Lecture notes for the course

Linear Algebra with Application

LAWA - 2020

CZECH TECHNICAL UNIVERSITY IN PRAGUE

Francesco Dolce

`francesco.dolce@fjfi.cvut.cz`

May 13, 2020

These notes are available at the address `http://irif.fr/~dolce/ens/1920/lawa2020.pdf`.

Contents

Chapter 1

Basic notions

Let us start by fixing some notation and giving some basic example.

1.1 Numerical sets

We denote by $\mathbb{N}$ the set of *natural numbers*, that is

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

and with $\mathbb{Z}$ the set of *integers*

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \ldots\}.$$

Two important subsets of $\mathbb{Z}$ are the sets $\mathbb{Z}^+$ of *positive integers* and $\mathbb{Z}^-$ of *negative integers*:

$$\mathbb{Z}^+ = \{1, 2, 3, \ldots\}, \quad \text{and} \quad \mathbb{Z}^- = \{-1, -2, -3, \ldots\}.$$

We can also use the notations Z_0^+ and $\mathbb{Z}_0^-$ to denote respectively the sets of *non-negative integers* and the one of *non-positive integers*, that is:

$$\mathbb{Z}_0^+ = \{0, 1, 2, 3, \ldots\} \quad \text{and} \quad \mathbb{Z}_0^- = \{0, -1, -2, -3, \ldots\}.$$

Note that $Z_0^+ = \mathbb{Z}^+ \cup \{0\} = \mathbb{N}$.

Other important numerical sets are the set of *rational numbers* $\mathbb{Q}$ defined us

$$\mathbb{Q} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z},\ n \neq 0 \right\} = \left\{ 0, \frac{1}{2}, -\frac{3}{5}, \ldots \right\}$$

and the one of *real numbers* $\mathbb{R}$ (a formal definition is beyond the scope of this course)

$$\mathbb{R} = \left\{ 0, 1, -5, \frac{7}{4}, \sqrt{2}, \pi, e, \ldots \right\}.$$

The last numerical set we will consider is the set of *complex numbers* $\mathbb{C}$ defined us

$$\mathbb{C} = \{a + ib \mid a, b \in R\},$$

where i is called the *imaginary unit* and satisfies $i^2 = -1$.

We have the following chain of inclusions:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

1.2 Algebraic structures: one operation

A *binary operation* on a set M is a map f from the cartesian product $M \times M$ to M, that is

$$\begin{array}{rl} f: & M \times S \to M \\ & (x, y) \mapsto f(x, y). \end{array}$$

Example 1.1 The *sum* on $\mathbb{N}$ is the binary operation defined as

$$\begin{array}{rl} +: & \mathbb{N} \times \mathbb{N} \to \mathbb{N} \\ & (a, b) \mapsto +(a, b) \end{array}.$$

Instead of $+(a, b)$, we usually denote the image of the sum by $a + b$. For istance, $2 + 3 = 5$.

Some sets, when equipped with a binary operation (or more than one), have particular properties. According to which properties are satisfied, we use different names. In this section we consider the main algebraic structures.

1.2.1 Groupoids

Let us consider a set M and a binary operation $\circ : M \times M \to M$.

The pair $(M, \circ)$ is called a *groupoid* whenever the set M is *closed under* the operation $\circ$. That is, we have

$$\forall\, a, b \in M \quad a \circ b \in M.$$

Example 1.2 The pair $(\mathbb{N}, +)$ is a groupoid, since for every two natural numbers a, b one has $a + b \in \mathbb{N}$.

On the other hand, if we consider the set $M = \{0, 1, 2, \ldots, 9\}$, we have that M is not closed under the sum, since, for instance, $2 + 9 \notin M$.

1.2.2 Semigroups

A groupoid $(M, \circ)$ is called a *semigroup* if the operation $\circ$ is *associative*, that is if

$$\forall\, a, b, c \in M \quad (a \circ b) \circ c = a \circ (b \circ c).$$

Example 1.3 The set $\mathbb{N}$ with the usual *multiplication* is a semigroup, since for every $a, b, c \in \mathbb{N}$ one has $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. The same is not true if we consider the *exponentiation* as binary operation. Indeed

$$(2^3)^2 = 64 \neq 512 = 2^{(3^2)}.$$

1.2.3 Monoids

A semigroup $(M, \circ)$ is called a *monoid* whenever there exists a *neutral element*, that is an element $e \in M$ such that

$$\forall\, a \in M \quad a \circ e = e \circ a = e.$$

Note that in the definition we want e to be a neutral element both on the left and on the right. In some tricky case we could have only a *left neutral element* but no *right neutral element*, or viceversa, or we could have both but distinct.

Example 1.4 The set $\mathbb{N}$ provided with the sum is a monoid. The neutral element is the element $0 \in \mathbb{N}$.

Proposition 1.5 *The neutral element of a monoid is unique.*

Proof. Let $(M, \circ)$ be a monoid and e some neutral element (from the definition we know that at least one exists!). We prove by contradiction that e is the only neutral element. By contradiction, assume that in the monoid there exists another neutral element e' different from e. It holds that

$$e' = e' \circ e = e,$$

using the property of the neutral element from the definition. We get a contradiction with the statement that $e' \neq e$. ■

1.2.4 Groups

Let us consider a monoid $(M, \circ)$ with neutral element e. We say that an element $a \in M$ is *invertible*, whenever there exists another element $b \in M$ such that $a \circ b = e = b \circ a$. Such an element is called an *inverse* of a and it is usually denoted $-a$ (whenever we use the additive notation) or a^{-1} (whenever we use the multiplicative notation). Note that, again, we consider an inverse as both a *left inverse* and a *right inverse.*

A monoid $(M, \circ)$, with neutral element e, is called a *group* if every element is invertible, that is if

$$\forall\, a \in M\ \exists\, a^{-1} \in M \text{ such that } \quad a \circ a^{-1} = e \text{ and } a^{-1} \circ a = e.$$

Example 1.6 The monoid $(\mathbb{Z}, +)$ is a group, since the inverse of any element $a \in \mathbb{Z}$ is $-a$. On the other hand $(\mathbb{N}, +)$ is not a group since, for istance, the element $2 \in \mathbb{N}$ has no inverse.

Proposition 1.7 *Each element of a group has exactly one inverse.*

Proof. Let $(M, \circ)$ be a group, a an arbitrary element of the group and a^{-1} one of its inverse elements (from the definition we know that there exists at least one!) Let us suppose, by contradiction, that there exists another element $\bar{a}$, different from a^{-1} such that $\bar{a} \circ a = e = a \circ \bar{a}$, where e is the neutral element of the group. Hence, it holds that

$$\bar{a} = \bar{a} \circ e = \bar{a} \circ (a \circ a^{-1}) = (\bar{a} \circ a) \circ a^{-1} = e \circ a^{-1} = a^{-1},$$

contradicting the fact that $\bar{a} \neq a^{-1}$. ■

A group $(M, \circ)$ is called *Abelian*, or *commutative*, if its elements commute, that is if

$$\forall\, a, b \in M \quad a \circ b = b \circ a.$$

Example 1.8 Both the additive group $(\mathbb{Q}, +)$ and the multiplicative group $(\mathbb{Q}, \cdot)$ are Abelian. The first has neutral element 0, while the second has neutral element 1.

A very important example of non Abelian group will be given later in this chapter.

1.3 Algebraic structures: two operations

Let us consider now triplets $(M, +\cdot)$, with M a nonempty set and $+, \cdot$ two binary operations on M.

1.3.1 Rings

We say that a triplet $R = (M, +, \cdot)$ is a *ring* if the following hold:

- $(M, +)$ is an Abelian group;
- $(M, \cdot)$ is a monoid;
- both left and right *distributive laws* hold, i.e. $\forall\, a, b, c \in M$ we have
  - $a \cdot (b + c) = a \cdot b + a \cdot c$ and
  - $(b + c) \cdot a = b \cdot a + c \cdot a$.

We respect the standard convention that multiplication has a higher priority than addition, so we can write $a \cdot b + a \cdot c$ instead of $(a \cdot b) + (a \cdot c)$. standard

Moreover, when it is clear from the context we replace $\cdot$ by the simple juxstaposition, that is we write ab instead of $a \cdot b$.

Example 1.9 Both the triplets $(\mathbb{Z}+,\cdot)$ and $(\mathbb{Q},+,\cdot)$ are rings. On the other hand the triplet $(\mathbb{N},+,\cdot)$ is not a ring since $(\mathbb{N},+)$ is not a group.

The triplet $(\{0\},+,\cdot)$, with $0+0=0$ and $0\cdot 0=0$ is a ring called the *trivial ring.*

We say that a ring $R=(M,+,\cdot)$ is a *commutative ring* whenever $\cdot$ is commutative. The group $(M,+)$ is called the *additive group* of R, while the monoid $(M,\cdot)$ is the *multiplicative monoid* of R.

The neutral element of the additive group is called the *zero element* of the ring, and it is denoted by 0, and the inverse element of $a\in M$ is denoted by $-a$. We can also define the *subtraction* of two elements $a,b\in M$ by

$$a-b:=a+(-b).$$

Proposition 1.10 *Let* $(M,+,\cdot)$ *be a ring. Left and right distributive laws hold for the subtraction, that is:*

$$\forall\, a,b,c\in M \quad a(b-c)=ab-ac \quad \textit{and} \quad (a-b)c=ac-bc.$$

Proof. Let us prove the left distributive law, the right one being proved symmetrically. Since the distributive law hold for the sum, we have

$$ac+a(b-c)=a(c+b-c)=ab.$$

Thus, by subtracting ac to both members, we have

$$a(b-c)=ab-ac.$$

■

Example 1.11 The set of polynomials with real coefficients $\mathbb{R}[x]$ is a ring. The zero element is the zero polynomial $p(x)=0$. We will talk more of this example in a following chapter.

1.3.2 Integral domains

Let $(M,+,\cdot)$ be a ring. Two non-zero elements $a,b\in M$ are called *zero divisors* if $a\cdot b=0$.

A commutative ring without zero divisors is called an *integral domain*

Example 1.12 The ring $(\mathbb{Z},+,\cdot)$ is an integral domain. On the other hand the ring $(\mathbb{Z}_6,+_6,\cdot_6)$, where the sum and the product are defined *modulo* 6, is not an integral domain, since $2,3\neq 0$ but $2\cdot_6 3=0$.

1.3.3 Fields

A ring $(M, +, \cdot)$ is a *field* if $(M \setminus \{0\}, \cdot)$ is an Abelian group. This group is called the *multiplicative group* of the field.

Example 1.13 The ring of integers $(\mathbb{Z}, +, \cdot)$ is not a field, since $(\mathbb{Z} \setminus \{0\}, \cdot)$ misses some inverse elements.

On the other hand, $(\mathbb{Q}, +, \cdot)$ is a field. Moreover, this is the smallest number field with the common arithmetical operations.

Example 1.14 The smallest field is the so-called *trivial field* $(\{0, 1\}, +, \cdot)$, with operations

$$0 + 0 = 0, \quad 0 + 1 = 1 + 0 = 1, \quad 1 + 1 = 0,$$

and

$$0 \cdot 0 = 0, \quad 0 \cdot 1 = 1 \cdot 0 = 0, \quad 1 \cdot 1 = 1.$$

Proposition 1.15 *Each field is an integral domain.*

Proof. Since the multiplicative group $(M \setminus \{0\}, \cdot)$ is closed under multiplication, for all non-zero elements $a, b \in M$ it holds that their product $a \cdot b \in M \setminus \{0\}$ is again non-zero. ■

Chapter 2

Matrices: Definitions

In this chapter we will define and study matrices over $\mathbb{R}$. We will see, in the next chapters, that it is possible to consider matrices using other classes of numbers.

2.1 Definition of matrix

A *matrix* A over $\mathbb{R}$ is a rectangular array of real numbers of the form

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

with $m, n \in \mathbb{N}$ and $a_{ij} \in \mathbb{R}$ for all $1 \leq i \leq m$ and $1 \leq j \leq m$.

The (i, j)-*entry* of A is the number a_{ij}, while its i-*row* and its j-*column* are respectively

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

A matrix having m rows and n columns is called a $m \times n$-matrix. When $m = n$, the matrix is called a *square matrix.*

Example 2.1 The matrix

$$A = \begin{pmatrix} 1 & -2 & 0 & \pi \\ 0 & 2 & -3 & 0 \\ \sqrt{5} & -1 & 0 & 7 \\ 2 & 6 & 7 & 9 \end{pmatrix}$$

is a square matrix of size 4×4, that is it has 4 rows and 4 columns.

The matrix

$$B = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 \end{pmatrix}$$

is a 2×4-matrix.

The $(3,2)$-entry of A is -1 and the $(1,4)$-entry of B is 2. The 2-row of A is $\begin{pmatrix} 0 & 2 & -3 & 0 \end{pmatrix}$, while the 1-column of B is $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$.

We say that two matrices A and B are *equal*, and we write $A = B$, if and only if A and B have the same size and the corresponding entries are equal.

The set of matrices over $\mathbb{R}$ of size $m \times n$ is denoted by $\mathcal{M}_{m,n}(\mathbb{R})$, that is

$$\mathcal{M}_{m,n}(\mathbb{R}) = \left\{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \mid a_{ij} \in \mathbb{R} \text{ for all } 1 \leq i \leq m,\ 1 \leq j \leq n \right\}.$$

Sometimes, when it is clear from the context, we can also write the matrix A as (a_{ij}).

2.2 Matrix addition

Given two matrices $A = (a_{ij})$ and $B = (b_{ij})$ of the same size we can add them just by adding the corresponding entries, that is the *sum* of A and B is the matrix

$$A + B = (a_{ij} + b_{ij}).$$

. Similarly, we can define the *difference*

$$A - B = (a_{ij} - b_{ij}).$$

The *negative* of A is the matrix $-A$ obtained as

$$-A = (-a_{ij}).$$

Example 2.2 Let un consider the tree matrices

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

We have

$$A+B = \begin{pmatrix} 1 & 1 & 0 \\ 3 & 5 & 3 \end{pmatrix}, \quad A-B = \begin{pmatrix} 1 & -1 & -4 \\ 7 & -1 & -3 \end{pmatrix} \quad \text{and} \quad -A = \begin{pmatrix} -1 & 0 & 2 \\ -5 & -2 & 0 \end{pmatrix}.$$

We can not define $A + C$ or $A - C$ since A and C have different size.

Let us denote by $O_{m,n}$ the $m \times n$-matrix having each entry equal to 0. Such a matrix is called the *zero matrix* of size $m \times n$. When $m = n$ we will simply write O_n instead of $O_{n,n}$. When the size is clear from the context, it is simply denoted by O.

Example 2.3 The zero matrices of order 2×3 and 2×2 are respectively

$$O_{2,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathcal{M}_{2,3}(\mathbb{R}) \quad \text{and} \quad O_{2,2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R}).$$

Proposition 2.4 *The set* $\mathcal{M}_{m,n}(\mathbb{R})$ *with the sum operation is an Abelian group.*

Proof. (`Exercise`) ■

2.3 Matrix multiplication

Let $n \in \mathbb{N}$. Let us consider a n-row-matrix, that is a matrix

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \in \mathcal{M}_{1,n}(\mathbb{R})$$

and an n-column-matrix, that is a matrix

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R}).$$

We define the *dot product* of A and B as the number

$$\sum_{k=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots a_n b_n \in \mathbb{R}.$$

The *product* of a $m \times n$-matrix A and an $n \times p$-matrix B is the $m \times p$-matrix AB having as (i, j)-entry the dot product of the i-row of A and the j-column of B. So, if the two matrices are

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{np} & \end{pmatrix}$$

then the (i, j)-entry of AB is

$$\sum_{k=1}^{n} a_{ik} bkj = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}$$

and the matrix AB has the form

$$AB = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}b_{k1} & \sum_{k=1}^{n} a_{1k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kp} \\ \sum_{k=1}^{n} a_{2k}b_{k1} & \sum_{k=1}^{n} a_{2k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{2k}b_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{mk}b_{k1} & \sum_{k=1}^{n} a_{mk}b_{k2} & \cdots & \sum_{k=1}^{n} a_{mk}b_{kp} \end{pmatrix}.$$

Note that the product AB is defined if and only if the number of columns of A is equal as the number of columns of B.

Example 2.5 Let $A = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$, $B = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$ and $C = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 2 & -1 \end{pmatrix}$. Then we have

$$AB = \begin{pmatrix} 1 & -5 \\ 1 & -1 \end{pmatrix}, \quad BA = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}, \quad BC = \begin{pmatrix} -1 & -1 & 4 \\ 2 & 4 & -2 \end{pmatrix}, \quad AC = \begin{pmatrix} 2 & 5 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

We can not define CA because of of the size.

Proposition 2.6 *The matrix multiplication is associative, i.e., that for all matrices A, B, C having the right sizes, we have $(AB)C = A(BC)$.*

Proof. (`Exercise`) ■

Proposition 2.7 *The distributive laws hold both for the sum and for the subtraction, i.e., for all matrices A, B, C having the right sizes, we have*

- $A(B + C) = AB + AC$;
- $A(B - C) = AB - AC$;
- $(A + B)C = AC + BC$;
- $(A - B)C = AC - BC$.

Proof. (`Exercise`) ■

The $n \times n$ *identity matrix* is the matrix $I_n \in \mathcal{M}_{n,n}(\mathbb{R})$ with 1s on the *main diagonal*, i.e., the entries of the form (i, i), and zero elsewhere. When the size is clear from the context, it is simply denoted by I.

Example 2.8 One has

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Proposition 2.9 *Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$. Then*

$$I_m A = A \quad \text{and} \quad A I_n = A.$$

Proof. (`Exercise`) ∎

Corollary 2.10 *The set of $m \times n$-matrices over $\mathbb{R}$, i.e., $\mathcal{M}_{m,n}(\mathbb{R})$ with the matrix addition and the matrix multiplication is a ring.*

In the context, when it is clear from the context, we will use the same notation for the set and the ring of matrices, i.e., we will write $\mathcal{M}_{m,n}(\mathbb{R})$ instead of $(\mathcal{M}_{m,n}(\mathbb{R}), +, \cdot)$.

Example 2.11 Let A, B as in Example 2.5. We have

$$AB = \begin{pmatrix} 1 & -5 \\ 1 & -1 \end{pmatrix} \neq \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} = BA$$

The previous example shows that commutativity does not hold in general for matrices.

Corollary 2.12 *The ring $\mathcal{M}_{m,n}(\mathbb{R})$ is not a commutative ring.*

Example 2.13 Let us consider the matrix $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. One has

$$A^2 = A \cdot A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The last example shows that the ring $\mathcal{M}_{m,n}(\mathbb{R})$ has zero divisors.

Corollary 2.14 *The ring $\mathcal{M}_{m,n}(\mathbb{R})$ is not an integral domain (nor a field).*

2.4 Scalar multiplication

Let $A = (a_{ij}) \in \mathcal{M}_{m,n}(\mathbb{R})$ be a matrix and $\lambda \in \mathbb{R}$ a real number. The *scalar product* λA it defined as the matrix of the form (λa_{ij}), that is

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

Example 2.15 Let $A = \begin{pmatrix} 1 & 0 \\ -1 & 2 \\ 3 & -3 \end{pmatrix}$. Then

$$2A = \begin{pmatrix} 2 & 0 \\ -2 & 4 \\ 6 & -6 \end{pmatrix}.$$

Proposition 2.16 *Let A, B be two matrices of the same size and let λ, μ be two real numbers. Then*

1. $\lambda(A + B) = \lambda A + \lambda B$*;*
2. $(\lambda + \mu)A = \lambda A + \mu A$*;*
3. $\lambda(\mu A) = (\lambda\mu)A$*;*
4. $1A = A$*;*
5. $\lambda(AB) = (\lambda A)B = A(\lambda B)$*.*

Proof. (`Exercise`) ■

2.5 Transposition

Let $A = (a_{ij}) \in \mathcal{M}_{m,n}(\mathbb{R})$ be a matrix. The *transpose* of A is the matrix $A^T \in \mathcal{M}_{n,m}(\mathbb{R})$ defined as

$$A^T = (b_{ij}) \quad \text{with} \quad b_{ij} = a_{ji} \text{ for all } i, j.$$

A matrix A is called symmetric if $A^T = A$.

Example 2.17 Let

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

Then we have

$$A^T = \begin{pmatrix} 3 & 1 \\ 2 & 0 \end{pmatrix}, \quad B^T = \begin{pmatrix} 1 & 5 \\ 0 & -1 \\ -2 & 0 \end{pmatrix} \quad \text{and} \quad C^T = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

Moreover, the matrix C is symmetric.

Proposition 2.18 *Let $A, B \in \mathcal{M}_{m,n}(\mathbb{R})$ and let $\lambda \in \mathbb{R}$. Then*

1. *If A is symmetric then $m = n$;*
2. $(A^T)^T = A$*;*

3. $(\lambda A)^T = \lambda A^T$*;*

4. $(A+B)^T = A^T + B^T$*;*

5. $(AB)^T = B^T A^T$*.*

Proof. (`Exercise`) ∎

2.6 Matrix inverse

Let us consider a square matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$. A matrix B is called an *inverse* of A if $AB = I$ and $BA = I$. Note that if such a matrix B exists, then it has the same size as A.

A square matrix having an inverse is called an *invertible matrix.*

Example 2.19 Let us consider the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$. The matrix A is invertible and the matrix $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$ one of its inverses. Indeed,

$$AB = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

and

$$BA = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Similarly to what we have done in Proposition 1.7 we can prove that the inverse of matrix, when it exists, is unique.

Proposition 2.20 *Let* $A \in \mathcal{M}_{n,n}(\mathbb{R})$ *an invertible matrix. Then its inverse is unique.*

Proof. Let $B, C \in \mathcal{M}_{n,n}(\mathbb{R})$ and let us suppose that both matrices are inverses of A. Then

$$B = BI = B(AC) = (BA)C = IC = C.$$

∎

Whenever a matrix A is invertible we denote by A^{-1} its unique inverse.

Note that not all matrices have inverses.

Example 2.21 Let us conside the matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$. This matrix is not invertible. Indeed, if we suppose by contradiction that there exists a matrix $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $AB = I$. Then, one has

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ a & b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

which implies both that $a = 1$ and $a = 0$, a contradiction.

Also, non-square matrices do not have inverses and the zero matrix O_n is not invertible neither (`Exercise`).

Let us now give some properties about inverses.

Theorem 2.22 *The following properties hold for square matrices.*

1. *The identity matrix I is invertible and its inverse is I itself.*
2. *If A is invertible, then A^{-1} is invertible as well and $\left(A^{-1}\right)^{-1} = A$.*
3. *If A and B are invertible, then so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.*
4. *Let $A_1, A_2, \dots A_k$ be invertible matrices, then their product $A_1A_2 \cdots A_k$ is also invertible, and $(A_1A_2 \cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_1^{-1}$*
5. *If A is invertible, then for all $k \geq 1$ the matrix A^k is invertible as well and $(A^k)^{-1} = (A^{-1})^k$.*
6. *If A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.*
7. *If A^T is invertible, then so is A and $A^{-1} = ((A^T)^{-1})^T$.*
8. *If A is invertible and $\lambda \neq 0$ is a real number, then λA is also invertible and $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$.*

Proof.

1. This easily follows from the fact that $II = I$.
2. The second item also follows from the fact that $A^{-1}A = I$ and $AA^{-1} = I$.
3. It holds because
$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$
and
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$
4. Let us prove it by induction on k. The case $k = 1$ is trivial and the case $k = 2$ follows from point 3. So, let us suppose that the property holds for $k - 1$, that is that $A_1A_2 \cdots A_{k-1}$ is invertible and that its inverse is $A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}$. Then
$$(A_1A_2 \cdots A_{k-1}A_k) = (A_1A_2 \cdots A_{k-1})A_k$$
is a product of two invertible matrices and, by the previous point, is invertible itself. Moreover, its inverse is exactly
$$A_k^{-1}(A_1A_2 \cdots A_{k-1})^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}.$$

5. This easily follows from the previous item by chosing $A_i = A$ for all $1 \leq i \leq k$.

6. Using the last point of Proposition 2.18 we have

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$$

and

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I.$$

7. Using both the fact that $(A^T)^{-1} = (A^{-1})^T$, proved in the previous point, and that $(A^T)^T = A$, proved in Proposition 2.18, we have

$$A((A^T)^{-1})^T = A((A^{-1})^T)^T = AA^{-1} = I$$

and

$$((A^T)^1)^T A = ((A^{-1})^T)^T A = A^{-1}A = I.$$

8. To prove the last point we use Proposition 2.16 and show that

$$(\lambda A)\left(\frac{1}{\lambda}A^{-1}\right) = \left(\lambda\frac{1}{\lambda}\right)\left(AA^{-1}\right) = 1I = I$$

and

$$\left(\frac{1}{\lambda}A^{-1}\right)(\lambda A) = \left(\frac{1}{\lambda}\lambda\right)\left(A^{-1}A\right) = 1I = I.$$

■

Corollary 2.23 *Let $A, B \in \mathcal{M}_{n,n}(\mathbb{R})$. If A and AB are both invertible, then B is also invertible.*

Proof. (`Exercise`) ■

2.7 Diagonal and triangular matrices

A square matrix $A = (a_{ij}) \in \mathcal{M}_{n,n}(\mathbb{R})$ is called *diagonal* if every entry not in the main diagonal is 0, that is if for every i, j with $1 \leq i, j \leq n$ and $i \neq j$ one has $(a_{ij}) = 0$.

Example 2.24 The matrices

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -5 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

are diagonal.

When a matrix $A = (a_{ij})$ is diagonal, we can also denote it simply as $A = \operatorname{diag}(a_{11}, a_{22}, \dots, a_{nn})$.

Example 2.25 Let us consider again the three matrices of Example 2.24. We have

$$A = \operatorname{diag}(1, 2, -5), \quad B = \operatorname{diag}(-1, 4, 0, 1) \quad \text{and} \quad C = \operatorname{diag}(0, 0).$$

Proposition 2.26 *Let $A, B \in \mathcal{M}_{n,n}(\mathbb{R})$ be two diagonal matrices. Then*

1. *$A + B$ is diagonal;*
2. *AB is diagonal.*

Proof. `(Exercise)` ■

A square matrix $A = (a_{ij})$ is called *upper triangular* if every entry below the main diagonal is zero, that if for every i, j with $i > j$ one has $a_{ij} = 0$. Un upper triangular matrix is called *strictly upper triangular* if the entries on the main diagonal are zero as well.

In a symmetric way we define *lower triangular* and *strictly lower triangular* matrices.

Example 2.27 Let us consider the four matrices

$$A = \begin{pmatrix} 6 & 9 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix} \text{ and } D = \begin{pmatrix} 0 & 0 & 0 \\ 5 & 0 & 0 \\ 7 & 6 & 0 \end{pmatrix}.$$

The matrix A is upper triangular, the matrix B is strictly upper triangular. Simarly C and D are respectively lower triangular and strictly lower triangular.

Proposition 2.28 *Let $A, B \in \mathcal{M}_{n,n}(\mathbb{R})$ be two upper triangular matrices. Then*

1. *$A + B$ is upper triangular.*
2. *AB is upper triangular.*

Proof. `(Exercise)` ■

A similar result also holds by replacing the condition "upper triangular" with "strictly upper triangular", "lower triangular" or "strictly lower triangular".

Chapter 3

Linear equations

In this chapter we will study linear equations and systems of linear equations. We will also discuss how to use matrices to represent and solve such equations.

3.1 Variables, coefficients and solutions

We call a *linear equation* in the *variables* $x_1, x_2, \ldots, x_n$, with $n \in \mathbb{N}$, an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \tag{3.1}$$

where $a_1, a_2, \ldots, a_n, b \in \mathbb{R}$. The numbers $a_1, a_2, \ldots, a_n$ are called *coefficients* of the variables $x_1, x_2, \ldots, x_n$, while the number b is called the *constant term* of the equation.

We can represent the variables using the column matrix

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

called the *matrix of variables* of the equation.

A row-matrix

$$\begin{pmatrix} s_1 & s_2 & \ldots & s_n \end{pmatrix}$$

is called a *solution* of the linear Equation (3.1) if

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n = b$$

that is, if replacing x_i with s_i for every $1 \leq i \leq n$ on the right side, we obtain exactly b.

Example 3.1 Let us consider the linear equation

$$2x_1 + x_2 - x_3 = 3$$

The coefficients of the variables x_1, x_2, x_3 are respectively $2, 1$ and -1, while the constant term of the equation is 3. The matrix of variables is

$$X = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

and a solution of the linear equation is

$$\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$$

since

$$2 \cdot 1 + 1 \cdot 1 - 1 \cdot 0 = 3.$$

Another possible solution for the equation is $\begin{pmatrix} 1 & 0 & -1 \end{pmatrix}$.

From the previous example, we see that the solution of a linear equation is, in general, not unique. We call one possible solution a *particular solution* of the equation. A way to find all the solutions of the equation is fixing two variables and compute the third one with respect to the previous.

Example 3.2 Let us consider the equation of Example 3.1. By setting $x_1 = s$ and $x_2 = t$, we find that $2s + t - x_3 = 3$, which implies that $x_3 = s2 + t - 3$. So, all solutions have the form

$$X = \begin{pmatrix} s & t & 2s + t - 3 \end{pmatrix}$$

for certain $s, t \in \mathbb{R}$.

Following the terminology of the previous example, we call X the *general solution* of the equation and s, t the *parameters* of the solution.

3.2 Systems of linear equations

A *system of linear equations* is a finite collection of linear equations. Its general form is

$$\left\{ \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \right. \tag{3.2}$$

A solution satisfying every equation of a system is called a *solution of the system.*

Example 3.3 The system of two linear equations in the variables x, y

$$\left\{ \begin{array}{rcl} x + y & = & 10 \\ 2x - y & = & 5 \end{array} \right. \tag{3.3}$$

has an unique solution $X = \begin{pmatrix} 5 & 5 \end{pmatrix}$.

Note that some system may have no solution. In this case we say that the system is *inconsistent*.

Example 3.4 The system

$$\begin{cases} x + y & = & 1 \\ 2x + 2y & = & 3 \end{cases} \tag{3.4}$$

has no solution (`Exercise`), so it is inconsistent.

When a system has (at least) one solution, we call it *consistent*.

Example 3.5 The system

$$\begin{cases} x + y + z & = & 2 \\ x - y + z & = & 0 \end{cases} \tag{3.5}$$

has infintely many solution (`Exercise`). Thus it is consistent.

Let us now show how we can represent the system in Equation 3.2 using matrices. The *coefficient matrix* and the *constant matrix* for this system are respectively the $m \times n$-matrix and the $m \times 1$-matrix defined as

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

We can also combine the two to obtain the *augmented matrix* defined as the $m \times (n + 1)$-matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

Example 3.6 Let us consider the system 3.3. Its matrix of variable is $\begin{pmatrix} x \\ y \end{pmatrix}$ while its constant matrix and its coefficient matrix are respectively

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 10 \\ 5 \end{pmatrix}.$$

Its augmented matrix is the (2×3)-matrix

$$\begin{pmatrix} 1 & 1 & 10 \\ 2 & -1 & 5 \end{pmatrix}.$$

Note that, sing the coefficient matrix, the variable matrix and the constant matrix, we can represent the system of linear equations 3.2 as a single matrix equation

$$AX = B.$$

Example 3.7 Using Example 3.6 we can represent the system 3.3 as the matrix equation

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ 5 \end{pmatrix}.$$

Example 3.8 Les us consider the system of three linear equations in four variables

$$\left\{ \begin{array}{rcl} x_1 - x_4 + x_3 & = & 5 \\ x_3 - 3x_4 & = & -1 \\ x_4 & = & 2 \end{array} \right. . \tag{3.6}$$

The previous system is in a very special form and it can be solved by using *back-substitution*. From the last equation we get $x_4 = 2$. Then we substitute 2 for the variable x_4 into the second last equation to solve for $x_3 = 5$. Finally, we substitute $x_3 = 5$ and $x_4 = 2$ and we replace x_2 with a parameter s. The *general solution* of the original system has thus the form

$$X = \begin{pmatrix} 4s & s & 5 & 2 \end{pmatrix}$$

where s is the parameter of the solution. That means that every solution can be obtained by replacing s with a real number.

The system of linear equations presented in Example 3.8 was in a very special form. In the next sections we'll see how to use this technique to solve a system in a more general form.

3.3 Equivalent systems

Two systems of linear equations having the same set of solutions are called *equivalent*.

Example 3.9 Let us consider the system in Example 3.5 and let us swap the two equations:

$$\left\{ \begin{array}{rcl} x - y + z & = & 0 \\ x + y + z & = & 2 \end{array} \right. . \tag{3.7}$$

It is clear that this new system is equivalent ot the system 3.5.

Example 3.10 Starting again from the system in Example 3.5, let us multiply the left and the right term of the second equation by 2:

$$\left\{ \begin{array}{rcl} x + y + z & = & 2 \\ 2x - 2y + 2z & = & 0 \end{array} \right. . \tag{3.8}$$

One can see that this system is also equivalent to the system 3.5. (`Exercise`)

Example 3.11 Using one more time the system in Example 3.5, let us replace the second equation by the sum of the two original equations.

$$\begin{cases} x+y+z &= 2 \\ 2x+2z &= 2 \end{cases}. \tag{3.9}$$

Also in this case it can be shown that the system is equivalent to 3.5. (`Exercise`)

Following the previous examples we define the three *elementary operations* on a system of linear equation as:

i) interchange two equations;

ii) multiply one of the equations by a nonzero number;

iii) add a multiple of one equation to a different equation.

A similar set of operations can be defined also on matrices. We call *elementary row operations* on a matrix the following operations:

i) interchange two rows;

ii) multiply one of the rows by a nonzero number;

iii) add a multiple of one row to a different row.

These row operations can be seen as the respective of the elementary operations on system on the related augmented matrices.

Example 3.12 Let us consider the system of linear equations 3.5. Its augmented matrix is

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

The augmented matrices of the systems 3.7, 3.8 and 3.9 are respectively:

$$\begin{pmatrix} 1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & -2 & 2 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & 0 & 2 & 2 \end{pmatrix}.$$

They are obtained by the first matrix by applying an elementary row operation of type, respectively, *i*), *ii*) and *iii*).

Theorem 3.13 *Let us consider a system of linear equations. The system obtained by applying an elementary operation is equivalent to the original system.*

The previous theorem tells us that in order to find the solution of asystem we can apply a series of elementary operations to reduce a system to one which is easier to solve.

3.4 Gaussian elimination

In this section we'll see how to find the solutions of a general system of linear equations. Before giving the algorithm let us start with an example.

Example 3.14 Let us consider the following system of linear equations

$$\left\{\begin{array}{rcl} -x_3+3x_4 & = & 1 \\ x_1-4x_2+x_3 & = & 5 \\ 2x_1-8x_2+2x_3-3x_4 & = & 4 \end{array}\right. . \tag{3.10}$$

The augmented matrix of the system 3.10 is

$$\begin{pmatrix} 0 & 0 & -1 & 3 & 1 \\ 1 & -4 & 1 & 0 & 5 \\ 2 & -8 & 2 & -3 & 4 \end{pmatrix}.$$

Starting fom the 1-row, we find the first column from the left containing a non-zero entry, in our case the $(2,1)$-entry 1. Let us put the the 2-row on the top, that is let us interchange the 2-row with the 1-row (we use an elementary row operation of type *i*)). We get the matrix

$$\begin{pmatrix} 1 & -4 & 1 & 0 & 5 \\ 0 & 0 & -1 & 3 & 1 \\ 2 & -8 & 2 & -3 & 4 \end{pmatrix}.$$

The first non-zero entry 1, in the 1-row is called the *leading 1* for the first row. By subtracting 2 times the 1-row from the 3-row (i.e., applying an elementary row operation of type *iii*)), we obtain the matrix

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

Now, let us forget the first row and let us modify the matrix in a similar way we have done so far, starting from the second row. That is, starting from the 2-row, we find the first column from the left containing a non-zero entry, in our case the $(2,3)$-entry -1. Since this non-zero element is already in the good position, we don't need to interchange rows (elementary row operation of type *i*). On the other hand, to obtain a leading 1 in the second row, we can multiply the whole 2-row by -1 (elementary operation of type *ii*). This way we obtain the matrix

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

Now that we have the two first leading ones, et us do the same operation starting from the third row. Starting from the 3-row, we find the first column from the left containing a non-zero entry, in pour case the $(3,4)$-entry -3. As before, let us multiply the 3-row by $-\frac{1}{3}$ in order to create the leading 1 in the third column, obtaining the matrix

$$\begin{pmatrix} 1 & -4 & 1 & 0 & 5 \\ 0 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}. \tag{3.11}$$

The matrix 3.11 is the augmented matrix of the system of linear equations 3.6. From what we have seen in the previous section, that means that the system of linear equations in Equation (3.10) is equivalent to the system of linear equations in Equation (3.6), and thus the two have the same set of solutions.

A matrix is said to be in *row-echelon form*, and it will be called a *row-echelon matrix* if the following conditions are satisfied:

1. All zero rows are at the bottom;
2. The first non-zero entry from the left in each non-zero row is a 1, and we call it the *leading 1* of the row;
3. Each leading 1 is to the right of all leading 1s in the rows above it.

A row-echelon matrix is said to be in *reduced row-echelon form*, or is it called a *reduced row-echelon matrix* if it satisfies as well the condition

4. Each leading 1 is the only non-zero entry in its column.

Example 3.15 The matrix 3.11 has is in row-echelon form but non in reduced row-echelon form. All the other matrices in Example 3.14 are not in row-echelon form.

The following matrices are in reduced row-echelon form

$$\begin{pmatrix} 1 & 0 & 0 & 10 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The following result shown that given a system of linear equations we can always find an equivalent *easier* system which can be solved using the back-substitution (when a solution exists). The algorithm in the proof is called the *Gaussian Algorithm*

Theorem 3.16 *Every matrix can be carried, in a finite number of steps to a row-echelon form (reduced, if desiderd), using a sequence of elementary row operations.*

Proof.[Gaussian Algorithm] Do the following steps until you obtain a (reduced) row-echelon matrix.

Step 1-1 Starting from the 1-row, find the first non-zero entry in the first column from the left and interchange the corresponding row with the 1-row;

Step 1-2 Multiply the 1-row by a costant to create the first leading 1 in the 1-row;

Step 1-3 Make each entry below the leading 1 to be zero by subtracting multiples of the 1-row from lower rows.

Step 1-3b (to obtain a reduced row-echelon matrix) Make each entry above the leading 1 to be zero by subtracting multiples of the 1-row from upper rows.

Step 2-1 Starting from the 2-row, find the first non-zero entry in the first column from the left and interchange the corresponding row with the 2-row;

Step 2-2 Multiply the 2-row by a costant to create the first leading 1 in the 2-row;

Step 2-3 Make each entry below the leading 1 to be zero by subtracting multiples of the 2-row from lower rows.

Step 2-3b (to obtain a reduced row-echelon matrix) Make each entry above the leading 1 to be zero by subtracting multiples of the 2-row from upper rows.

$\vdots$

Step k-1 Starting from the k-row, find the first non-zero entry in the first column from the left and interchange the corresponding row with the k-row;

Step k-2 Multiply the k-row by a costant to create the first leading 1 in the k-row;

Step k-3 Make each entry below the leading 1 to be zero by subtracting multiples of the k-row from lower rows.

Step k-3b (to obtain a reduced row-echelon matrix) Make each entry above the leading 1 to be zero by subtracting multiples of the k-row from upper rows.

■

The algorithm presented in the proof of Theorem 3.16 can be implemented in your favorite program language (`Exercise`).

Note that the way to carry a matrix to its reduced form is not unique. Indeed one can obtain the same result by changing the order of some of the steps. That means that, even though the algorithm always work, it is not necessarily the most efficient one.

Example 3.17 Let us prove that the following system has no solution

$$\begin{cases} x+y &= 1 \\ x-2 &= 2 \\ y+z &= 1 \end{cases} \tag{3.12}$$

The reduction of the augmented matrix is

$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & 1 \end{pmatrix} \xrightarrow[R_2 \to R_2 - R_1]{iii)} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$
$$\xrightarrow[R_2 \to -R_2]{ii)} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$
$$\xrightarrow[R_3 \to R_3 - R_2]{iii)} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$
$$\xrightarrow[R_3 \to \frac{1}{2}R_3]{ii)} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where at every passage we show which elementary row operation we are applying: for instance $\xrightarrow[R_2 \to R_2 - R_1]{iii)}$ means that we are using the elementary row operation of type *iii)* by subtracting once the first row to the second row.

The last row of the last matrix corresponds to the equation

$$0x + 0y + 0z = 1,$$

which is clearly never satisfied, no matter the choice of x, y and z. Since the solution of the system must satisfy all equations, the system of linear equation corresponding to this augmented matrix, and thus the equivalent original system, has no solution.

Example 3.18 Let us find all solutions of the following system of linear equations

$$\begin{cases} x+y-z &= 3 \\ -2x-y &= -4 \\ 4x+2y+3z &= -1 \end{cases}.$$

(`Exercise`)

Looking at the row-echelon reduction of the augmented matrix of a system of linear equations, we can also determine if the original system has no solution, a unique solution or infinitely many solution. Indeed, let us suppose that we have a system of m linear equations in n variables, like the one in Equation (3.2), and let A be its augmented matrix. If we reduce A to a row-echelon form R, then we have the following cases

1. If there is a leading 1 in the last column, then the system of linear equation has <u>no</u> solution (as seen in Example 3.17);
2. If there is no leading 1 in the last column, then the system has at least one solution. We call the number of leading 1s the *rank* of the matrix, and we denote it by rank (A). Note that the rank does not change under elementary row operations, so rank $(A) =$ rank (R). Moreover, since there are no leading 1s in the last column, we have rank $(a) \leq n$. We can thus distinguish two cases:
   - (a) If there is at least one solution and rank $(A) = n$, then the solution is <u>unique</u>, and it can be found simply by back-substitution.
   - (b) If there is at least one solution and rank $(A) < n$, then the system has <u>infinitely many</u> solutions (as in Example 3.14). In this case we assign $n -$ rank (A) parameters to the variable corresponding to the columns without leading 1s, and we solve, again, by back-substitution (see also Example 3.8).

This method is also called *Gaussian elimination.*
We can summarize what seen so far in the following theorem.

Theorem 3.19 *For any system of linear equations there are exactly three possibilities:*

1. *The system has no solution.*
2. *The system has a unique solution.*
3. *The system has infinitely many solutions. Moreover, if* n *is the number of variables and* r *is the rank of the augmented matrix, then the set of solutions has exactly* $(n - r)$ *parameters.*

When it is clear from the context we will also call *rank* of the system the rank of the associated augmented matrix.

3.5 Homogeneous systems

Let us consider a system of linear equations in which all the constant terms are zero, such as the following one

$$\left\{ \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0 \end{array} \right. \tag{3.13}$$

Such a system is called *homogeneous.* It is clear that when choosing $x_1 = x_2 = \cdots = x_n = 0$ the equations are satisfied. Thus a homogeneous system has always (at least) one solution, namely $X = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}$. We call this solution the *trivial* solution of the homogeneous system. Any other possible solution with at least one of the variables non-zero is called a *non-trivial* solution.

From what seen in the previous section we can prove the following result.

Theorem 3.20 *If a homogeneous system of linear equations has more variables than equations, then it has nontrivial solutions.*

Proof. Let us consider a system of m linear equations in n variables and let us suppose that $n > m$. Let A be the augmented matrix. We know that the system has at least one solution, the trivial one. Since $\operatorname{rank}(A) \leq m < n$, we have, using Theorem 3.19, that the system has infinitely many solutions. ∎

Example 3.21 Let us consider the following homogeneous system

$$\left\{\begin{array}{rcl} x_1 - 2x_2 + 4x_3 - x_4 + 5x_6 & = & 0 \\ -2x_1 + 4x_2 - 7x_3 + x_4 + 2x_5 - 8x_6 & = & 0 \\ 3x_1 - 6x_2 + 12x_3 - 3x_4 + x_5 + 15x_6 & = & 0 \\ 2x_1 - 4x_2 + 9x_3 - 3x_4 + 3x_5 + 12x_6 & = & 0 \end{array}\right. .$$

The augmented matrix of this system is

$$\begin{pmatrix} 1 & -2 & 4 & -1 & 0 & 5 & 0 \\ -2 & 4 & -7 & 1 & 2 & -8 & 0 \\ 3 & -6 & 12 & -3 & 1 & 15 & 0 \\ 2 & -4 & 9 & -3 & 3 & 12 & 0 \end{pmatrix}.$$

The $(1,1)$-entry being the first leading 1, we proceed us in the previous section to clean the rest of the 1-column and, after that, to find the other leading 1s and continue with elementary row operations until we obtain a reduced row-echelon matrix (`Exercise`)

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

Thus the original system is equivalent to the following one:

$$\left\{\begin{array}{rcl} x_1 - 2x_2 + x_4 - 3x_6 & = & 0 \\ x_3 - x_4 + 2x_6 & = & 0 \\ x_5 & = & 0 \end{array}\right. .$$

The leading 1s in the augmented matrix correspond to the variables x_1, x_3 and x_5, and the rank of the system is 3. The other variables, i.e. x_2, x_4 and x_6 are called *non-leading variables*. To find the general solution we will associate some parameters, let us call them s, t and u, to the non-leading variables. The general solution has thus the form

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 2s-3t+3u \\ s \\ t-2u \\ t \\ 0 \\ u \end{pmatrix} = s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + u \begin{pmatrix} 3 \\ 0 \\ -2 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Let us denote

$$X_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} -3 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} 3 \\ 0 \\ -2 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

We call X_1, X_2 and X_3 *basic solutions* of the system. The general solution X is a *linear combination* of the three basic solutions, since

$$X = sX_1 + tX_2 + uX_3$$

for any choice of s, t and u.

We can generalize the previous example in the following theorem.

Theorem 3.22 *Let us consider a system of homogeneous linear equations in n variables and let us suppose that its rank is r. Then*

- *The Gaussian algorithm produces exactly $n - r$ basic solutions;*
- *Every solution is a linear combination of these basic solutions.*

Given a general system of linear equations we can associate to it a homogeneous system called simply by replacing the constant terms with zero. We will refer to this system as the *associated homogeneous system* of the original one.

Example 3.23 Let us consider the following system of 3 linear equations in 4 variables

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 &= 2 \\ -x_1 + 2x_2 + x_4 &= 1 \\ 2x_1 - 4x_2 + x_3 &= 1 \end{cases}. \tag{3.14}$$

A possible solution of the system 3.14 is

$$X_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Note that this particular solutionis not, in general, unique and it is not always easy to find.

As seen in the first section of the Chapter, we can rewrite this sytem of linear equations as a single matrix equation

$$AX = B$$

where

$$A = \begin{pmatrix} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

are respectively the coefficient matrix, the matrix of variables and the matrix of constants of the system.

The associated homogeneous system is represented by the matrix equation

$$\begin{pmatrix} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.15}$$

By reduction of the augmented matrix of the homogeneous system in a reduced row-echelon form

$$\begin{pmatrix} 1 & -2 & 1 & 1 & 0 \\ -1 & 2 & 0 & 1 & 0 \\ 2 & -4 & 1 & 0 & 0 \end{pmatrix} \xrightarrow[\substack{R_2 \to R_2 + R_1 \\ R_3 \to R_3 - 2R_1}]{iii)} \begin{pmatrix} 1 & -2 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & -1 & -2 & 0 \end{pmatrix}$$

$$\xrightarrow[\substack{R_1 \to R_1 - R_2 \\ R_3 \to R_3 - +2R_2}]{iii)} \begin{pmatrix} 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

we find that the general solution of the system 3.15 is

$$X = s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix}$$

where s and t are parameters representing arbitrary numbers.

The general solution of the system 3.14 is

$$X = X_0 + sX_1 + tX_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix},$$

with s and t arbitrary numbers, X_0 the particular solution seen before and X_1, X_2 the basic solutions computed above.

The previous example illustrate the following theorem.

Theorem 3.24 *Let us consider the system of linear equations $AX = B$, and let us suppose that X_0 is a particular solution. Then*

1. *if X' is a solution to the associated homogeneous system $AX = O$, then $X = X_0 + X'$ is a solution to the system $AX = B$.*
2. *Every solution to the system $AX = B$ has the form $X = X_0 + X'$ for some solution X' to the associated homogeneous system $AX = O$.*

Example 3.25 Let us consider the system of linear equations

$$\left\{\begin{array}{rcl} x_1 - 2x_2 + 2x_3 - x_4 & = & 1 \\ 2x_1 - 4x_2 + 3x_3 + x_4 & = & 2 \\ 3x_1 - 6x_2 + 5x_3 & = & 3 \end{array}\right. .$$

Using the Gaussian elimination and Theorem 3.24, we can write the general solution to the system as the sum of a particular solution and the general solution to the associated homogeneous system (`Exercise`).

Chapter 4

Inverse of a Matrix and Elementary matrices

As we did in Chapter 2, we will consider here matrices over $\mathbb{R}$.

4.1 The Matrix Inverse algorithm

In Section 2.6 we defined the inverse of a square matrix A as the matrix B such that

$$AB = I \quad \text{and} \quad BA = I,$$

where B has the same size of A and I is the identity matrix.

Example 4.1 Let us consider the matrix

$$A = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R}).$$

Let us prove that the inverse of A is its square A^2. Indeed, the matrix A^2 is given by

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

while the matrix A^3 is

$$A^3 = A^2 A = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Thus $AA^2 = A^3 = I$ and $A^2A = A^3 = I$, which prove the claim.

Finding the inverse of a given square matrix, when this exists, is not generally a trivial task. In the following example we show how to use the tools from the previous chapter in order to find the inverse of a matrix.

Example 4.2 Let us consider the 2×2 matrix

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

and let us suppose that its inverse exists and has the form

$$B = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$$

for certains $x_1, x_2, x_3, x_4 \in \mathbb{R}$.

Since $BA = I$, we have

$$\begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \begin{pmatrix} 2 & -5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 & -5x_1 + 2x_2 \\ 2x_3 + x_4 & -5x_3 + 2x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is equivalent to the system of four linear equations in four variables

$$\begin{cases} 2x_1 + x_2 &= 1 \\ -5x_1 + 2x_2 &= 0 \\ 2x_3 + x_4 &= 0 \\ -5x_3 + 2x_4 &= 1 \end{cases}.$$

Using the Gaussian algorithm on the augmented matrix of the system, we can find the equivalent matrix in reduced row-echelon form as follows:

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 1 \\ -5 & -2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix} \xrightarrow[R_1 \to \frac{1}{2}R_1]{ii)} \begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ -5 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2 \to R_2 + 5R_1]{iii)} \begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{9}{2} & 0 & 0 & \frac{5}{2} \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2 \to \frac{2}{9}R_2]{ii)} \begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow[R_1 \to R_1 - \frac{1}{2}R_2]{iii)} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{2}{9} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow[R_3 \to \frac{1}{2}R_3]{ii)} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{2}{9} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & -5 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow[R_4 \to R_4 + 5R_3]{iii)} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{2}{9} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{9}{2} & 1 \end{pmatrix}$$

$$\xrightarrow[R_4 \to \frac{2}{9} R_4]{ii)} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{2}{9} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & \frac{2}{9} \end{pmatrix}$$

$$\xrightarrow[R_3 \to R_3 - \frac{1}{2} R_4]{iii)} \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{2}{2} \\ 0 & 1 & 0 & 0 & \frac{5}{9} \\ 0 & 0 & 1 & 0 & -\frac{1}{9} \\ 0 & 0 & 0 & 1 & \frac{2}{9} \end{pmatrix}.$$

Thus we get the solution

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & \frac{5}{9} & -\frac{1}{9} & \frac{2}{9} \end{pmatrix}.$$

One can check that considering the system of linear equations associated to the matrix equation

$$AB = \begin{pmatrix} 2 & -5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} = \begin{pmatrix} 2x_1 - 5x_3 & 2x_2 - 5x_4 \\ x_1 + 2x_3 & x_2 + 2x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

one find the same solution (`Exercise`).

The (unique) inverse of A is thus the matrix

$$B = \begin{pmatrix} \frac{2}{9} & \frac{5}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 2 & 5 \\ -1 & 2 \end{pmatrix} \in \mathcal{M}_{2,2}(\mathbb{R}).$$

The following result gives us a method to compute the inverse of an invertible matrix using the Gaussian Algorithm.

Theorem 4.3 (Matrix Inverse Algorithm) *Let A be a square matrix. If there exists a sequence of elementary row operations that carry $A \to I$, then A is invertible and this same sequence carries $I \to A^{-1}$. Thus, applying the same sequence of row operations on the matrix $\begin{pmatrix} A & I \end{pmatrix}$, one has the reduction*

$$\begin{pmatrix} A & I \end{pmatrix} \to \begin{pmatrix} I & A^{-1} \end{pmatrix}.$$

Example 4.4 Let A be the matrix defined in Example 4.2. The reduction to the reduced row-echelon form of the matrix $\begin{pmatrix} A & I \end{pmatrix}$ is the following

$$\begin{pmatrix} 2 & -5 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{pmatrix} \xrightarrow[R_1 \leftrightarrow R_2]{i)} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 2 & -5 & 1 & 0 \end{pmatrix}$$

$$\xrightarrow[R_2 \to R_2 - 2R_1]{iii)} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & -9 & 1 & -2 \end{pmatrix}$$

$$\xrightarrow[R_2 \to -\frac{1}{9} R_2]{ii)} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$$

$$\xrightarrow[R_1 \to R_1 - 2R_2]{iii)} \begin{pmatrix} 1 & 0 & \frac{2}{9} & \frac{5}{9} \\ 0 & 1 & -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$$

Thus the inverse of the matrix $\begin{pmatrix} 2 & -5 \\ 1 & 2 \end{pmatrix}$ is the matrix $\begin{pmatrix} \frac{2}{9} & \frac{5}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$ (which we already known from Example 4.2).

Using Theorem 4.3 we can find a formula to the inverse of an invertible 2×2-matrix.

Example 4.5 Let us consider the 2×2-matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $a, b, c, d \in \mathbb{R}$ and $ad - bc \neq 0$. Then A is invertible and its inverse is the matrix

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Indeed, by using the algorithm described in Theorem 4.3 we have the reduction (`Exercise`)

$$\begin{pmatrix} a & b & & 1 & 0 \\ c & d & & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & & \frac{d}{\Delta} & -\frac{b}{\Delta} \\ 0 & 1 & & -\frac{c}{\Delta} & \frac{a}{\Delta} \end{pmatrix}$$

where $\Delta = ad - bc$ is called the *determinant* of A. We also call the matrix

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

the *adjoint* of A (we will discuss more about determinants and adjoints later).

To double check we can also verify that

$$\begin{aligned} AA^{-1} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tfrac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= \tfrac{1}{ad-bc} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= \tfrac{1}{ad-bc} \begin{pmatrix} ad - bc & 0 \\ 0 & -cb + ad \end{pmatrix} \\ &= I \end{aligned}$$

and

$$\begin{aligned} A^{-1}A &= \tfrac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \tfrac{1}{ad-bc} \begin{pmatrix} ad - bc & 0 \\ 0 & -bc + ad \end{pmatrix} \\ &= I. \end{aligned}$$

Example 4.6 Let us consider the matrix

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

Using the matrix inverse algorithm we can show that (`Exercise`) the inverse of A is the matrix

$$B = \frac{1}{26}\begin{pmatrix} -8 & 3 & 7 \\ 22 & -5 & -3 \\ 10 & -7 & 1 \end{pmatrix}.$$

To make sure that the answer is right it is enough to verify that $AB = I$ and $BA = I$.

If a matrix A is not invertible, then no sequence of row operations can carry $A \to I$. Hence the algorithm breaks down because a row of zeros is encountered.

Example 4.7 The matrix

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 3 \\ 4 & 7 & 1 \end{pmatrix}$$

has no inverse. Indeed, let us try the matrix inverse algorithm on A.

$$\begin{pmatrix} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 3 & 3 & 0 & 1 & 0 \\ 4 & 7 & 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[\substack{R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 4R_1}]{iii)} \begin{pmatrix} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -1 & 5 & -2 & 1 & 0 \\ 0 & -1 & 5 & -4 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2 \to -R_2]{ii)} \begin{pmatrix} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -5 & 2 & -1 & 0 \\ 0 & -1 & 5 & -4 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[\substack{R_1 \to R_1 - 2R_2 \\ R_3 \to R_3 + R_2}]{iii)} \begin{pmatrix} 1 & 0 & 9 & -3 & 2 & 0 \\ 0 & 1 & -5 & 2 & -1 & 0 \\ 0 & 0 & 0 & -2 & -1 & 1 \end{pmatrix}.$$

Since A will never be transformed to the identity matrix by elementary row operations, A is not invertible.

4.2 Inverses and systems of linear equations

As we have seen at the end of Section 3.4, some systems of linear equations have a unique solution. Here we show how to find such a solution.

Theorem 4.8 *Let us consider a system of n linear equations in n variables and let us suppose that we can write this system in matrix form as*

$$AX = B.$$

If the n-square matrix A is invertible, the system has the unique solution

$$X = A^{-1}B.$$

Proof. Note that, since the system of linear equations has n equations and n variables, then A has size $n \times n$. Since A^{-1} is well defined, and both A^{-1} and B has size $n \times 1$, then X is also well defined and it has size $n \times 1$. Thus the system has at least one solution, namely X.

Moreover, since A is invertible, then we can use the matrix inverse algorithm to reduce the matrix $\begin{pmatrix} A & I_n \end{pmatrix}$ to $\begin{pmatrix} I_n & A^{-1} \end{pmatrix}$. Using the same sequence of row operations we can thus reduce

$$A \to I_n,$$

which implies that $\operatorname{rank}(A) = \operatorname{rank}(I_n) = n$. Thus, from what we have seen in Section 3.4, the solution X is unique. ■

Example 4.9 Let us consider the system of linear equations

$$\begin{cases} 2x - 5y &= 1 \\ x + 2y &= 2 \end{cases}.$$

The system can be written the a matrix equation

$$AX = B \tag{4.1}$$

where

$$A = \begin{pmatrix} 2 & -5 \\ 1 & 2 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

From Example 4.2 we know that A is invertible and that its inverse is

$$A^{-1} = \begin{pmatrix} \frac{2}{9} & \frac{5}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix}.$$

Thus. multiplying both sides of Equation 4.1 by A^{-1} we obtain

$$\begin{aligned} A^{-1}AX &= A^{-1}B \\ IX &= \begin{pmatrix} \frac{2}{9} & \frac{5}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ X &= \begin{pmatrix} \frac{2}{9} + \frac{10}{9} \\ -\frac{1}{9} + \frac{4}{9} \end{pmatrix} \\ \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{4}{3} \\ \frac{1}{3} \end{pmatrix}. \end{aligned}$$

Hence, the solution of our system is $x = \frac{4}{3}$ and $y = \frac{1}{3}$.

Example 4.10 Let us consider the system of 3 linear equations in 3 variables

$$\begin{cases} x + 2y - z &= 1 \\ 2x + 3y - 5z &= 2 \\ 4x + y + z &= -1 \end{cases}.$$

We can represent the system as the matrix equation

$$AX = B$$

where

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & -5 \\ 4 & 1 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

From Example 4.6 we know that A is invertible and that its inverse is the matrix

$$A^{-1} = \frac{1}{26} \begin{pmatrix} -8 & 3 & 7 \\ 22 & -5 & -3 \\ 10 & -7 & 1 \end{pmatrix}$$

Then the unique solution of the system is

$$X = A^{-1}B = \frac{1}{26} \begin{pmatrix} -8 & 3 & 7 \\ 22 & -5 & -3 \\ 10 & -7 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} -9 \\ 15 \\ -5 \end{pmatrix},$$

that is, we have

$$x = -\frac{9}{26}, \quad y = \frac{15}{26} \quad \text{and} \quad z = -\frac{5}{26}.$$

4.3 Conditions for invertibility

The following result (given without proof) summarizes the relation between an invertible matrix and the associated system of linear equations.

Theorem 4.11 *Let A be a n-square matrix. The following conditions are equivalent:*

1. *The matrix A is invertible.*
2. *There exists a matrix C such that $AC = I$.*
3. *The matrix A can be carried to the identity matrix I by elementary row operations.*
4. *The system $AX = B$ has a solution X for every choice of column B.*
5. *The homogeneous system $AX = O$ has only the trivial solution $X = O$.*

Some of the equivalence in the previous theorem can be proved by using the definition of inverse and the results in Sections 4.1 and 4.2.

We can also give an extra result, without proof, for invertible matrices.

Theorem 4.12 *Let A, C be two square matrices. If $AC = I$ then $CA = I$ also. Moreover, in this case, A and C are both invertible, $C = A^{-1}$ and $A = C^{-1}$*

Using the previous theorem we can show that the only invertible matrices are square. That is, if A is an $m \times n$ matrix, and $AC = I_m$ and $CA = I_n$ hold for some $n \times m$ matrix C, then $m = n$.

This is false if A and C are not square matrices.

Example 4.13 Let us consider the two matrices

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

One has

$$AC = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

but

$$CA = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3.$$

Example 4.14 Let A, B be two square matrices and let us suppose that $A^3 = B$ and that B is invertible. Then, using Theorem 4.12, we can prove that A is invertible too (`Exercise`).

4.4 Elementary matrices

In Section 3.3 we defined the three types of elementary row operations on a matrix. Similary, we call *elementary column operations* on a matrix the following operations:

i) interchange two columns;

ii) multiply one of the columns by a nonzero number;

iii) add a multiple of one column to a different column.

Example 4.15 Let us consider the same matrix

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

of Example 3.12. The three matrices

$$\begin{pmatrix} 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 5 & 1 & 1 & 2 \\ 5 & -1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 & 2 \\ 1 & -1 & 0 & 0 \end{pmatrix}$$

are obtained from A using respectively an elementary row operation of type *i*) (interchanging column 2 with column 4), type *ii*) (multiplying the first column by 5), and type *iii*) (adding the second column to the third one).

A square matrix E that is obtained by doing a single elementary row operation or a single elementary column operation to the identity matrix I is called an *elementary matrix*.

We say that E is of type *i*), *ii*) or *iii*) when the correspoding row or column operation is of type *i*), *ii*) or *iii*).

Example 4.16 The matrices

$$E_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

are elementary matrices of type *i*), *ii*) and *iii*) respectively, obtained by performing the following row operations on the 3×3 identity matrix I_3:

$$I \xrightarrow[R_1 \leftrightarrow R_3]{i)} E_1, \quad I \xrightarrow[R_2 \to \frac{1}{3}R_2]{ii)} E_2 \quad \text{and} \quad I \xrightarrow[R_3 \to R_3 - 2R_1]{iii)} E_3.$$

Theorem 4.17 *Every elementary matrix E is invertible, and E^{-1} is the elementary matrix (of the same type of E) obtained from I by the inverse of the operation that produces E from I.*

Example 4.18 Let us consider the three elementary matrices E_1, E_2 and E_3 seen in Example 4.16. Then we can find their inverses E_1^{-1}, E_2^{-1} and E_3^{-1} (`Exercise`).

The left multiplication by an elementary matrix of a certain type is equivalent to a corresponding elementary row operation of the same type.

Example 4.19 Let us consider the matrix

$$A = \begin{pmatrix} 2 & 2 & 2 & 0 \\ 0 & 3 & -2 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Interchanging the 1-row and the 3-row (elementary row operation of type *i*)) can be performed by multiplying A by the matrix E_1 in Example 4.16. Indeed

$$E_1 A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 2 & 2 & 0 \\ 0 & 3 & -2 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & -2 & 0 \\ 2 & 2 & 2 & 0 \end{pmatrix}.$$

Subtracting 2 times the 1-row from the 3-row in the previous matrix (elementary row operation of type *iii*)) can be done by multiplying $E_1 A$ by the matrix E_3 in Example 4.16. Indeed

$$E_3(E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & -2 & 0 \\ 2 & 2 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Multiplying the 2-row by $\frac{1}{3}$ in the previous matrix can be done by left multiplication by the matrix E_2 in Example 4.16. Indeed

$$E_2(E_3 E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus we can get an equivalent matrix to A in row echelon form by multiplying on the left by $(E_2E_3E_1)$. We can also obtain a reduced row-echelon form of A by left multiplying $E_2E_3E_1A$ by the elementary matrix of type *iii*)

$$\widetilde{E}_3 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

indeed

$$\widetilde{E}_3E_2E_3E_1A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{5}{3} & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Similarly, the right multiplication by an elementary matrix of a certain type is equivalent to a correspoding elementary column operation of the same type.

Example 4.20 Let us consider the matrix

$$A = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 2 & -1 \end{pmatrix}.$$

Interchanging the first column with the third column (elementary column operation of type *i*) corresponds to multiply A on the right with the matrix E_1 of Example 4.16. Indeed

$$AE_1 = \begin{pmatrix} 5 & 0 & 1 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 5 \\ -1 & 2 & 0 \end{pmatrix}.$$

Subtracting 5 times the 1-column from the 3-column in the previous matrix (elementary operation of type *iii*)) can be performed by right multiplication by the elmentary matrix of type *iii*)

$$E_3 = \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Indeed

$$(AE_1)E_3 = \begin{pmatrix} 1 & 0 & 5 \\ -1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 5 \end{pmatrix}.$$

Multiplying the 2-column of the previous matrix by $\frac{1}{2}$ is done by right multiplying AE_1E_3 by the elementary matrix of type *ii*)

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Indeed

$$(AE_1E_3)E_2 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 5 \end{pmatrix}.$$

By right multiplying $AE_1E_3E_2$ by the elmentary matrices of type *iii*)

$$\widetilde{E_3} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \hat{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix},$$

we obtain the matrix

$$\begin{aligned} (AE_1E_3E_2)\widetilde{E}_3\hat{E}_3 &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{E}_3 \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

This matrix is called a *reduced column-echelon form* of A. Its transposition

$$(AE_1E_3E_2\widetilde{E}_3\hat{E}_3)^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

is in reduced row-echelon form.

The previous examples can be generalized in the following result.

Theorem 4.21 *Let us consider two matrices A, B and let us suppose that there exists a series of row operations carrying $A \to B$. Then*

1. *There exists an invertible matrix U such that $B = UA$.*
2. *U can be constructed by performing the same row operations carrying A to B on the double matrix $\begin{pmatrix} A & I \end{pmatrix}$, that is*

$$\begin{pmatrix} A & I \end{pmatrix} \longrightarrow \begin{pmatrix} B & U \end{pmatrix}.$$

3. *$U = E_k \cdots E_2E_1$, where $E_1, E_2, \ldots, E_k$ are the elementary matrices corresponding in order to the row operations carrying A to B.*

Example 4.22 Let us consider the matrix

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

seen in Example 4.19. We have seen that this matrix can be carried to the matrix

$$B = \begin{pmatrix} 1 & 0 & \frac{5}{3} & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

by a series of elementary row operations, and that $B = UA$ where

$$U = \widetilde{E}_3 E_2 E_3 E_1$$

is an invertible matrix with $_1, E_2, E_2$ and $\widetilde{E}_3$ elementary row matrices.

Example 4.23 Let us consider the matrix

$$A = \begin{pmatrix} 3 & -2 & 5 \\ 1 & -1 & 0 \end{pmatrix}.$$

Using Theorem 4.21 we can find an invertible matrix U (with its decomposition in elementary matrices) and a matrix B in reduced row-echelon form such that $B = UA$ (`Exercise`).

4.5 Elementary matrices and rank

Combining Theorem 4.21 with Theorem 4.3 we obtain that the inverse of an invertible matrix can be written as a product of elementary matrices.

Example 4.24 Let us consider the invertible matrix

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

A possible reduction of A is reduced row-echelon form is the following:

$$\begin{aligned} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} & \xrightarrow[R_1 \leftrightarrow R2]{i)} \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \\ & \xrightarrow[R_2 \to R2 - 2R_1]{iii)} \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \\ & \xrightarrow[R_2 \to \frac{1}{3} R2]{ii)} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \\ & \xrightarrow[R_1 \to R_1 + R2]{iii)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

The elementary matrices corresponding to the previous elementary operations are, in order:

$$E_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \quad \text{and} \quad E_4 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

From Theorem 4.21 we thus have $I = A^{-1}A$, where

$$A^{-1} = E_4E_3E_2E_1 = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \frac{1}{3}\end{pmatrix}\begin{pmatrix}1 & 0\\ -2 & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \frac{1}{3}\begin{pmatrix}1 & 1\\ 1 & -2\end{pmatrix}.$$

Note that we can also write A as a product of elementary matrices. Indeed, since $A = \left(A^{-1}\right)^{-1}$, we have

$$\begin{array}{rcl} A & = & (E_4E_3E_2E_1)^{-1} = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1} \\ & = & \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix}\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} \end{array}.$$

If we combine row operations and column operations, we can get a simpler for of any matrix.

Theorem 4.25 *Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ be a matrix of rank r. Then there exit two invertible matrices $U \in \mathcal{M}_{m,m}(\mathbb{R})$ and $V \in \mathcal{M}_{n,n}(\mathbb{R})$ such that*

$$UAV = \begin{pmatrix} I_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{pmatrix}$$

or, for short and when the size is clear from the context, $\begin{pmatrix} I_r & O \\ O & O\end{pmatrix}$. Moreover, the matrices U and V can be computed using the Gaussian Algorithm as follows:

$$\begin{pmatrix} A & I_m \end{pmatrix} \to \begin{pmatrix} R & U \end{pmatrix},$$

where R is a reduced row-echelon matrix; and

$$\begin{pmatrix} R^T & I_n \end{pmatrix} \to \begin{pmatrix} \begin{pmatrix} I_r & O \\ O & O\end{pmatrix}^T & V^T \end{pmatrix}.$$

Proof.[Idea] Let us give an idea of the proof.

First, we use a similar idea as the matrix inversion algorithm. We add the $m \times m$ identity matrix to the right side of A to get the $m \times (m+n)$ matrix $\begin{pmatrix} A & I_m \end{pmatrix}$. Using the Gaussian algorithm we can perform a sequence of elementary row operation and obtain $\begin{pmatrix} R & U \end{pmatrix}$, where R is a reduced row-echelon matrix, equivalent to A, and U is the multiplication of the elementary matrices corresponding to the elementary row operations, according to Theorem 4.21.

If R is not already in the form $\begin{pmatrix} I_r & O \\ O & O\end{pmatrix}$, we consider its transpose R^T and we procede in a similar way. We add the $n \times n$ identity matrix to the right side of R^T, and doing a sequence of elementary row operations we obtain a matrix of the form $\begin{pmatrix} \begin{pmatrix} I_r & O \\ O & O\end{pmatrix}^T & V^T \end{pmatrix}$. From Theorem 4.21 it follows that V corresponds to the multiplication of the elementary matrices corresponding to the elementary column operations.

Since we have that $R = UA$ in the first step of the theorem, then we also have (always using Theorem 4.21)

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}^T = V^T R^T = V^T (UA)^T = V^T A^T U^T = (UAV)^T,$$

Recalling that for any matrix B we have $(B^T)^T = B$, we can conclude that

$$UAV = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}.$$

$\blacksquare$

Example 4.26 Let us consider the matrix

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

Let us use Theorem 4.25 to show that $\operatorname{rank}(A) = 2$ and that there exist two matrices U, V such that

$$UAV = \begin{pmatrix} I_2 & O \\ O & O \end{pmatrix}.$$

Let us first consider the reduction $\begin{pmatrix} A & I_3 \end{pmatrix} \to \begin{pmatrix} R & U \end{pmatrix}$ as in the first step of Theorem 4.25.

$$\begin{pmatrix} 1 & -2 & 3 & 1 & & 1 & 0 & 0 \\ -1 & 2 & -1 & 1 & & 0 & 1 & 0 \\ 2 & -4 & 5 & 1 & & 0 & 0 & 1 \end{pmatrix} \xrightarrow[\substack{R_2 \to R_2 + R_1 \\ R_3 \to R_3 - 2R_1}]{iii)} \begin{pmatrix} 1 & -2 & 3 & 1 & & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 & & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 & & -2 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2 \to \frac{1}{2}R_2]{ii)} \begin{pmatrix} 1 & -2 & 3 & 1 & & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & -1 & -1 & & -2 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[\substack{R_1 \to R_1 - 3R_2 \\ R_3 \to R_3 + R_2}]{iii)} \begin{pmatrix} 1 & -2 & 0 & -2 & & -\frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & 0 & 1 & 1 & & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & & -\frac{3}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

Thus we have

$$R = \begin{pmatrix} 1 & -2 & 0 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} -\frac{1}{2} & -\frac{3}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{3}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

Note that here the reduced row-echelon matrix R has a unique form, while U may have different forms.

Moreover, since R has two leadings ones, we have $\operatorname{rank}(A) = \operatorname{rank}(R) = 2$.

Using the second step of Theorem 4.25, that is the reduction of $\begin{pmatrix} R^T & I_4 \end{pmatrix}$, we obtain:

$$\begin{pmatrix} 1 & 0 & 0 & & 1 & 0 & 0 & 0 \\ -2 & 0 & 0 & & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 & 1 & 0 \\ -2 & 1 & 0 & & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[\substack{R_2 \to R_2 + 2R_1 \\ R_4 \to R_4 + 2R_1}]{iii)} \begin{pmatrix} 1 & 0 & 0 & & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & & 2 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2 \leftrightarrow R_3]{i)} \begin{pmatrix} 1 & 0 & 0 & & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & & 2 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[R_4 \to R_4 - R_1]{iii)} \begin{pmatrix} 1 & 0 & 0 & & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & & 2 & 0 & -1 & 1 \end{pmatrix}.$$

$$= \quad \left(\begin{pmatrix} I_2 & O_{2,1} \\ O_{2,2} & O_{2,1} \end{pmatrix} \quad V^T \right).$$

where

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

Note that the reduced row-echelon matrix $\begin{pmatrix} I_2 & O_{2,1} \\ O_{2,2} & O_{2,1} \end{pmatrix}$ equivalent to R has a unique form, while V may have different forms.

Finally, one can check that we actually have

$$\begin{pmatrix} -\frac{1}{2} & -\frac{3}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{3}{2} & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 & 1 \\ -1 & 2 & -1 & 1 \\ 2 & -4 & 5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

that is

$$UAV = \begin{pmatrix} I_2 & O_{2,2} \\ O_{1,2} & O_{1,2} \end{pmatrix}.$$

Example 4.27 Following the previous example, let us show that given the matrix

$$A = \begin{pmatrix} 3 & -3 & 6 \\ 1 & -1 & 1 \end{pmatrix}$$

we can write

$$UAV = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

for two invertible matrices U, V (`Exercise`).

Chapter 5

Determinant and Diagonalization

5.1 Determinant

In Example 4.5 we defined the *determinant* of a generic 2 × 2-matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

as

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

In this section we define the determinant of a generic square matrix and we show how to compute it.

To define the determinant, we give a recursive definition, that is we give a definition for a base case, here for a 1×1-matrix, and then we define the determinant of a $n \times n$-matrix using the determinant of a $(n-1) \times (n-1)$-matrix.

- Let $A = \begin{pmatrix} a \end{pmatrix} \in \mathcal{M}_{1,1}(\mathbb{R})$. Then $\det(A) = a$.
- Let $A = \begin{pmatrix} a_{ij} \end{pmatrix} \in \mathcal{M}_{n,n}(\mathbb{R})$. Then

  $$\det(A) = a_{11}C_{11}(A) + a_{12}C_{12}(A) + \cdots + a_{1n}C_{1n}(A), \tag{5.1}$$

  where $C_{ij}(A)$ is called the (i,j)-*cofactor* of A and it's defined as

  $$C_{ij}(A) = (-1)^{i+j}\det(A_{ij}),$$

  for each i and j, where A_{ij} is the $(n-1) \times (n-1)$-matrix obtained from A by delating the i-row and the j-column. We also call $(-1)^{i+j}$ the *sign* of the (i,j)-position in A.

Equation (5.1) is called the *Laplace expansion*, or *cofactor expansion*, of A along the 1-row.

Example 5.1 The definition of determinant is consistent for 2×2-matrices. Indeed we have

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

Example 5.2 Let us find the determinant of the 3×3-matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Using Equation (5.1) we have

$$\det A = a_{11}C_{11}(A) + a_{12}C_{12}(A) + a_{13}C_{13}(A).$$

From the definition of cofactor, it follows that

$$C_{11}(A) = (-1)^{1+1}\det(A_{11}) = (-1)^2 \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} = a_{22}a_{33} - a_{23}a_{32},$$

$$C_{12}(A) = (-1)^{1+2}\det(A_{12}) = (-1)^3 \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} = -(a_{21}a_{33} - a_{23}a_{31}),$$

$$C_{13}(A) = (-1)^{1+3}\det(A_{13}) = (-1)^4 \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = a_{21}a_{32} - a_{22}a_{31}.$$

Thus, we have

$$\begin{aligned} \det A &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Example 5.3 Following the previous example we can find the determinant of the matrix

$$A = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 1 & 2 \\ 5 & 0 & 3 \end{pmatrix}$$

(`Exercise`).

When we have a lot of zeros, especially in the first row, the computation is easier.

Example 5.4 Let us compute the determinant of the matrix

$$A = \begin{pmatrix} 0 & a_{12} & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{pmatrix}.$$

By definition we have

$$\det(A) = a_{12}(-1)^{1+2}\det\begin{pmatrix} a_{21} & 0 & 0 \\ 0 & a_{33} & 0 \\ 0 & 0 & a_{44} \end{pmatrix}.$$

Let us now consider the 3×3-matrix

$$B = \begin{pmatrix} a_{21} & 0 & 0 \\ 0 & a_{33} & 0 \\ 0 & 0 & a_{44} \end{pmatrix}$$

Its determinant is given by

$$\det(B) = a_{21}(-1)^{1+1}\det\begin{pmatrix} a_{33} & 0 \\ 0 & a_{44} \end{pmatrix} = a_{21} \cdot 1 \cdot (a_{33}a_{44}) = a_{21}a_{33}a_{44}.$$

Thus, we finally have

$$\det A = a_{12} \cdot (-1) \cdot (a_{21}a_{33}a_{44}) = -a_{12}a_{21}a_{33}a_{44}.$$

In the definition of determinant we considered the cofactor expansion along the first row. The following important result show us that we can compute it in a different way.

Theorem 5.5 (Laplace Expansion Theorem) *Let A be a square matrix. The determinant of A is equal to the cofactor expension along any row or column of A.*

Example 5.6 Let us consider the matrix

$$A = \begin{pmatrix} 2 & -1 & 0 & 3 \\ 1 & 0 & 5 & 7 \\ 7 & 9 & 0 & 2 \\ 4 & 0 & 0 & 8 \end{pmatrix}.$$

A smart way to compute $\det(A)$ using a cofactor expansion along the rows and the column having the bigger number of zeros. Let us thus start by doing the cofactor expansion along the 4-row.

$$\begin{aligned} \det(A) &= 4 \cdot (-1)^{4+1}\det\begin{pmatrix} -1 & 0 & 3 \\ 0 & 5 & 7 \\ 9 & 0 & 2 \end{pmatrix} + 0 \cdot (-1)^{4+2}\det\begin{pmatrix} 2 & 0 & 3 \\ 1 & 5 & 7 \\ 7 & 0 & 2 \end{pmatrix} \\ &\quad +0 \cdot (-1)^{4+3}\det\begin{pmatrix} 2 & -1 & 3 \\ 1 & 0 & 7 \\ 7 & 9 & 2 \end{pmatrix} + 8 \cdot (-1)^{4+4}\det\begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 5 \\ 7 & 9 & 0 \end{pmatrix} \\ &= -4\det\begin{pmatrix} -1 & 0 & 3 \\ 0 & 5 & 7 \\ 9 & 0 & 2 \end{pmatrix} + 8\det\begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 5 \\ 7 & 9 & 0 \end{pmatrix}. \end{aligned}$$

Applying the cofactor expansion along the 2-column of the first matrix and along the 3-column of the second matrix we obtain respectively:

$$\det\begin{pmatrix}-1 & 0 & 3\\ 0 & 5 & 7\\ 9 & 0 & 2\end{pmatrix} = 5\cdot(-1)^{2+2}\det\begin{pmatrix}-1 & 3\\ 9 & 2\end{pmatrix} = 5(-2-27) = -145$$

and

$$\det\begin{pmatrix}2 & -1 & 0\\ 1 & 0 & 5\\ 7 & 9 & 0\end{pmatrix} = 5\cdot(-1)^{2+3}\det\begin{pmatrix}2 & -1\\ 7 & 9\end{pmatrix} = -5(18+7) = -125.$$

Thus

$$\det(A) = -4\cdot(-145) + 8\cdot(-125) = -420.$$

The following results easily follows from Theorem 5.5

Corollary 5.7 *If a square matrix A has a row or a column of zeros, then $det(A) = 0$.*

Example 5.8 Let us consider the matrix

$$A = \begin{pmatrix}1 & 2 & 3\\ 0 & 0 & 0\\ 1 & -1 & 1\end{pmatrix}.$$

Then, using the cofactor expansion along the second row, we find

$$\det(A) = -0\cdot\det\begin{pmatrix}2 & 3\\ -1 & 1\end{pmatrix} + 0\cdot\det\begin{pmatrix}1 & 3\\ 1 & 1\end{pmatrix} - 0\cdot\det\begin{pmatrix}1 & 2\\ 1 & -1\end{pmatrix} = 0.$$

5.2 Elementary operations and determinants

Because of its recursive definition, it is often hard to compute the determinant. Using elementary operations we can create more zeros in a matrix, but these operations will change the determinant. Let us see how.

Theorem 5.9 *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$ be square matrix.*

(1) If B is obtained from A by interchanging two different rows (elementary row operation of type i)) or two different columns (elementary column operation of type i)), then

$$det(B) = -det(A).$$

(2) If B is obtained from A by multiplying a row (elementary row operation of type ii)) or a column (elementary column operation of type ii)) by a number k, then

$$det(B) = k\cdot det(A).$$

(3) If B is obtained from A by adding a multiple of some row of A to a different row (elementary row operation of type iii)) or a multiple of some column of A to a different column (elementary column operation of type iii)), then

$$det(B) = det(A).$$

The proof of the previous theorem is not hard, but its out of the scope of this course.

Example 5.10 Let us consider the matrix

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 3 \\ 4 & 7 & 0 \end{pmatrix}.$$

Because of point (3) of Theorem 5.9, we know that the determinant does not change if we subtract twice the 1-row from the 3-row and the 2-row from the 3-row, so

$$\det(A) = \det \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 3 \\ 2 & 3 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 3 \\ 0 & 0 & -1 \end{pmatrix}.$$

Thus, using Theorem 5.5 we have

$$\det(A) = -1 \cdot (-1)^{3+3} \det \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = -(1 \cdot 3 - 2 \cdot 2) = 1.$$

Example 5.11 Following the previous result we can compute the determinant of the following matrices:

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 3 \\ 4 & 7 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & a & a & a \\ a & 1 & a & a \\ a & a & 1 & a \\ a & a & a & 1 \end{pmatrix}$$

(`Exercise`).

5.3 Some properties on determinants

The following result easily follows from point (2) of Theorem 5.9.

Theorem 5.12 *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. Then for any number $k \in \mathbb{R}$*

$$det(kA) = k^n det(A).$$

Example 5.13 Let us consider the matrix

$$A = \begin{pmatrix} 2 & 4 \\ 6 & 2 \end{pmatrix} = 2 \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}.$$

Then

$$\det(A) = 4 - 24 = -20 = 2^2 \cdot (1 - 6).$$

The determinant of diagonal and triangular matrices is quite easy to compute.

Theorem 5.14 *If a square matrix is triangular, then its determinant is the product of the entries of the main diagonal.*

Proof.[Idea] Let us consider an upper triangular matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$, that is a matrix of the form

$$A = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 & 0 \\ * & a_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & a_{n-1,n-1} & 0 \\ * & * & \cdots & * & a_{n,n} \end{pmatrix}$$

where the $*$ represent arbitrary real numbers. Then, by considering recursively the Laplace expansion along the first row, we find that

$$\det(A) = a_{1,1}a_{1,2}\cdots a_{n,n}.$$

■

Example 5.15 Let us consider the matrix

$$A = \begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & -1 & 7 & 4 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

The matrix is (lower) triangular. Then, by Theorem 5.14 we have

$$\det(A) = 1 \cdot (-1) \cdot 2 \cdot (-3) = 6.$$

Column operations from a matrix A a to a matrix B can be accomplished by doing the corresponding row operations from $A^T 4$ to B^T, then take the transpose to B^T back to B. The following theorem tell us that transposing a matrix does not change its determinatn.

Theorem 5.16 *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. Then*

$$det\left(A^T\right) = det(A).$$

Example 5.17 Let us consider a matrix

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

and its transpose

$$A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

Then

$$\det\left(A^T\right) = 1 \cdot 4 - 3 \cdot 2 = -2 = 1 \cdot 4 - 2 \cdot 3 = \det(A).$$

Theorem 5.18 *Let E be an elementary matrix.*

1. *If E is of type $i)$ then $det(E) = -1$.*
2. *If E is of type $ii)$ and is obtained from I by multiplying a row (or a column) by a number k, then $det(E) = k$.*
3. *If E is of type $iii)$, then $det(E) = 1$.*

Proof. In this proof we will consider elementary matrices obtained from I_n, with $n \in \mathbb{N}$, using a row elementary operation. The case of matrices obtained by elementary column operations can be proved in a symmetric way.

1. Let us first consider the case of an elementary matrix of type $i)$. Let i, j with $1 \leq i, j \leq n$ be such that $I \xrightarrow[R_1 \leftrightarrow R_j]{i)} E$. Let us prove our result by induction on n. If $n = 2$, then necessarily we

$$E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and $\det(E) = -1$.

If $n > 2$ then, we have

$$E = \begin{array}{ccccccccc}
1 & & & i & & j & & n & \\
1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 \\
0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\
0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & i \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\
0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & j \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\
0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & n
\end{array}$$

Let $k \neq i, j$. Then the (k, k)-entry of the matrix is 1 while for all $h \neq k$, the (k, h)-entry of E is 0 (that is, the k-row has zeros everywhere except at its k-th position). Using Theorem 5.5, we can consider the cofactor expansion of E along the k-row and obtain

$$\begin{aligned} \det(E) &= (-1)^{k+k} \cdot 1 \cdot \det(A_{k,k}) + \textstyle\sum_{h \neq k} (-1)^{k+h} \cdot 0 \cdot \det(A_{k,h}) \\ &= \det(A_{k,k}), \end{aligned}$$

where $A_{k,k}$ is of the form

$$A_{k,k} = \begin{array}{cccccccc c} 1 & & & i' & & j' & & n-1 & \\ 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & i' \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & j' \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & n-1 \end{array}$$

for certains $1 \leq i', j' \leq n-1$. Note that $A_{k,k}$ is an elmentary matrix of size $(n-1) \times (n-1)$ and that it can be obtain from I_{n-1} by the elementary row-operation

$$I_{n-1} \xrightarrow[R_{i'} \leftrightarrow R_{j'}]{i)} A_{k,k}.$$

By inductive hypothesis $\det(A_{k,k} = 1)$. Thus $\det(E) = 1$ as well.

2. Let us now consider the case of an elementary matrix of type *ii*). Let i and k, with $1 \leq i \leq n$ and $k \neq 0$ be such that $I \xrightarrow[R_i \to kR_i]{ii)} E$. Then, we can prove that $\det(E) = k$ (`Exercise`).

3. Let us finally consider the case of an elementary matrix of type *iii*). Let i, j and k with $1 \leq i, j \leq n$ and $k \in \mathbb{R}$ be such that $I \xrightarrow[R_j \to R_j + kR_i]{ii)} E$. Then, we can prove that $\det(E) = 1$ (`Exercise`).

■

Example 5.19 Let us consider the three elementary matrices of Example 4.16. We have

$$\det(E_1) = -1, \quad \det(E_2) = \frac{1}{3} \quad \text{and} \quad \det(E_3) = 1.$$

The following important result tell us how to compute the determinant of the product of two matrices.

Theorem 5.20 (Product Theorem) *Let* $A, B \in \mathcal{M}_{n,n}(\mathbb{R})$. *Then*

$$det(AB) = det(A)\, det(B)\,.$$

Proof. Let us first consider an elementary matrix $E \in \mathcal{M}_{n,n}(\mathbb{R})$. We know, from Section 4.4 that the matrix $B = EA$ represents the matrix A after we apply an elementary row operation and that the matrix $C = AE$ represents the matrix A after we apply an elementary column operation. Combining Theorem 5.9 andand Theorem 5.18 we have that:

- If E is of type $i)$, then $\det(E) = -1$. Moreover, B is obtained from A by interchanging two different rows and the matrix C is obtained from A by interchanging two columns. Thus
$$\det(EA) = \det(B) = -\det(A) = \det(E)\det(A)$$
and
$$\det(AE) = \det(C) = -\det(A) = \det(A)\det(E).$$
- If E is of type $ii)$, then $\det(E) = k$ for a certain number $k \neq 0$. The matrix B is obtained from A by multiplying all elements of a certain row by k, while the matrix C is obtained from A by multiplying all elements of a certain column by k. Thus
$$\det(EA) = \det(B) = k \cdot \det(A) = \det(E)\det(A)$$
and
$$\det(AE) = \det(C) = k \cdot \det(A) = \det(A)\det(E)$$
- If E is of type $iii)$, then $\det(E) = 1$. Moreover B is obtained from A by adding a multiple of some row of A to a different row, while C is obtained from A by adding a multiple of some column of A t a differenc column. Thus
$$\det(EA) = \det(B) = \det(A) = \det(E)\det(A)$$
and
$$\det(AE) = \det(C) = \det(A) = \det(A)\det(E).$$

By Theorem 4.25, we know that there exists two invertible matrices $U, V \in \mathcal{M}_{n,n}(\mathbb{R})$ such that
$$UAV = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix},$$
with $r = \operatorname{rank}(A)$. Moreover, from Section 4.5, we know that there exists elementary matrices $E_1, E_2, \ldots, E_p, \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_q$ such that
$$U = E_1 E_2 \cdots E_p \quad \text{and} \quad V = \widetilde{E}_1 \widetilde{E}_2 \cdots \widetilde{E}_q.$$
Thus one has
$$\begin{aligned} A &= U^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} V^{-1} \\ &= E_p^{-1} \cdots E_2^{-1} E_1^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1}. \end{aligned}$$
Note that $E_1^{-1}, E_2^{-1}, \ldots, E_p^{-1}, \widetilde{E}_1^{-1}, \widetilde{E}_2^{-1}, \ldots, \widetilde{E}_q^{-1}$ are also elementary matrices. If $\operatorname{rank}(A) < n$, then
$$\begin{aligned} \det(AB) &= \det\left(E_p^{-1} \cdots E_2^{-1} E_1^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1} B \right) \\ &= \det\left(E_p^{-1}\right) \cdots \det\left(E_2^{-1}\right) \det\left(E_1^{-1}\right) \det\left(\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1} B \right) \\ &= \det\left(E_p^{-1}\right) \cdots \det\left(E_2^{-1}\right) \det\left(E_1^{-1}\right) \cdot 0 \\ &= 0, \end{aligned}$$

where the third equality follows from Corollary 5.7 and the fact that in the matrix

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1} B$$

the last $n-r$ rows (and the last $n-r$ columns) are zero rows (zero columns).

If $\operatorname{rank}(A) = n$, then

$$\begin{aligned} \det(AB) &= \det\left(E_p^{-1} \cdots E_2^{-1} E_1^{-1} I_n \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1} B\right) \\ &= \det\left(E_p^{-1}\right) \cdots \det\left(E_1^{-1}\right) \det(I_n) \det\left(\widetilde{E}_q^{-1}\right) \cdots \det\left(\widetilde{E}_1^{-1}\right) \det(B) \\ &= \det\left(E_p^{-1} \cdots E_2^{-1} E_1^{-1} I_n \widetilde{E}_q^{-1} \cdots \widetilde{E}_2^{-1} \widetilde{E}_1^{-1}\right) \det(B) \\ &= \det(A) \det(B). \end{aligned}$$

■

Example 5.21 Let us consider the two matrices

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

Then

$$\det(AB) = \det \begin{pmatrix} 7 & 8 \\ 4 & 1 \end{pmatrix} = -25 = -5 \cdot 5 = \det(A) \det(B)$$

We give the following result without proof. However the second point is easy to prove (`Exercise`).

Theorem 5.22 *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. Then*

1. *A is invertible if and only if $det(A) \neq 0$.*
2. *If A is invertible, then $det\left(A^{-1}\right) = \frac{1}{det(A)}$.*

Example 5.23 Let $A, B \in \mathcal{M}_{n,n}(\mathbb{R})$ and let us suppose that

$$\det(A) = 2 \quad \text{and} \quad \det(B) = -3.$$

Using the previous theorems we can compute $\det\left(2A^3B^{-1}A^TB^2\right)$ (`Exercise`).

Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $B \in \mathcal{M}_{p,q}(\mathbb{R})$ be two matrices. We call *block upper triangular matrix* a matrix of the form

$$\begin{pmatrix} A & X \\ O & B \end{pmatrix},$$

where X is is a matrix of size $m \times q$ and $O = O_{p,n}$. A *blok lower triangular matrix* will be defined in a symmetrical way.

Theorem 5.24 *Let A and B be two square matrices, possibly of different size. Then*

$$\det\begin{pmatrix} A & X \\ O & B \end{pmatrix} = det(A)\, det(B)$$

and

$$\det\begin{pmatrix} A & O \\ X & B \end{pmatrix} = det(A)\, det(B)\,.$$

Proof. Let us consider rhe block upper triangular matrix. We can decompose it as

$$\begin{pmatrix} A & X \\ O & B \end{pmatrix} = \begin{pmatrix} I & O \\ O & B \end{pmatrix} \begin{pmatrix} A & X \\ O & I \end{pmatrix}.$$

From the Product Theorem we know that

$$\det\begin{pmatrix} A & X \\ O & B \end{pmatrix} = \det\begin{pmatrix} I & O \\ O & B \end{pmatrix} \det\begin{pmatrix} A & X \\ O & I \end{pmatrix}.$$

By repeated cofactor expansions it is easy to see that

$$\det\begin{pmatrix} I & O \\ O & B \end{pmatrix} = \det(B) \quad \text{and} \quad \det\begin{pmatrix} A & X \\ O & I \end{pmatrix} = \det(A)\,.$$

To prove the result for the block lower triangular matrices we can take the transpose and use Theorem 5.16. ■

Example 5.25 Let

$$A = \begin{pmatrix} 1 & 2 & a & b & c \\ 2 & 5 & d & e & f \\ 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

We have

$$\det(A) = \det\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \det\begin{pmatrix} 2 & -1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} = 1 \cdot 8 = 8.$$

5.4 Cramer's Rule

Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. We define the *adjoint* of A the transpose of the matrix of cofactors:

$$\operatorname{adj}(A) = \left(C_{i,j}(A)\right)^T.$$

Note that $\operatorname{adj}(A)$ is also a $n \times n$-matrix. When it is clear from the context, we will write $C_{i,j}$ instead of $C_{i,j}(A)$.

Example 5.26 Let us consider the matrix

$$A = \begin{pmatrix} 3 & 0 & -1 \\ 4 & 7 & 3 \\ -2 & 8 & 5 \end{pmatrix}.$$

Its adjoint is the matrix:

$$\begin{aligned} \operatorname{adj}(A) &= \begin{pmatrix} C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}^T \\ &= \begin{pmatrix} C_{1,1} & C_{2,1} & C_{3,1} \\ C_{1,2} & C_{2,2} & C_{3,2} \\ C_{1,3} & C_{2,3} & C_{3,3} \end{pmatrix} \\ &= \begin{pmatrix} 11 & -8 & 7 \\ -26 & 13 & -13 \\ 46 & -24 & 21 \end{pmatrix}. \end{aligned}$$

Example 5.27 Let us consider a generic 2×2-matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Its adjoint is the matrix

$$\operatorname{adj}(A) = \begin{pmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{pmatrix}^T = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}^T = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Using the Laplace Expansion Theorem, we can obtain the following result.

Theorem 5.28 (Adjoint Formula) *Let A be a square matrix. Then*

1. $A \cdot adj(A) = det(A)\, I = adj(A) \cdot A$.
2. *If* $det(A) \neq 0$, *then* $A^{-1} = \frac{1}{det(A)} adj(A)$.

Example 5.29 Let us consider the matrix

$$A = \begin{pmatrix} 3 & 0 & -1 \\ 4 & 7 & 3 \\ -2 & 8 & 5 \end{pmatrix}.$$

Its determinant is

$$\det(A) = 3 \det \begin{pmatrix} 7 & 3 \\ 8 & 5 \end{pmatrix} - \det \begin{pmatrix} 4 & 7 \\ -2 & 8 \end{pmatrix} = 3 \cdot 11 - 1 \cdot 46 = -13,$$

while its adjoint is

$$\operatorname{adj}(A) = \begin{pmatrix} 11 & -26 & 46 \\ -8 & 13 & -24 \\ 7 & -13 & 21 \end{pmatrix}^T = \begin{pmatrix} 11 & -8 & 7 \\ -26 & 13 & -13 \\ 46 & -24 & 21 \end{pmatrix}.$$

Thus the inverse of A is the matrix

$$A^{-1} = -\frac{1}{13}\begin{pmatrix} 11 & -8 & 7 \\ -26 & 13 & -13 \\ 46 & -24 & 21 \end{pmatrix}.$$

If A is a $n \times n$-matrix and B is a $n \times 1$-column, then let us set $A_i(B)$ the $n \times n$-matrix obtained from A by replacing the i-column by B.

Theorem 5.30 (Cramer's Rule) *Let us consider the system of linear equations $AX = B$, where A is invertible. If*

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

where $x_1, x_2, \ldots, x_n$ are the variables, then

$$x_i = \frac{det\,(A_i(B))}{det\,(A)} \quad \textit{for each } i = 1, 2, \ldots, n.$$

Example 5.31 Let us consider the system of linear equations

$$\left\{ \begin{array}{rcl} 5x_1 - 7x_2 + 8x_3 & = & 23 \\ 2x_1 + 6x_2 - 9x_3 & = & 61 \\ -x_1 - 4x_2 + 3x_3 & = & 19 \end{array} \right. .$$

The matrix

$$A = \begin{pmatrix} 5 & -7 & 8 \\ 2 & 6 & -9 \\ -1 & -4 & 3 \end{pmatrix}$$

is invertible since its determinant is

$$\det(A) = -127 \neq 0.$$

To find x_2 let us consider the matrix

$$A_2(B) = \begin{pmatrix} 5 & 23 & 8 \\ 2 & 61 & -9 \\ -1 & -19 & 3 \end{pmatrix}.$$

Since $\det(A_2(B)) = 313$, we have

$$x_2 = \frac{\det(A_2(B))}{\det(A)} = -\frac{313}{127}.$$

Example 5.32 Let us consider the system of linear equations

$$\begin{cases} 3x_1 - x_3 & = & 1 \\ 4x_1 + 7x_2 + 3x_3 & = & 0 \\ -2x_1 + 8x_2 + 5x_3 & = & 1 \end{cases}.$$

Using the Cramer's Rule we can find the solution (`Exercise`)

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

5.5 Diagonalization

Let us recall from Section 2.7 that a $n \times n$-matrix D is called a *diagonal matrix* if all its entries off the main diagonal are zeros; that is, fi D has the form

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are numbers. These numbers are not necessarly reals (we will see an example later), so, instead that in $\mathbb{R}$, let us suppose that $\lambda_i \in \mathcal{C}$ for all i.

A n-square matrix A is called *diagonalizable* if there exists an invertible matrix $P \in \mathcal{M}_{n,n}(\mathbb{R})$ such that $P^{-1}AP = D$ is diagonal. In this case, the invertible matrix P is called a *diagonalizing matrix* for A.

Diagonalization is one of the most important ideas inlinear algebra. One of its uses is to give us an efficient method to calculate powers $A, A^2, A^3, \ldots$ of a square matrix A.

Theorem 5.33 *Let A be a diagonalizable matrix. Let us suppose that P is a diagonalizing matrix and $D = P^{-1}AP$. Then, for any $k \in \mathbb{N}$ one has*

$$A^k = PD^kP^{-1}.$$

Proof. (`Exercise`) ■

Example 5.34 Let us diagonalize the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & 0 \end{pmatrix}.$$

By definition, we need to find an invertible matrix P such that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3),$$

for certain numbers $\lambda_1, \lambda_2, \lambda_3$. Let us set

$$P = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix}$$

where

$$X_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad X_2 = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}.$$

This is equivalent to find X_1, X_2 and X_3 such that

$$\begin{aligned} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} &= \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 x_1 & \lambda_2 y_1 & \lambda_3 z_1 \\ \lambda_1 x_2 & \lambda_2 y_2 & \lambda_3 z_2 \\ \lambda_1 x_3 & \lambda_2 y_3 & \lambda_3 x_3 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 X_1 & \lambda_2 X_2 & \lambda_3 X_3 \end{pmatrix}. \end{aligned}$$

Comparing columns, it shows that $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$ if and only if $AX_i = \lambda_i X_i$ for $i = 1, 2, 3$. Moreover, if we want that $P = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix}$ is invertible, we need to make sure that $X_i \neq O$.

In the following, we begin to find λ and $X \neq O$ such that $AX = \lambda X$. This is equivalent to asking that the homogenous linear system

$$(A - \lambda I_3) X = 0 \tag{5.2}$$

has a nontrivial solution $X \neq O$. Using the Gaussian algorithm we reduce the matrix $(A - \lambda I_3)$ into a (reduced) row-echelon form B which is equivalent to left multiplication by a certain invertible matrix, say U, that is we have

$$U (A - \lambda I_3) = B.$$

By the Product Theorem we have

$$\det(U) \det(A - \lambda I_3) = \det(B).$$

Since $\det(U) \neq 0$ (the matrix is invertible), we have

$$\operatorname{rank}(A - \lambda I_3) = \operatorname{rank}(B) \leq n \quad \Leftrightarrow \quad \det(B) \neq 0 \quad \Leftrightarrow \quad \det(A - \lambda I_3) = 0.$$

Then we compute the determinant of $A - \lambda I_3$

$$\begin{aligned} \det(A - \lambda I_3) &= \det \begin{pmatrix} 1-\lambda & 1 & 1 \\ 0 & 2-\lambda & -1 \\ 0 & -3 & -\lambda \end{pmatrix} \\ &= (1-\lambda) \det \begin{pmatrix} 2-\lambda & -1 \\ -3 & -\lambda \end{pmatrix} \\ &= (1-\lambda)\left(-\lambda(2-\lambda) - 3\right) \\ &= (1-\lambda)(\lambda-3)(\lambda+1). \end{aligned}$$

For the equation $\det(A - \lambda I_3)$, we obtain three solutions which are

$$\lambda_1 = 1, \quad \lambda_2 = -1 \quad \text{and} \quad \lambda_3 = 3.$$

Then, we substitute each λ_i into the Equation (5.2) to fin a basic solution for each equation. For example, we solve

$$(A - \lambda_1 I_3) X = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -3 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = O,$$

which is equivalent to solve

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We thus get

$$X = s \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

where s is an arbitrary number. We can use the basic solution

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

which is not a trivial solution as our solution corresponding to $\lambda_1 = 1$. Similarly, we can get

$$X_2 = \begin{pmatrix} -2 & 1 & 3 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} 0 & -1 & 1 \end{pmatrix}$$

corresponding respectively to to $\lambda_2 = -1$ and $\lambda_3 = 3$. Note that here X_1, X_2, X_3 can be arbitrary nonzero solutions corresponding to $\lambda_1, \lambda_2, \lambda_3$.

Thus we can solve the equation

$$P^{-1} A P = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) \quad \Leftrightarrow \quad AP = P \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$$

by obtaining

$$P \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & -1 \\ 0 & 3 & 1 \end{pmatrix},$$

and

$$\operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) = \operatorname{diag}(1, -1, 3).$$

Using the Matrix Inverse Algorithm we can find

$$\begin{aligned}
\begin{pmatrix} 1 & -2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 3 & 1 & 0 & 0 & 1 \end{pmatrix}
&\xrightarrow[\substack{R_1 \to R_1 + 2R_2 \\ R_3 \to R_3 - 3R_2}]{iii)} \begin{pmatrix} 1 & 0 & -2 & 1 & 2 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 4 & 0 & -3 & 1 \end{pmatrix} \\
&\xrightarrow[R_3 \to \frac{1}{4} R_3]{ii)} \begin{pmatrix} 1 & 0 & -2 & 1 & 2 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -\frac{3}{4} & \frac{1}{4} \end{pmatrix} \\
&\xrightarrow[\substack{R_1 \to R_1 + 2R_3 \\ R_2 \to R_2 + R_3}]{iii)} \begin{pmatrix} 1 & 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & -\frac{3}{4} & \frac{1}{4} \end{pmatrix}.
\end{aligned}$$

Thus P is invertible and

$$P^{-1} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & -\frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

In conclusion, we have

$$P^{-1}AP = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & -\frac{3}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & -1 \\ 0 & 3 & 1 \end{pmatrix} = \operatorname{diag}(1, -1, 3).$$

We can generalize the previous example to an $n \times n$-matrix. Finding P such that $P^{-1}AP$ is a diagonal matrix is equivalent to find n column vectors $X_1, X_2, \ldots, X_n$ and n numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$AX_i = \lambda_i X_i \quad \text{for each } \ i = 1, 2, \ldots n.$$

Moreover, if $P = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}$ is invertible, A is diagonalizable.

5.6 Eigenvalues and Eigenvectors

Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. A number λ is called an *eigenvalue* of A if

$$AX = \lambda X$$

for some column $X \neq O$. Such a nonzero column X is called an *eigenvector* of A corresponding to the eigenvalue λ.

Note that the condition $AX = \lambda X$ is automatically satisfied if $X = O$, so the requirement that $X \neq O$ is critical.

The *characteristic polynomia* $c_A(x)$ is defined by

$$c_A(x) = \det(xI - A).$$

A number λ is called a *root* of the characteristic polynomial $c_A(x)$ if $c_A(\lambda) = 0$.

Note that $c_A(\lambda) = 0$ if and only if $-c_A(\lambda) = 0$. For this reason, in the following we will work indifferently with both equations $\det(xI - A) = 0$ and $\det(A - \lambda I) = 0$.

Example 5.35 Let us consider the matrix

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

Its characteristic polynomial is

$$\begin{aligned} c_A(x) &= \det(xI - A) \\ &= \det\left(x\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}5 & -2\\ 4 & -1\end{pmatrix}\right) \\ &= \det\left(\begin{pmatrix}x & 0\\ 0 & x\end{pmatrix} - \begin{pmatrix}5 & -2\\ 4 & -1\end{pmatrix}\right) \\ &= \det\begin{pmatrix}x-5 & 2\\ -4 & x+1\end{pmatrix} \\ &= (x-5)(x+1) - 2(-4) \\ &= x^2 - 4x + 3 \\ &= (x-1)(x-3). \end{aligned}$$

The two roots of the characteristic polynomial are thus $\lambda_1 = 1$ and $\lambda_2 = 3$.

Theorem 5.36 *Let A be a $n \times n$-matrix.*

1. *The eigenvalues of A are the roots of the characteristic polynomial $c_A(x)$ of A.*
2. *The eigenvectors X corresponding to the eigenvalues λ are the nonzero solutions to the homogenous system of linear equations $(\lambda I - A)X = O$.*

Note that there are many eigenvectors of a square matrix A associated with a given eigenvalue λ. In fact every nonzero solution X of $(\lambda I - A)X = O$ is an eigenvector. Of course the eigenvalue λ is chosen so that there must be nonzero solutions.

The eigenvalues of a real matrix need not to be real numbers.

Example 5.37 Let us find the eigenvalues of the matrix

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

The characteristic polynomial of the matrix is $\det(xI - A) = x^2 + 1$. So by Theorem 5.36, the eigenvalues of A are the nonreal complex roots $\lambda_1 = i$ and $\lambda_2 = -i$.

A $n \times n$-matrix has n (possibly complex) eigenvalues, but they may not be distinct.

Example 5.38 Let us find the eigenvalues of the matrix

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

Its characteristic polynomial is $c_A(x) = (x-1)^2$. So there is only one eigenvalue of A, namely $\lambda_1 = 1$. However, λ_1 is a double root of $c_A(x)$ and we say that $\lambda_1 = 1$ has *multiplicity* 2.

The following result illustrate the previous example

Theorem 5.39 *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$.*

1. *A is diagonalizable if and only if it has eigenvectors $X_1, X_2, \ldots, X_n$ such that the matrix*
$$P = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}$$
is invertible.

2. *When this is the case, we have*
$$P^{-1}AP = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$$
where, for each i, λ_i is the eigenvalue of A corresponding to X_i.

Example 5.40 Let us show that the matrix
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$
is not diagonalizable.

We know from Example 5.38 that A has only one eigenvalue $\lambda_1 = 1$, which is of multiplicity 2. But the system of linear equations $(\lambda_1 I - A)X = O$ has general solution
$$X = s\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$
so there is only one basic solution:
$$X_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$
Hence we can only choose
$$P = \begin{pmatrix} s & t \\ 0 & 0 \end{pmatrix}$$
which is never invertible no matter the choice of s and t.

5.7 The Diagonalization Algorithm

In this section we give an altorithm to diagonalize a square matrix.

Diagonalization Algorithm. *Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$ be a square matrix. To diagonalize A we apply the following steps:*

Step 1. *Find all the eigenvalues of A, which are the roots of the characteristic polynomial $c_A(x)$;*

Step 2. *For each eigenvalue λ compute an eigenvector, by finding the basic solution of the homogenous system $(\lambda I - A)X = O$;*

Step 3. *The matrix A is diagonalizable if and only if there are n basic eigenvectors in total;*

Step 4. *If A is diagonalizable, the $n \times n$-matrix P having these eigenvectors as columns is a diagonalizing matrix for A; that is, P is invertible and $P^{-1}AP$ is diagonal.*

Example 5.41 Let us apply the previous algorithm to the matrix

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

Step 1. Let us first compute the characteristic polynomial

$$\begin{aligned} c_A(x) &= \det(xI - A) \\ &= \det \begin{pmatrix} x & -1 & -1 \\ -1 & x & -1 \\ -1 & -1 & x \end{pmatrix} \\ &= \det \begin{pmatrix} x-2 & x-2 & x-2 \\ -1 & x & -1 \\ -1 & -1 & x \end{pmatrix} \\ &= \det \begin{pmatrix} x-2 & 0 & 0 \\ -1 & x+1 & 0 \\ -1 & 0 & x+1 \end{pmatrix} \\ &= (x-2)(x+1)^2, \end{aligned}$$

where to compute the determinant we first added the second and the third row to the first row, and then we subtracted the first column from the second and from the third column.

Hence, the equation $c_A(x) = 0$ has two solutions: $\lambda_1 = 2$ and $\lambda_2 = -1$, with the last one having multiplicity two.

Step 2. For $\lambda_1 = 2$, the system

$$(\lambda_1 I - A)X = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} X = O$$

solution

$$X = s \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

where t is an arbitrary number. So the basic solution

$$X_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = 2$.

For $\lambda_2 = -1$, the system

$$(\lambda_2 I - A)X = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} X = O$$

has general solution

$$X = s\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

where s and t are arbitrary numbers. Hence there are two basic solutions

$$X_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

corresponding to $\lambda_2 = -1$.

Step 3. Since there are three eigenvectors, X_1, X_2 and X_3, we can deduce that A is diagonalizable.

Step 4. If we take

$$P = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

we find that P is invertible and

$$P^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Thus

$$P^{-1}AP = \operatorname{diag}(2, -1, -1) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In a general case, an eigenvalue λ of a square matrix A is said to have multiplicity m if it occurs m times as a root of the characteristic polynomial $c_A(x)$. When the homogenous system $(\lambda I - A)X = O$ is solved, any set of basic solutions is called a set of *basic eigenvectors* corresponding to λ. Here the number of basic eigenvectors equals the number of parameters involved in the solution of the system $(\lambda I - A)X = O$.

Theorem 5.42 *A square matrix A is diagonalizable if and only if it the multiplicity of every eigenvalue λ of A equals the number of basic eigenvectors corresponding to λ (which is the number of parameters in the solution of $(\lambda I - A)X = O$).*

In this case, the basic solutions of the system $(\lambda I - A)X = O$ become columns in the invertible diagonalizing matrix P such that $P^{-1}AP$ is diagonal.

Since for each eigenvalues there is at least a basic eigenvector, we have the following immediate consequence of the previous theorem.

Corollary 5.43 *If A is a $n \times n$-matrix with n distinct eigenvalues, then A is diagonalizable.*

A good example which illustrate an application of diagonalization is given in the following example.

Example 5.44 Let us compute A^{100} for

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & 0 \end{pmatrix}.$$

As we have already seen in Example 5.34, the matrix A has eigenvalues

$$\lambda_1 = 1, \quad \lambda_2 = -1 \quad \text{and} \quad \lambda_3 = 3,$$

with corresponding eigenvectors

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

A diagonalizing matrix for A is thus given by the invertible matrix

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

having inverse

$$P^{-1} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & -\frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

We thus have

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix} P^{-1}.$$

Thus, using Theorem 5.33, we have

$$\begin{aligned}
A^{100} &= P\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 3\end{pmatrix}^{100} P^{-1}\\
&= P\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 3^{100}\end{pmatrix} P^{-1}\\
&= \begin{pmatrix}1 & -2 & 0\\ 0 & 1 & -1\\ 0 & 3 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 3^{100}\end{pmatrix}\begin{pmatrix}1 & \frac{1}{2} & \frac{1}{2}\\ 0 & \frac{1}{4} & \frac{1}{4}\\ 0 & -\frac{3}{4} & \frac{1}{4}\end{pmatrix}\\
&= \begin{pmatrix}1 & 0 & 0\\ 0 & \frac{1+3^{101}}{4} & \frac{1-3^{100}}{4}\\ 0 & \frac{3-3^{101}}{4} & \frac{3+3^{100}}{4}\end{pmatrix}.
\end{aligned}$$

Example 5.45 Let us consider the matrix

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

We can compute A^{20} (to do so we need first to diagonalize the matrix). (`Exercise`)

5.8 Similar matrices

Let us consider two square matrices A and B of the same size. We say that A and B are *similar* if

$$B = P^{-1}AP$$

for some invertible matrix P. When this is the case, we write $A \sim B$.

Using this terminology, we can say that a square matrix A is diagonalizable if and only if it is similar to a diagonal matrix.

Here are some simply properties of similarity.

Proposition 5.46 *Let* $A, B, C \in \mathcal{M}_{n,n}(\mathbb{R})$.

1. $A \sim A$ *for all square matrix* A.
2. *If* $A \sim B$ *then* $B \sim A$.
3. *If* $A \sim B$ *and* $B \sim C$, *then* $A \sim C$.

Proof.

- The first point is clear since $A = I^{-1}AI$, and I is invertible.
- If $A \sim B$ then there exists an invertible matrix P such that $B = P^{-1}AP$. Thus $A = PBP^{-1}$, with P^{-1} invertible. That is, $B \sim A$.

- Let P and Q be two invertible matrices such that $B = P^{-1}AP$ and $C = Q^{-1}BQ$. Thus

  $$C = Q^{-1}\left(P^{-1}AP\right)Q = \left(Q^{-1}P^{-1}\right)A\,(PQ) = (PQ)^{-1}\,A\,(PQ)\,,$$

  with PQ invertible. Hence $A \sim C$.

■

The properties in the previous proposition are often expressed by saying that the similarity relation $\sim$ is an *equivalente relation* on the set of $n \times n$-matrices.

Proposition 5.47 *Let A, B be two square matrices such that $A \sim B$. Then*

1. $A^{-1} \sim B^{-1}$.
2. $A^T \sim B^T$.
3. $A^k \sim B^k$ *for all* $k \geq 0$.

Example 5.48 Let A, B be two square matrices such that $A \sim B$. If A is diagonalizable, then B is also diagonalizable. (`Exercise`)

Following the previous example, it is possible to prove that if A is diagonalizable, then so are also the matrices A^T, A^{-1} (if it exists) and A^k for all $k \geq 0$.

The following theorem easily follows from the Product Theorem and the Diagonalization Algorithm.

Theorem 5.49 *Let A, B be two similar matrices. Then*

1. $det(A) = det(B)$.
2. $c_A(x) = c_B(x)$.
3. *A and B have the same eigenvalues.*

Example 5.50 Let

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

Then $c_A(A) = 0$.

Indeed, we saw in Example 5.41 that the characteristic polynomial is

$$c_A(x) = (x-2)(x+1)^2$$

When we *evaluate* this polynomial at A, we obtain $c_A(A) = (A - 2I)(A + I)^2$.

Let us prove that this evaluation equals wero. Recall from Example 5.41 that $A = PDP^{-1}$ with

$$P = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

So

$$\begin{aligned}
c_A(A) &= (A - 2I)\,(A + I)^2 \\
&= \left(PDP^{-1} - 2I\right)\left(PDP^{-1} + I\right)^2 \\
&= \left(PDP^{-1} - 2PIP^{-1}\right)\left(PDP^{-1} + PIP^{-1}\right)^2 \\
&= \left(P(D - 2I)P^{-1}\right)\left(P(D + I)P^{-1}\right)^2 \\
&= P(D - 2I)P^{-1}P(D + I)P^{-1}P(D + I)P^{-1} \\
&= P(D - 2I)(D + I)(D + I)P^{-1} \\
&= P \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^2 P^{-1} \\
&= P \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 9 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^{-1} \\
&= P \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^{-1} \\
&= 0.
\end{aligned}$$

We can generalize the previous example in the following importan theorem

Theorem 5.51 (Cayley-Hamilton Theorem) *Let A be a square matrix. Thus $c_A(A) = 0$.*