Chapter 2

Matrices and linear equations

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

2.1 Matrices

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

We also define a product for matrices:

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

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

Examples 2.1.8. *1.*

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

2.

$$ \underbrace{\begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}}_{\in M_{1n}(\mathbb{R})} \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}}_{\in M_{n1}(\mathbb{R})} = a_1b_1 + a_2b_2 + \dots + a_nb_n \ \in M_{11}(\mathbb{R}) $$

3.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(ii) One observes that

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

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

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

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

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

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

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

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

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

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

More generally, one has

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

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

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

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

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

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

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

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

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

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

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

2.3 Homogeneous linear equations

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

Example 2.3.1. *Let us consider the equation*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 2.3.4. *Let*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.4 Row operations and Gauss elimination

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

from which one deduces the new system of equations

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

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

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

Based on this example, let us formalize the procedure.

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

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

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

(iii) interchanging two rows.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

we consider the augmented matrix and some elementary row operations:

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

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

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

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

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

2.5 Elementary matrices

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

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

These matrices satisfy the relation

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

See Exercise 2.21 for the proof of this statement.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which proves the statement. □

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

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

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

2.6 Exercises

Exercise 2.1. *Let us consider*

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

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

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

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

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

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

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

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

Exercise 2.7.
1. *Find some* $\mathcal{A} \in M_2(\mathbb{R})$ *such that* $\mathcal{A}^2 = -\mathbf{1}_2$.
2. *Determine all* $\mathcal{A} \in M_2(\mathbb{R})$ *such that* $\mathcal{A}^2 = \mathcal{O}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

Draw a picture of your results.

Exercise 2.12. *Let*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1. $\mathbf{1}_m - I_{rr} + cI_{rr}$, *the matrix obtained from the identity matrix by multiplying the* r*-th diagonal component by* c,

2. *For* $r \neq s$, $(\mathbf{1}_m + I_{rs} + I_{sr} - I_{rr} - I_{ss})$, *the matrix obtained from the identity matrix by interchanging the* r*-th row with the* s*-th row,*

3. *For* $r \neq s$, $(\mathbf{1}_m + cI_{rs})$, *the matrix having the* rs*-th component equal to* c, *all other components* 0 *except the diagonal components which are equal to* 1.

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

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

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

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

Exercise 2.24. *Consider the equation*

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

where k is an arbitrary constant.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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