrices $\mathcal{B}_j$ such that $\mathcal{A}' := \mathcal{B}_p\mathcal{B}_{p-1}\dots\mathcal{B}_1\mathcal{A}$ with $\mathcal{A}'$ in the standard form, one gets from (6.2.7) that

$$\begin{aligned}\mathrm{Det}(\mathcal{A}') &= \mathrm{Det}\big(\mathcal{B}_p\mathcal{B}_{p-1}\dots\mathcal{B}_1\mathcal{A}\big)\\ &= \mathrm{Det}(\mathcal{B}_p)\,\mathrm{Det}\big(\mathcal{B}_{p-1}\dots\mathcal{B}_1\mathcal{A}\big)\\ &= \dots\\ &= \underbrace{\mathrm{Det}(\mathcal{B}_p)\,\mathrm{Det}(\mathcal{B}_{p-1})\dots\mathrm{Det}(\mathcal{B}_1)}_{\neq 0}\,\mathrm{Det}(\mathcal{A}).\end{aligned}$$

Thus, one infers that $\mathrm{Det}(\mathcal{A}) \neq 0$ if and only if $\mathrm{Det}(\mathcal{A}') \neq 0$. Since by the above observations 2) and 3) one already knows that $\mathrm{Det}(\mathcal{A}') \neq 0$ if and only if $\mathcal{A}$ is invertible, one then concludes that $\mathrm{Det}(\mathcal{A}) \neq 0$ if and only if $\mathcal{A}$ is invertible. This corresponds to the equivalence between (i) and (ii).

For the second equivalence, consider $\mathrm{L}_{\mathcal{A}} : \mathbb{F}^n \to \mathbb{F}^n$ be the linear map defined by $\mathrm{L}_{\mathcal{A}}X = \mathcal{A}X$ for any $X \in \mathbb{F}^n$. By definition of the rank, $\mathcal{A}^1, \dots, \mathcal{A}^n$ are linearly independent if and only if $\mathrm{rank}(\mathcal{A}) = n$. However, by Corollary 4.4.2, Theorem 4.3.5 and Lemma 4.7.8 one has

$$\mathrm{rank}(\mathcal{A}) = n \Leftrightarrow \dim\big(\mathrm{Ran}(\mathrm{L}_{\mathcal{A}})\big) = n \Leftrightarrow \dim\big(\mathrm{Ker}(\mathrm{L}_{\mathcal{A}})\big) = 0 \Leftrightarrow \mathrm{L}_{\mathcal{A}} \text{ is invertible.}$$

Finally, from Example 4.7.2 one also infers that $\mathrm{L}_{\mathcal{A}}$ is invertible if and only if $\mathcal{A}$ is invertible. Summing up these information, one has obtained that $\mathcal{A}^1, \dots, \mathcal{A}^n$ are linearly independent if and only if $\mathcal{A}$ is invertible, which corresponds to the equivalence between (ii) and (iii).

The equivalence between (iii) and (iv) corresponds to a reformulation of Corollary 6.2.7. □

Corollary 6.2.9. *Let* $\mathbb{F}$ *be any field and let* $X_1, \dots, X_n$ *be* n *elements of* $\mathbb{F}^n$. *Then* $X_1, \dots, X_n$ *are linearly independent if and only if* $\mathrm{Det}(X_1\,X_2\dots X_n) \neq 0$.

Let us now prove an extension of (6.2.7) valid for arbitrary matrices.

Proposition 6.2.10. *For any field* $\mathbb{F}$ *and any* $\mathcal{A}, \mathcal{C} \in M_n(\mathbb{F})$ *one has*

$$\mathrm{Det}(\mathcal{A}\mathcal{C}) = \mathrm{Det}(\mathcal{A})\,\mathrm{Det}(\mathcal{C}). \tag{6.2.8}$$

Proof. Since there exist elementary matrices $\mathcal{B}_j$ such that $\mathcal{A} := \mathcal{B}_1^{-1}\mathcal{B}_2^{-1}\dots\mathcal{B}_p^{-1}\mathcal{A}'$ with $\mathcal{A}'$ in the standard form, one gets from (6.2.7) that

$$\begin{aligned}\mathrm{Det}(\mathcal{A}\mathcal{C}) &= \mathrm{Det}\big(\mathcal{B}_1^{-1}\mathcal{B}_2^{-1}\dots\mathcal{B}_p^{-1}\mathcal{A}'\mathcal{C}\big)\\ &= \mathrm{Det}(\mathcal{B}_1^{-1})\,\mathrm{Det}\big(\mathcal{B}_2^{-1}\dots\mathcal{B}_p^{-1}\mathcal{A}'\mathcal{C}\big)\\ &= \dots\\ &= \underbrace{\mathrm{Det}(\mathcal{B}_1^{-1})\,\mathrm{Det}(\mathcal{B}_2^{-1})\dots\mathrm{Det}(\mathcal{B}_p^{-1})}_{\neq 0}\,\mathrm{Det}(\mathcal{A}'\mathcal{C}).\end{aligned}$$

Thus, if $\mathcal{A}' = \mathbf{1}_n$, one deduces that

$$\mathrm{Det}(\mathcal{A}\mathcal{C}) = \underbrace{\mathrm{Det}(\mathcal{B}_1^{-1})\,\mathrm{Det}(\mathcal{B}_2^{-1})\dots\mathrm{Det}(\mathcal{B}_p^{-1})}_{=\mathrm{Det}(\mathcal{A})}\,\mathrm{Det}(\mathcal{C}) = \mathrm{Det}(\mathcal{A})\,\mathrm{Det}(\mathcal{C}).$$

On the other hand, if $\mathcal{A}' \neq \mathbf{1}_n$, then the last row of $\mathcal{A}'$ is filled with 0 and one has $\mathrm{Det}(\mathcal{A}') = 0 = \mathrm{Det}(\mathcal{A})$, where we have used an argument from the previous proof for the last equality. However, one also deduces from the formula (2.2.4) on the product of two matrices that the last row of $\mathcal{A}'\mathcal{C}$ is also filled only with 0, and this implies that $\mathrm{Det}(\mathcal{A}'\mathcal{C}) = 0$ as well. As a consequence, one has both

$$\mathrm{Det}(\mathcal{A}\mathcal{C}) = \mathrm{Det}(\mathcal{B}_1^{-1})\,\mathrm{Det}(\mathcal{B}_2^{-1})\dots\mathrm{Det}(\mathcal{B}_p^{-1})\,\mathrm{Det}(\mathcal{A}'\mathcal{C}) = 0$$

and $\mathrm{Det}(\mathcal{A})\,\mathrm{Det}(\mathcal{C}) = 0\,\mathrm{Det}(\mathcal{C}) = 0$. Again, the equality (6.2.8) holds. □

6.3 Cramer's rule and the inverse of a matrix

The next proposition is usually referred as Cramer's rule.

Proposition 6.3.1. *Let $\mathcal{A} \in M_n(\mathbb{F})$ with $\mathrm{Det}(\mathcal{A}) \neq 0$, and consider the system of equations $\mathcal{A}X = B$ with $B \in \mathbb{F}^n$. Then its solution $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$ is given by*

$$x_j = \frac{1}{\mathrm{Det}(\mathcal{A})}\mathrm{Det}\big(\mathcal{A}^1\,\mathcal{A}^2\dots B\dots\mathcal{A}^n\big),$$

where B is replacing the column $\mathcal{A}^j$.

The proof of this statement is provided in Exercise 6.12.

Corollary 6.3.2. *If $\mathcal{A} \in M_n(\mathbb{F})$ is invertible, then its inverse is given by the following formula*

$$(\mathcal{A}^{-1})_{ij} = (-1)^{i+j}\frac{\mathrm{Det}\big(\mathcal{A}(j,i)\big)}{\mathrm{Det}(\mathcal{A})}.$$

Proof. For fixed $j \in \{1, \dots, n\}$, consider the equation $\mathcal{A}X = E_j$ with $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$ and with $E_j \in \mathbb{F}^n$ the vector consisting in 1 at the position j and 0 everywhere else. Since $\mathcal{A}$ is invertible this equation is equivalent to $\mathcal{A}^{-1}E_j = X$, or more precisely $x_i = \sum_{k=1}^n (\mathcal{A}^{-1})_{ik}(E_j)_k$ for any $i \in \{1, \dots, n\}$. Since $(E_j)_k = 0$ whenever $j \neq k$ one gets $x_i = (\mathcal{A}^{-1})_{ij}$.

On the other hand, from the previous proposition with $B = E_j$ one also gets

$$x_i = \frac{1}{\mathrm{Det}(\mathcal{A})}\mathrm{Det}\big(\mathcal{A}^1\,\mathcal{A}^2 \dots E_j \dots \mathcal{A}^n\big) = \frac{1}{\mathrm{Det}(\mathcal{A})}(-1)^{i+j}\,\mathrm{Det}\big(\mathcal{A}(j,i)\big)$$

where formula (6.2.4) with respect to the column i has been used. By identifying the two expressions for x_i one gets the stated equality. □

6.4 Exercises

Exercise 6.1. *Let us define the map* $\mathrm{F} : \underbrace{M_n(\mathbb{R}) \times \cdots \times M_n(\mathbb{R})}_{m \text{ arguments}} \to \mathbb{R}$ *by*

$$\mathrm{F}\big(\mathcal{A}_1, \ldots, \mathcal{A}_m\big) = \mathrm{Tr}\big(\mathcal{A}_1 \ldots \mathcal{A}_m\big).$$

Show that F *is a* m*-linear map.*

Exercise 6.2. *Show that the the cross product in* $\mathbb{R}^3$ *is a bilinear alternating map.*

Exercise 6.3. *Exhibit 3 different alternating bilinear maps on* $\mathbb{R}^3$*.*

Exercise 6.4. *For* $r \in \{1, \ldots, m\}$ *and* $s \in \{1, \ldots, m\}$*, let* $I_{rs} \in M_m(\mathbb{F})$ *be the matrix whose* rs*-component is* 1 *and all the other ones are equal to* 0*. For* $c \neq 0$*, consider the following 3 types of elementary matrices :*

1. $\mathbf{1}_m - I_{rr} + cI_{rr}$*, the matrix obtained from the identity matrix by multiplying the* r*-th diagonal component by* c*,*
2. *For* $r \neq s$*,* $(\mathbf{1}_m + I_{rs} + I_{sr} - I_{rr} - I_{ss})$*, the matrix obtained from the identity matrix by interchanging the* r*-th row with the* s*-th row,*
3. *For* $r \neq s$*,* $(\mathbf{1}_m + cI_{rs})$*, the matrix having the* rs*-th component equal to* c*, all other components* 0 *except the diagonal components which are equal to* 1*.*

Compute the determinant of these elementary matrices.

Exercise 6.5. *For an arbitrary field* $\mathbb{F}$ *let* $\mathcal{A} = (a_{ij}) \in M_n(\mathbb{F})$ *and recall the formula:*

$$\begin{aligned} \mathrm{Det}(\mathcal{A}) &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \,\mathrm{Det}\big(\mathcal{A}(i,j)\big) \qquad \text{for any fixed } j \in \{1, \ldots, n\} \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \,\mathrm{Det}\big(\mathcal{A}(i,j)\big) \qquad \text{for any fixed } i \in \{1, \ldots, n\}. \end{aligned}$$

Show that $\mathrm{Det}(\mathbf{1}_n) = 1$ *for any* n*. For* $n = 2$*, show that*

(i) the determinant is linear as a function of the columns of $\mathcal{A}$*,*

(ii) the determinant is alternating as a function of the columns of $\mathcal{A}$*.*

Can you do it for $n = 3$*? For arbitrary* n *(a proof by induction over the dimension* n *is recommended).*

Exercise 6.6. *Compute the determinant of the following matrices:*

$$a) \begin{pmatrix} 4 & -1 & 1 \\ 2 & 0 & 0 \\ 1 & 5 & 7 \end{pmatrix}, \quad b) \begin{pmatrix} 1 & 4 & 6 \\ 0 & 0 & 1 \\ 0 & 0 & 8 \end{pmatrix} \quad c) \begin{pmatrix} 3 & 1 & 1 \\ 2 & 5 & 5 \\ 8 & 7 & 7 \end{pmatrix} \quad d) \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & -1 & 1 & 0 \\ 3 & 0 & 0 & 5 \end{pmatrix}$$

Exercise 6.7. *Let* $\mathcal{A} = (a_{jk}) \in M_n(\mathbb{R})$ *be an upper triangular matrix. Compute* $\mathrm{Det}(\mathcal{A})$.

Exercise 6.8. *Show that two similar square matrices share the same determinant.*

Exercise 6.9. *Show that if* $\mathcal{A} \in M_n(\mathbb{F})$ *is invertible then the following equality holds:*

$$\mathrm{Det}\big(\mathcal{A}^{-1}\big) = \frac{1}{\mathrm{Det}(\mathcal{A})}.$$

Exercise 6.10. *Compute the determinant of the matrix* $\begin{pmatrix} x+1 & x-1 \\ x & 2x+5 \end{pmatrix}$.

Exercise 6.11. *Consider the matrix* $\mathcal{A} = \begin{pmatrix} \lambda & 1 & 1 \\ -1 & \lambda & 1 \\ 1 & 1 & \lambda \end{pmatrix}$ *with* $\lambda \in \mathbb{R}$.

(i) Compute the determinant of $\mathcal{A}$,

(ii) For which values of λ *is* $\mathcal{A}$ *invertible ?*

Exercise 6.12. *Prove Cramer's rule,* i.e. *show that if* $\mathcal{A} \in M_n(\mathbb{F})$ *is invertible and if* $X \in \mathbb{F}^n$ *satisfies* $\mathcal{A}X = B$ *for some* $B \in \mathbb{F}^n$, *then*

$$x_j = \frac{1}{\mathrm{Det}(\mathcal{A})}\mathrm{Det}\big(\mathcal{A}^1\,\mathcal{A}^2 \dots B \dots \mathcal{A}^n\big),$$

where B *is replacing the column* $\mathcal{A}^j$. *For that purpose, one should first recall that* $\mathcal{A}X = B$ *is equivalent to* $x_1\mathcal{A}^1 + x_2\mathcal{A}^2 + \cdots + x_n\mathcal{A}^n = B$, *and insert this equality in the term* $\mathrm{Det}\big(\mathcal{A}^1\,\mathcal{A}^2 \dots B \dots \mathcal{A}^n\big)$.

Exercise 6.13. *By using determinants, find the inverse for the following matrices :*

$$a)\ \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 2 & 7 \end{pmatrix} \qquad b)\ \begin{pmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{pmatrix}$$

Exercise 6.14. *By using determinants, solve the following systems of equations :*

$$a)\ \begin{cases} x + 2y - z & = 1 \\ y + z & = 1 \\ 2y + 7z & = 1 \end{cases} \qquad b)\ \begin{cases} 2x + y + 2z & = 0 \\ 3y - z & = 1 \\ 4x + y + z & = 2 \end{cases}$$

Exercise 6.15. *Let* X, Y *be two vectors in* $\mathbb{R}^2$. *Check that the area of the parallelogram spanned by* X *and* Y *is equal to the absolute value of the determinant of the matrix* $(X\ Y) \in M_2(\mathbb{R})$. *More generally, if* $X_1, \dots, X_n$ *are* n *vectors of* $\mathbb{R}^n$, *one writes* $\mathsf{Vol}(X_1, \dots, X_n)$ *for the volume of the* n*-dimensional box spanned by* $X_1, \dots, X_n$. *Why is it natural to have*

$$\mathsf{Vol}(X_1, \dots, X_n) = |\mathrm{Det}(X_1 \dots X_n)| \text{ ?}$$

Exercise 6.16. *Let* $\{V_1, \dots, V_n\}$ *and* $\{V'_1, \dots, V'_n\}$ *be two bases of* $\mathbb{R}^n$, *and let* $\mathcal{B} \in M_n(\mathbb{R})$ *be the matrix of change of bases,* i.e. $V'_j = \mathcal{B}V_j$ *for any* $j = 1, 2, \dots, n$. *What is the geometric interpretation of* $|\mathrm{Det}(\mathcal{B})|$ *in this setting ? For that purpose, one should first check that if* $(V_1\ V_2 \dots V_n)$ *denotes the matrix with columns* V_j *and* $(V'_1\ V'_2 \dots V'_n)$ *denotes the matrix with columns* V'_j, *then one has*

$$(V'_1\ V'_2 \dots V'_n) = (\mathcal{B}V_1\ \mathcal{B}V_2 \dots \mathcal{B}V_n) = \mathcal{B}(V_1\ V_2 \dots V_n).$$

Chapter 7

Eigenvectors and eigenvalues

7.1 Eigenvalues and eigenvectors

We start with the main definition of this chapter.

Definition 7.1.1. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathrm{L} : V \to V$ be a linear map. An element $\lambda \in \mathbb{F}$ is* an eigenvalue *of* L *if there exists $X \in V$ with $X \neq \mathbf{0}$ such that*

$$\mathrm{L}(X) = \lambda X.$$

In such a case, X is called an eigenvector *or* an eigenfunction *associated with the eigenvalue λ.*

Examples 7.1.2. *(i) Consider $\mathrm{L}_\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2$ with $\mathcal{A} = \left(\begin{smallmatrix} 1 & 2 \\ 4 & 3 \end{smallmatrix}\right)$. Then one observes that*

$$\mathrm{L}_\mathcal{A}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix} = 5\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Thus, $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ is an eigenvector of $\mathrm{L}_\mathcal{A}$ associated with the eigenvalue 5. Similarly, one can check that $\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ is an eigenvector of $\mathrm{L}_\mathcal{A}$ associated with the eigenvalue -1.

(ii) If $\mathcal{A} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{mn} \end{pmatrix}$, then E_j is an eigenvector of $\mathrm{L}_\mathcal{A}$ associated with the eigenvalue a_{jj}.

(iii) If $V = C^1(\mathbb{R})$ and if $\mathrm{L} = \frac{\mathrm{d}}{\mathrm{d}x}$, then any $\lambda \in \mathbb{R}$ is an eigenvalue of L *since the function $x \mapsto e^{\lambda x}$ belongs to $C^1(\mathbb{R})$ and satisfies*

$$\big[\mathrm{L}\big(e^{\lambda\cdot}\big)\big](x) = \big(e^{\lambda\cdot}\big)'(x) = \lambda e^{\lambda x}.$$

Thus this function is an eigenvector associated with the eigenvalue λ.

Remark 7.1.3. *An eigenvector is never unique. Indeed, if X is an eigenvector associated with the eigenvalue λ of* L*, then for any $c \in \mathbb{F}$ with $c \neq 0$ the element $cX \in V$ is also an eigenvector of L associated with the eigenvalue λ. Indeed, one only has to observe that*

$$\mathrm{L}(cX) = c\mathrm{L}(X) = c\lambda X = \lambda(cX).$$

More generally one has:

Lemma 7.1.4. *The set of eigenvectors associated with the eigenvalue λ of* L *is a subspace of V.*

This vector space is called *the eigenspace* associated with the eigenvalue λ of L.

Proof. We have just seen that if X is an eigenvector of L associated with the eigenvalue λ, then cX is an eigenvector associated with the same eigenvalue. This corresponds to the second condition of the definition of a subspace of V, see Definition 3.1.5.

For the first condition of the same definition, observe that if X_1, X_2 satisfy $\mathrm{L}(X_1) = \lambda X_1$ and $\mathrm{L}(X_2) = \lambda X_2$, then one has

$$\mathrm{L}(X_1 + X_2) = \mathrm{L}(X_1) + \mathrm{L}(X_2) = \lambda X_1 + \lambda X_2 = \lambda(X_1 + X_2),$$

which corresponds to this condition. □

Example 7.1.5. *Let $\mathcal{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \in M_3(\mathbb{R})$ and consider the corresponding map $\mathrm{L}_\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$. Then 0 and 3 are eigenvalues of $\mathrm{L}_\mathcal{A}$, with E_1 an eigenvector associated with the eigenvalue 0, and any $cE_2 + dE_3$, with $c, d \in \mathbb{R}$, an eigenvector associated with the eigenvalue 3. Note that the eigenspace associated with the eigenvalue 0 is of dimension 1 while the eigenspace associated with the eigenvalue 3 is of dimension 2.*

The following result is important, especially in relation with quantum mechanics.

Theorem 7.1.6. *Let $\lambda_1, \dots, \lambda_m$ be eigenvalues of* L*, and let $X_1, \dots, X_m$ be corresponding eigenvectors. If $\lambda_i \neq \lambda_j$ for any $i \neq j$, then the vectors $X_1, \dots, X_m$ are linearly independent.*

Proof. This proof is a proof by induction. Clearly, if $m = 1$ then the only eigenvector $X_1 \neq \mathbf{0}$ is linearly independent. So, let us assume that the statement is true for a certain $m-1 \geq 1$, and let us prove it for m. Thus, let us assume that $X_1, \dots, X_{m-1}$ are linearly independent, and show that $X_1, \dots X_m$ are also linearly independent. For this purpose, consider the linear combination

$$c_1 X_1 + c_2 X_2 + \cdots + c_m X_m = \mathbf{0}, \tag{7.1.1}$$

for some coefficients $c_j \in \mathbb{F}$. By multiplying this equality by λ_m one gets

$$c_1 \lambda_m X_1 + c_2 \lambda_m X_2 + \cdots + c_m \lambda_m X_m = \mathbf{0}. \tag{7.1.2}$$

On the other hand, by applying L to (7.1.1) one gets

$$c_1\mathrm{L}(X_1)+c_2\mathrm{L}(X_2)+\cdots+c_m\mathrm{L}(X_m)=c_1\lambda_1X_1+c_2\lambda_2X_2+\cdots+c_m\lambda_mX_m=\mathbf{0}. \quad (7.1.3)$$

Finally, by subtracting (7.1.3) to (7.1.2) one obtains

$$c_1\underbrace{(\lambda_m-\lambda_1)}_{\neq 0}X_1+c_2\underbrace{(\lambda_m-\lambda_2)}_{\neq 0}X_2+\cdots+c_{m-1}\underbrace{(\lambda_m-\lambda_{m-1})}_{\neq 0}X_{m-1}=\mathbf{0}.$$

Since $X_1,\dots,X_{m-1}$ are linearly independent, it follows that $c_1=c_2=\cdots=c_{m-1}=0$. We then conclude from (7.1.1) that $c_m=0$ as well, meaning that $X_1,\dots,X_m$ are linearly independent. □

Corollary 7.1.7. *If $\mathcal{A}\in M_n(\mathbb{F})$, then the linear map $\mathrm{L}_\mathcal{A}:\mathbb{F}^n\to\mathbb{F}^n$ can have at most n distinct eigenvalues.*

Proof. If $\mathrm{L}_\mathcal{A}$ had $m>n$ eigenvalues, then the eigenvectors $X_1,\dots,X_m$ would be a family of m linearly independent elements of $\mathbb{F}^n$, which is impossible. □

7.2 The characteristic polynomial

If V is a vector space over a field $\mathbb{F}$, and if $\mathrm{L}:V\to V$ is a linear map, how can one find out the set of eigenvalues of L ? In this section, we shall answer this question.

Theorem 7.2.1. *Assume that V is a finite dimensional vector space over $\mathbb{F}$, and let $\mathrm{L}:V\to V$ be linear. Then $\lambda\in\mathbb{F}$ is an eigenvalue of L if and only if $\mathrm{L}-\lambda\mathbf{1}$ is not invertible.*

Proof. If λ is an eigenvalue of L, with $X\in V$ an associated eigenvector, then

$$[\mathrm{L}-\lambda\mathbf{1}](X)=\mathrm{L}(X)-\lambda X=\lambda X-\lambda X=\mathbf{0},$$

and therefore $X\in\mathrm{Ker}(\mathrm{L}-\lambda\mathbf{1})$. By Theorem 4.7.8, it follows that $\mathrm{L}-\lambda\mathbf{1}$ is not invertible.

Reciprocally, if $\mathrm{L}-\lambda\mathbf{1}$ is not invertible, it follows from the same theorem that there exists $X\in\mathrm{Ker}(\mathrm{L}-\lambda\mathbf{1})$ with $X\neq\mathbf{0}$. In other words, there exists $X\in V$ with $X\neq\mathbf{0}$ such that $\mathrm{L}(X)-\lambda X=\mathbf{0}$, which means that $\mathrm{L}(X)=\lambda X$. Thus, λ is an eigenvalue of L and X is an associated eigenvector. □

Let us consider a special case of the previous statement. If $V=\mathbb{F}^n$ and $\mathrm{L}=\mathrm{L}_\mathcal{A}$ for some $\mathcal{A}\in M_n(\mathbb{F})$ one infers that $\lambda\in\mathbb{F}$ is an eigenvalue of $\mathrm{L}_\mathcal{A}$ if and only if $\mathrm{L}_\mathcal{A}-\lambda\mathbf{1}$ is not invertible, *i.e.* if and only if $\mathcal{A}-\lambda\mathbf{1}_n$ is not invertible. However, we have seen that this holds if and only if $\mathrm{Det}(\mathcal{A}-\lambda\mathbf{1}_n)=0$. We have thus proved:

Corollary 7.2.2. *Let $\mathbb{F}$ be an arbitrary field, and let $\mathcal{A}\in M_n(\mathbb{F})$. Then λ is an eigenvalue of $\mathrm{L}_\mathcal{A}$ if and only if $\mathrm{Det}(\mathcal{A}-\lambda\mathbf{1}_n)=0$.*

Definition 7.2.3. *For any $\mathcal{A} \in M_n(\mathbb{F})$ and $\lambda \in \mathbb{F}$, one sets*

$$P_{\mathcal{A}}(\lambda) := \mathrm{Det}(\mathcal{A} - \lambda \mathbf{1}_n)$$

and call it the characteristic polynomial *associated with $\mathcal{A}$.*

Note that some authors use the following definition: $P_{\mathcal{A}}(\lambda) = \mathrm{Det}(\lambda \mathbf{1}_n - \mathcal{A})$ which is equal to $\pm\mathrm{Det}(\mathcal{A} - \lambda \mathbf{1}_n)$, depending if n is even or odd. Note also that if $\mathcal{A} \in M_n(\mathbb{F})$, then $P_{\mathcal{A}}$ is a polynomial of degree n. As a consequence of the previous corollary, one has obtained:

Proposition 7.2.4. *For any $\mathcal{A} \in M_n(\mathbb{F})$, the scalar $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathrm{L}_{\mathcal{A}}$ if and only if $P_{\mathcal{A}}(\lambda) = 0$.*

Examples 7.2.5. *(i) Let $\mathcal{A} = \left(\begin{smallmatrix} 1 & 2 \\ 4 & 3 \end{smallmatrix}\right)$, then*

$$P_{\mathcal{A}}(\lambda) = \mathrm{Det}(\mathcal{A} - \lambda \mathbf{1}_2) = \mathrm{Det}\begin{pmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{pmatrix} = (1-\lambda)(3-\lambda) - 8 = (\lambda-5)(\lambda+1).$$

Thus, the eigenvalues of $\mathrm{L}_{\mathcal{A}}$ are -1 and 5.

(ii) For $\mathcal{A} = \left(\begin{smallmatrix} 1 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 7 \end{smallmatrix}\right)$ one has

$$P_{\mathcal{A}}(\lambda) = \mathrm{Det}\begin{pmatrix} 1-\lambda & 1 & 2 \\ 0 & 5-\lambda & -1 \\ 0 & 0 & 7-\lambda \end{pmatrix} = (1-\lambda)(5-\lambda)(7-\lambda),$$

and the eigenvalues of $\mathrm{L}_{\mathcal{A}}$ are 1, 5 and 7.

(iii) For $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ one has $P_{\mathcal{A}}(\lambda) = (1-\lambda)(1+\lambda)$, and the eigenvalues are -1 and 1.

(iv) For $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ one has $P_{\mathcal{A}}(\lambda) = \lambda^2 + 1$ and the eigenvalues are... ?

Note that once the eigenvalues have been determined, it is possible to find the eigenvectors (or the eigenspaces) by solving a linear system. Indeed, if λ is an eigenvalue of $\mathrm{L}_{\mathcal{A}}$ one looks for some $X \in \mathbb{F}^n$ such that $\mathcal{A}X = \lambda X \Leftrightarrow (\mathcal{A} - \lambda \mathbf{1}_n)X = \mathbf{0}$.

Examples 7.2.6. *(i) For $\mathcal{A} = \left(\begin{smallmatrix} 1 & 2 \\ 4 & 3 \end{smallmatrix}\right)$ and $\lambda = 5$, one has to solve*

$$\left[\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} - \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}\right]\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -4 & 2 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which is equivalent to

$$\begin{cases} -4x + 2y = 0 \\ 4x - 2y = 0 \end{cases} \Leftrightarrow \begin{cases} x \text{ arbitrary} \\ y = 2x \end{cases}.$$

Thus, the eigenspace associated with the eigenvalue 5 is given by $\{\left(\begin{smallmatrix} x \\ 2x \end{smallmatrix}\right) \mid x \in \mathbb{R}\}$ or equivalently $\{x\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right) \mid x \in \mathbb{R}\}$.

(ii) For $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ *and the eigenvalue* $\lambda = 1$ *one has to solve*

$$\begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

which is equivalent to the single equation $-x + y = 0$*, or equivalently to* $x = y$*. Thus, the eigenspace associated with the eigenvalue* 1 *is* $\{x\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) \mid x \in \mathbb{R}\}$*.*

Let us now come back to the matrix $\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$ with $P_{\mathcal{A}}(\lambda) = \lambda^2 + 1$. Assume for a while that there exists λ, solution of $\lambda^2 + 1 = 0$, or equivalently $\lambda^2 = -1$. One can then wonder about the corresponding eigenspace ? For that purpose, consider

$$\left[\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} - \lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right]\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}-\lambda & 1\\ -1 & -\lambda\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix},$$

which is equivalent to

$$\left\{\begin{array}{l}-\lambda x + y = 0\\ -x - \lambda y = 0\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}y + \lambda^2 y = 0\\ x = -\lambda y\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}y(1+\lambda^2) = 0\\ x = -\lambda y\end{array}\right. .$$

Since $1 + \lambda^2 = 0$, the element y can be chosen arbitrarily, and then one can define x by the relation $x = -\lambda y$. Thus, the eigenspace associated with the eigenvalue λ is $\{y\left(\begin{smallmatrix}-\lambda\\1\end{smallmatrix}\right) \mid y \in \mathbb{R}\}$ which is a one dimensional vector space. Everything looks fine, except that there is no $\lambda \in \mathbb{R}$ satisfying $\lambda^2 + 1 = 0$! At this point, it is necessary to introduce the notion of complex numbers, which will be done in the last chapter.

As a final example, one can consider the matrix $\mathcal{A} = \left(\begin{smallmatrix}2&1&0\\0&1&-1\\0&2&4\end{smallmatrix}\right)$ with corresponding characteristic polynomial $P_{\mathcal{A}}(\lambda) = (2-\lambda)^2(3-\lambda)$. Thus, the eigenvalues of $\mathrm{L}_{\mathcal{A}}$ are 2 and 3. It is good exercise to check this characteristic polynomial, and to determine the eigenspace corresponding to these eigenvalues, see Exercise 7.8.

We can now define an important set related to each linear map.

Definition 7.2.7. *Let* V *be a finite dimensional vector space, and let* $\mathrm{L} : V \to V$ *be a linear map. The set of all eigenvalues of* L *is called* the spectrum *of* L *and is denoted by* $\sigma(L)$*,* i.e. $\sigma(L) = \{\lambda_1, \lambda_2, \dots\}$ *with each* λ_j *an eigenvalue of* L*.*

Before the next statement, let us remind that if $\mathcal{B}$ is an invertible matrix, then one has

$$1 = \mathrm{Det}(\mathbf{1}_n) = \mathrm{Det}\big(\mathcal{B}\mathcal{B}^{-1}\big) = \mathrm{Det}(\mathcal{B})\mathrm{Det}\big(\mathcal{B}^{-1}\big)$$

which means that $\mathrm{Det}(\mathcal{B}^{-1}) = \mathrm{Det}(\mathcal{B})^{-1}$.

Lemma 7.2.8. *Let* $\mathcal{A} \in M_n(\mathbb{F})$ *and consider* $\mathrm{L}_{\mathcal{A}} : \mathbb{F}^n \to \mathbb{F}^n$ *the associated linear map. Let* $\mathcal{B} \in M_n(\mathbb{F})$ *be invertible. Then*

$$\sigma\big(\mathrm{L}_{\mathcal{B}\mathcal{A}\mathcal{B}^{-1}}\big) = \sigma\big(\mathrm{L}_{\mathcal{A}}\big).$$

Proof. One has

$$\begin{aligned}\mathrm{Det}\big(\mathcal{B}\mathcal{A}\mathcal{B}^{-1} - \lambda \mathbf{1}_n\big) &= \mathrm{Det}\big(\mathcal{B}\mathcal{A}\mathcal{B}^{-1} - \lambda \mathcal{B}\mathbf{1}_n\mathcal{B}^{-1}\big) = \mathrm{Det}\big(\mathcal{B}(\mathcal{A} - \lambda\mathbf{1}_n)\mathcal{B}^{-1}\big) \\ &= \mathrm{Det}(\mathcal{B})\mathrm{Det}(\mathcal{A} - \lambda\mathbf{1}_n)\mathrm{Det}(\mathcal{B}^{-1}) = \mathrm{Det}(\mathcal{A} - \lambda\mathbf{1}_n).\end{aligned}$$

Thus, λ is an eigenvalue of $\mathrm{L}_\mathcal{A}$ if and only if λ is an eigenvalue of $\mathrm{L}_{\mathcal{B}\mathcal{A}\mathcal{B}^{-1}}$. □

Lemma 7.2.9. *For any $\mathcal{A} \in M_n(\mathbb{F})$, $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathrm{L}_\mathcal{A}$ if and only if λ is an eigenvalue of $\mathrm{L}_{{}^t\mathcal{A}}$.*

Proof. It is sufficient to observe that $({}^t\mathcal{A} - \lambda\mathbf{1}_n) = {}^t(\mathcal{A} - \lambda\mathbf{1}_n)$ and to recall that $\mathrm{Det}(\mathcal{B}) = \mathrm{Det}({}^t\mathcal{B})$ for any $\mathcal{B} \in M_n(\mathbb{F})$, see Lemma 6.2.6. □

7.3 Eigenvalues and eigenfunctions for symmetric matrices

The aim of this section is to show that if $\mathcal{A} \in M_n(\mathbb{R})$ is symmetric, *i.e.* ${}^t\mathcal{A} = \mathcal{A}$, then the corresponding linear map $\mathrm{L}_\mathcal{A}$ has n eigenvalues $\lambda_1, \dots, \lambda_n$ (some of them can be equal) and n mutually orthogonal eigenvectors. In fact, we shall prove a slightly more general statement, valid for more general linear maps.

First of all, recall that if ${}^t\mathcal{A} = \mathcal{A}$, then the corresponding bilinear map $\mathrm{F}_\mathcal{A} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and defined by $\mathrm{F}_\mathcal{A}(X, Y) = {}^tX\mathcal{A}Y$ is symmetric. In other word, it means that $\mathrm{F}_\mathcal{A}(X, Y) = \mathrm{F}_\mathcal{A}(Y, X)$, see Exercise 5.6.

Lemma 7.3.1. *If $\mathcal{A} \in M_n(\mathbb{R})$ is symmetric, and if $\lambda_1, \lambda_2 \in \mathbb{R}$ are eigenvalues of $\mathrm{L}_\mathcal{A}$ with $\lambda_1 \neq \lambda_2$, then any associated eigenvectors X_1 and X_2 satisfy $X_1 \perp X_2$.*

Proof. One has

$$\mathrm{F}_\mathcal{A}(X_1, X_2) = {}^tX_1\mathcal{A}X_2 = {}^tX_1(\lambda_2 X_2) = \lambda_2{}^tX_1X_2 = \lambda_2(X_1 \cdot X_2)$$

since $\mathcal{A}X_2 = \lambda_2 X_2$. Here $(X_1 \cdot X_2)$ means the scalar product between the two vectors X_1 and X_2. However, since $\mathrm{F}_\mathcal{A}$ is symmetric one also has

$$\mathrm{F}_\mathcal{A}(X_1, X_2) = \mathrm{F}_\mathcal{A}(X_2, X_1) = {}^tX_2\mathcal{A}X_1 = {}^tX_2(\lambda_1 X_1) = \lambda_1{}^tX_2X_1 = \lambda_1(X_2 \cdot X_1)$$

since $\mathcal{A}X_1 = \lambda_1 X_1$. By comparing these expressions, one has thus obtained that

$$\lambda_2(X_1 \cdot X_2) = \lambda_1(X_2 \cdot X_1).$$

However, since $X_1 \cdot X_2 = X_2 \cdot X_1$ and since $\lambda_1 \neq \lambda_2$ one concludes that $X_1 \cdot X_2 = 0$, which means that the two vectors are orthogonal. □

Let us now observe that if $\mathcal{A} \in M_n(\mathbb{R})$ and if λ is an eigenvalue of $\mathrm{L}_\mathcal{A}$ with the corresponding eigenspace of dimension m, then one can always choose m mutually orthogonal elements $X_1, \dots, X_m$ which satisfy $\mathrm{L}_\mathcal{A}(X_j) = \lambda X_j$ for $j \in \{1, \dots, m\}$. Indeed, if we denote by V_λ the eigenspace associated with the eigenvalue λ, we can apply Graham-Schmidt to this subspace and obtain a basis of V_λ containing m elements. Each of these elements still satisfies $\mathrm{L}_\mathcal{A}(X_j) = \lambda X_j$. Note that the dimension of the eigenspace V_λ is called *the geometric multiplicity* of the eigenvalue λ.

Theorem 7.3.2. *Let $\mathcal{A} \in M_n(\mathbb{R})$, and assume that there exists $X_1, \dots, X_n \in \mathbb{R}^n$, with $X_j \neq 0$ and such that $\mathrm{L}_\mathcal{A}(X_j) = \lambda_j X_j$ for some $\lambda_j \in \mathbb{R}$ and all $j \in \{1, \dots, n\}$. Assume also that $\mathrm{Vect}(X_1, \dots, X_n) = \mathbb{R}^n$. Then if one defines the matrix $\mathcal{B}$ with the column $\mathcal{B}^j$ given by $\mathcal{B}^j = X_j$, it follows that $\mathcal{B}$ is invertible and that*

$$\mathcal{B}^{-1}\mathcal{A}\mathcal{B} = \mathrm{diag}(\lambda_1, \dots, \lambda_n),$$

where $\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ corresponds to the diagonal matrix with entries $\lambda_1, \dots, \lambda_n$ on its diagonal.

Remark 7.3.3. *(i) We shall prove subsequently that the assumptions of this theorem are satisfied whenever $\mathcal{A}$ is symmetric. The assumptions are also satisfied if $\mathcal{A}$ is arbitrary but $\mathrm{L}_\mathcal{A}$ has n distinct eigenvalues, see Theorem 7.1.6.*

(ii) If we consider $\mathcal{B}$ as a change of bases, then the statement means that in the basis defined by the vectors $X_1, \dots, X_n$, the linear map $\mathrm{L}_{\mathcal{B}^{-1}\mathcal{A}\mathcal{B}}$ is diagonal.

Proof. Since $X_1, \dots, X_n$ are linearly independent, it follows that $\mathrm{Det}(\mathcal{B}) \neq 0$ and thus that $\mathcal{B}$ is invertible, with inverse denoted by $\mathcal{B}^{-1}$.

Let us now compute

$$\begin{aligned}
\mathcal{B}^{-1}\mathcal{A}\mathcal{B} &= \mathcal{B}^{-1}\mathcal{A}\big(X_1\ X_2\ \dots\ X_n\big) = \mathcal{B}^{-1}\big(\mathcal{A}X_1\ \mathcal{A}X_2\ \dots\ \mathcal{A}X_n\big) \\
&= \mathcal{B}^{-1}\big(\lambda_1 X_1\ \lambda_2 X_2\ \dots\ \lambda_n X_n\big) = \mathcal{B}^{-1}\mathcal{B}\ \mathrm{diag}(\lambda_1, \dots, \lambda_n) = \mathrm{diag}(\lambda_1, \dots, \lambda_n).
\end{aligned}$$

Indeed, observe that

$$\begin{aligned}
\big((X_1\ X_2\ \dots\ X_n)\ \mathrm{diag}(\lambda_1, \dots, \lambda_n)\big)_{ij} &= \sum_{k=1}^n (X_1\ X_2\ \dots\ X_n)_{ik}\ \mathrm{diag}(\lambda_1, \dots, \lambda_n)_{kj} \\
&= (X_1\ X_2\ \dots\ X_n)_{ij}\ \lambda_j \\
&= (\lambda_1 X_1\ \lambda_2 X_2\ \dots\ \lambda_n X_n)_{ij}
\end{aligned}$$

since $\mathrm{diag}(\lambda_1, \dots, \lambda_n)_{kj} = \lambda_j$ if $k = j$ and 0 otherwise. □

From now on, we shall establish a link between the eigenvalues/eigenvectors and a geometric construction. For that purpose and for any symmetric matrix $\mathcal{A} \in M_n(\mathbb{R})$ let us define $f_\mathcal{A} : \mathbb{R}^n \to \mathbb{R}$ by

$$f_\mathcal{A}(X) := \mathrm{F}_\mathcal{A}(X, X) = {}^tX\mathcal{A}X,$$

and call it *the quadratic form* associated with $\mathcal{A}$.

Examples 7.3.4. *(i) If* $\mathcal{A} = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$*, then*

$$f_{\mathcal{A}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1\ x_2)\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3x_1^2 - 2x_1x_2 + 3x_2^2.$$

(ii) More generally, if $\mathcal{A} = (a_{ij}) \in M_n(\mathbb{R})$ *with* $\mathcal{A}$ *symmetric, then*

$$f_{\mathcal{A}}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1\ \dots\ x_n)\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j=1}^{n} a_{ij}\, x_i\, x_j.$$

Let us now consider the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, *i.e.*

$$\mathbb{S}^{n-1} = \{X \in \mathbb{R}^n \mid \|X\| = 1\},$$

and for a symmetric matrix $\mathcal{A} \in M_n(\mathbb{R})$ we consider $f_{\mathcal{A}}(X)$ with $X \in \mathbb{S}^{n-1}$.

Definition 7.3.5. *A point* $X \in \mathbb{S}^{n-1}$ *is a maximum for* $f_{\mathcal{A}}$ *on* $\mathbb{S}^{n-1}$ *if* $f_{\mathcal{A}}(X) \geq f_{\mathcal{A}}(Y)$ *for any* $Y \in \mathbb{S}^{n-1}$.

Note that such a maximum always exists, but it can be non-unique. For example if $\mathcal{A} = \mathbf{1}_n$, then

$$f_{\mathcal{A}}(X) = f_{\mathbf{1}_n}(X) = {}^t X \mathbf{1}_n X = X \cdot X = \|X\|^2 = 1$$

and thus $f_{\mathbf{1}_n}$ is constant on the sphere. It means that any $X \in \mathbb{S}^{n-1}$ is a maximum for $f_{\mathbf{1}_n}$ on $\mathbb{S}^{n-1}$.

The following result establishes a link between the eigenvalues of $\mathrm{L}_{\mathcal{A}}$ and the maximum points of $f_{\mathcal{A}}$.

Theorem 7.3.6. *If* $\mathcal{A} \in M_n(\mathbb{R})$ *is symmetric and if* X *is a maximum for* $f_{\mathcal{A}}$ *on* $\mathbb{S}^{n-1}$*, then the value* $f_{\mathcal{A}}(X)$ *is an eigenvalue for* $\mathrm{L}_{\mathcal{A}}$ *with a corresponding eigenvector* X*,* i.e.

$$\mathrm{L}_{\mathcal{A}}(X) = \mathcal{A}X = f_{\mathcal{A}}(X)\,X.$$

Proof. Let $H_{\mathbf{0},X} = \{Y \in \mathbb{R}^n \mid Y \cdot X = 0\}$ be the hyperplane perpendicular to X, of dimension $n-1$, and let us choose any $Y \in H_{\mathbf{0},X}$ with $\|Y\| = 1$. For any $t \in \mathbb{R}$, one sets

$$C(t) := \cos(t)\,X + \sin(t)\,Y \quad \in \mathbb{R}^n.$$

Observe that since $X \cdot Y = 0$ one has

$$\|C(t)\|^2 = \|\cos(t)\,X\|^2 + \|\sin(t)\,Y\|^2 = \cos^2(t)\,\|X\|^2 + \sin^2(t)\,\|Y\| = \cos^2(t) + \sin^2(t) = 1.$$

It follows that for any $t \in \mathbb{R}$ the point $C(t)$ belongs to $\mathbb{S}^{n-1}$, and in addition one has $C(0) = X$. In more precise words, the map

$$\mathbb{R} \ni t \mapsto C(t) \in \mathbb{S}^{n-1}$$

is a curve on $\mathbb{S}^{n-1}$ passing through X for $t = 0$. Let us also observe that

$$C'(t) = -\sin(t)\,X + \cos(t)\,Y$$

and that $C'(0) = Y$. Note that this latter quantity corresponds to the direction of the curve at $t = 0$

Consider now the map $\mathbb{R} \ni t \mapsto f_{\mathcal{A}}\big(C(t)\big) \equiv {}^tC(t)\mathcal{A}C(t) \quad \in \mathbb{R}$. Since $f_{\mathcal{A}}(X)$ is maximal and since $C(0) = X$, this map $t \mapsto f_{\mathcal{A}}\big(C(t)\big)$ is (locally) maximal at $t = 0$, and thus $f_{\mathcal{A}}\big(C(t)\big)'\big|_{t=0} = 0$. Since one has

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\big({}^tC(t)\mathcal{A}C(t)\big)\big|_{t=0} &= \big({}^tC'(t)AC(t) + {}^tC(t)\mathcal{A}C'(t)\big)\big|_{t=0} \\ &= {}^tY\mathcal{A}X + {}^tX\mathcal{A}Y \\ &= 2{}^tY\mathcal{A}X,\end{aligned}$$

where we have used that ${}^tY\mathcal{A}X = {}^tX\mathcal{A}Y$ (see Exercise 5.6), it follows that ${}^tY\mathcal{A}X = 0$ for any $Y \in H_{\mathbf{0},Y}$. In addition, since ${}^tY\mathcal{A}X = Y \cdot (\mathcal{A}X)$, one infers that $\mathcal{A}X \in H_{\mathbf{0},X}^{\perp}$, and consequently that $\mathcal{A}X = \lambda X$ for some $\lambda \in \mathbb{R}$ (recall that $H_{\mathbf{0},X}$ is of dimension $n-1$ and thus that only $\mathrm{Vect}(X)$ is perpendicular to it).

Finally, one observes that since $\|X\| = 1$ one has

$$f_{\mathcal{A}}(X) = {}^tX\mathcal{A}X = X \cdot (\mathcal{A}X) = X \cdot (\lambda X) = \lambda\|X\|^2 = \lambda$$

which means that $\mathrm{L}_{\mathcal{A}}(X) = \mathcal{A}X = f_{\mathcal{A}}(X)X$, as expected. □

Let us observe that by using the notation introduced in Chapter 5 one has

$$f_{\mathcal{A}}(X) = {}^tX\mathcal{A}X = X \cdot (\mathcal{A}X) = \langle X, \mathcal{A}X\rangle = \langle X, \mathrm{L}_{\mathcal{A}}(X)\rangle$$

and that

$$H_{\mathbf{0},X} = \{Y \in \mathbb{R}^n \mid Y \cdot X = 0\} = \{Y \in \mathbb{R}^n \mid \langle Y, X\rangle = 0\}.$$

Thus, what really matters in the previous statement and its proof is the existence of a scalar product, and that $\langle Y, \mathrm{L}_{\mathcal{A}}(X)\rangle = \langle \mathrm{L}_{\mathcal{A}}(Y), X\rangle$ (which is a more general formulation of the equality ${}^tY\mathcal{A}X = {}^tX\mathcal{A}Y$). By using this observation, one can easily generalize the previous proof and statement. For that purpose, let us first provide a new definition.

Definition 7.3.7. *Let V be a vector space and let $\langle\cdot,\cdot\rangle$ be a scalar product on V. A linear map $\mathrm{L} : V \to V$ is* symmetric with respect to the scalar product *if it satisfies*

$$\langle Y, \mathrm{L}(X)\rangle = \langle \mathrm{L}(Y), X\rangle \qquad \forall X, Y \in V.$$

Theorem 7.3.8. *Let V be a finite dimensional vector space endowed with a scalar product, and let $\mathrm{L} : V \to V$ be a linear map which is symmetric with respect to the scalar product. Then L possess an eigenvalue, with eigenvector $X \neq \mathbf{0}$.*

Definition 7.3.9. *Let V be a vector space and $\mathrm{L} : V \to V$ be a linear map. A subspace $W \subset V$ is* stable for L *if* $\mathrm{L}(W) \subset W$, i.e. *if whenever $X \in W$ then* $\mathrm{L}(X) \in W$.

Examples 7.3.10. *(i)* $\{\mathbf{0}\}$ *and V are always stable for any linear map* $\mathrm{L} : V \to V$,

(ii) $\mathrm{Ker(L)}$ *is stable since for any $X \in \mathrm{Ker(L)}$ one has* $\mathrm{L}(X) = \mathbf{0} \in \mathrm{Ker(L)}$,

(iii) If W is the eigenspace associated with an eigenvalue λ of L, *then W is stable.*

For the next statement, recall that if W is a subspace of a vector space V endowed with a scalar product, then

$$W^\perp = \{Y \in V \mid \langle Y, X\rangle = 0\ \forall X \in W\}.$$

Lemma 7.3.11. *Let V be a finite dimensional vector space and let $\langle\cdot,\cdot\rangle$ be a scalar product on V. Let $\mathrm{L} : V \to V$ be a linear map which is symmetric with respect to the scalar product. If W is stable for* L, *then $W^\perp$ is stable for* L.

Proof. Let $Y \in W^\perp$ and $X \in W$, then $\langle \mathrm{L}(Y), X\rangle = \langle Y, \mathrm{L}(X)\rangle = 0$ since $\mathrm{L}(X) \in W$. Thus $\mathrm{L}(Y) \in W^\perp$ for any $Y \in W^\perp$, which means precisely that $W^\perp$ is stable. □

We can now state and prove the most important result of this section.

Theorem 7.3.12. *Let V be a vector space of dimension n and endowed with a scalar product $\langle\cdot,\cdot\rangle$. Let $\mathrm{L} : V \to V$ be a linear map which is symmetric with respect to the scalar product. Then V possesses an orthonormal basis of eigenvectors of* L. *In other words there exist $Y_1, \dots, Y_n$ mutually orthogonal and with $\|Y_j\|^2 = \langle Y_j, Y_j\rangle = 1$ such that $V = \mathrm{Vect}(Y_1, \dots, Y_n)$ and such that $\mathrm{L}(Y_j) = \lambda_j Y_j$ for some λ_j.*

Proof. By Theorem 7.3.8 there exists $X_1 \neq \mathbf{0}$ such that $\mathrm{L}(X_1) = \lambda_1 X_1$ for some λ_1. If one sets $W_1 = \mathrm{Vect}(X_1)$, then W is stable for L, and the same property holds for $W_1^\perp$. Thus $W_1^\perp$ is a subspace of V of dimension $n-1$, and L is a symmetric linear map in $W_1^\perp$ (endowed with the scalar product inherited from V). Thus, we can again apply the previous theorem in $W_1^\perp$ instead of in V, and there exists $X_2 \in W_1^\perp$ with $X_2 \neq \mathbf{0}$, such that $\mathrm{L}(X_2) = \lambda_2 X_2$. Then, by defining $W_2 := \mathrm{Vect}(X_2)$, one obtains that $W_2^\perp$ (the subspace orthogonal to W_2 in W_1) is of dimension $n-2$, and is stable for L. Since L is a symmetric linear map in $W_2^\perp$ one can go on iteratively in the procedure, up to W_n.

Finally, by fixing $Y_j := X_j/\|X_j\|$ one gets that $Y_j \in W_j$, that $\|Y_j\| = 1$ and by construction Y_j is orthogonal to Y_k whenever $j \neq k$. One has thus obtained a basis of V which satisfies the stated properties. □

Remark 7.3.13. *In the basis $\{Y_1, \dots, Y_k\}$ the linear map* L *is diagonal. Whenever there exists a basis such that a linear map* L *is diagonal is this basis, one says that* L *is* diagonalizable.

Let us summarize our findings: One has obtained that in a vector space of finite dimension and endowed with a scalar product, any symmetric linear map is diagonalizable. Equivalently, if $\mathcal{A} \in M_n(\mathbb{R})$ is symmetric, then the linear map $\mathrm{L}_\mathcal{A}$ is diagonalizable. In particular it means that if $\mathcal{A} \in M_n(\mathbb{R})$ is symmetric, then there exists $\lambda_1, \dots \lambda_n \in \mathbb{R}$ such that

$$P_\mathcal{A}(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots(\lambda_n - \lambda). \tag{7.3.1}$$

Note that all λ_j need not be different. For example, one could have $\lambda_2 = \lambda_1$ but $\lambda_3 \neq \lambda_1$. The number of times a value λ_j appears in this decomposition is called *the algebraic multiplicity* of the eigenvalue λ_j. What the previous theorem says is that if $\mathcal{A}$ is symmetric, the algebraic multiplicity of an eigenvalue is equal to the geometric multiplicity of this eigenvalue (*i.e.* to the dimension of the corresponding eigenspace). Note that this equality holds for symmetric matrices, but it is not true in general.

7.4 Complex vector spaces

In Chapter 9, the field $\mathbb{C}$ of complex numbers is recalled. Thus, one can speak about complex vector spaces, as for example $\mathbb{C}^n$, which is of dimension n. One can also freely speak about $M_n(\mathbb{C})$, *i.e.* matrices with each entry in $\mathbb{C}$.

For any $\mathcal{A} \in M_n(\mathbb{C})$, let us consider $\mathrm{L}_\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^n$ defined by $\mathrm{L}_\mathcal{A}(X) = \mathcal{A}X$ which is obviously a linear map. Then, *the fundamental theorem of algebra* says that there exist $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ such that

$$P_\mathcal{A}(\lambda) = \mathrm{Det}(\mathcal{A} - \lambda\mathbf{1}_n) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots(\lambda_n - \lambda).$$

Note that we have already seen such a factorization in equation (7.3.1), but it was only for symmetric matrices. Here, there is no restriction on $\mathcal{A}$, but the eigenvalues λ_j can be complex. In other words, this fundamental theorem of algebra claims that counting multiplicity there always exist n solutions to the equation $P_\mathcal{A}(\lambda) = 0$. However, be careful that this factorization does not imply that any matrix $\mathcal{A}$ is diagonalizable, even on $\mathbb{C}^n$. For example, for the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, one has $P_\mathcal{A}(\lambda) = \lambda^2$ (which means that $\lambda_1 = \lambda_2 = 0$), but this matrix can not be diagonalized in any basis.

Another natural question when dealing with $\mathbb{C}^n$ is how to endow it with a scalar product ? Let us recall that a scalar product was used for defining a norm by the relation $\|X\|^2 = \sqrt{\langle X, X\rangle}$, see Definition 5.1.5. For example, if $x \in \mathbb{R}$, it is necessary that $\langle x, x\rangle \geq 0$. Thus, let us consider two complex numbers z_1, z_2 and set

$$\langle z_1, z_2\rangle := z_1\,\overline{z_2}. \tag{7.4.1}$$

Then one observes that if $z = x + iy$ with $x, y \in \mathbb{R}$ one has

$$\langle z, z\rangle = (x + iy)\overline{(x + iy)} = (x + iy)(x - iy) = x^2 + y^2 \geq 0.$$

In fact, this corresponds to the (square of the) norm of z when one identifies $\mathbb{C}$ with the plane $\mathbb{R}^2$. Similarly, if $Z = {}^t(z_1, \dots, z_n) \in \mathbb{C}^n$ and $Z' = {}^t(z'_1, \dots, z'_n) \in \mathbb{C}^n$, one sets

$$\langle Z, Z' \rangle := \sum_{j=1}^{n} z_j \overline{z'_j} \tag{7.4.2}$$

and observes again that $\langle Z, Z \rangle \geq 0$.

In Chapter 5, the abstract notion of a scalar product was defined for real vector space. Let us complement this definition in the case of a complex vector space (but observe that the real scalar product is a special case of the following definition).

Definition 7.4.1. A scalar product *on a complex vector space V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ such that for any $X, Y, Z \in V$ and $\lambda \in \mathbb{C}$ one has*

(i) $\langle X, Y \rangle = \overline{\langle Y, X \rangle}$,

(ii) $\langle X + Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$,

(iii) $\langle \lambda X, Y \rangle = \lambda \langle X, Y \rangle = \langle X, \overline{\lambda} y \rangle$,

(iv) $\langle X, X \rangle \geq 0$ *and* $\langle X, X \rangle = 0$ *if and only if* $X = \mathbf{0}$.

It is then easily observed that the definition provided in (7.4.1) and in (7.4.2) are indeed scalar product on $\mathbb{C}$ and $\mathbb{C}^n$ respectively.

Let us now consider $\mathcal{A} = (a_{ij}) \in M_n(\mathbb{C})$ and let $Z, Z' \in \mathbb{C}^n$. Then one has

$$\begin{aligned}
\langle \mathrm{L}_{\mathcal{A}}(Z), Z' \rangle &= \langle \mathcal{A}Z, Z' \rangle = \sum_{j=1}^{n} (\mathcal{A}Z)_j \overline{Z'_j} = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} Z_k \overline{Z'_j} \\
&= \sum_{k=1}^{n} \sum_{j=1}^{n} Z_k {}^t a_{kj} \overline{Z'_j} = \sum_{k=1}^{n} Z_k \Big(\sum_{j=1}^{n} \overline{\overline{{}^t a_{kj}} Z'_j} \Big) = \langle Z, \overline{{}^t\mathcal{A}} Z' \rangle \\
&= \langle Z, \mathrm{L}_{\overline{{}^t\mathcal{A}}}(Z') \rangle.
\end{aligned}$$

For simplicity, let us set $\mathcal{A}^* := \overline{{}^t\mathcal{A}}$. We have thus shown that $\langle \mathrm{L}_{\mathcal{A}}(Z), Z' \rangle = \langle Z, \mathrm{L}_{\mathcal{A}^*}(Z') \rangle$.

In the next statement, we rephrase in this more precise setting what has already been obtained in Theorem 7.3.12.

Theorem 7.4.2. *If $\mathcal{A} \in M_n(\mathbb{C})$ satisfies $\mathcal{A}^* = \mathcal{A}$, then $\mathrm{L}_{\mathcal{A}}$ is diagonalizable, with n real eigenvalues λ_j.*

For completeness, let us check that the eigenvalues of $L_{\mathcal{A}}$ are real, provided $\mathcal{A}^* = \mathcal{A}$. Thus, assume that λ_j is an eigenvalue of $\mathrm{L}_{\mathcal{A}}$ with corresponding eigenvector $X_j \neq 0$ and observe that

$$\begin{aligned}
\lambda_j \|X_j\|^2 &= \langle \lambda_j X_j, X_j \rangle = \langle L_{\mathcal{A}}(X_j), X_j \rangle = \langle X_j, L_{\mathcal{A}}(X_j) \rangle \\
&= \langle X_j, \lambda_j X_j \rangle = \overline{\lambda_j} \langle X_j, X_j \rangle = \overline{\lambda_j} \|X_j\|^2.
\end{aligned}$$

Since $\|X_j\| \neq 0$ it follows that $\lambda_j = \overline{\lambda_j}$, which implies that λ_j is real.

Example 7.4.3. *If* $\mathcal{A} = \left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right) \in M_2(\mathbb{C})$*, then* $\mathcal{A}^* = \mathcal{A}$ *and one observes that* $P_{\mathcal{A}}(\lambda) = \mathrm{Det}\left(\begin{smallmatrix} -\lambda & i \\ -i & -\lambda \end{smallmatrix}\right) = (\lambda+1)(\lambda-1)$*. Thus the eigenvalue of* $\mathcal{A}$ *are real, even so* $\mathcal{A}$ *looks rather complex !*

Remark 7.4.4. *Let us stress that Theorem 7.4.2 is at the root of quantum mechanics. Indeed, in a suitable framework it says that "the observables have real spectrum".*

7.5 Exercises

Exercise 7.1. *Let* $\mathrm{P} : V \to V$ *be a linear map on a vector space* V *and assume that* P *is a projection. Show that* P *can only have two possible eigenvalues, namely* 0 *and* 1.

Exercise 7.2. *For any* $\theta \in [0, 2\pi)$, *consider the matrix* $\mathcal{A}(\theta) := \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$ *and show that the corresponding linear map* $\mathrm{L}_{\mathcal{A}(\theta)} : \mathbb{R}^2 \to \mathbb{R}^2$ *always admits the eigenvalue* 1.

Exercise 7.3. *Consider the matrix* $\mathcal{A} := \begin{pmatrix} 2 & 0 \\ 3 & 4 \end{pmatrix}$ *and show that* 2 *and* 4 *are eigenvalues of the associated linear map* $\mathrm{L}_{\mathcal{A}}$. *What are all corresponding eigenvectors ? Similarly, consider the matrix* $\mathcal{B} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 7 \end{pmatrix}$ *and show that* 1, 5 *and* 7 *are eigenvalues of the associated linear map. Determine the corresponding eigenspaces.*

Exercise 7.4. *Let* $\mathcal{A} \in M_n(\mathbb{R})$ *be invertible, and assume that* $\lambda \in \mathbb{R}$ *is an eigenvalue of* $\mathrm{L}_{\mathcal{A}}$ *with* $X \in \mathbb{R}^n$ *a corresponding eigenvector.*

1. *Is* X *an eigenvector of* $\mathrm{L}_{\mathcal{A}^3}$ *? If so, what is the corresponding eigenvalue ?*
2. *Is* X *an eigenvector of the linear map associated with* $\mathcal{A} + 2\mathbf{1}_n$ *? If so, what is the corresponding eigenvalue ?*
3. *Is* X *an eigenvector of* $\mathrm{L}_{4\mathcal{A}}$ *? If so, what is the corresponding eigenvalue ?*
4. *Can* λ *be equal to* 0 *?*
5. *Is* X *an eigenvector of* $\mathrm{L}_{\mathcal{A}^{-1}}$ *? If so, what is the corresponding eigenvalue ?*
6. *What can you say about* $\mathrm{Ker}(\mathrm{L}_{\mathcal{A}} - \lambda \mathbf{1})$ *?*
7. *What can you say about* $\mathrm{Det}(\mathcal{A} - \lambda \mathbf{1}_n)$ *?*

Exercise 7.5. *For any* $\mathcal{A} \in M_2(\mathbb{R})$, *show the following equality*

$$P_{\mathcal{A}}(\lambda) = \lambda^2 - \lambda \mathrm{Tr}(\mathcal{A}) + \mathrm{Det}(\mathcal{A}).$$

Exercise 7.6. *Let* $\mathcal{A} \in M_n(\mathbb{R})$ *and assume that* $\mathcal{A}$ *has* n *eigenvalues* $\lambda_1, \ldots, \lambda_n$. *Then, show the following equalities:*

(i) $\mathrm{Det}(\mathcal{A}) = \lambda_1 \lambda_2 \ldots \lambda_n$ *(product of the eigenvalues)*

(ii) $\mathrm{Tr}(\mathcal{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$ *(sum of the eigenvalues)*

Exercise 7.7. *Let* $\mathcal{A} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 7 \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_{\mathcal{A}} : \mathbb{R}^3 \to \mathbb{R}^3$. *Determine the eigenvalues of* $L_{\mathcal{A}}$ *and the corresponding eigenspaces.*

Exercise 7.8. *Let* $\mathcal{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$. *Determine the eigenvalues of* $\mathrm{L}_\mathcal{A}$ *and the corresponding eigenspaces.*

Exercise 7.9. *Let* $\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2$. *Determine the eigenvalues of* $\mathrm{L}_\mathcal{A}$ *and the corresponding eigenspaces. Consider then the matrix* $\mathcal{B} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$ *and compute the product* $\mathcal{B}^{-1}\mathcal{A}\mathcal{B}$. *What do you observe, and how do you understand your result ?*

Exercise 7.10. *Let* $\mathcal{A} = \begin{pmatrix} -2 & -7 \\ 1 & 2 \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2$. *Determine the eigenvalues of* $\mathrm{L}_\mathcal{A}$ *and the corresponding eigenspaces.*

Exercise 7.11. *Let* $\mathcal{A} \in M_n(\mathbb{R})$ *and consider the linear maps* $\mathrm{L}_\mathcal{A}$ *and* $\mathrm{L}_{{}^t\mathcal{A}}$. *Show that these linear maps have the same eigenvalues.*

Exercise 7.12. *Show that if* $\mathcal{A} \in M_n(\mathbb{R})$ *is orthogonal (*i.e. ${}^t\mathcal{A} = \mathcal{A}^{-1}$*), then the (real) eigenvalues of* $\mathrm{L}_\mathcal{A}$ *can only be* 1 *or* -1.

Exercise 7.13. *For* $\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ *consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$. *Determine the eigenvalues of* $\mathrm{L}_\mathcal{A}$ *and the corresponding eigenspaces. Find the change of bases such that in the corresponding new basis this linear map becomes diagonal.*

Exercise 7.14. *Let* $\mathcal{A} \in M_n(\mathbb{R})$ *be symmetric. Show that there exists* $\mathcal{B} \in M_n(\mathbb{R})$ *such that* $\mathcal{B}^3 = \mathcal{A}$.

Exercise 7.15. *For a symmetric matrix* $\mathcal{A} \in M_n(\mathbb{R})$, *one says that* $\mathcal{A}$ *is* positive definite *if* $\langle \mathcal{A}X, X\rangle > 0$ *for any* $X \in \mathbb{R}^n$ *with* $X \neq \mathbf{0}$. *In fact, this is precisely the condition which makes the bilinear map* $\mathrm{F}_\mathcal{A}$ *define a scalar product, see Exercise 5.6. If* $\mathcal{A}$ *is symmetric and positive definite, show that*

1. *All eigenvalues of* $\mathrm{L}_\mathcal{A}$ *are strictly positive,*
2. $\mathcal{A}^2$ *is symmetric and positive definite,*
3. $\mathcal{A}^{-1}$ *is symmetric and positive definite.*

Exercise 7.16. *Let* $\mathcal{A} = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2$. *Compute* $\mathcal{A}^n$ *for* $n = 2$, $n = 3$, $n = 25$ *and* $n = \infty$. *You are allowed to use the result of Exercise 7.9.*

Exercise 7.17. *Let* $\mathcal{A} = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$, *and consider the associated linear map* $\mathrm{L}_\mathcal{A} : \mathcal{C}^2 \to \mathcal{C}^2$. *Determine the eigenvalues of* $\mathrm{L}_\mathcal{A}$ *and the corresponding eigenspaces. Show that these eigenspaces are orthogonal.*

Exercise 7.18. *Do there exist* $\mathcal{A}, \mathcal{B} \in M_n(\mathbb{R})$ *such that* $\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A} = \mathbf{1}_n$ *? Justify your answer. Note that the notion of trace can be useful for this exercise.*

Chapter 8

Applications

In this chapter, we provide two examples of applications of the various concepts and results we have developed in the previous chapters.

8.1 Discrete dynamical systems

8.1.1 Coyotes and roadrunners

Let us consider two species in a desert: coyotes and roadrunners. We denote by $c(n)$ the population of coyotes at year n, and by $r(n)$ the population of roadrunners at year n. The following equation models the transformation of the system, as a function of the year:

$$\begin{cases} c(n+1) = 0.86\,c(n) + 0.08\,r(n) \\ r(n+1) = -0.12\,c(n) + 1.14\,r(n) \end{cases} \tag{8.1.1}$$

Note that each of these coefficients has an interpretation: 0.86 and 1.14 correspond to the birth rate of the coyotes and or the roadrunners, respectively. The coefficient 0.08 can be interpreted as a favorable factor on the population of coyotes whenever roadrunners can be eaten, while -0.12 corresponds to the decrease in the population of roadrunners due to coyotes' appetite. Now, if one sets $X(n) = \begin{pmatrix} c(n) \\ r(n) \end{pmatrix}$ one can then rewrite this system as

$$X(n+1) = \begin{pmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{pmatrix} X(n).$$

For later use, we set $\mathcal{A} := \left(\begin{smallmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{smallmatrix}\right)$.

Question: If at year $n = 0$ we observe the populations $X(0) = \begin{pmatrix} c(0) \\ r(0) \end{pmatrix}$, what about the populations $X(n)$ for large n, *i.e.* in the far future ?

For example, it $X(0) = \left(\begin{smallmatrix} 100 \\ 100 \end{smallmatrix}\right)$, then $X(1) = \mathcal{A}X(0) = \left(\begin{smallmatrix} 96 \\ 102 \end{smallmatrix}\right)$, $X(2) = \mathcal{A}^2X(0) = \mathcal{A}X(1)$, and $X(10) = \mathcal{A}^{10}X(0) \cong \left(\begin{smallmatrix} 80 \\ 170 \end{smallmatrix}\right)$. Here the computations are rather lengthy. On the other hand, if $X(0) = \binom{100}{300}$, then $X(1) = \binom{110}{330} = 1.1X(0)$, $X(2) = \mathcal{A}^2X(0) = \mathcal{A}X(1) = \mathcal{A}\big(1.1X(0)\big) = 1.1^2X(0)$, and thus $X(10) = 1.1^{10}X(0)$. In such a case, the computations are easier, but note that the populations are continuously increasing (there is no correlation between these two facts). Note also that a similar computation shows that if $X(0) = \binom{200}{100}$, then one has $X(n) \to \binom{0}{0}$ as n tends to infinity.

Now, a natural question is how can one organize these information more systematically ? In fact, this is possible with the help of the eigenvalues and the eigenvectors of the matrix $\mathcal{A}$. For that purpose, let us first observe that

$$P_{\mathcal{A}}(\lambda) = \lambda^2 - 2\lambda + 0.99 = (\lambda - 1.1)(\lambda - 0.9).$$

Thus, the eigenvalues of $\mathcal{A}$ are 1.1 and 0.9. Moreover, if X_1 and X_2 are eigenvectors of $\mathcal{A}$ associated with the eigenvalues 1.1 and 0.9 respectively, recall that we can define an invertible matrix $\mathcal{B}$ by $\mathcal{B} = \big(X_1\ X_2\big)$, and then that

$$\mathcal{B}^{-1}\mathcal{A}\mathcal{B} = \begin{pmatrix} 1.1 & 0 \\ 0 & 0.9 \end{pmatrix} \quad \text{or equivalently} \quad \mathcal{A} = \mathcal{B}\begin{pmatrix} 1.1 & 0 \\ 0 & 0.9 \end{pmatrix}\mathcal{B}^{-1}.$$

With this rather simple information at hands, some computations simplify a lot. For example, one directly obtains that

$$\mathcal{A}^n = \mathcal{B}\begin{pmatrix} 1.1^n & 0 \\ 0 & 0.9^n \end{pmatrix}\mathcal{B}^{-1}$$

which requires much less efforts than multiplying n times $\mathcal{A}$ by itself.

Alternatively, since $\{X_1, X_2\}$ generate a basis of $\mathbb{R}^2$, any initial condition $X(0)$ can be decomposed with respect to this basis and one has $X(0) = c_1X_1 + c_2X_2$, with $c_1, c_2 \in \mathbb{R}$. Then, one has

$$\begin{aligned} \mathcal{A}^nX(0) &= \mathcal{A}^n(c_1X_1 + c_2X_2) = c_1\mathcal{A}^nX_1 + c_2\mathcal{A}^nX_2 \\ &= c_1(1.1)^nX_1 + c_2(0.9)^nX_2 = 1.1^nc_1X_1 + 0.9^nc_2X_2. \end{aligned}$$

Thus, knowing the eigenvalues of $\mathcal{A}$, one can better understand why, depending on the initial populations, the populations can either increase as n tends to infinity (due to the eigenvalue 1.1), or vanish as n tends to infinity (due to the eigenvalue 0.9).

Question: Can one find $X(0)$ such that both populations remain constant as n goes to infinity ?

8.1.2 Discrete dynamical systems with real eigenvalues

More generally, suppose that we consider discrete evolution system given by the equation $X(n+1) = \mathcal{A}X(n)$ for some $\mathcal{A} \in M_m(\mathbb{R})$. Assume in addition that $\mathcal{A}$ is diagonalizable, with its m eigenvalues real, *i.e.* there exists an invertible matrix $\mathcal{B} \in M_m(\mathbb{R})$ such that $\mathcal{A} = \mathcal{B}\ \mathrm{diag}(\lambda_1, \dots, \lambda_m)\ \mathcal{B}^{-1}$, with $\lambda_j \in \mathbb{R}$ for any $j \in \{1, \dots, m\}$. Then, the family of eigenvectors associated with the eigenvalues of $\mathcal{A}$ generates a basis of $\mathbb{R}^m$ and any initial state $X(0)$ can be decomposed with respect to this basis. Thus, if X_j denotes an eigenvector associated with the eigenvalue λ_j one has $X(0) = c_1X_1 + \cdots + c_mX_m$ for some $c_j \in \mathbb{R}$, and again

$$\mathcal{A}^n X(0) = \lambda_1^n c_1 X_1 + \lambda_2^n c_2 X_2 + \cdots + \lambda_m^n c_m X_m.$$

Then, depending on λ_j, the large n behavior of the system can be predicted. Note that in the following description, the index j is an arbitrary element of $\{1, \ldots, m\}$.

(i) If $\lambda_j > 1$, the corresponding part of the system grows infinitely,

(ii) If $\lambda_j = 1$, the corresponding part of the system remains the same forever,

(iii) If $0 < \lambda_j < 1$, the corresponding part of the system tends to vanish as n tends to infinity,

(iv) If $\lambda_j = 0$, the corresponding part of the system disappears already for $n = 1$ (this part of the system belongs to the kernel of $\mathcal{A}$),

(v) If $-1 < \lambda_j < 0$, the corresponding part of the system tends to vanish as n tends to infinity, but its sign is alternating for n even or n odd,

(vi) If $\lambda_j = -1$, the corresponding part of the system alternates between two states with a different sign,

(vii) If $\lambda_j < -1$, the corresponding grows infinitely in norm, but its sign is alternating for n even or n odd.

8.1.3 Discrete dynamical systems with complex eigenvalues

Let us now consider a discrete evolution system given by the equation $X(n+1) = \mathcal{A}X(n)$ with $\mathcal{A} \in M_2(\mathbb{R})$, but with $P_{\mathcal{A}}(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)$ and $\lambda_1 = x + iy$, $\lambda_2 = x - iy$ and $y \neq 0$. In such a case, there again exists an invertible matrix $\mathcal{B}$ such that $\mathcal{A} = \mathcal{B}\begin{pmatrix} x+iy & 0 \\ 0 & x-iy \end{pmatrix}\mathcal{B}^{-1}$ and therefore

$$\mathcal{A}^n = \mathcal{B}\begin{pmatrix} (x+iy)^n & 0 \\ 0 & (x-iy)^n \end{pmatrix}\mathcal{B}^{-1}.$$

In order to compute $(x \pm iy)^n$, let us first observe that

$$x \pm iy = |x \pm iy| \, \frac{x \pm iy}{|x \pm iy|} = \sqrt{x^2+y^2}\Big(\frac{x}{\sqrt{x^2+y^2}} \pm i \frac{y}{\sqrt{x^2+y^2}}\Big).$$

As a consequence, one can write $x + iy = \sqrt{x^2+y^2}\big(\cos(2\pi\theta) + i\sin(2\pi\theta)\big)$ for a unique $\theta \in [0, 1)$. Then, by using de Moivre's formula, as shown in Exercise 9.6, one gets that

$$(x+iy)^n = (x^2+y^2)^{n/2}\big(\cos(2\pi n\theta) + i\sin(2\pi n\theta)\big).$$

Note that a similar formula holds for $(x - iy)$ and for $(x - iy)^n$. Then, depending if $\sqrt{x^2 + y^2}$ is bigger, equal or smaller than 0, and depending if θ is rational or irrational (*i.e.* if $\theta = p/q$ for some $p, q \in \mathbb{Z}$ or not), the asymptotic behavior of $(x + iy)^n$ changes drastically. For example, if $\sqrt{x^2 + y^2} > 1$, the $|(x + iy)^n|$ goes to infinity as n goes to infinity, while if $\sqrt{x^2 + y^2} < 1$, then $|(x + iy)^n|$ goes to 0 as n tends to infinity. If $\sqrt{x^2 + y^2} = 1$ and if $\theta = p/q$ for some $p, q \in \mathbb{Z}$, then $(x + iy)^n$ is periodic, with a period depending on p and q, while if $\theta \neq p/q$ for any $p, q \in \mathbb{Z}$, then $(x + iy)^n$ takes different values for any n. A notion of aperiodicity appears in fact in such a situation. Note that an enumeration of all possible behaviors as in the previous subsection could also be established in the present setting.

8.2 The \$ 25'000'000'000 eigenvector

This section is inspired from the paper "The \$25'000'000'000 eigenvector: the linear algebra behind Google"[1], which gives an another opportunity to use the concepts introduced in the previous chapters.

Let us first list what a search engine for internet has to do:

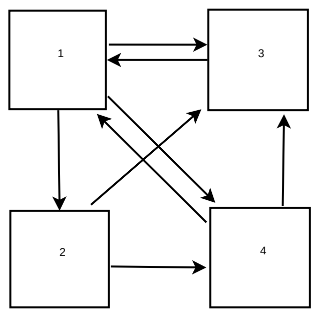
(i) Locate all web pages with public access,

(ii) Index these data with keywords,

(iii) Rate the importance of each page.

Question: How can one define and quantify the "importance" of a web page ?

We shall call *the importance score* or simply *the score* such a quantitative rating. The main idea behind Google page ranking is derived from the links to that page (called backlinks). For example, let us look at the world wide web which contains only 4 pages as represented in the following figure. Each arrow $A \longrightarrow B$ represents a link from A to



B (a backlink for B) Let us also set $x_k \geq 0$ for the importance of the page k, with the convention that $x_j > x_k$ means that the page j is more important than the page k.

[1]Kurt Bryan, Tanya Leise, *The \$25'000'000'000 eigenvector: the linear algebra behind Google*, SIAM REVIEW, Vol. 48, No. 3, pp. 569–581.

A first idea for the ranking could be to assign for x_k the number of backlinks. In this case, one would obtain $x_1 = 2$, $x_2 = 1$, $x_3 = 3$ and $x_4 = 2$, and the page 3 would then be the most important one. However, one also would like that a link to page j from an important page boosts the score of page j more than a link to page j from an unimportant page. For example, a link from BBC.com to your web page is certainly more important than a link from the web page of your neighbour to your web page. Thus, a more refined way to compute the score should be implemented.

A second idea for the ranking could be to score the page j with the sum of the score of the pages linking to page j. With this approach, an important page with a link to page j would boost the score of page j. Of course, the procedure becomes more complicated because it is self-referential. In addition, a single page should not gain any importance by containing too many links. For that purpose, we shall impose that each page has a total vote of 1, as in a democracy.

Taking these remarks into account, let us set

$$x_j := \sum_{\text{pages } k \text{ linking to } j} \frac{x_k}{n_k}$$

where n_k is the number of links emerging from page k. Note that if a page has a link to itself, this link is ignored. For the example shown above, we then obtain

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

which is equivalent to $X = \mathcal{A}X$ with $X = {}^t(x_1, x_2, x_3, x_4)$ and $\mathcal{A}$ the matrix shown above. In other words, computing the scores x_j corresponds to finding an eigenvector X for the linear map $\mathrm{L}_{\mathcal{A}}$ associated with the eigenvalue 1, if such an eigenvalue exists. In this setting, the matrix $\mathcal{A}$ is called *the link matrix* for a given web.

Note that in the example shown above, any multiple of the vector ${}^t(12, 4, 9, 6)$ is an eigenvector of $\mathrm{L}_{\mathcal{A}}$ associated with the eigenvalue 1. In we impose in addition that $\sum_j x_j = 1$, then we get $x_1 = 12/31 \cong 0.387$, $x_2 = 4/31 \cong 0.129$, $x_3 = 9/31 \cong 0.290$ and $x_4 = 6/31 \cong 0.194$. Let us observe that page 3 is no more the most important one. Indeed, this important page has only one link to page 1, and this single link boosts the score of page 1 which then becomes the most important one.

Let us now try to think a little bit more generally. Assume that the web has no page with 0 outgoing link, which means that $n_k \neq 0$ for any k. With such an assumption, the entries of any column of a link matrix sum up to 1. Indeed, each page j gives $\frac{1}{n_j}$ of its vote to n_j different pages. With this observation, let's come back to mathematics.

Lemma 8.2.1. *If $n_k \neq 0$ for any k, then the linear map associated with any link matrix possesses the eigenvalue* 1.

Proof. Observe that 1 is an eigenvalue of $L_{{}^t\mathcal{A}}$ with eigenvector ${}^t(1, 1, \dots 1)$. Indeed, the entries in each row of ${}^t\mathcal{A}$ sum to 1, and therefore

$$ {}^t\mathcal{A} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. $$

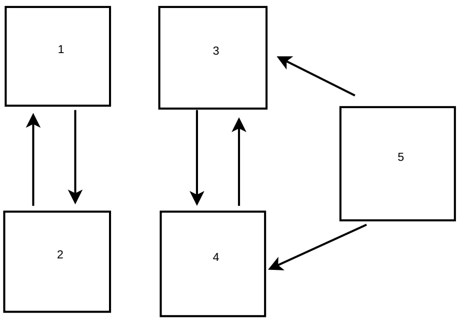
Since $L_{{}^t\mathcal{A}}$ and $L_{\mathcal{A}}$ share the same spectrum (see Lemma 7.2.9), it follows that 1 also belongs to the spectrum of $L_{\mathcal{A}}$. □

Since the spectrum of the linear map associated with any link matrix contains the value 1, let us denote by V_1 the corresponding eigenspace. We can then wonder if this eigenspace is of dimension one (in which case the ranking is unique) or if it is of dimension higher than one (in which case there exist different rankings which can not be distinguished with our criteria) ? Another question is how to get rid of the assumption $n_k \neq 0$, since this assumption is not always satisfied (there exist interesting web pages without any outgoing link) ?

Let us now observe that the non-uniqueness of the ranking is possible if the web is disconnected. For example, consider the link matrix

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

associated with the following world wide web Such a disconnected web gives rise to a



matrix $\mathcal{A}$ which is block diagonal, see Exercises 4.28 and 4.29. In the above example, and more generally in any situation with a link matrix which is block diagonal, it is not difficult to see that the eigenspace associated with the eigenvalue 1 is of dimension 2 or higher.

One way to solve this problem is to replace the link matrix $\mathcal{A}$ by a slightly improved version of it. More precisely, consider the matrix

$$ \mathcal{A}_m = (1 - m)\mathcal{A} + m \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{pmatrix} $$

for some $m \in [0,1]$ and with n the number of web pages of the world wide web. Then, for any $m \in (0,1]$ the matrix $\mathcal{A}_m$ is no more block diagonal. In fact, this procedure corresponds to adding artificial links between each web pages on the web, which becomes no more disconnected. If $m = 1$, all pages are rated equally. At a certain time, Google used the value $m = 0.15$. In this setting the following statement can then be proved.

Proposition 8.2.2. *If $n_k \neq 0$ for any k, and if $m \in (0,1]$, then the dimension of the eigenspace associated with the eigenvalue 1 of $\mathrm{L}_{\mathcal{A}_m}$ is of dimension 1.*

The proof of this statement as well as much more information on how Google ranks the web pages of the world wide web are available in the paper of Kurt Bryan and Tanya Leise. Updated and more precise information are also available on internet. Just use Google to find them !

Chapter 9

Complex numbers

9.1 Basic introduction

The aim of this chapter is to provide a very short introduction to complex numbers. One use of complex numbers is to find solutions of the equations $x^2 = -1$, or more generally to find solutions of the equation $ax^2 + bx + c = 0$ for arbitrary $a, b, c \in \mathbb{R}$.

The first step in the construction is based on an analogy with $\mathbb{R}^2$. Note that for simplicity we shall denote the elements of $\mathbb{R}^2$ by (x, y) instead of ${}^t(x, y)$. Let us consider $\mathbb{R}^2$ endowed with the usual addition: $(x, y) + (x', y') = (x + x', y + y')$ for any (x, y) and (x', y') in $\mathbb{R}^2$. We now define a complex multiplication $*$ for these two elements:

$$(x, y) * (x', y') = (xx' - yy', xy' + yx') \quad \in \mathbb{R}^2 \tag{9.1.1}$$

Let us stress that up to now, we had not defined any product of elements of $\mathbb{R}^2$: the scalar product is also taking two elements of $\mathbb{R}^2$ but the result of the scalar product is an element of $\mathbb{R}$, not of $\mathbb{R}^2$!

Since (9.1.1) is rather complicated to remember, let us introduce a symbol i with the only rule that

$$i\,i = i^2 = -1. \tag{9.1.2}$$

We also rewrite (x, y) as $x + iy$. Then, one can again multiply $x + iy$ and $x' + iy'$ by using the common rule of multiplication. One gets

$$\begin{aligned}(x + iy)(x' + iy') &= xx' + (iy)x' + x(iy') + (iy)(iy')\\ &= xx' + i^2yy' + ixy' + iyx'\\ &= (xx' - yy') + i(xy' + yx').\end{aligned} \tag{9.1.3}$$

Note that by comparing (9.1.1) with (9.1.3), one observes that the same result is obtained, but (9.1.3) is certainly easier to remember since only usual multiplications are involved. The key point in the construction is the equality mentioned in (9.1.2). Let us mention that the notation z^2 is also used for $z\,z$ (the product of z by itself), and that with this notation, the usual addition can be rewritten as

$$(x + iy) + (x' + iy') = (x + x') + i(y + y'). \tag{9.1.4}$$

We are now ready for introducing the set of complex numbers:

Definition 9.1.1. *One defines*

$$\mathbb{C} := \{z = x + iy \mid x, y \in \mathbb{R}\}$$

endowed with the addition recalled in (9.1.4) *and with the multiplication introduced in* (9.1.3). *This set is called* the set of complex numbers.

Let us stress that we write indifferently $x + iy$ or $x + yi$.

Example 9.1.2. $1+1i$, $7-2i$, $-3+i$, 3, $2i$ *are elements of* $\mathbb{C}$, *where we have identified* 3 *with* $3+0i$, $2i$ *with* $0+2i$ *and* $-3+i$ *with* $-3+1i$.

By taking into account the identification of x with $x + 0i$, it is clear that $\mathbb{R}$ is included in $\mathbb{C}$. It corresponds to the elements on the horizontal axis in the mentioned analogy of $\mathbb{C}$ with $\mathbb{R}^2$.

Let us still add some examples of multiplications or additions:

Examples 9.1.3. *(i)* $(3 + 2i) + (1 + 1i) = 4 + 3i$,

(ii) $(3 + 2i) + (1 - 3i) = 4 - 1i$,

(iii) $(2 + 2i)(1 + 3i) = 2 + 2i + 6i - 6 = -4 + 8i$,

(iv) $(1 + 2i)(-3 - 2i) = -3 - 6i - 2i + 4 = 1 - 8i$.

Let us now prove an important result about complex numbers. We recall that the notion of a field has been introduced in Definition 3.1.1.

Theorem 9.1.4. $\mathbb{C}$ *is a field.*

Proof. The proof consists in checking the various properties mentioned in Definition 3.1.1. For that purpose, let us set $z = x + iy$ and $z_j = x_j + iy_j$ for $j \in \{1, 2, 3\}$ and with $x, y, x_j, y_j \in \mathbb{R}$. Then one has

(i) $z_1 + z_2 \in \mathbb{C}$ and $z_1 z_2 \in \mathbb{C}$, which means that these operations are internal,

(ii) $(z_1+z_2)+z_3 = z_1+(z_2+z_3)$ and $(z_1 z_2)z_3 = z_1(z_2 z_3)$, as shown in Exercise 9.1. This corresponds to the associativity of the addition and of the complex multiplication

(iii) $z_1 + z_2 = z_2 + z_1$ and $z_1 z_2 = z_2 z_1$, as shown in Exercise 9.2. This means that the addition and the complex multiplication are commutative,

(iv) Let us set $0 \equiv 0 + 0i$ and $1 \equiv 1 + 0i$, which correspond to the usual 0 and 1 of $\mathbb{R}$. Then it is easily observed that $z + 0 = z$ and that $1\,z = z$. This property corresponds to the existence of identity elements for the addition and for the complex multiplication,

(v) Observe that if $z = x + iy \in \mathbb{C}$, then $-x - iy$ also belongs to $\mathbb{C}$ and one has $(x + iy) + (-x - iy) = 0$. Thus $-x - iy$ is the inverse of $x + iy$ for the addition. For the inverse of $x + iy$ with respect to the addition, let us assume that $x + iy \neq 0$, which means that $(x, y) \neq (0, 0)$, and let us consider the complex number $\frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}$ $\mathbb{C}$. This element is well defined since its denominator is different from 0. Then one observes that

$$(x + iy)\Big(\frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}\Big) = \frac{x^2 + y^2}{x^2 + y^2} + i\frac{-xy + xy}{x^2 + y^2} = 1$$

Thus one has $(x + iy)^{-1} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}$, when the inverse with respect to the complex multiplication is considered.

(vi) The distributivity of complex multiplication with respect to the addition of complex numbers is shown in Exercise 9.1.

□

In addition to $\mathbb{R}$ we have thus a second field at our disposal, the field $\mathbb{C}$ of complex numbers. The corresponding complex vector spaces and linear maps on complex vector spaces are briefly studied in Section 7.4. Let us still emphasize one formula which has been derived in the previous proof: for any $z = x + iy \in \mathbb{C}$ with $z \neq 0$ one has

$$(x + iy)^{-1} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}. \tag{9.1.5}$$

Let us now introduce some notations. For any $z = x + iy \in \mathbb{C}$, one sets $\Re(z) := x$ and $\Im(z) := y$ for the *real part* and the *imaginary part* of z. We also introduce the *complex conjugate* $\overline{z}$ of z by

$$\overline{z} = \overline{x + iy} := x - iy.$$

Note that in the mentioned analogy of $\mathbb{C}$ with $\mathbb{R}^2$, it corresponds to taking the image of z by a symmetry along the horizontal axis. Then, with this concept of complex conjugate, it is easily observed that

$$\Re(z) = \frac{z + \overline{z}}{2} \qquad \text{and} \qquad \Im(z) = \frac{z - \overline{z}}{2}.$$

For any complex number $z = x + iy$ we also define $|z| := \sqrt{x^2 + y^2}$ and set

$$z = r\big(\cos(\theta) + i\sin(\theta)\big)$$

with $r = |z|$, $x = r\cos(\theta)$ and $y = r\sin(\theta)$. This is called *the polar coordinate representation* of the complex number z. The number $r \equiv |z|$ is called the *norm* or the *modulus* of z, and θ its *argument*, *i.e.* $\theta = \arg(z)$. We also introduce the notation

$$e^z = e^{x+iy} := e^x\big(\cos(y) + i\sin(y)\big).$$

These notations will be used in the Exercises, and they are very useful tools for complex numbers.

Remark 9.1.5. *Let us emphasize that $\mathbb{C}$ has no ordering. Indeed, even if $\mathbb{R}$ has a ordering (one says for example that $-2 < 4$), it is impossible to compare two complex numbers as for example $3 - 2i$ and $4 + i$.*

Let us now provide one of the basic result for complex numbers, which is part of the motivation for introducing them.

Proposition 9.1.6. *For any $a + ib \in \mathbb{C}$, there exists $z_1, z_2 \in \mathbb{C}$ with $z_1 \neq z_2$ (except if $a + ib = 0$) such that $z_1^2 = z_2^2 = a + ib$. In other words, every complex number has two distinct square roots.*

Proof. Let us first observe that for any $a, b \in \mathbb{R}$, one has

$$a + \sqrt{a^2 + b^2} \geq 0 \qquad \text{and} \qquad -a + \sqrt{a^2 + b^2} \geq 0.$$

Thus, one can define $x := \sqrt{\frac{a+\sqrt{a^2+b^2}}{2}}$ with the usual square root of positive numbers, and also $y := \sqrt{\frac{-a+\sqrt{a^2+b^2}}{2}}$ with the usual square root. We then set

$$z_1 := x + i\mu y \qquad \text{and} \qquad z_2 := -x - i\mu y$$

with $\mu = 1$ if $b \geq 0$ and $\mu = -1$ if $b < 0$. It only remains to check with the definition of the complex multiplication that $z_1^2 = a + ib$ and that $z_2^2 = a + ib$ as well. □

By using the well-known formula for the solutions of a second degree equation, one infers that:

Corollary 9.1.7. *The equation $az^2 + bz + c = 0$ has always two solutions in $\mathbb{C}$.*

Let us finally mention that this corollary is at the root of the fundamental theorem of algebra asserting that any polynomial of degree n has n solutions in $\mathbb{C}$.

9.2 Exercises

Exercise 9.1. *Let z_1, z_2, z_3 be three complex numbers. Show that*

$$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$$

and that

$$(z_1 z_2) z_3 = z_1 (z_2 z_3).$$

These properties correspond to the associativity *of the addition and of the complex multiplication. In addition, check that $z_1(z_2+z_3) = z_1 z_2 + z_1 z_3$. This property corresponds to* the distributivity *of the complex multiplication with respect to the addition of complex numbers.*

Exercise 9.2. *For $z_1, z_2 \in \mathbb{C}$, show that $z_1 + z_2 = z_2 + z_1$ and that $z_1\, z_2 = z_2\, z_1$. These properties correspond to the commutativity of the addition and of the complex multiplication.*

Exercise 9.3. *Compute the real part and the imaginary part of the number $\frac{3+2i}{2-3i}$. Same question with the number $\frac{1}{i} + \frac{3}{1+i}$ and the number $\sqrt{1+i}$.*

Exercise 9.4. *Find all solutions of the equation $z^4 = -1$.*

Exercise 9.5. *For any $z_1, z_2 \in \mathcal{C}$, show that $|z_1 z_2| = |z_1||z_2|$ and that*

$$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2).$$

Exercise 9.6. *Deduce from the previous exercise* de Moivre's formula*: for any $n \in \mathbb{N}$ and for $z = r\big(\cos(\theta) + i \sin(\theta)\big)$ one has*

$$z^n = r^n \big(\cos(n\theta) + i \sin(n\theta)\big).$$

Exercise 9.7. *Deduce that for any complex number $z = r\big(\cos(\theta) + i \sin(\theta)\big)$, the n-th roots of z are given by*

$$z_j := \sqrt[n]{r} \left[\cos\left(\frac{\theta + 2\pi j}{n}\right) + i \sin\left(\frac{\theta + 2\pi j}{n}\right)\right]$$

for $j \in \{0, 1, \dots, n-1\}$.

Exercise 9.8. *Show the following properties:*

1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$,
2. $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$,
3. $z\,\overline{z} = |z|^2$,
4. $z^{-1} = \overline{z}/|z|^2$ *whenever* $z \neq 0$,

5. *$\Re(z) = (z + \overline{z})/2$ and $\Im(z) = (z - \overline{z})/(2i)$, where $\Re(z)$ and $\Im(z)$ are the real and the imaginary part of z.*

Exercise 9.9. *Show also that $|\overline{z}| = |z|$ and that $\arg(\overline{z}) = -\arg(z)$.*

Exercise 9.10. *Show the following properties:*

1. *$e^{z_1+z_2} = e^{z_1} e^{z_2}$ for any $z_1, z_2 \in \mathcal{C}$,*
2. *e^z is never equal to 0,*
3. *$|e^{x+iy}| = e^x$,*
4. *$e^{i\pi} = -1$ (Euler's identity, and "one of the most beautiful formula in mathematics").*