Lecture Notes on Mathematics I for the Bachelor's in Economics, in Finance, and in Management

Introduction to Differential and Integral Calculus in $\mathbb{R}$
and to Linear Algebra

Telmo Peixe

Mathematics Department

November 2025

Contents

Introduction

These notes were written for the Mathematics I course in the Bachelor's degrees in Economics, in Finance, and in Management at ISEG - Lisbon School of Economics & Management, Universidade de Lisboa, with the aim of serving as a study aid for students in these programmes.

However, for a thorough understanding of the material, these notes should be complemented with fundamental bibliographic references, such as those listed in the References (in the end of this lecture notes).

The content is organised into two main parts: Part I on **Mathematical Analysis** and Part II on **Linear Algebra**. Part I consists of the following chapters: 1 Topology, induction, and geometric series; 2 Real functions of a real variable; 3 Limits and continuity; 4 Differential calculus; 5 Integral calculus. Part II consists of the following chapters: 6 Vectors; 7 Matrices; 8 Determinants; 9 Systems of linear equations.

At the end, an index is provided to assist in locating terms and concepts throughout the text.

These notes should be regarded as a work in progress. Any corrections of typos, or suggestions for improvement, are most welcome.

This english version of the lecture notes is an adaptation of the original notes written in Portuguese by *Filipa de Carvalho* and *Telmo Peixe* [4, 5]. While the content remains largely faithful to the original, some modifications and adjustments have been made to ensure clarity and coherence.

Part I

Mathematical Analysis

Chapter 1

Topology, induction and geometrical series

In this chapter, we will study three topics that may initially seem unrelated. We begin by exploring topology in the set of real numbers. Next, we provide a brief introduction to numerical series, with a particular focus on the geometric series. Finally, we examine an important method for proving mathematical statements: the Principle of Mathematical Induction.

1.1 Topology in $\mathbb{R}$

Before delving into the fundamental concepts of topology, let us first review some general notions of order relations in $\mathbb{R}$.

1.1.1 Preliminary concepts

Consider a non-empty subset $A \subset \mathbb{R}$. We define:

- $M \in \mathbb{R}$ as an **upper bound** of A if $x \leq M$ for any $x \in A$;
- $m \in \mathbb{R}$ as a **lower bound** of A if $m \leq x$ for any $x \in A$;
- A as a **bounded above set** if A has upper bounds;
- A as a **bounded below set** if A has lower bounds;
- if A is bounded above, the smallest upper bound of A is called the **supremum** of A and is denoted by $L = \sup A$;
- if A is bounded below, the greatest lower bound of A is called the **infimum** of A and is denoted by $l = \inf A$.

Remark 1.1.1. *Note that the supremum and/or infimum of a non-empty subset of $\mathbb{R}$ may or may not belong to the set itself. In this regard, consider the following definitions.*

- If $L = \sup A$ belongs to the set A, then L is called the **maximum** of the set A and is denoted by $L = \max A$;
- If $l = \inf A$ belongs to the set A, then l is called the **minimum** of the set A and is denoted by $l = \min A$.

Example 1.1.2. *Let $A = [1, 5[$. Then:*

- *The upper bounds of A are $[5, +\infty[$, as for any $M \in [5, +\infty[$, it holds that $x \leq M$ for all $x \in A$;*
- *The lower bounds of A are $]-\infty, 1]$, as for any $m \in]-\infty, 1]$, it holds that $m \leq x$ for all $x \in A$;*
- $5 = \sup A$, *since it is the smallest of all upper bounds of A;*
- $1 = \inf A$, *since it is the greatest of all lower bounds of A;*
- *Since* $\sup A = 5 \notin A$, *the set A has no maximum;*
- *Since* $\inf A = 1 \in A$, *we have* $\min A = 1$.

Theorem 1.1.3 (Supremum and Infimum Principle). *Let A be a non-empty subset of $\mathbb{R}$. If A is bounded above (resp., bounded below), then it admits a supremum (resp., infimum).*

Proof. See [10]. □

Definition 1.1.4. *Given a non-empty subset A of $\mathbb{R}$, we define:*

- *A as **bounded above** if it has upper bounds;*
- *A as **bounded below** if it has lower bounds;*
- *A as **bounded** if it is both bounded above and below, i.e., there exist m and M, respectively a lower bound and an upper bound of A, such that*

  $$m \leq x \leq M, \quad \textit{for any } x \in A,$$

  or equivalently,

  $$A \subset [m, M].$$

1.1.2 Definitions - Topology

Definition 1.1.5. *Let $a, \varepsilon \in \mathbb{R}$, with $\varepsilon > 0$. The **neighbourhood** of center a and radius ε, denoted by $\mathcal{B}_\varepsilon(a)$, is the open interval $]a-\varepsilon, a+\varepsilon[$.*

Remark 1.1.6. *Observe that*

$$\begin{aligned} x \in \mathcal{B}_\varepsilon(a) =]a-\varepsilon, a+\varepsilon[\quad &\Leftrightarrow \quad a-\varepsilon < x < a+\varepsilon \\ &\Leftrightarrow \quad x < a+\varepsilon \ \wedge \ x > a-\varepsilon \\ &\Leftrightarrow \quad x-a < \varepsilon \ \wedge \ x-a > -\varepsilon \\ &\Leftrightarrow \quad |x-a| < \varepsilon \,, \end{aligned}$$

thus, $\mathcal{B}_\varepsilon(a)$ represents the points whose distance to the point $a \in \mathbb{R}$ is less than ε.

Let A be a subset of $\mathbb{R}$. A point $a \in \mathbb{R}$ is said to be:

- **interior** to the set A if there exists a neighbourhood $\mathcal{B}_\varepsilon(a) \subseteq A$;
- **exterior** to the set A if there exists a neighbourhood $\mathcal{B}_\varepsilon(a)$ such that $\mathcal{B}_\varepsilon(a) \cap A = \emptyset$;
- **boundary** of the set A if it is neither interior nor exterior to A, i.e., for any neighbourhood $\mathcal{B}_\varepsilon(a)$, we have $\mathcal{B}_\varepsilon(a) \cap A \neq \emptyset$ and $\mathcal{B}_\varepsilon(a) \cap (\mathbb{R} \setminus A) \neq \emptyset$;
- **accumulation point** of A if for any neighbourhood $\mathcal{B}_\varepsilon(a)$, we have $(\mathcal{B}_\varepsilon(a) \setminus \{a\}) \cap A \neq \emptyset$;
- **isolated point** of A if a belongs to A and there exists a neighbourhood $\mathcal{B}_\varepsilon(a)$ such that $(\mathcal{B}_\varepsilon(a) \setminus \{a\}) \cap A = \emptyset$.

The set of interior points of A is called the **interior** of A and is denoted by $int(A)$; the set of exterior points of A is called the **exterior** of A and is denoted by $ext(A)$; the set of boundary points of A is called the **boundary** of A and is denoted by $bd(A)$; and the set of accumulation points of A is called the **derived set** of A and is denoted by A'.

Remark 1.1.7. *The sets $int(A)$, $ext(A)$, and $bd(A)$ are pairwise disjoint, and*

$$int(A) \cup ext(A) \cup bd(A) = \mathbb{R}.$$

Example 1.1.8. *1. Let $a, b \in \mathbb{R}$ such that $a < b$, and consider $A =]a, b[$. Then*

$$int(A) = A\,, \quad ext(A) = \mathbb{R} \setminus [a,b]\,, \quad bd(A) = \{a, b\},\,, \quad A' = [a,b].$$

The set A has no isolated points.

2. *Let n distinct points, $x_1, x_2, \dots, x_n \in \mathbb{R}$, and consider*

$$B = \{x_1, x_2, \dots, x_n\}.$$

   Then

$$int(B) = \emptyset\,, \quad ext(B) = \mathbb{R} \setminus B\,, \quad bd(B) = B \quad \text{and} \quad B' = \emptyset.$$

   The points of B are all isolated points.

3. *For the set $\mathbb{Q}$ of rational numbers, we have*

$$int(\mathbb{Q}) = \emptyset\,, \quad ext(\mathbb{Q}) = \emptyset\,, \quad bd(\mathbb{Q}) = \mathbb{R} \quad \text{and} \quad \mathbb{Q}' = \mathbb{R}.$$

   The set $\mathbb{Q}$ has no isolated points.

Consider A a subset of $\mathbb{R}$.

- The set A is called **open** if $A = int(A)$.
- The **closure** of A, denoted by $cl(A)$, is the union of its interior with its boundary, i.e., $cl(A) = int(A) \cup bd(A)$.
- The set A is called **closed** if $A = cl(A)$.
- The set A is called **compact** if A is closed and bounded.

Proposition 1.1.9. *For any subset A of $\mathbb{R}$, we have*

$$cl(A) = A' \cup \{\text{isolated points of } A\}.$$

Remark 1.1.10. *The notions of open and closed sets are not mutually exclusive, i.e., some sets can be both open and closed, while others are neither. For example, $\mathbb{Q}$ is neither open nor closed.*

Exercise 1.1.11. *Show that the sets $\emptyset$ and $\mathbb{R}$ are both open and closed.*

Example 1.1.12. *Let $a, b \in \mathbb{R}$ such that $a < b$. We have:*

- *$A = [a, b]$ is closed and bounded, hence compact;*
- *$B = [b, +\infty[$ is closed but not bounded, hence not compact;*
- *$C = [a, b[$ is bounded but not closed, hence not compact.*

1.2 Mathematical Induction

Imagine that you have an infinite number of domino pieces standing vertically in a sequential manner. If it can be proven that

1. the first piece falls;
2. if any given piece falls, then the next piece also falls,

then it is proven that all (infinitely many) domino pieces will fall. This represents one of the most elementary mathematical strategies, known as the **mathematical induction**, used to prove statements formulated in terms of the natural number n.

As illustrated with the domino pieces, the principle of mathematical induction essentially consists of two steps. Suppose we aim to prove that a property $P(n)$, expressed in terms of the natural number n, is valid for all natural numbers n, starting from a given natural number p. By proving that:

1. the property holds for the first natural number p;
2. if the property holds for any given natural number $k \geq p$ (**induction hypothesis**), then it also holds for the next natural number, $k+1$ (**induction step**),

it follows that the property $P(n)$ is valid for all natural numbers $n \geq p$. Consider the following examples.

Example 1.2.1. *We prove by mathematical induction that for any natural number n, the sum of the first n odd natural numbers is n^2, that is,*

$$P(n) : 1 + 3 + 5 + \ldots + (2n - 1) = n^2\,, \quad \forall n \in \mathbb{N}\,.$$

1. *The property $P(n)$ holds for the first natural number $n = 1$, since $1 = 1^2$.*
2. *Now, we show that if the property holds for some $k \geq 1$, then it also holds for the next natural number, i.e., for $k+1$. In other words, assuming that:*

(I.H.) Induction hypothesis:

$$1 + 3 + 5 + \ldots + (2k - 1) = k^2,$$

for some natural number $k \geq 1$,

we want to prove that

(I.S.) Induction step:

$$1 + 3 + 5 + \ldots + (2k-1) + (2(k+1)-1) = (k+1)^2.$$

Now, we have

$$\begin{aligned} \underbrace{1 + 3 + 5 + \ldots + (2k-1)}_{=k^2} + (2(k+1)-1) &\overset{(I.H.)}{=} k^2 + (2k+2-1) \\ &= k^2 + 2k + 1 \\ &= (k+1)^2\,, \end{aligned}$$

which completes the proof.

Example 1.2.2. *We prove by mathematical induction that for any natural number n,*

$$P(n) : \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^n} = 1 - \frac{1}{2^n}\,.$$

1. *The property $P(n)$ holds for the first natural number $n = 1$, since $\frac{1}{2} = 1 - \frac{1}{2^1}$.*
2. *Now, we show that if the property holds for some $k \geq 1$, then it also holds for the next natural number, i.e., for $k+1$. Assuming that*

(I.H.) Induction hypothesis:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^k} = 1 - \frac{1}{2^k},$$

for some natural number $k \geq 1$,

we want to prove that

(I.S.) Induction step:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^k} + \frac{1}{2^{k+1}} = 1 - \frac{1}{2^{k+1}}.$$

Now, we have

$$\begin{aligned} \underbrace{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^k}}_{=1-\frac{1}{2^k}} + \frac{1}{2^{k+1}} &\overset{(I.H.)}{=} \left(1 - \frac{1}{2^k}\right) + \frac{1}{2^{k+1}} \\ &= 1 - \frac{1}{2^k}\left(1 - \frac{1}{2}\right) \\ &= 1 - \frac{1}{2^{k+1}}\,, \end{aligned}$$

which completes the proof.

1.3 Sequences and the Geometric Series

Let us begin this section by recalling general concepts about sequences and studying some important examples.

1.3.1 General Definitions

A **sequence in** $\mathbb{R}$ or **real sequence** is a function

$$\begin{aligned} u : \mathbb{N} &\longrightarrow \mathbb{R} \\ n &\longmapsto u(n) \end{aligned}$$

usually represented, for example, by $(u_n)_{n\in\mathbb{N}}$ or simply by u_n, where the term u_k represents the element of the sequence of order k, and the set

$$\{u_1, u_2, u_3, \ldots, u_k, \ldots\}$$

is called the **set of sequence terms**.

A sequence can be defined in various ways. For example:

- using a formula, which is called the **general term of the sequence**;
- or by **recurrence**, which consists of defining one or more initial terms, and a rule for obtaining a term from previous ones.

Example 1.3.1.

- *The set of terms of the sequence* u_n *with general term* $u_n = 2n + 3$ *is* $\{5, 7, 9, 11, 13, \ldots\}$.
- *The set of terms of the sequence* v_n *defined recursively by*

$$\begin{cases} v_1 = 5 \\ v_{n+1} = v_n + 2 \end{cases},$$

is $\{5, 7, 9, 11, 13, \ldots\}$.

Definition 1.3.2 (**Limit, Convergent and Divergent Sequence**)**.** *A sequence* u_n *is said to have a* ***limit*** L *if, for any real number* $\varepsilon > 0$ *there exists an order* $m \in \mathbb{N}$ *such that, for all* $n \geq m$*,*

$$|u_n - L| < \varepsilon\,.$$

*In this case, the sequence u_n is said to be **convergent**, and that it converges to L, written as*

$$\lim_{n\to\infty} u_n = L\,, \quad \textit{or} \quad \lim u_n = L\,, \quad \textit{or simply} \quad u_n \xrightarrow[n\to\infty]{} L\,.$$

*A sequence that is not convergent is said to be **divergent**.*

Theorem 1.3.3. *Let A be a non-empty subset of $\mathbb{R}$. A point $c \in \mathbb{R}$ is an accumulation point of A if and only if c is the limit of a sequence of points of A distinct from c.*

Definition 1.3.4 (**Bounded Sequence**)**.** *A sequence is said to be **bounded** if the set of its terms is a bounded set.*

Theorem 1.3.5. *The product of a bounded sequence by an infinitesimal is an infinitesimal.*

1.3.2 Topology of the Set $A = \{\frac{1}{n} : n \in \mathbb{N}\}$

Consider the sequence $u_n = \frac{1}{n}$. Then $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots\}$ is the set of terms of the sequence u_n. It follows that:

1. $int(A) = \emptyset \neq A$, hence A is not an open set;
2. $ext(A) = \mathbb{R} \setminus (A \cup \{0\})$;
3. $bd(A) = A \cup \{0\}$;
4. $cl(A) = int(A) \cup bd(A) = A \cup \{0\} \neq A$, hence A is not a closed set. Therefore, it is also not compact, although it is bounded, as $A \subset [0, 1]$, for example;
5. $A' = \{0\}$, meaning that the set A has a unique accumulation point, which is 0 (noting that Theorem 1.3.3 applies, as the terms of the sequence u_n belong to set A and $\lim\limits_{n\to\infty} u_n = 0$);
6. all elements of A are isolated points.

1.3.3 Study of the Sequence with General Term $u_n = r^n$

Let $r \in \mathbb{R}$. The sequence $u_n = r^n$ is called a **geometric progression** with ratio r, since for any $n \in \mathbb{N}$, $\frac{u_{n+1}}{u_n} = r$.

Regarding the parameter r, we have the following cases:

- if $r > 1$, then $\lim\limits_{n\to\infty} u_n = +\infty$, hence u_n is divergent;
- if $r = 1$, then $u_n = 1$, and the sequence is constant, thus u_n is convergent;
- if $0 < r < 1$, then $\lim\limits_{n\to\infty} u_n = 0$, hence u_n is convergent;
- if $r = 0$, then $u_n = 0$ is a constant sequence, hence u_n is convergent;
- if $-1 < r < 0$, then $\lim\limits_{n\to\infty} u_n = 0$ (alternating between negative and positive values), hence u_n is convergent;
- if $r = -1$, then $u_n = (-1)^n$ alternates between the terms -1 and 1, hence u_n is divergent;
- if $r < -1$, then $\lim\limits_{n\to\infty} u_n$ does not exist, as positive terms tend to $+\infty$ and negative terms tend to $-\infty$.

Example 1.3.6.

- *The sequence whose terms are* $\{1, 3, 9, 27, 81, \dots\}$ *is a geometric progression with ratio* $r = 3$ *and can be defined as*

$$\begin{cases} u_1 = 1 \\ u_{n+1} = 3u_n \end{cases}.$$

- *The sequence whose terms are* $\left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots\right\}$ *is a geometric progression with ratio* $r = \frac{1}{2}$ *and can be defined as*

$$\begin{cases} u_1 = 1 \\ u_{n+1} = \frac{1}{2}u_n \end{cases}.$$

Proposition 1.3.7. *For each* $n \in \mathbb{N}_0$*, we have* $u_n = r^{n-1}u_1$*.*

Proof. Exercise. □

Proposition 1.3.8. *The sum of the first terms of the geometric progression* $u_n = r^n$ *is*

$$S_n = u_1 \frac{1 - r^n}{1 - r}.$$

Proof. Exercise. □

1.3.4 Zeno's Paradox

Zeno of Elea (495 BC ? – 435 BC ?) was a Greek philosopher who liked to challenge thinkers of his time by formulating various paradoxes. One of these paradoxes, known as **Zeno's paradox** or **the Runner's paradox**, states that an athlete can never reach the finish line in a race because he always has to run half of each remaining distance before completing the total distance. This means that, having run the first half, he still has to run the second half. When he has run half of this, he still has a quarter of the total left. When he has run half of this quarter, he still has an eighth of the initial distance left, and so on indefinitely.

These fractions subdivide the total course into an infinite number of ever smaller segments (as represented in Figure 1.1).

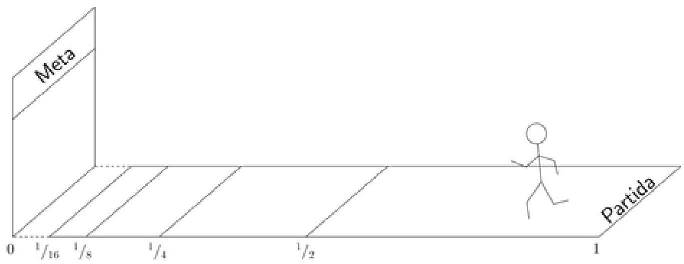

Figure 1.1: Representation of the subdivision of the total course of a race into an infinite number of increasingly smaller segments, in the proportion of $^1/_2$.

To traverse each segment, a certain amount of time is required, and the total time needed to run the entire course is the sum of the times spent on each partial interval.

Saying that the runner never reaches the finish line means that he cannot reach that point within a finite time interval, that is, that the sum of an infinite number of positive values (times spent on each interval) cannot be finite.

The theory of infinite series, developed mainly in the 16th-18th centuries, helped clarify such problems, as exemplified by Zeno's paradox. Another relatively well-known paradox is that of **Achilles and the Tortoise**: the tortoise, in a race against Achilles, starts with an initial advantage. It is impossible for Achilles to catch the tortoise because when Achilles reaches the tortoise's initial position, it has already moved to B. When Achilles reaches B, the tortoise has moved further to C, and so on indefinitely.

In reality, we know that the situations described in both Zeno's paradox and Achilles and the Tortoise do not hold, which is why they are called paradoxes. In a race, in general, the runner reaches the finish line, just as a faster runner can overtake a slower one, even if the latter started ahead.

Mathematicians began to consider that it might be possible to generalise the idea of ordinary addition of finite sets of numbers to infinite sets, so that, under certain conditions, the sum of an infinite set of numbers could be finite.

Suppose the runner moves at a constant speed and requires T minutes to run the first half of the course. In the next quarter, he will need $\frac{T}{2}$ minutes, in the next eighth $\frac{T}{4}$ minutes, and so on. Thus, the total time required to complete the course is

$$T + \frac{T}{2} + \frac{T}{4} + \frac{T}{8} + \frac{T}{16} + \ldots$$

This is an example of what is called a **numerical series**, and the problem is to determine whether there is a number that can represent this infinite sum or whether, on the contrary, the sum diverges to infinity.

Experience shows that, in the case of a runner moving at a constant speed, he will reach the finish line after $2T$ minutes. Is it then true that

$$T + \frac{T}{2} + \frac{T}{4} + \frac{T}{8} + \frac{T}{16} + \cdots = 2T \ ?$$

The theory of infinite series tells us how to interpret this equality. One approach is:

1st) Summing a finite number of terms, starting with:

- the first term,

$$S_1 = T = \left(2 - \frac{1}{2^0}\right) T \, ;$$

- the first plus the second,

$$S_2 = T + \frac{T}{2} = \frac{3}{2}T = \left(2 - \frac{1}{2^1}\right) T \, ;$$

- the first three terms,

$$S_3 = T + \frac{T}{2} + \frac{T}{4} = \frac{7}{4}T = \left(2 - \frac{1}{2^2}\right) T \, ;$$

- the first four terms,

$$S_4 = T + \frac{T}{2} + \frac{T}{4} + \frac{T}{8} = \frac{15}{8}T = \left(2 - \frac{1}{2^3}\right) T \, ;$$

- ...

– up to the n-th term,

$$S_n = \left(2 - \frac{1}{2^{n-1}}\right) T \,.$$

2nd) Taking the limit as $n \to \infty$, we obtain $S_n \xrightarrow[n\to\infty]{} 2T$.

1.3.5 General Definitions of Series

Definition 1.3.9 (**Numerical Series**)**.** *Given a sequence* $(u_n)_{n\in\mathbb{N}}$ *of real numbers, the sum*

$$u_1 + u_2 + ... + u_n + ...$$

obtained by adding all the (infinite) terms of the sequence u_n *is called a* ***numerical series*** *with general term* u_n*. The numerical series with general term* u_n *can be abbreviated as*

$$\sum_{n\geq 1} u_n \,, \quad \sum_{n\in\mathbb{N}} u_n \,, \quad or \quad \sum_{n=1}^{+\infty} u_n \,.$$

Definition 1.3.10. *Given a numerical series* $\sum_{n\geq 1} u_n$ *with general term* u_n*, the sequence*

$$S_n = u_1 + u_2 + ... + u_n$$

defined by the sum of the first n terms of u_n *is called the* ***sequence of partial sums*** *of the series. The series* $\sum_{n\geq 1} u_n$ *is said to be* ***convergent*** *if the sequence* S_n *of partial sums is convergent. In this case,*

$$\sum_{n\geq 1} u_n = \lim S_n = S,$$

where the limit S *of the sequence of partial sums is called the* ***sum*** *of the series. If the sequence* S_n *of partial sums is divergent, the series* $\sum_{n\geq 1} u_n$ *is said to be* ***divergent****.*

Proposition 1.3.11 (**Necessary Condition for Convergence**)**.** *If the series* $\sum_{n=1}^{+\infty} u_n$ *is convergent, then* $\lim_{n\to\infty} u_n = 0$.

Proof. See [10]. □

Remark 1.3.12.

The condition $\lim_{n\to\infty} u_n = 0$ *is necessary but not sufficient, meaning that a series* $\sum_{n=1}^{+\infty} u_n$ *may still be divergent even if* $\lim_{n\to\infty} u_n = 0$*. For example, consider the so-called* ***harmonic series***

$$\sum_{n=1}^{+\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots ,$$

which is divergent, despite $\lim_{n\to\infty} \frac{1}{n} = 0$*.*

A particularly useful application of Proposition 1.3.11 is its contrapositive:

Corolary 1.3.13. *If* $\lim_{n\to\infty} u_n \neq 0$*, then the series* $\sum_{n=1}^{+\infty} u_n$ *is divergent.*

Example 1.3.14. *Since* $\lim_{n\to\infty} 2^n = +\infty$*, then, by the contrapositive of Proposition 1.3.11, we can conclude that the series* $\sum_{n=1}^{+\infty} 2^n$ *is divergent.*

Proposition 1.3.15. *Let* $\sum_{n=1}^{+\infty} u_n$ *and* $\sum_{n=1}^{+\infty} v_n$ *be two convergent series, with sums* u *and* v*, respectively. Then:*

1. *the series* $\sum_{n=1}^{+\infty} (u_n + v_n)$ *is convergent, and its sum is* $u + v$*;*

2. *for each* $\alpha \in \mathbb{R}$*, the series* $\sum_{n=1}^{+\infty} \alpha u_n$ *is convergent, and its sum is* αu*.*

Proof. Exercise. □

Example 1.3.16. *Consider the series* $\sum_{n=1}^{+\infty} \frac{1}{2^n} = 1$ *and* $\sum_{n=1}^{+\infty} \frac{1}{3^n} = \frac{1}{2}$*. We have:*

1. *the series* $\sum_{n=1}^{+\infty} \left(\frac{1}{2^n} + \frac{1}{3^n}\right)$ *is convergent, and its sum is* $1 + \frac{1}{2} = \frac{3}{2}$*;*

2. *the series* $\sum_{n=1}^{+\infty} \left(\frac{-5}{3^n}\right)$ *is convergent, and its sum is* $(-5)\frac{1}{2} = -\frac{5}{2}$*.*

1.3.6 Geometric series

In Example 1.3.16, we considered series in which the ratio between any term of the sequence and the previous term is constant, i.e., the general term of these series forms a geometric progression. For this reason, such series are called geometric series.

Definition 1.3.17 (**Geometric series**)**.** *Given two constants $k, r \in \mathbb{R}$, with $k \neq 0$, the series*

$$k + kr + kr^2 + kr^3 + \ldots + kr^n + \ldots = \sum_{n=0}^{+\infty} kr^n$$

is called a ***geometric series with ratio r*** *and first term k.*

Remark 1.3.18. *In the case $r = 0$, we conventionally set $0^0 = 1$ for the first term.*

Next, we study the convergence of the geometric series with ratio r and first term k. As we will see, the convergence of this series depends on the ratio r.

The sequence of partial sums of the geometric series with ratio r and first term k is given by:

$$S_n = k + kr + kr^2 + \ldots + kr^{n-1}.$$

If $r = 1$, then $S_n = \underbrace{k + k + \ldots + k}_{n \text{ times}} = kn$. Since the sequence with general term kn diverges to $+\infty$ or $-\infty$, the geometric series is divergent in this case.

Now, consider the case $r \neq 1$. We have:

$$S_n = k + kr + kr^2 + \ldots + kr^{n-1}$$

and

$$rS_n = kr + kr^2 + \ldots + kr^{n-1} + kr^n.$$

Thus,

$$S_n - rS_n = k - kr^n \Leftrightarrow S_n(1-r) = k(1-r^n) \Leftrightarrow S_n = \frac{k(1-r^n)}{1-r}.$$

Hence, if $r \neq 1$, we get

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{k(1-r^n)}{1-r} = \begin{cases} \infty & , \text{ if } r > 1 \\ \frac{k}{1-r} & , \text{ if } |r| < 1 \\ \text{does not exist} & , \text{ if } r \leq -1 \end{cases}.$$

Remark 1.3.19.

1. *If* $r \geq 1$, *then* $\lim_{n\to\infty} r^n = +\infty$, *hence the series* $\sum_{n=1}^{+\infty} r^n$ *is divergent.*

2. *If the series* $\sum_{n=1}^{+\infty} r^n$ *is convergent, then* $\lim_{n\to\infty} r^n = 0$, *which holds if* $|r| < 1$.

Proposition 1.3.20. *Let* $r \in \mathbb{R}$ *such that* $|r| < 1$ *and let* $p \in \mathbb{N}$. *Then the series* $\sum_{n=p}^{+\infty} kr^n$ *is convergent, and its sum is* $\frac{kr^p}{1-r}$.

Proof. Exercise. □

Exercise 1.3.21. *Using the method of mathematical induction, show that for any natural number* n, *if* $r \neq 1$, *then*

$$k + kr + kr^2 + \ldots + kr^{n-1} = \frac{k(1-r^n)}{1-r}.$$

Example 1.3.22. *The series* $\sum_{n=0}^{+\infty} \left(-\frac{1}{2}\right)^n$ *is a geometric series with ratio* $r = -\frac{1}{2}$. *Since* $|r| = \left|-\frac{1}{2}\right| < 1$, *we conclude that the series is convergent and that its sum is*

$$\frac{1}{1-\left(-\frac{1}{2}\right)} = \frac{2}{3}.$$

Chapter 2

Real functions of a real variable

A real function of a real variable is a mapping from a subset $D_f \subseteq \mathbb{R}$ into $\mathbb{R}$ such that for each $x \in D_f$, its image is the unique element $f(x) \in \mathbb{R}$, which can be schematically represented by

$$\begin{array}{rcl} f : D_f \subseteq \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & f(x)\,. \end{array}$$

2.1 Domain and codomain

The **domain** of a function f is the subset D_f of $\mathbb{R}$ formed by all real numbers that, when placed in the position of the variable x, transform the given expression into a real number. The **codomain** of a function f is the set of all real numbers that f assumes (the image of all elements in its domain under f), which can be written as

$$f(D_f) = \{f(x) \in \mathbb{R} : x \in D_f\}\,.$$

Given a function f, for each $x \in D_f$, we can graphically represent in $\mathbb{R}^2$ the points $(x, f(x))$, which is usually referred to as "drawing the graph of the function f", that is, representing in $\mathbb{R}^2$ the set

$$Gr(f) = \{(x, y) \in \mathbb{R}^2 : x \in D_f \wedge y = f(x)\}\,.$$

2.2 Polynomial function and series

A function P defined from $\mathbb{R}$ to $\mathbb{R}$ is called a **polynomial function** if for each $x \in \mathbb{R}$, its image is given by

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 \,,$$

for some $n \in \mathbb{N}_0$ and where $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$, called coefficients, belong to $\mathbb{R}$. If $a_n \neq 0$, we say that P is a **polynomial** of degree n.

All these series have the particular form $\sum_{n=0}^{+\infty} a_n x^n$, and are known as **power series** or **entire series**. The numbers $a_0, a_1, a_2, \ldots, a_n, \cdots \in \mathbb{R}$ are called the **coefficients** of the power series.

From the geometric series, we know that if $|x| < 1$, then

$$1 + x + x^2 + x^3 + \cdots = \sum_{n=0}^{+\infty} x^n = \frac{1}{1-x} \,. \tag{2.1}$$

Continuing to assume that $|x| < 1$, if we replace x with x^2 in (2.1), we obtain (sum of even powers)

$$1 + x^2 + x^4 + x^6 \cdots = \sum_{n=0}^{+\infty} x^{2n} = \frac{1}{1-x^2} \,. \tag{2.2}$$

Observe that if $|x| < 1$, then $|x^2| < 1$.

Multiplying now (2.2) by x, we get (sum of odd powers)

$$x + x^3 + x^5 + x^7 \cdots = \sum_{n=0}^{+\infty} x^{2n+1} = \frac{x}{1-x^2} \,. \tag{2.3}$$

Returning to (2.1), if we replace x with $-x$, we obtain

$$1 - x + x^2 - x^3 + \cdots = \sum_{n=0}^{+\infty} (-1)^n x^n = \frac{1}{1+x} \,. \tag{2.4}$$

2.2.1 Exponential series

The geometric series is a particular case of a power series, where all coefficients are equal to 1, i.e., $a_n = 1$ for all $n \in \mathbb{N}$. Another particular case is the **exponential series**

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \,. \tag{2.5}$$

Considering $f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$, it is easy to verify that $f'(x) = f(x)$ for any $x \in \mathbb{R}$, so a well-known function satisfying this property is the exponential function. That is, for any constant $K \in \mathbb{R}$, we have $f(x) = Ke^x$ verifying $f'(x) = f(x)$. However, note that if $f(0) = 1$, then $Ke^0 = 1$, hence $K = 1$, and in this case, $f(x) = e^x$. Thus, we obtain the following result.

Proposition 2.2.1. *The exponential series converges for any $x \in \mathbb{R}$, and its sum is e^x, i.e.,*

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = e^x \,, \ \forall x \in \mathbb{R} \tag{2.6}$$

Remark 2.2.2. *A justification that $\lim\limits_{x\to 0} \dfrac{e^x - 1}{x} = 1$ can be made using Proposition 2.2.1 and Equation* (2.6)*. From* (2.6)*, subtracting* 1 *from both sides of the equation and, for $x \neq 0$, dividing both sides by x, we obtain*

$$\frac{e^x - 1}{x} = 1 + \frac{x}{2!} + \frac{x^2}{3!} + \frac{x^3}{4!} \cdots \,, \tag{2.7}$$

from which,

$$\lim_{x\to 0} \frac{e^x - 1}{x} = \lim_{x\to 0} \left(1 + \frac{x}{2!} + \frac{x^2}{3!} + \frac{x^3}{4!} \cdots \right) = 1 \,.$$

2.3 General Properties of Functions

2.3.1 Injectivity

Definition 2.3.1 (**One-to-one Function**)**.** *A function $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ is said to be **one-to-one** or **injective** if for any $x_1, x_2 \in D_f$ such that $x_1 \neq x_2$, we have $f(x_1) \neq f(x_2)$.*

Remark 2.3.2. *It is often advantageous to use the contrapositive of this definition, that is, the function f is one-to-one if, given $x_1, x_2 \in D_f$ such that $f(x_1) = f(x_2)$, then $x_1 = x_2$.*

Example 2.3.3.

1) *The function* $f(x) = 2x + 3$ *is one-to-one because, given* $x_1, x_2 \in D_f = \mathbb{R}$,

$$f(x_1) = f(x_2) \Rightarrow 2x_1 + 3 = 2x_2 + 3 \Rightarrow x_1 = x_2 .$$

2) *The function* $f(x) = x^2$ *is not one-to-one because* $-2 \neq 2$ *but* $f(-2) = f(2)$.

2.3.2 Monotonicity

Definition 2.3.4 (Monotonicity of a Function). *A function* $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ *is said to be* ***increasing*** *(resp.* ***strictly increasing****) if for any* $x_1, x_2 \in D_f$ *such that* $x_1 > x_2$*, we have* $f(x_1) \geq f(x_2)$ *(resp.* $f(x_1) > f(x_2)$*).*

Similarly, the function f *is said to be* ***decreasing*** *(resp.* ***strictly decreasing****) if for any* $x_1, x_2 \in D_f$ *such that* $x_1 > x_2$*, we have* $f(x_1) \leq f(x_2)$ *(resp.* $f(x_1) < f(x_2)$*).*

A function is said to be ***strictly monotonic*** *if it is either strictly increasing or strictly decreasing.*

Proposition 2.3.5. *If the function* f *is strictly monotonic in* $A \subseteq D_f$*, then* f *is one-to-one in* A.

Proof. Exercise. □

2.3.3 Bounded Function

Definition 2.3.6 (Bounded Function). *A function* $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ *is said to be* ***bounded*** *if there exists* $M \in \mathbb{R}^+$ *such that* $|f(x)| \leq M$ *for any* $x \in D_f$.

Remark 2.3.7. *Equivalently, a function* f *is said to be* ***bounded*** *if its image set,* $f(D_f)$*, is a bounded set.*

2.3.4 Parity

Definition 2.3.8 (Even Function). *A function* $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ *is said to be* ***even*** *if for any* $x \in D_f$, $f(x) = f(-x)$.

Example 2.3.9. *The function* $f(x) = x^2$ *is even because for any* $x \in D_f = \mathbb{R}$, $f(x) = x^2 = f(-x)$.

Definition 2.3.10 (Odd Function). *A function* $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ *is said to be* ***odd*** *if for any* $x \in D_f$, $f(x) = -f(-x)$.

Example 2.3.11. *The function* $f(x) = x^3$ *is odd because for any* $x \in D_f = \mathbb{R}$, $f(-x) = -x^3 = -f(x)$.

2.3.5 Periodicity

Definition 2.3.12 (Periodic Function). *A function $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ is said to be **periodic** if there exists $T \in \mathbb{R}^+$ such that, for any $x \in D_f$, $x + T \in D_f$ and $f(x+T) = f(x)$.*

*The smallest positive real number T such that $f(x+T) = f(x)$ is called the **period** of the function f.*

Example 2.3.13. *The sine, cosine, and tangent functions are periodic with periods 2π, 2π, and π, respectively.*

2.4 Composition of Functions

Consider two real functions of a real variable $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ and $g : D_g \subseteq \mathbb{R} \longrightarrow \mathbb{R}$.

One can consider the composition of the function f with the function g, denoted by $f \circ g$, which is read as the function "f composed with g" or "f after g", defined by

$$\begin{array}{rcl} f \circ g : D_{f\circ g} \subseteq \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & (f \circ g)(x) = f(g(x)) \end{array}$$

where

$$D_{f\circ g} = \{x \in D_g \,:\, g(x) \in D_f\}\,.$$

This operation of function composition can be generalised to more than two functions. For example, given another real function of a real variable $h : D_h \subseteq \mathbb{R} \longrightarrow \mathbb{R}$, one can define the composite function $h \circ f \circ g$, where

$$D_{h\circ f\circ g} = \{x \in D_{f\circ g} \,:\, (f \circ g)(x) \in D_h\}\,,$$

and so on.

2.5 Inverse Function

Given a real function of a real variable $f : A \subseteq \mathbb{R} \longrightarrow B \subseteq \mathbb{R}$, where $A = D_f$ and $B = f(D_f)$, for each $x \in D_f$, let $y = f(x)$. We wish to determine whether it is possible, and under what conditions, to find a function $g : B \subseteq \mathbb{R} \longrightarrow A \subseteq \mathbb{R}$ such that $g(y) = x$, that is, such that $g \circ f = id$ (i.e., $(g \circ f)(x) = x$) or $f \circ g = id$ (i.e., $(f \circ g)(y) = y$).

If such a function g exists, it is called the inverse function of f and is denoted by f^{-1}.

Let us examine the conditions that a given function f must satisfy to guarantee the existence of its inverse function f^{-1}. For instance, if a certain function f maps two distinct objects to the same image, i.e., for $x_1, x_2 \in D_f$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2)$, then it would not be meaningful for a function to map a single object to two different images. That is, considering $y = f(x_1) = f(x_2)$, it would not make sense to have a function g such that $g(y) = x_1$ and $g(y) = x_2$. This leads to the following property:

Proposition 2.5.1. *A function admits an inverse if and only if it is one-to-one.*

Example 2.5.2. *Consider the real function of a real variable*

$$f : \mathbb{R} \setminus \{1\} \longrightarrow \mathbb{R} \setminus \{2\}$$

defined by

$$f(x) = \frac{2x+3}{x-1} = 2 + \frac{5}{x-1}\,.$$

It can be verified that this function f is one-to-one over its entire domain, and hence, by Proposition 2.5.1, it admits an inverse. In this case, the inverse function of f is given by

$$f^{-1} : \mathbb{R} \setminus \{2\} \longrightarrow \mathbb{R} \setminus \{1\}$$

defined by

$$f^{-1}(x) = \frac{x+3}{x-2} = 1 + \frac{5}{x-2}\,.$$

Example 2.5.3. *The real function of a real variable*

$$g : \mathbb{R} \longrightarrow \mathbb{R}^+$$

defined by

$$g(x) = e^x$$

admits an inverse, which is the function

$$g^{-1} : \mathbb{R}^+ \longrightarrow \mathbb{R}$$

defined by

$$f^{-1}(x) = \ln x\,.$$

Remark 2.5.4. *1) If a given function is not one-to-one, it is possible to consider the same function but defined only on a subset of its domain so that this new function, typically referred to as a restriction of the original function, is one-to-one in this "new" domain. Examples of this situation will appear in the definition of inverse trigonometric functions presented below.*

2) If a continuous function is strictly monotonic, then it is one-to-one.

3) The graphs of a function and its inverse are symmetric with respect to the graph of the function $f(x) = x$ (the bisector of the first and third quadrants).

2.5.1 Inverse Trigonometric Functions

Arcsine

Consider the function

$$f : \mathbb{R} \longrightarrow [-1, 1]$$

defined by

$$f(x) = \sin x\,.$$

It is known that the function f is not one-to-one; however, a restriction can be considered, for example,

$$f_{\big|\left[-\frac{\pi}{2},\frac{\pi}{2}\right]} : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow [-1, 1]\,,$$

whose graph is represented in Figure 2.1.

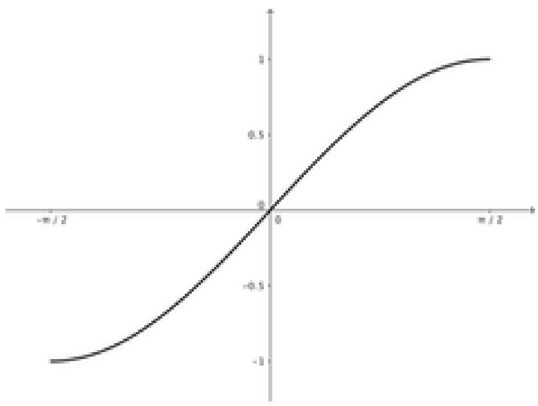

Figure 2.1: Graph of the function $\sin x$ restricted to the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Thus, the function $f\big|_{[-\frac{\pi}{2},\frac{\pi}{2}]}$ is now one-to-one, so (as stated in Observation 2.5.4) its inverse can be defined. The inverse function of the sine function is

$$\left(f\big|_{[-\frac{\pi}{2},\frac{\pi}{2}]}\right)^{-1} : [-1,1] \longrightarrow \left[-\frac{\pi}{2},\frac{\pi}{2}\right],$$

defined by

$$f^{-1}(x) = \arcsin x\,,$$

and is called "arcsine," whose graph can be seen in Figure 2.2.

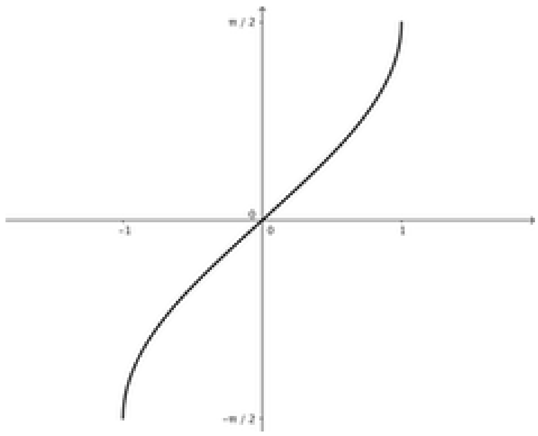

Figure 2.2: Graph of the function $\arcsin x$.

Proposition 2.5.5. *Under the conditions in which the functions are defined, we have*

$$\sin(\arcsin x) = x \qquad \textit{and} \qquad \arcsin(\sin x) = x\,.$$

Arc Cosine

Consider the function

$$f : \mathbb{R} \longrightarrow [-1,1]$$

defined by

$$f(x) = \cos x\,.$$

It is known that the function f is not one-to-one; however, one can consider its restriction, for example,

$$f\big|_{[0,\pi]} : [0,\pi] \longrightarrow [-1,1]\,,$$

whose graph is shown in Figure 2.3.

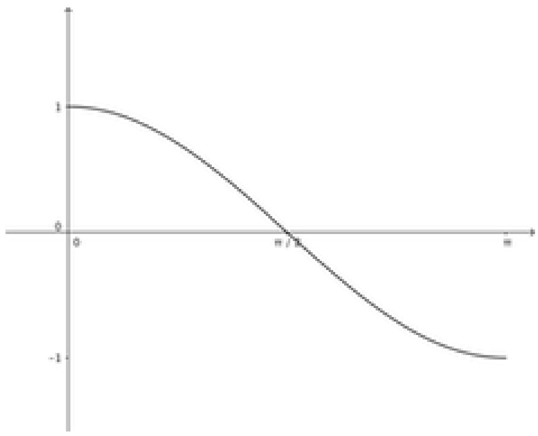

Figure 2.3: Graph of the function $\cos x$ restricted to the interval $[0, \pi]$.

Thus, the function $f\big|_{[0,\pi]}$ is already one-to-one, so (as seen in Observation 2.5.4) its inverse can be defined. The inverse function of the cosine function is

$$\left(f\big|_{[0,\pi]}\right)^{-1} : [-1, 1] \longrightarrow [0, \pi] \ ,$$

defined by

$$f^{-1}(x) = \arccos x \, ,$$

and is read as "arc cosine," whose graphical representation can be seen in Figure 2.4.

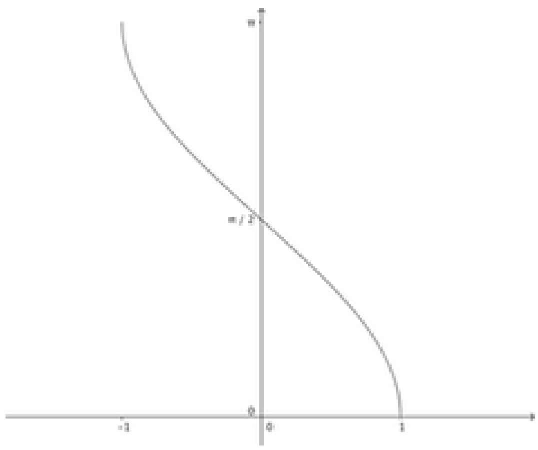

Figure 2.4: Graph of the function $\arccos x$.

Remark 2.5.6. *The function* $f^{-1}(x) = \arccos x$ *should be interpreted as "the* *angle whose cosine is x." For example,* $f^{-1}\left(\frac{\sqrt{3}}{2}\right) = \arccos\frac{\sqrt{3}}{2} = \frac{\pi}{6}$, *that is, "the angle whose cosine is* $\frac{\sqrt{3}}{2}$ *is* $\frac{\pi}{6}$*."*

Proposition 2.5.7. *Under the conditions where the functions are defined, we have*

$$\cos(\arccos x) = x \qquad \text{and} \qquad \arccos(\cos x) = x\,.$$

Arc Tangent

Consider the function

$$f : \mathbb{R} \setminus \left\{\frac{\pi}{2} + k\pi \;:\; k \in \mathbb{Z}\right\} \longrightarrow \mathbb{R}$$

defined by

$$f(x) = \tan x\,.$$

It is known that the function f is not one-to-one; however, one can consider its restriction, for example,

$$f_{\left|\left]-\frac{\pi}{2},\frac{\pi}{2}\right[\right.} : \left]-\frac{\pi}{2},\frac{\pi}{2}\right[\longrightarrow \mathbb{R}\,,$$

whose graph is shown in Figure 2.5.

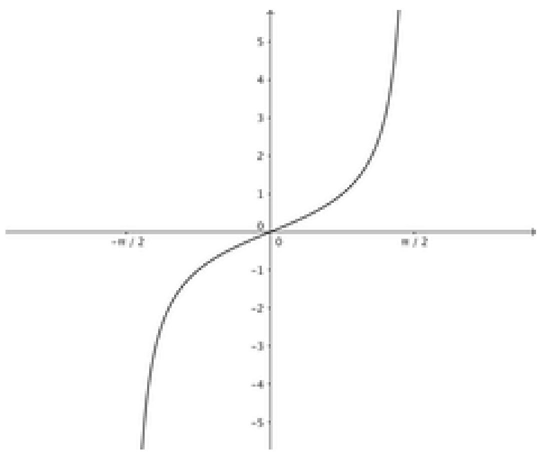

Figure 2.5: Graph of the function $\tan x$ restricted to the interval $\left]-\frac{\pi}{2},\frac{\pi}{2}\right[$.

Thus, the function $f_{\left|\left]-\frac{\pi}{2},\frac{\pi}{2}\right[\right.}$ is already one-to-one, so (as seen in Observation 2.5.4) its inverse can be defined. The inverse function of the tangent function is

$$\left(f_{|\left]-\frac{\pi}{2},\frac{\pi}{2}\right[}\right)^{-1} : \mathbb{R} \longrightarrow \left]-\frac{\pi}{2},\frac{\pi}{2}\right[\,,$$

defined by

$$f^{-1}(x) = \arctan x \,,$$

and is read as "arc tangent," whose graphical representation can be seen in Figure 2.6.

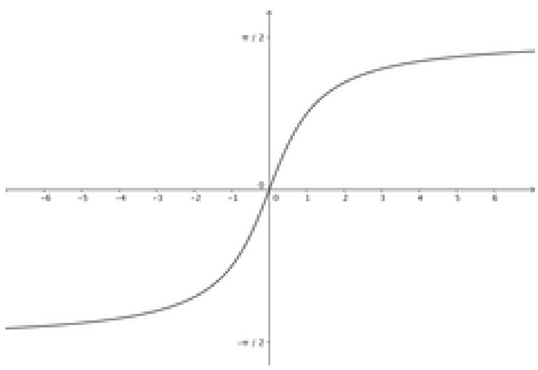

Figure 2.6: Graph of the function $\arctan x$.

Remark 2.5.8. *The function $f^{-1}(x) = \arctan x$ should be interpreted as "the angle whose tangent is x." For example, $f^{-1}\left(\sqrt{3}\right) = \arctan\sqrt{3} = \frac{\pi}{3}$, that is, "the angle whose tangent is $\sqrt{3}$ is $\frac{\pi}{3}$".*

Proposition 2.5.9. *Under the conditions where the functions are defined, we have*

$$\tan(\arctan x) = x \qquad \textit{and} \qquad \arctan(\tan x) = x \,.$$

Chapter 3

Limits and continuity

Consider a real-valued function of a real variable $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$.

3.1 Limits

Definition 3.1.1 (Cauchy's Definition of Limit). *Let $a \in \mathbb{R}$ be an accumulation point of the domain of f, D_f. We say that the* ***limit of f at the point a*** *is $b \in \mathbb{R}$, and write $\lim_{x\to a} f(x) = b$, if*

$$\forall \delta > 0 \quad \exists \varepsilon > 0 \quad \forall x \in D_f \quad 0 < |x-a| < \varepsilon \Rightarrow |f(x) - b| < \delta\,.$$

3.2 Continuity

Definition 3.2.1 (Continuity of a function at a point). *Let a be a point in the domain of f, i.e., let $a \in D_f$. We say that f* ***is continuous at the point*** *a if*

$$\text{there exists} \quad \lim_{x\to a} f(x) \quad \text{and} \quad \lim_{x\to a} f(x) = f(a)\,.$$

Definition 3.2.2 (Continuity of a function on a set $A \subseteq D_f$). *We say that f* ***is a continuous function*** *on a set $A \subseteq D_f$ if f is continuous at every point of A.*

Example 3.2.3. *The real-valued function of a real variable $f : \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R}$ defined by $f(x) = \frac{\sin x}{x}$ is continuous, as it is continuous at every point of its domain.*

3.2.1 Properties

In this section, we state a number of general results (properties) regarding the continuity of real functions. Proofs of these results can be found, for example, in [2], [6], [10], or [11].

Theorem 3.2.4. *If f and g are continuous at the point x_0, then the functions $f+g$, $f-g$, and fg are also continuous at x_0. Moreover, if $g(x_0) \neq 0$, then the function $\frac{f}{g}$ is also continuous at x_0.*

Theorem 3.2.5. *Let f and g be functions such that g is continuous at x_0 and f is continuous at $g(x_0)$. Then the composition $f \circ g$ is continuous at x_0.*

Theorem 3.2.6 (Weierstrass Theorem). *If $f : D_f \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous function and its domain D_f is a compact set (i.e., closed and bounded), then f attains both a maximum and a minimum on D_f.*

Theorem 3.2.7 (Bolzano's Theorem or Intermediate Value Theorem). *Let $a, b \in \mathbb{R}$ such that $a < b$, and let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. Then f takes all values between $f(a)$ and $f(b)$, i.e., for every $d \in [f(a), f(b)]$ or $[f(b), f(a)]$, there exists at least one $c \in [a, b]$ such that $f(c) = d$.*

Corolary 3.2.8. *Let $a, b \in \mathbb{R}$ such that $a < b$, and let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. If $f(a) \times f(b) < 0$, then there exists $c \in [a, b]$ such that $f(c) = 0$.*

3.2.2 Extension by Continuity to a Point

Definition 3.2.9 (Extension by continuity). *We say that f **can be extended by continuity** to a point $a \in fr\,(D_f)$ which does not belong to D_f if $\lim\limits_{x \to a} f(x)$ exists. In this case, if $\lim\limits_{x \to a} f(x) = b$, the extension of f is given by the function $\tilde{f} : D_f \cup \{a\} \longrightarrow \mathbb{R}$ defined by*

$$\tilde{f}(x) = \begin{cases} f(x) & , \quad x \in D_f \\ b & , \quad x = a \end{cases} .$$

Example 3.2.10. *Since $\lim\limits_{x \to 0} \dfrac{\sin x}{x} = 1$, the function f defined in Example 3.2.3, can be extended by continuity to the point $x = 0$, and its extension $\tilde{f} : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by*

$$\tilde{f}(x) = \begin{cases} \frac{\sin x}{x} & , \quad x \neq 0 \\ 1 & , \quad x = 0 \end{cases} .$$

Chapter 4

Differential calculus

4.1 Derivative of a Function at a Point

Geometric Definition

Given a real function of a real variable, the derivative of this function at a given point, if it exists, is the slope of the tangent line to the graph of the function at that point.

Analytical Definition

Given a real function of a real variable $f : D_f \subseteq \mathbb{R} \to \mathbb{R}$, the derivative of f at a point $a \in D_f$, which can be written as $f'(a)$, is

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}\,. \tag{4.1}$$

Remark 4.1.1. *Note that, considering the change of variable $x - a = h$, when $x \to a$ then $h \to 0$, and the limit in* (4.1) *can be rewritten as*

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}\,.$$

Definition 4.1.2 (Differentiability of the function at a point). *Let $a \in D_f$ be an accumulation point of D_f. It is said that the function f* ***is differentiable*** *or* ***differentiable at point*** *a if the limit $\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$ exists and is finite. This limit, when it exists, is called the* ***derivative of*** *f* ***at point*** *a and is denoted by $f'(a)$ (Lagrange notation) or $\frac{df}{dx}(a)$ (Leibniz notation).*

Definition 4.1.3 (Differentiability of the function on a set $A \subseteq D_f$). *It is said that the function f is* ***differentiable*** *on a set $A \subseteq D_f$ if f is differentiable at every point of A.*

The equation of the tangent line to the graph of f at the point $(a, f(a))$ is given by

$$y - f(a) = f'(a)\,(x - a)\ ,$$

that is,

$$y = f(a) + f'(a)\,(x - a)\ ,$$

which is equivalent to its reduced form

$$y = f'(a)\,x + (f(a) - a\,f'(a))\ .$$

Theorem 4.1.4. *If $f : D_f \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at the point $a \in D_f$, then f is continuous at that point.*

Proof. See [6]. □

4.2 Derivative Function

Given a real function of a real variable $f : D_f \subseteq \mathbb{R} \to \mathbb{R}$ that is differentiable, a new function can be considered, defined by

$$\begin{array}{rcl} f' : A \subseteq D_f & \to & \mathbb{R} \\ x & \longmapsto & f'(x) \end{array}$$

where A is the set of points in D_f where the function f has a derivative.

The function f' is called the **derivative function** of f. The function f' can also be represented by $\frac{df}{dx}$.

Example 4.2.1. *The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is differentiable on $\mathbb{R}$ and $f' : \mathbb{R} \to \mathbb{R}$ with $f'(x) = 2x$ is the derivative function of f.*

4.3 Differentiation

Theorem 4.3.1. *Let $f, g : D \subseteq \mathbb{R} \to \mathbb{R}$ be two functions differentiable at $a \in D$. Then:*

1) $f + g$ is differentiable at a, and $(f + g)'(a) = f'(a) + g'(a)$;

2) for any $k \in \mathbb{R}$, kf is differentiable at a, and $(kf)'(a) = kf'(a)$;

3) $f \times g$ is differentiable at a, and $(f \times g)'(a) = f'(a)\,g(a) + f(a)\,g'(a)$;

4) for any $n \in \mathbb{N}$, f^n is differentiable at a, and $(f^n)'(a) = nf'(a)f^{n-1}(a)$;

5) if $g(a) \neq 0$, then $\frac{f}{g}$ is differentiable at a, and

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)\,g(a) - f(a)\,g'(a)}{g^2(a)}.$$

Differentiation Rules

Let $k \in \mathbb{R}$. For the following differentiable functions, the differentiation rules are:

1) if $f(x) = k\,x^n$, with $n \in \mathbb{N}$, then $f'(x) = n\,k\,x^{n-1}$;

2) if $f(x) = \sin x$, then $f'(x) = \cos x$;

3) if $f(x) = \cos x$, then $f'(x) = -\sin x$;

4) for $k > 0$, if $f(x) = k^x$, then $f'(x) = k^x \ln k$;

   4.1) if $k = e$, $f(x) = e^x$ and $f'(x) = e^x$;

5) for $k > 0$, if $f(x) = \log_k x$, then $f'(x) = \frac{1}{x \ln k}$;

   5.1) if $k = e$, $f(x) = \ln x$ and $f'(x) = \frac{1}{x}$.

4.4 Derivative of the Composed Function - Chain rule

Given the composition of two functions that are differentiable at a given point, the **derivative of the composed function**, usually designated as **the chain rule**, can be calculated by the derivative of each of these functions. Consider the following theorem.

Theorem 4.4.1 (**Derivative of the Composed Function - chain rule**). *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ and $g : E \subseteq \mathbb{R} \to \mathbb{R}$ be functions such that $g(E) \subseteq D$. If g is differentiable at $b \in E$ and f is differentiable at $a = g(b) \in D$, then $f \circ g : E \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at b and we have the so-called* ***chain rule***

$$(f \circ g)'(b) = f'(g(b))\; g'(b)\,.$$

Proof. See [6]. □

Example 4.4.2. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be the functions defined by $f(x) = x^{10}$ and $g(x) = x^2+3x+5$. The function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = (x^2+3x+5)^{10}$ can be seen as the composition of the functions f and g, with $h = f \circ g$. By Theorem 4.4.1, h is differentiable at $x \in \mathbb{R}$ and*

$$h'(x) = (f \circ g)'(x) = f'(g(x)) \times g'(x) = 10\left(x^2+3x+5\right)^9 (2x+3)\ .$$

4.5 Derivative of the Inverse Function

Given a one-to-one function, by Proposition 2.5.1, that function has an inverse. In this case, if the function is also differentiable, then the derivative of its inverse at a given point can be calculated without explicitly calculating its inverse function. Consider the following theorem.

Theorem 4.5.1 (**Derivative of the Inverse Function**). *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable and one-to-one function, and let $a \in D$ such that $f'(a) \neq 0$. Then its inverse function f^{-1} is differentiable at $f(a)$ and, considering $b = f(a)$, we have the so-called **derivative rule for the inverse function***

$$\left(f^{-1}\right)'(b) = \frac{1}{f'(a)} = \frac{1}{f'\left(f^{-1}(b)\right)}\ .$$

Proof. See [6]. □

Example 4.5.2. *The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x+3$ is differentiable and one-to-one. Its inverse is $f^{-1} : \mathbb{R} \to \mathbb{R}$ where $f^{-1}(x) = \frac{x-3}{2}$. For any $x \in \mathbb{R}$, we have $f'(x) = 2 \neq 0$, so by Theorem 4.5.1, we have that f^{-1} is differentiable and $\left(f^{-1}\right)'(x) = \frac{1}{f'(x)} = \frac{1}{2}$.*

Exercise 4.5.3. *By Theorem 4.5.1, one can deduce the derivatives of the inverse trigonometric functions and thus obtain:*

- $(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$;
- $(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$;
- $(\arctan x)' = \frac{1}{1+x^2}$.

Exercise 4.5.4. *By Theorems 4.4.1 and 4.5.1, one can deduce the derivatives of the inverse trigonometric functions where the argument is a function f that depends on x, and thus obtain:*

- $(\arcsin f(x))' = \frac{f'(x)}{\sqrt{1-f^2(x)}}$;
- $(\arccos f(x))' = -\frac{f'(x)}{\sqrt{1-f^2(x)}}$;
- $(\arctan f(x))' = \frac{f'(x)}{1+f^2(x)}$.

4.6 Cauchy's Rule

In the calculation of limits, when faced with certain indeterminate forms, such as $\frac{0}{0}$ or $\frac{\infty}{\infty}$, in some cases one can apply a differentiation rule, known as **Cauchy's Rule**, which causes the indeterminacy to "disappear," thus facilitating the calculation of the limit. Consider the following theorem.

Theorem 4.6.1 (**Cauchy's Rule**). *Let $D \subseteq \mathbb{R}$ be an interval and let $a \in ad(D)$. Let $f, g : D \setminus \{a\} \to \mathbb{R}$ be differentiable functions and assume that $g(x) \neq 0$ for all $x \in D \setminus \{a\}$. If*

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0 \quad \text{or} \quad \lim_{x\to a} g(x) = \pm\infty$$

and

$$\text{there exists} \quad \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

then $\lim_{x\to a} \frac{f(x)}{g(x)}$ exists and we have

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Proof. See [6]. ☐

Remark 4.6.2. *Note that if $\lim_{x\to a} g(x) = \pm\infty$ and $\lim_{x\to a} f(x)$ is a constant, then $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$, so it is not necessary to apply Cauchy's Rule. Therefore, in the case where $\lim_{x\to a} g(x) = \pm\infty$, Cauchy's Rule is only applied if $\lim_{x\to a} f(x) = \pm\infty$. See item 2) in the following example.*

Example 4.6.3.

*1) The $\lim_{x\to 0} \frac{e^x - 1}{x} = \frac{0}{0}$, where, under the conditions of Theorem 4.6.1, one can apply **Cauchy's Rule**, obtaining*

$$\lim_{x\to 0} \frac{e^x - 1}{x} = \lim_{x\to 0} e^x = 1.$$

2) The $\lim_{x\to 0} x\ln x = \lim_{x\to 0} \frac{\ln x}{\frac{1}{x}} = \frac{\infty}{\infty}$*, where, under the conditions of Theorem 4.6.1, one can apply* ***Cauchy's Rule****, obtaining*

$$\lim_{x\to 0} \frac{\ln x}{\frac{1}{x}} = \lim_{x\to 0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x\to 0} -x = 0\,.$$

4.7 Local and Global Maxima and Minima

Definition 4.7.1. *Let* $f : D \subseteq \mathbb{R} \to \mathbb{R}$ *be a function and let* $a \in D$*. We say that:*

1) f *has a* ***local maximum*** *(or* ***relative maximum****) at* a *or that* $f(a)$ *is a* ***local maximum*** *(or* ***relative maximum****) of the function* f*, if there exists* $\varepsilon > 0$ *such that* $f(x) \leq f(a)$ *for all* $x \in \mathcal{B}_\varepsilon(a) \cap D$ *;*

2) f *has a* ***local minimum*** *(or* ***relative minimum****) at* a *or that* $f(a)$ *is a* ***local minimum*** *(or* ***relative minimum****) of the function* f*, if there exists* $\varepsilon > 0$ *such that* $f(x) \geq f(a)$ *for all* $x \in \mathcal{B}_\varepsilon(a) \cap D$ *.*

One can use the expression ***local extremes*** *(or* ***relative extremes****) of the function* f *when referring (indistinctly) to its local maxima or minima.*

3) f *has a* ***global maximum*** *(or* ***absolute maximum****) at* a *or that* $f(a)$ *is a* ***global maximum*** *(or* ***absolute maximum****) of the function* f*, if* $f(x) \leq f(a)$ *for any* $x \in D$ *;*

4) f *has a* ***global minimum*** *(or* ***absolute minimum****) at* a *or that* $f(a)$ *is a* ***global minimum*** *(or* ***absolute minimum****) of the function* f*, if* $f(x) \geq f(a)$ *for any* $x \in D$ *.*

One can use the expression ***global extreme*** *(or* ***absolute extreme****) of the function* f *when referring (indistinctly) to its global maximum or minimum.*

The points where f *has a maximum (resp. minimum) are called* ***maximising points*** *(resp.* ***minimising points****)). A maximising or minimising point is called an* ***extremising point****.*

Remark 4.7.2. *If* $a \in D$ *is a global extreme point of a function* f*, then* a *is a local extreme point of* f*. Note that the converse of this statement is not true.*

Theorem 4.7.3 (**Fermat's Theorem**). *Let* $f : D \subseteq \mathbb{R} \to \mathbb{R}$ *be a differentiable function at* $a \in \operatorname{int}(D)$*. If* f *has a local extreme at* a*, then* $f'(a) = 0$ *.*

Proof. See [10]. □

Remark 4.7.4. *The converse of Fermat's Theorem is not true. Consider the following example: for the function $f(x) = x^3$, we have $f'(0) = 0$, but f does not have any extreme point.*

Definition 4.7.5 (**Critical Point**). *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable function at $a \in D$. We say that a **is a critical point of** f or that f **has a critical point at** a if $f'(a) = 0$.*

Theorem 4.7.6 (**Rolle's Theorem**). *Let f be a continuous function on $[a, b]$ (with $a, b \in \mathbb{R}$ and $a < b$) and differentiable on $]a, b[$. If $f(a) = f(b)$, then there exists a point $c \in]a, b[$ such that $f'(c) = 0$.*

Proof. Since f is continuous on $[a, b]$, by Weierstrass' Theorem (see Theorem 3.2.6), f has a maximum and a minimum on $[a, b]$.

If the maximum and minimum are achieved at the endpoints of $[a, b]$, as we assume that $f(a) = f(b)$, then f is constant on $[a, b]$, so $f'(c) = 0$ for any $c \in]a, b[$.

Otherwise, either the maximum or the minimum is achieved at $c \in]a, b[$, so by Fermat's Theorem (see Theorem 4.7.3), $f'(c) = 0$. □

Corolary 4.7.7. *Between two zeros of a differentiable function in an interval, there is at least one zero of its derivative.*

Proof. Exercise. □

Theorem 4.7.8 (**Lagrange's Theorem**). *Let f be a function defined and continuous on $[a, b]$ (with $a, b \in \mathbb{R}$ and $a < b$) and differentiable on $]a, b[$. Then there exists at least one point $c \in]a, b[$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

Proof. Let $\lambda = \frac{f(b)-f(a)}{b-a}$. We have $f(b) - \lambda b = f(a) - \lambda a$. The function $h(x) = f(x) - \lambda x$ is continuous on $[a, b]$, differentiable on $]a, b[$ and satisfies $h(a) = h(b)$. Thus, by Rolle's Theorem (see Theorem 4.7.6), there exists $c \in]a, b[$ such that $h'(c) = 0$, i.e., $f'(c) = \frac{f(b)-f(a)}{b-a}$. □

4.8 Higher-order derivatives

Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable function on $D_1 \subseteq D$. It is known that $f' : D_1 \subseteq \mathbb{R} \to \mathbb{R}$ is a real function defined on $D_1 \subseteq D$. If the function f' is differentiable at $a \in D_1$, then the function f is twice differentiable at a. The second derivative of f at a is denoted by $f''(a)$ or $\frac{d^2 f}{dx^2}(a)$ and is naturally defined by

$$f''(a) = \lim_{x \to a} \frac{f'(x) - f'(a)}{x - a} ,$$

or, considering the change of variable $x - a = h$,

$$f''(a) = \lim_{h \to 0} \frac{f'(a+h) - f'(a)}{h}\,.$$

More generally, if the derivatives $f', f'', \ldots, f^{(n-1)}$ exist in D, where $f^{(n-1)}$ represents the $(n-1)$-th derivative of f for $n \in \mathbb{N}$ with $n \geq 2$, and if the function $f^{(n-1)}$ is differentiable at $a \in D_n \subseteq D_{n-1} \subseteq \ldots \subseteq D_1 \subseteq D$, then it is said that f has a derivative of order n at the point $a \in D_n$. It holds that

$$f^{(n)}(a) = \lim_{x \to a} \frac{f^{(n-1)}(x) - f^{(n-1)}(a)}{x - a}\,.$$

If f is a differentiable function on D and f' is continuous also on D, then it is said that f is a C^1 function on D, and it is written $f \in C^1(D)$. Similarly, if f is twice differentiable on D and f' and f'' are continuous functions also on D, then it is said that f is a C^2 function on D, and it is written $f \in C^2(D)$. More generally, if f is n times differentiable on D and the functions $f', f'', \ldots, f^{(n)}$ are continuous also on D, then it is said that f is an C^n function on D, and it is written $f \in C^n(D)$.

One can write $f \in C^0(D)$ to denote that f is a continuous function on D.

In the case where the function f has derivatives of all orders on D, it is said that f is indefinitely differentiable on D or that f is of class $C^\infty(D)$, and it is written $f \in C^\infty(D)$.

Remark 4.8.1. *If $f \in C^n(D)$, then $f \in C^k(D)$ for every $k \leq n$.*

Example 4.8.2.

1) $f(x) = e^x$ is a $C^\infty(\mathbb{R})$ function.

2) $f(x) = x^n$, with $n \in \mathbb{N}$, is a $C^\infty(\mathbb{R})$ function.

The easiest functions to study from the point of view of mathematical analysis are polynomial functions. Given any function, there is a method that allows, under certain conditions, to find a polynomial that, in the neighbourhood of a point in its domain, has a graph that is "practically the same" as the graph of the given function. This method, given by the following theorem, is known as *Taylor's Formula*, and the resulting polynomial is known as the *Taylor Polynomial*.

Theorem 4.8.3 (**Taylor's Formula**). *Let $f :]a,b[\to \mathbb{R}$ be a function that is $n+1$ times differentiable in $]a,b[$, with $a, b \in \mathbb{R}$ and $a < b$. If $x, x_0 \in]a,b[$, then there exists at least one s between x and x_0 such that*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_n(x)\,,$$

where

$$R_n(x) = \frac{f^{(n+1)}(s)}{(n+1)!}(x - x_0)^{(n+1)}.$$

Definition 4.8.4 (**Taylor Polynomial**). *The polynomial*

$$P^n_{x_0}(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

is called the ***Taylor polynomial of order*** n ***of*** f ***at the point*** x_0*, and the function* $R_n(x)$ *is called the* ***Lagrange remainder of order*** n*.*

Remark 4.8.5.

1) *In the proof of Theorem 4.8.3, it is shown that* $\lim_{n\to+\infty} R_n(x) = 0$*.*

2) *If* $x_0 = 0$*, the Taylor formula of order* n *of* f *at the origin is written as*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x),$$

and is called the ***Maclaurin formula of order*** n ***of*** f*.*

Example 4.8.6. 1) *The Taylor polynomial of order* n *of* $f(x) = e^x$ *at the point* $x_0 = 0$ *is given by*

$$P^n_0(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}.$$

2) *The Taylor polynomial of order* n *of* $f(x) = \sin x$ *at the point* $x_0 = 0$ *is given by*

$$P^n_0(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} \cdots + (-1)^k \frac{x^{2k+1}}{2k+1!},$$

with $n = 2k + 1$ *and* $k \in \mathbb{N}$*.*

3) *The Taylor polynomial of order* n *of* $f(x) = \cos x$ *at the point* $x_0 = 0$ *is given by*

$$P^n_0(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \cdots + (-1)^k \frac{x^{2k}}{2k!},$$

with $n = 2k$ *and* $k \in \mathbb{N}$*.*

4.9 Classification of Critical Points

Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable function, and let $a \in \text{int}(D)$ be a local extremum of f. Then a is a critical point of f, i.e., $f'(a) = 0$. However, not all critical points of a function are necessarily points of extremum.

Proposition 4.9.1. *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a twice differentiable function at a point $a \in \text{int}(D)$, a critical point of f. Then:*

1) if $f''(a) > 0$, a is a local minimum point;

2) if $f''(a) < 0$, a is a local maximum point.

Proof. See [10]. □

More generally, the following proposition holds:

Proposition 4.9.2. *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be an n times differentiable function at a point $a \in \text{int}(D)$, a critical point of f. Denoting by $f^{(n)}$, with $n > 1$, the first of the successive derivatives of f that do not vanish at a, i.e.,*

$$f'(a) = f''(a) = \cdots = f^{(n-1)}(a) = 0 \quad \text{and} \quad f^{(n)}(a) \neq 0\,,$$

then:

1) if n is even:

1.1) if $f^{(n)}(a) > 0$, then a is a local minimum point;

1.2) if $f^{(n)}(a) < 0$, then a is a local maximum point;

2) if n is odd, then a is not a local extremum point.

Proof. See [10]. □

4.10 Concavities and Inflections

Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable function at $a \in \text{int}(D)$. The graph of f then has a tangent line at the point $(a, f(a))$, where, intuitively, it makes sense to say that the graph is concave down (respectively, concave up) at the point a if in some neighbourhood of a the graph is "below" (respectively, "above") the tangent line at $(a, f(a))$.

The equation of the **tangent line** to the graph of the function f at the point $(a, f(a))$ is given by

$$f(x) = f(a) + f'(a)(x - a)\,.$$

Definition 4.10.1. *It is said that:*

1) *the graph of f* ***is concave up*** *at the point a if there exists $\varepsilon > 0$ such that $f(x) > f(a) + f'(a)(x - a)$ for all $x \in \mathcal{B}_\varepsilon(a) \setminus \{a\}$;*

2) *the graph of f* ***is concave down*** *at the point a if there exists $\varepsilon > 0$ such that $f(x) < f(a) + f'(a)(x - a)$ for all $x \in \mathcal{B}_\varepsilon(a) \setminus \{a\}$;*

3) *the graph of f has a* ***point of inflection*** *at the point a if there exists $\varepsilon > 0$ such that in one of the intervals $]a-\varepsilon, a[$ or $]a, a+\varepsilon[$ the condition $f(x) > f(a) + f'(a)(x - a)$ holds, and in the other $f(x) < f(a) + f'(a)(x - a)$.*

The question then arises as to how to analyse the concavity of the graph of the function f at the point $a \in \operatorname{int}(D)$.

If f is at least twice differentiable at the point a and assuming that $f''(a) \neq 0$, by the first-order Taylor expansion of f at the point a, we have

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(s)}{2!}(x - a)^2 \,,$$

with s between a and x, and $f''(s)$ has the same sign as $f''(a)$.

Since $\frac{(x-a)^2}{2!} > 0$ for all x in a neighbourhood $\mathcal{B}_\varepsilon(a) \setminus \{a\}$, it follows that:

1) if $f''(a) > 0$, then $f(x) > f(a) + f'(a)(x - a)$;

2) if $f''(a) < 0$, then $f(x) < f(a) + f'(a)(x - a)$,

from which we can deduce the following proposition.

Proposition 4.10.2. *If $f : D \subseteq \mathbb{R} \to \mathbb{R}$ is twice differentiable at $a \in \operatorname{int}(D)$, then:*

1) *if $f''(a) > 0$, the graph of f is concave up at the point a;*

2) *if $f''(a) < 0$, the graph of f is concave down at the point a.*

Proof. See [10]. □

Corolary 4.10.3. *If $f : D \subseteq \mathbb{R} \to \mathbb{R}$ is twice differentiable at $a \in \operatorname{int}(D)$ and a is an inflection point of f, then $f''(a) = 0$.*

Proof. See [10]. □

Remark 4.10.4. *The converse of Corollary 4.10.3 is not true. Consider the example of the function $f(x) = x^4$. We have that $f'(x) = 4x^3$ and $f''(x) = 12x^2$, so $f''(0) = 0$. However, $x = 0$ is not an inflection point but a minimum point. In such cases, one can analyse the function around the critical point. In this example, it can be checked that $f''(x) > 0$ for all $x \in \mathbb{R} \setminus \{0\}$, from which it can be concluded that $x = 0$ is the global minimum point of f.*

More generally, the following proposition holds.

Proposition 4.10.5. *Let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be an n times differentiable function at a point $a \in \operatorname{int}(D)$. Denoting by $f^{(n)}$, with $n > 2$, the first of the successive derivatives of f that do not vanish at a, i.e.,*

$$f''(a) = \cdots = f^{(n-1)}(a) = 0 \quad \text{and} \quad f^{(n)}(a) \neq 0\,,$$

then:

1) if n is even, we have:

1.1) if $f^{(n)}(a) > 0$, the graph of f is concave up at the point a;

1.2) if $f^{(n)}(a) < 0$, the graph of f is concave down at the point a;

2) if n is odd, then a is an inflection point.

Proof. See [6]. □

Example 4.10.6. *The function $f(x) = x^5$ is an example where Proposition 4.10.5 can be applied.*

Chapter 5

Integral calculus

5.1 The indefinite integral

Definition 5.1.1. *Let $I \subseteq \mathbb{R}$ be an interval containing more than one point, and let $f : I \longrightarrow \mathbb{R}$ be a real-valued function of a real variable. An* ***indefinite integral of*** f ***on*** I *is any function $F : I \longrightarrow \mathbb{R}$ such that*

$$F'(x) = f(x) \quad \forall x \in I\,.$$

Proposition 5.1.2. *A continuous function on an interval I (with more than one point) has an indefinite integral on I.*

Proof. See [6]. □

Remark 5.1.3. *If F is an indefinite integral of f on I, then for any constant $C \in \mathbb{R}$, $F+C$ is another indefinite integral of f on I, since $(F(x)+C)' = f(x)$ for all $x \in I$. From this, it can be deduced that if f has an indefinite integral on I, then f has infinitely many indefinite integrals on I.*

The symbol $\int$ is used to denote the integral, and can be written as

$$\int f(x)\,dx\,,$$

where $\displaystyle\int$ is the integral sign, f is the integrand, and dx indicates the integration variable x.

Proposition 5.1.4. *If F_1 and F_2 are indefinite integrals of f on I, then there exists a constant $C \in \mathbb{R}$ such that $F_1(x) - F_2(x) = C$ for all $x \in I$.*

Proof. We have that

$$(F_1(x) - F_2(x))' = F_1'(x) - F_2'(x) = f(x) - f(x) = 0\,,$$

so there exists a constant $C \in \mathbb{R}$ such that $F_1(x) - F_2(x) = C$. □

Proposition 5.1.5. *Let f and g be functions with indefinite integrals on $I \subseteq \mathbb{R}$ and let $\alpha \in \mathbb{R}$. Then:*

1) $\displaystyle\int f(x) + g(x)\, dx = \int f(x)\, dx + \int g(x)\, dx$;

2) $\displaystyle\int \alpha f(x)\, dx = \alpha \int f(x)\, dx$.

5.2 Immediate Antiderivatives

1) $\displaystyle\int x^\alpha\, dx = \frac{x^{\alpha+1}}{\alpha+1} + C$ for all $\alpha \in \mathbb{R} \setminus \{-1\}$ and any $C \in \mathbb{R}$.

   Example 5.2.1. $\displaystyle\int x^5\, dx = \frac{x^6}{6} + C$, *for any $C \in \mathbb{R}$.*

2) Let f be a non-zero and differentiable function on I. We have that

$$\ln|f(x)| = \begin{cases} \ln f(x) & , \text{ if } f(x) > 0 \\ \ln(-f(x)) & , \text{ if } f(x) < 0 \end{cases}.$$

   Then

$$(\ln|f(x)|)' = \begin{cases} \frac{f'(x)}{f(x)} & , \text{ if } f(x) > 0 \\ \frac{-f'(x)}{-f(x)} & , \text{ if } f(x) < 0 \end{cases} = \frac{f'(x)}{f(x)}.$$

   From this, it can be deduced that

$$\int \left(\frac{f'(x)}{f(x)}\right) dx = \ln|f(x)| + C, \quad \text{for any } C \in \mathbb{R}.$$

   Example 5.2.2.

   i) $\displaystyle\int \left(\frac{1}{x}\right) dx = \ln|x| + C$ *in $I \subseteq \mathbb{R}$ such that $0 \notin I$, for any $C \in \mathbb{R}$.*

   ii) $\displaystyle\int \left(\frac{x}{1-2x^2}\right) dx = -\frac{1}{4}\int \left(\frac{-4x}{1-2x^2}\right) dx = -\frac{1}{4}\ln|1-2x^2| + C$ *in $I \subseteq \mathbb{R}$ such that $1 - 2x^2 \neq 0$, for any $C \in \mathbb{R}$.*

3) Let $\alpha \in \mathbb{R} \setminus \{-1\}$ and let f be a differentiable function on I. We have that

$$\left(f^{\alpha+1}(x)\right)' = (\alpha+1)\, f^\alpha(x) f'(x),$$

so

$$\int f'(x)f^{\alpha}(x)\,dx = \frac{f^{\alpha+1}(x)}{\alpha+1} + C,$$

for any $C \in \mathbb{R}$.

Example 5.2.3.

i) $\displaystyle\int \left(\frac{\ln x}{x}\right) dx = \int (\ln x)\left(\frac{1}{x}\right) dx = \frac{(\ln x)^2}{2} + C$ *in* $]0, +\infty[$, *for any* $C \in \mathbb{R}$.

ii) $\displaystyle\int x\sqrt{2x^2+1}\,dx = \frac{1}{4}\int 4x\left(2x^2+1\right)^{\frac{1}{2}}\,dx = \frac{1}{4}\frac{\left(2x^2+1\right)^{\frac{1}{2}+1}}{\frac{1}{2}+1} + C$
$\displaystyle= \frac{1}{6}\left(2x^2+1\right)^{\frac{3}{2}} + C$ *in* $\mathbb{R}$, *for any* $C \in \mathbb{R}$.

4) If f is differentiable on I, we have that $(\arcsin f(x))' = \frac{f'(x)}{\sqrt{1-f^2(x)}}$, where

$$\int \frac{f'(x)}{\sqrt{1-f^2(x)}}\,dx = \arcsin f(x) + C,$$

for any $C \in \mathbb{R}$.

Analogously, it can be deduced that

$$\int \frac{-f'(x)}{\sqrt{1-f^2(x)}}\,dx = \arccos f(x) + C,$$

for any $C \in \mathbb{R}$, and

$$\int \frac{f'(x)}{1+f^2(x)}\,dx = \arctan f(x) + C,$$

for any $C \in \mathbb{R}$.

Example 5.2.4.

i) $\displaystyle\int \frac{1}{\sqrt{1-4x^2}}\,dx = \frac{1}{2}\int \frac{2}{\sqrt{1-(2x)^2}}\,dx = \frac{1}{2}\arcsin(2x) + C$, *in* $]-\frac{1}{2}, \frac{1}{2}[$, *for any* $C \in \mathbb{R}$.

ii) $\displaystyle\int \frac{e^x}{e^{2x}+4}\,dx = \int \frac{e^x}{4+(e^x)^2}\,dx = \frac{2}{4}\int \frac{\frac{e^x}{2}}{1+\left(\frac{e^x}{2}\right)^2}\,dx = \frac{1}{2}\arctan\left(\frac{e^x}{2}\right) + C$, *in* $\mathbb{R}$, *for any* $C \in \mathbb{R}$.

5.3 Rational Functions

Recall that a **rational function** is any function that can be written in the form $\frac{P(x)}{Q(x)}$, where $P(x)$ and $Q(x)$ are polynomials with real coefficients.

Proposition 5.3.1. *If degree $(P(x)) \geq$ degree $(Q(x))$, then there exist polynomials $C(x)$ and $R(x)$ such that*

$$\frac{P(x)}{Q(x)} = C(x) + \frac{R(x)}{Q(x)}\,,$$

with degree $(R(x)) <$ degree $(Q(x))$.

Proof. See [6]. ☐

Example 5.3.2.

$$\frac{2x^3+3x^2}{x^2+2x} = 2x - 1 + \frac{2x}{x^2+2x}\,,$$

Remark 5.3.3. *1) Under the conditions of Proposition 5.3.1, we have*

$$\int \frac{P(x)}{Q(x)}\,dx = \int C(x)\,dx + \int \frac{R(x)}{Q(x)}\,dx\,.$$

Since the indefinite integral of the polynomial $C(x)$ is immediate, it suffices to compute the indefinite integral of rational functions $\frac{P(x)}{Q(x)}$ where degree $(P(x)) <$ degree $(Q(x))$.

2) If $Q'(x) = P(x)$, then $\displaystyle\int \frac{P(x)}{Q(x)}\,dx = \ln|Q(x)| + C$, for any $C \in \mathbb{R}$.

5.3.1 Factorisation of Polynomials

Proposition 5.3.4. *Any polynomial with real coefficients of degree greater than or equal to 1 can be written as a product of degree 1 polynomials (factors corresponding to real roots) and degree 2 polynomials (factors corresponding to complex roots).*

Example 5.3.5. *1) $x^2 + 2x = x(x+2)$.*

2) $x^3 + 2x^2 + 5x = x(x^2 + 2x + 5)$. Note that the polynomial $x^2 + 2x + 5$ cannot be factorised into degree 1 polynomials with real coefficients since it has two complex roots.

5.3.2 Partial Fractions

In computing the indefinite integral of rational functions, a very useful technique is the decomposition of the rational function into a sum of partial fractions.

Definition 5.3.6. *A* ***partial fraction*** *is a rational function of the form*

$$\frac{A}{(x-r)^s} \quad or \quad \frac{Bx+C}{x^2+bx+c},$$

where s is a natural number and A, B, C, r, b, c are real constants, such that x^2+bx+c has no real roots.

Example 5.3.7.

$$4\frac{2}{x^2+2x} = \frac{2}{x(x+2)} = \frac{A}{x} + \frac{B}{x+2}.$$

Looking at Example 5.3.7, one might ask how to determine the constants A and B such that the rational function $\frac{2}{x^2+2x}$ can be written as the sum of two partial fractions, $\frac{A}{x} + \frac{B}{x+2}$. Usually, the method called the **method of undetermined coefficients** is applied.

In this case, since the denominator of the rational function $\frac{2}{x^2+2x}$ is a degree 2 polynomial with two distinct real roots, we can factorise x^2+2x as the product of two degree 1 polynomials: $x^2+2x = x(x+2)$. We then want to find constants A and B such that

$$\frac{2}{x(x+2)} = \frac{A}{x} + \frac{B}{x+2}.$$

Bringing the right-hand side to a common denominator, we get

$$\frac{A(x+2)+Bx}{x(x+2)} = \frac{(A+B)x+2A}{x(x+2)},$$

so that

$$\frac{2}{x(x+2)} = \frac{(A+B)x+2A}{x(x+2)},$$

which holds if and only if

$$\begin{cases} A+B=0 \\ 2A=2 \end{cases} \iff \begin{cases} B=-1 \\ A=1 \end{cases},$$

and thus the decomposition is

$$\frac{2}{x^2+2x} = \frac{1}{x} - \frac{1}{x+2}.$$

With this decomposition, the indefinite integral of the rational function becomes immediate:

$$\int \frac{2}{x^2+2x}\,dx = \int \frac{1}{x}\,dx - \int \frac{1}{x+2}\,dx = \ln|x| - \ln|x+2| + C,$$

for any $C \in \mathbb{R}$.

Example 5.3.8.

$$\frac{2x^2+1}{x^3+x^2} = \frac{2x^2+1}{x^2(x+1)} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x+1}.$$

In this case, the denominator $x^3 + x^2$ has two real roots, $x = 0$ and $x = -1$, where $x = 0$ is a double root and $x = -1$ is a simple root. Thus, in the decomposition into partial fractions, three terms should appear: two associated with the root $x = 0$ and one associated with the root $x = -1$. Bringing the sum of the three terms to a common denominator, we get

$$\frac{2x^2+1}{x^2(x+1)} = \frac{Ax(x+1)+B(x+1)+Cx^2}{x^2(x+1)} = \frac{(A+C)x^2+(A+B)x+B}{x^2(x+1)},$$

which holds if and only if

$$\begin{cases} A+C=2 \\ A+B=0 \\ B=1 \end{cases} \iff \begin{cases} C=3 \\ A=-1 \\ B=1 \end{cases},$$

yielding the decomposition

$$\frac{2x^2+1}{x^3+x^2} = -\frac{1}{x} + \frac{1}{x^2} + \frac{3}{x+1}.$$

With this decomposition, the indefinite integral of the rational function becomes immediate:

$$\int \frac{2x^2+1}{x^3+x^2}dx = \int -\frac{1}{x}dx + \int \frac{1}{x^2}dx + \int \frac{3}{x+1}dx = -\ln|x| - \frac{1}{x} + 3\ln|x+1| + C$$

for any $C \in \mathbb{R}$.

Example 5.3.9.

$$\frac{4x^2+3x+5}{x^3+2x^2+5x} = \frac{4x^2+3x+5}{x(x^2+2x+5)} = \frac{A}{x} + \frac{Bx+C}{x^2+2x+5}.$$

In this case, the denominator $x^3 + 2x^2 + 5x$ has one real root, $x = 0$, and two complex roots. Thus, the decomposition into partial fractions should include two terms: one associated with the root $x = 0$ and one associated with the complex roots. For the term associated with the complex roots, i.e., the polynomial x^2+2x+5, the numerator is the degree 1 polynomial $Bx+C$. Bringing the sum of the two terms to a common denominator, we get

$$\frac{4x^2+3x+5}{x^3+2x^2+5x} = \frac{A(x^2+2x+5)+(Bx+C)x}{x(x^2+2x+5)} = \frac{(A+B)x^2+(2A+C)x+5A}{x(x^2+2x+5)},$$

which holds if and only if

$$\begin{cases} A+B=4 \\ 2A+C=3 \\ 5A=5 \end{cases} \iff \begin{cases} B=3 \\ C=1 \\ A=1 \end{cases},$$

so the decomposition is

$$\frac{4x^2+3x+5}{x^3+2x^2+5x} = \frac{1}{x} + \frac{3x+1}{x^2+2x+5}.$$

Thus, the indefinite integral of the rational function is

$$\int \frac{4x^2+3x+5}{x^3+2x^2+5x}\,dx = \int \frac{1}{x}\,dx + \int \frac{3x+1}{x^2+2x+5}\,dx = \ln|x| + \int \frac{3x+1}{x^2+2x+5}\,dx,$$

and we still need to compute the indefinite integral of the partial fraction associated with the complex roots. Let us then see how to compute the indefinite integral of such partial fractions. Start by observing the following proposition.

Proposition 5.3.10. *If the polynomial x^2+bx+c has no real roots, then there exist real constants α and β such that*

$$x^2+bx+c = (x-\alpha)^2+\beta^2.$$

Proof. Exercise. □

Considering the partial fraction $\frac{Bx+C}{x^2+bx+c}$ where x^2+bx+c has no real roots, we have two cases: $B=0$ (and $C \neq 0$), or $B \neq 0$.

1) If $B=0$, by Proposition 5.3.10, we have

$$\frac{C}{x^2+bx+c} = \frac{C}{\beta^2+(x-\alpha)^2},$$

hence

$$\int \frac{C}{x^2+bx+c}\,dx = \int \frac{C}{\beta^2+(x-\alpha)^2}\,dx = \frac{C}{\beta}\int \frac{1}{1+\left(\frac{x-\alpha}{\beta}\right)^2}\,dx = \frac{C}{\beta}\arctan\left(\frac{x-\alpha}{\beta}\right)+K,$$

for any $K \in \mathbb{R}$.

2) If $B \neq 0$, since $(x^2+bx+c)' = 2x+b$, we write

$$\frac{Bx+C}{x^2+bx+c} = \frac{\frac{B}{2}(2x+b)-\frac{Bb}{2}+C}{x^2+bx+c} = \frac{B}{2}\frac{2x+b}{x^2+bx+c} + \frac{C-\frac{Bb}{2}}{x^2+bx+c},$$

and so

$$\begin{aligned}\int \frac{Bx+C}{x^2+bx+c}\,dx &= \frac{B}{2}\int \frac{2x+b}{x^2+bx+c}\,dx + \int \frac{C-\frac{Bb}{2}}{x^2+bx+c}\,dx \\ &= \frac{B}{2}\ln\left(x^2+bx+c\right) + \frac{C-\frac{Bb}{2}}{\beta}\arctan\left(\frac{x-\alpha}{\beta}\right)+K,\end{aligned}$$

for any $K \in \mathbb{R}$.

Example 5.3.11. *This example completes the calculation of the indefinite integral of the rational function given in Example 5.3.9. Namely, from what has been presented previously, we have*

$$\frac{3x+1}{x^2+2x+5} = \frac{3}{2}\frac{2x+2}{x^2+2x+5} - \frac{2}{x^2+2x+5},$$

so that

$$\begin{aligned}\int \frac{3x+1}{x^2+2x+5}\,dx &= \frac{3}{2}\int \frac{2x+2}{x^2+2x+5}\,dx - \int \frac{2}{x^2+2x+5}\,dx \\ &= \frac{3}{2}\ln\left(x^2+2x+5\right) - \arctan\left(\frac{x+1}{2}\right)+K,\end{aligned}$$

for any $K \in \mathbb{R}$, since the roots of x^2+2x+5 are $x=-1\pm 2i$, and we take $\alpha=-1$ and $\beta=2$.

Exercise 5.3.12. *Compute the indefinite integral of the following rational* *functions:*

1) $\dfrac{1}{x^2-2x+1}$ *Sol:* $-\dfrac{1}{x-1}+C,$ *for any* $C\in\mathbb{R}$
2) $\dfrac{1}{x^2+x-2}$ *Sol:* $-\dfrac{1}{3}\ln|x+2|+\dfrac{1}{3}\ln|x-1|+C,$ *for any* $C\in\mathbb{R}$
3) $\dfrac{1}{x^2+2x+3}$ *Sol:* $\dfrac{\sqrt{2}}{2}\arctan\left(\dfrac{x+1}{\sqrt{2}}\right)+C,$ *for any* $C\in\mathbb{R}$
4) $\dfrac{1}{x(1+x^2)}$ *Sol:* $\ln|x|-\dfrac{1}{2}\ln(1+x^2)+C,$ *for any* $C\in\mathbb{R}$
5) $\dfrac{x+1}{x(1+x^2)}$ *Sol:* $\ln|x|-\dfrac{1}{2}\ln(1+x^2)+\arctan x+C,$ *for any* $C\in\mathbb{R}$
6) $\dfrac{x}{x-2}$ *Sol:* $x+2\ln|x-2|+C,$ *for any* $C\in\mathbb{R}$
7) $\dfrac{x^2+2x+3}{x^2-1}$ *Sol:* $x+\ln|x^2-1|+2\ln|x-1|-2\ln|x+1|+C,$ *for any* $C\in\mathbb{R}$.

5.4 Integration by Parts

The rule usually referred to as **integration by parts** applies when the computation of the indefinite integral of certain functions becomes significantly easier when the integrand can be viewed as the product of two functions. This rule is presented in the following theorem.

Theorem 5.4.1 (**Integration by Parts**). *Let f and g be differentiable functions on I. The product $f'g$ is integrable on I if and only if the product fg' is has an indifenite integral on I. In this case, we have the so-called **integration by parts formula**:*

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int f(x)g'(x)\,dx. \tag{5.1}$$

Proof. Exercise. Hint: use the product rule for differentiation. □

Example 5.4.2.

1) *In computing the indefinite integral of $x\sin x$, taking $g(x)=x$, $f'(x)=\sin x$ and applying (5.1), we have*

$$\begin{aligned}\int x\sin x\,dx &= -x\cos x - \int -\cos x\,dx\\ &= -x\cos x + \int \cos x\,dx\\ &= -x\cos x + \sin x + C\,,\end{aligned}$$

for any $C\in\mathbb{R}$.

2) In computing the indefinite integral of $\ln x$*, taking* $g(x) = \ln x$*,* $f'(x) = 1$ *and applying* (5.1)*, we have*

$$\begin{aligned}\int \ln x \, dx &= \int 1. \ln x \, dx \\ &= x \ln x - \int x . \frac{1}{x} \, dx \\ &= x \ln x - x + C \,,\end{aligned}$$

for any $C \in \mathbb{R}$*.*

3) In computing the indefinite integral of $\sin^2 x$*, taking* $g(x) = \sin x$*,* $f'(x) = \sin x$ *and applying* (5.1)*, we have*

$$\begin{aligned}\int \sin^2 x \, dx &= -\sin x \cos x + \int \cos^2 x \, dx \\ &= -\sin x \cos x + \int 1 - \sin^2 x \, dx \\ &= -\sin x \cos x + x - \int \sin^2 x \, dx \,,\end{aligned}$$

from which, by adding $\int \sin^2 x \, dx$ *to both sides, we obtain*

$$2 \int \sin^2 x \, dx = x - \sin x \cos x + C,$$

for any $C \in \mathbb{R}$*, that is,*

$$\int \sin^2 x \, dx = \frac{x}{2} - \frac{1}{2} \sin x \cos x + C,$$

for any $C \in \mathbb{R}$*.*

5.5 Integration by Substitution

The computation of the indefinite integral of certain functions can become significantly simpler by performing a change of variable in the integrand. This strategy for computing indefinite integrals is commonly referred to as the **integration by substitution rule**, and it is presented in the following theorem.

Theorem 5.5.1 (**Integration by Substitution**). *Let* $f : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ *be a function, and let* $\varphi : J \subseteq \mathbb{R} \longrightarrow I$ *be a bijective and differentiable function* *such that for $t \in J$, $\varphi'(t) \neq 0$. If the function $f(\varphi(t))\,\varphi'(t)$ has an indefinite integral, then $f(x)$ also has an indefinite integral, and in that case we have the so-called* ***integration by substitution rule***

$$\int f(x)\,dx = \int f(\varphi(t))\,\varphi'(t)\,dt, \tag{5.2}$$

where $t = \varphi^{-1}(x)$.

Proof. Consider $g(t) = f(\varphi(t))\,\varphi'(t)$. Let G be an indefinite integral of g. Observing that $\varphi(t) = x$, we can view g as a function of x, so that $g(x) = f(x)\varphi'(\varphi^{-1}(x))$, and hence G can also be seen as a function of x. Then, applying the chain rule (Theorem 4.4.1),

$$\left(G \circ \varphi^{-1}\right)'(x) = G'\left(\varphi^{-1}(x)\right)\left(\varphi^{-1}\right)'(x),$$

from which, noting that $G' = g$ and applying the derivative rule for the inverse function (Theorem 4.5.1),

$$\left(G \circ \varphi^{-1}\right)'(x) = g(t)\frac{1}{\varphi'(t)} = f(\varphi(t)) = f(x),$$

and thus

$$\int f(x)\,dx = \int \left(G \circ \varphi^{-1}\right)'(x)\,dx = \left(G \circ \varphi^{-1}\right)(x) = G(t) = \int g(t)\,dt,$$

which proves the result. □

Example 5.5.2.

1) *In computing the indefinite integral of $\frac{x}{1+\sqrt{x}}$, whith $x \in \mathbb{R}_0^+$, consider $\varphi : \mathbb{R}_0^+ \longrightarrow \mathbb{R}$ defined by $\varphi(t) = t^2$, and applying (5.2), we have*

$$\begin{aligned}
\int \frac{x}{1+\sqrt{x}}\,dx &= \int \frac{t^2}{1+t}\,2t\,dt = \int \frac{2t^3}{1+t}\,dt \\
&= \int 2t^2 - 2t + 2 - \frac{2}{1+t}\,dt \\
&= 2\frac{t^3}{3} - t^2 + 2t - 2\ln(1+t) + C \\
&= \frac{2}{3}\left(\sqrt{x}\right)^3 - x + 2\sqrt{x} - 2\ln\left(1+\sqrt{x}\right) + C,
\end{aligned}$$

for any $C \in \mathbb{R}$, since

$$x = \varphi(t) \;\Rightarrow\; x = t^2 \;\Rightarrow\; t = \sqrt{x},$$

as $t \in \mathbb{R}_0^+$.

2) *In computing the indefinite integral of* $\sqrt{a^2 - x^2}$, *with* $a \in \mathbb{R}^+$ *and* $x \in I = [-a, a]$, *consider* $\varphi : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow [-a, a]$ *defined by* $\varphi(t) = a \sin t$, *and applying* (5.2), *we have*

$$\begin{aligned}
\int \sqrt{a^2 - x^2}\, dx &= \int \sqrt{a^2 - (a\sin(t))^2}\, a\cos(t)\, dt = \int a^2 \cos^2(t)\, dt \\
&= a^2 \int \cos^2(t)\, dt = a^2 \int 1 - \sin^2(t)\, dt = a^2 t - \int \sin^2(t)\, dt \\
&= a^2 \left(t - \frac{t}{2} + \frac{1}{2}\sin(t)\cos(t)\right) + C \\
&= a^2 \left(\frac{t}{2} + \frac{1}{2}\sin(t)\cos(t)\right) + C \\
&= \frac{a^2}{2}\left(\arcsin\left(\frac{x}{a}\right) + \frac{x}{a}\frac{\sqrt{a^2 - x^2}}{a}\right) + C \\
&= \frac{a^2}{2}\arcsin\left(\frac{x}{a}\right) + \frac{x}{2}\sqrt{a^2 - x^2} + C,
\end{aligned}$$

for any $C \in \mathbb{R}$, *since*

$$\begin{aligned}
x = \varphi(t) \quad &\Rightarrow \quad x = a\sin(t) \;\Rightarrow\; \sin(t) = \frac{x}{a} \\
&\Rightarrow \quad t = \arcsin\left(\frac{x}{a}\right) \quad \textit{and} \quad \cos(t) = \sqrt{1 - \left(\frac{x}{a}\right)^2} = \frac{\sqrt{a^2 - x^2}}{a}.
\end{aligned}$$

5.6 The definite integral

The integral of a non-negative function f, i.e., such that $f(D_f) \subseteq \mathbb{R}_0^+$, computed over an interval $[a, b] \subseteq D_f$, can be interpreted as the area of the region in the plane bounded by the graph of f, the x-axis, and the vertical lines $x = a$ and $x = b$, as illustrated in Figure 5.1.

Definition 5.6.1. *Let* $f : [a, b] \longrightarrow \mathbb{R}$ *be a continuous function, with* $a, b \in \mathbb{R}$ *and* $a < b$. *The* ***definite integral*** *of the function* f *over the interval* $[a, b]$, *denoted by* $\int_a^b f(x)\, dx$, *is the real number* $F(b) - F(a)$, *where* F *is an indefinite integral of* f *on* $[a, b]$.

This gives the so-called ***Barrow's formula***

$$\int_a^b f(x)\, dx = [F(x)]_a^b = F(b) - F(a)\,,$$

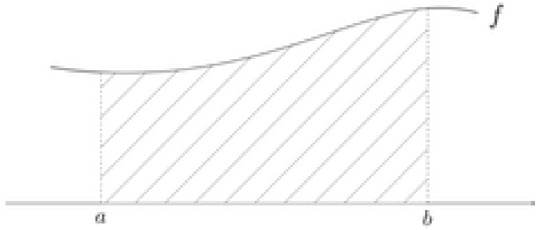

Figure 5.1: Shaded representation of the region bounded by the graph of a function f, the x-axis, and the vertical lines $x = a$ and $x = b$.

where $[F(x)]_a^b$ *denotes the evaluation of* $F(b) - F(a)$.

Proposition 5.6.2. *Let* $f, g : [a, b] \longrightarrow \mathbb{R}$ *be continuous functions, let* $k \in \mathbb{R}$ *and* $c \in]a, b[$. *Then:*

1. $\displaystyle\int_a^b k f(x)\, dx = k \int_a^b f(x)\, dx\,.$

2. $\displaystyle\int_a^b (f + g)(x)\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx\,.$

3. $\displaystyle\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx\,.$

Proof. See [10]. □

Remark 5.6.3.

1. $\displaystyle\int_a^a f(x)\, dx = 0\,.$

   Proof. Exercise. □

2. $\displaystyle\int_a^c f(x)\, dx = -\int_c^a f(x)\, dx\,.$

   Proof. Exercise. □

Theorem 5.6.4 (Definite integral by parts). *Let* $f, g : [a, b] \longrightarrow \mathbb{R}$ *be functions with continuous derivatives on* $[a, b]$ *(i.e.,* $f, g \in C^1([a, b])$*). Then*

$$\int_a^b f'(x)\, g(x)\, dx = [f(x)\, g(x)]_a^b - \int_a^b f(x)\, g'(x)\, dx\,.$$

Proof. See [10]. □

Example 5.6.5.

$$\begin{aligned}\int_0^{\frac{\pi}{2}} x\,\sin x\,dx &= [-x\,\cos x]_0^{\frac{\pi}{2}} + \int_0^{\frac{\pi}{2}} \cos x\,dx \\ &= [-x\,\cos x + \sin x]_0^{\frac{\pi}{2}} = 1\end{aligned}$$

Theorem 5.6.6 (Definite integral by substitution). *Let $f : [a,b] \longrightarrow \mathbb{R}$ and let $\varphi : [c,d] \longrightarrow [a,b]$ be a bijective and differentiable function such that for $t \in [c,d]$, $\varphi'(t) \neq 0$. If the function $f(\varphi(t))\,\varphi'(t)$ is integrable on $[c,d]$, then $f(x)$ is integrable on $[a,b]$, and in that case we have*

$$\int_a^b f(x)\,dx = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} f(\varphi(t))\,\varphi'(t)\,dt\,.$$

Proof. See [10]. □

Example 5.6.7.

$$\begin{aligned}\int_{-1}^{1} \sqrt{1-x^2}\,dx &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1-\sin^2 t}\,\cos t\,dt \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 t\,dt \\ &= \left[\frac{t}{2} + \frac{1}{2}\sin t\cos t\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{\pi}{2}\end{aligned}$$

5.7 Area calculation

If $f(x) = c$, with $c \in \mathbb{R}^+$, and if $a < b$, we have

$$\int_a^b f(x)\,dx = \int_a^b c\,dx = [cx]_a^b = c\,[x]_a^b = c(b-a) > 0\,.$$

In fact, this integral of the constant function $f(x) = c$ calculated over the interval $[a,b]$ corresponds to the area of the region of the plane bounded by the graph of f, the x-axis, and the vertical lines $x = a$ and $x = b$, as shown in Figure 5.2.

And if $c < 0$? Let $g(x) = c$, with $c \in \mathbb{R}^-$. If $a < b$, then

$$\int_a^b g(x)\,dx = c(b-a) < 0\,.$$

Figure 5.2: Dashed representation of the region bounded by the graph of the constant function $f(x) = c > 0$, the x-axis, and the vertical lines $x = a$ and $x = b$.

In this case, the integral of the constant function $g(x) = c$, with $c < 0$, calculated over the interval $[a, b]$ corresponds to the area of the region of the plane bounded by the graph of g, the x-axis, and the vertical lines $x = a$ and $x = b$, but with a negative sign, as illustrated in Figure 5.3.

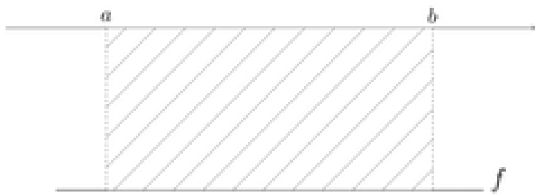

Figure 5.3: Dashed representation of the region bounded by the graph of the constant function $f(x) = c < 0$, the x-axis, and the vertical lines $x = a$ and $x = b$.

Now suppose that $f(x) = d$ and $g(x) = c$ with $d > c > 0$. In these conditions,

$$\int_a^b f(x) - g(x)\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx = d(b-a) - c(b-a) = (d-c)(b-a),$$

which corresponds to the area of the region of the plane bounded by the graphs of f and g, and the vertical lines $x = a$ and $x = b$, as represented in Figure 5.4.

More generally:

- If f is a continuous and positive function on $[a, b]$, the definite integral

  $$\int_a^b f(x)\,dx$$

  is the area of the region of the plane bounded by the graph of f, the x-axis, and the vertical lines $x = a$ and $x = b$, as shown in Figure 5.5.

Figure 5.4: Dashed representation of the region bounded between the graphs of the constant functions $f(x) = d$ and $g(x) = c$, the x-axis and the vertical lines $x = a$ and $x = b$.

- If g is a continuous and negative function on $[a, b]$, the definite integral

$$\int_c^d g(x)\, dx$$

is the area of the region of the plane bounded by the graph of f, the x-axis, and the vertical lines $x = a$ and $x = b$, but with a negative sign, as shown in Figure 5.5.

- If f and g are continuous and positive functions on $[a, b]$ such that $f > g$ over the interval $[a, b]$, then the definite integral

$$\int_a^b f(x) - g(x)\, dx$$

is the area of the region of the plane bounded between the graphs of f and g, the x-axis, and the vertical lines $x = a$ and $x = b$, as shown in Figure 5.6.

And what if the function f changes sign over the interval $[a, b]$? Let f be a continuous function defined on the interval $[a, b]$ such that $f(x) > 0$ for $x \in [a, c[$ and $f(x) < 0$ for $x \in]c, b]$. In that case, to calculate the area of the region of the plane bounded by the graph of the function f and the vertical lines $x = a$ and $x = b$, we need to compute the integral of the function $|f(x)|$ over the interval $[a, b]$, that is,

$$\int_a^b |f(x)|\, dx = \int_a^c f(x)\, dx - \int_c^b f(x)\, dx\,,$$

as illustrated in Figure 5.7.

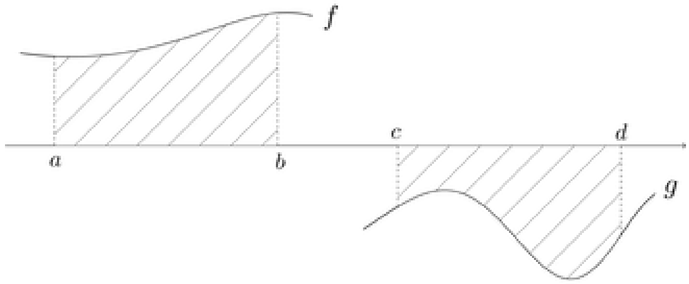

Figure 5.5: Dashed representation of the region bounded by the graph of the positive function f, the x-axis, and the vertical lines $x = a$ and $x = b$ (on the left), and the region bounded by the graph of the negative function g, the x-axis, and the vertical lines $x = a$ and $x = b$ (on the right).

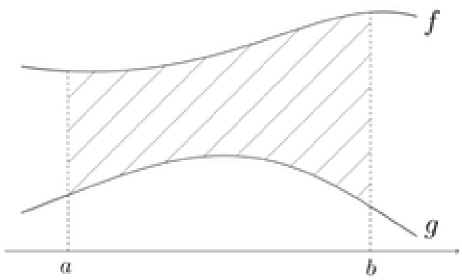

Figure 5.6: Dashed representation of the region bounded between the graphs of the positive functions f and g such that $f > g$ over the interval $[a, b]$, the x-axis, and the vertical lines $x = a$ and $x = b$.

Example 5.7.1. *Let $f(x) = \sin x$. Then*

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin x \, dx = [-\cos x]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 0\,,$$

and

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} |\sin x| \, dx = \int_{-\frac{\pi}{2}}^{0} -\sin x \, dx + \int_{0}^{\frac{\pi}{2}} \sin x \, dx = 2\,.$$

Example 5.7.2. *The area of the region of the plane defined by the set*

$$A = \{(x, y) \in \mathbb{R}^2 \,:\, x^2 \leq y \leq |x|\}\,,$$

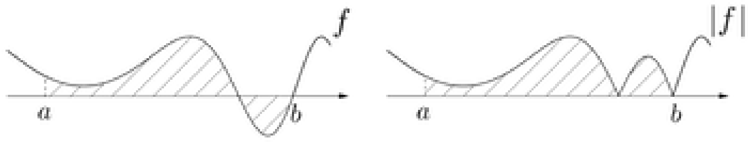

Figure 5.7: Dashed representation of the region bounded by the graph of the function f, the x-axis, and the vertical lines $x = a$ and $x = b$ (on the left), and the region bounded by the graph of the function $|f|$, the x-axis, and the vertical lines $x = a$ and $x = b$ (on the right).

geometrically represented in Figure 5.8, is

$$\begin{aligned}\int_{-1}^{1} |x| - x^2\,dx &= \int_{-1}^{0} |x| - x^2\,dx + \int_{0}^{1} |x| - x^2\,dx \\ &= \int_{-1}^{0} -x - x^2\,dx + \int_{0}^{1} x - x^2\,dx = \frac{1}{3}\,.\end{aligned}$$

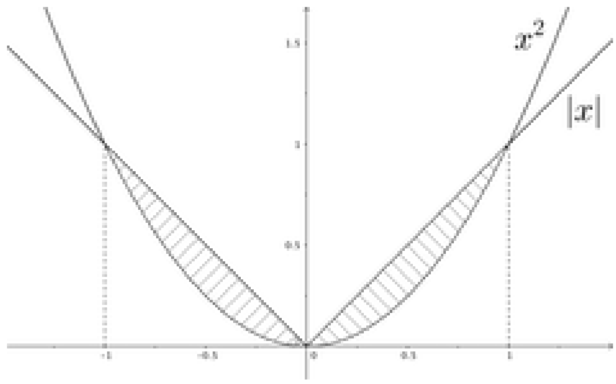

Figure 5.8: Dashed representation of the region of the plane defined by the set A.

5.8 Fundamental Theorem of Integral Calculus

Theorem 5.8.1 (**Fundamental Theorem of Integral Calculus**). *Let f be a continuous function on $[a, b]$. Then the function $F : [a, b] \longrightarrow \mathbb{R}$ defined by*

$$F(x) = \int_{a}^{x} f(t)\,dt$$

is differentiable, and for each $x \in [a,b]$, we have

$$F'(x) = f(x)\,.$$

Remark 5.8.2. *Thus, functions of the form $F(x) = \int_a^x f(t)\,dt$, where f is continuous, can be differentiated with $F'(x) = f(x)$ in any interval containing the point a.*

Example 5.8.3.

1) *The function $F(x) = \int_0^x \frac{t^3+1}{t^2+1}\,dt$ has derivative $F'(x) = \frac{x^3+1}{x^2+1}$.*

2) *The function $F(x) = \int_x^1 \sin\left(t^2\right)\,dt$ can be written as $F(x) = \int_1^x -\sin\left(t^2\right)\,dt$, hence its derivative is $F'(x) = -\sin x^2$.*

3) *What is the derivative of the function $F(x) = \int_2^{x^2} e^{-t^2}\,dt$? Considering $h(x) = x^2$ and $G(y) = \int_2^y e^{-t^2}\,dt$, we have $F(x) = (G \circ h)(x)$. Thus, by the Chain Rule (see Theorem 4.4.1), we obtain $F'(x) = G'(h(x))\,h'(x) = 2xe^{-x^4}$.*

4) *What is the derivative of the function $F(x) = \int_x^{x^2} \ln\left(\frac{1}{1+t^2}\right)\,dt$? Using a property of integrals, we have*

$$F(x) = \int_x^0 \ln\left(\frac{1}{1+t^2}\right)\,dt + \int_0^{x^2} \ln\left(\frac{1}{1+t^2}\right)\,dt\,,$$

hence $F'(x) = 2x\ln\left(\frac{1}{1+x^4}\right) - \ln\left(\frac{1}{1+x^2}\right)$.

Remark 5.8.4. *If $F(x) = \int_{h_1(x)}^{h_2(x)} f(t)\,dt$ then*

$$F'(x) = h_2'(x) f\left(h_2(x)\right) - h_1'(x) f\left(h_1(x)\right)\,.$$

5.9 Improper Integrals

In the integration theory presented so far, it is required that the function f be continuous on the interval $[a,b]$, which implies that it is bounded on $[a,b]$.

It is also required that the interval $[a, b]$ be bounded, i.e., neither a nor b can be $+\infty$ or $-\infty$. But what if, for example, f is defined on $[a, b[$ or $]a, b]$ and is unbounded in that set? And what if a or b are $+\infty$ or $-\infty$? Consider the following definition.

Definition 5.9.1. ***Improper integrals*** *are those in which the integration interval is unbounded or the integrand is unbounded within the integration interval.*

Two types of improper integrals will be considered:

- integral with an *unbounded interval of integration*, when one of the bounds of the interval is $+\infty$ or $-\infty$. In this case, we have

  i) $\displaystyle\int_a^{+\infty} f(x)\,dx = \lim_{b\to+\infty}\int_a^b f(x)\,dx$, assuming $a \in \mathbb{R}$ and the function f defined on $[a, +\infty[$ is integrable on every interval $[a, t]$ with $t > a$;

  ii) $\displaystyle\int_{-\infty}^{b} f(x)\,dx = \lim_{a\to-\infty}\int_a^b f(x)\,dx$, assuming $b \in \mathbb{R}$ and the function f defined on $]-\infty, b[$ is integrable on every interval $[t, b]$ with $t < b$;

  iii) $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{c} f(x)\,dx + \int_{c}^{+\infty} f(x)\,dx$, ;

- integral with an *unbounded integrand*, when the function approaches $+\infty$ or $-\infty$ at one of the bounds of the integration interval. In this case:

  i) if $\displaystyle\lim_{x\to b^-} f(x) = \pm\infty$, $\displaystyle\int_a^b f(x)\,dx = \lim_{t\to b^-}\int_a^t f(x)\,dx$, assuming $a, b \in \mathbb{R}$, with $a < b$, and the function f defined on $[a, b[$ is integrable on every interval $[a, t]$ with $a < t < b$;

  ii) if $\displaystyle\lim_{x\to a^+} f(x) = \pm\infty$, $\displaystyle\int_a^b f(x)\,dx = \lim_{t\to a^+}\int_t^b f(x)\,dx$, assuming $a, b \in \mathbb{R}$, with $a < b$, and the function f defined on $]a, b]$ is integrable on every interval $[t, b]$ with $a < t < b$.

Definition 5.9.2 (**integral convergence**)**.** *An improper integral is said to be:*

- ***convergent****, when the corresponding limit exists and is finite;*
- ***divergent****, otherwise, i.e., when the limit is infinite or does not exist.*

Example 5.9.3.

1) Consider the improper integral

$$\begin{aligned}\int_0^{+\infty} \frac{1}{1+x^2}\,dx &= \lim_{t\to+\infty}\int_0^t \frac{1}{1+x^2}\,dx \\ &= \lim_{t\to+\infty}[\arctan(x)]_0^t \\ &= \lim_{t\to+\infty}\arctan(t) - \underbrace{\arctan(0)}_{=0} \\ &= \frac{\pi}{2}\,,\end{aligned}$$

so the improper integral $\displaystyle\int_0^{+\infty} \frac{1}{1+x^2}\,dx$ *is convergent.*

2) Consider the improper integral

$$\begin{aligned}\int_0^{1} \frac{1}{\sqrt{1-x}}\,dx &= \lim_{t\to1^-}\int_0^t \frac{1}{\sqrt{1-x}}\,dx \\ &= \lim_{t\to1^-}[-2\sqrt{1-x}]_0^t \\ &= \underbrace{\lim_{t\to1^-} -2\sqrt{1-t}}_{=0} + 2\sqrt{1-0} \\ &= 2\,,\end{aligned}$$

so the improper integral $\displaystyle\int_0^{1} \frac{1}{\sqrt{1-x}}\,dx$ *is convergent.*

3) Consider the improper integral

$$\begin{aligned}\int_1^{+\infty} \frac{1}{x}\,dx &= \lim_{t\to+\infty}\int_1^t \frac{1}{x}\,dx \\ &= \lim_{t\to+\infty}[\ln x]_1^t \\ &= \lim_{t\to+\infty}\ln t - \underbrace{\ln 1}_{=0} \\ &= +\infty\,,\end{aligned}$$

so the improper integral $\displaystyle\int_1^{+\infty} \frac{1}{x}\,dx$ *is not convergent and is therefore said to be divergent.*

Part II

Linear Algebra

Chapter 6

Vectors

Considering $\mathbb{R}$ the set of real numbers, we define $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as the set whose elements are called **ordered pairs** (or, more generally, **vectors of** $\mathbb{R}^2$) and which are of the form (x, y) where both elements x and y belong to $\mathbb{R}$.

Analogously, we define $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ as the set whose elements are called **ordered triples** (or, more generally, **vectors of** $\mathbb{R}^3$) and which are of the form (x, y, z) where all elements x, y and z belong to $\mathbb{R}$.

More generally, given a natural number n, we define $\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}}$ as the set whose elements are called **n-tuples** (or, more generally, **vectors**) and which are of the form $(x_1, x_2, \ldots, x_n)$ where all elements $x_1, x_2, \ldots, x_n$ belong to $\mathbb{R}$.

6.1 Operations with vectors

Let us consider two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$.

6.1.1 Equality of vectors

We say that $\mathbf{x}$ and $\mathbf{y}$ are **equal**, and we write $\mathbf{x} = \mathbf{y}$, if each element of vector $\mathbf{x}$ is equal to the corresponding element of vector $\mathbf{y}$, i.e., if $x_i = y_i$ for every $i \in \{1, 2, \ldots, n\}$.

6.1.2 Addition of vectors

We can define the **addition of two vectors in** $\mathbb{R}^n$ as the element-wise sum of each vector, i.e., $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$.

Properties of vector addition

Proposition 6.1.1. *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$*,* $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ *and* $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ *be vectors in* $\mathbb{R}^n$ *and let* $\mathbf{0} = (0, 0, \ldots, 0)$ *be the vector in* $\mathbb{R}^n$ *with all its elements equal to* 0*. Then:*

1. *Commutativity:* $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$*;*
2. *Associativity:* $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$*;*
3. *Identity element:* $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$*;*
4. *Additive inverse:* $\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0}$*, where* $-\mathbf{x}$ *denotes the vector obtained from* $\mathbf{x}$ *by multiplying each element by* -1*.*

Proof. Exercise. □

6.1.3 Scalar multiplication

Given $\lambda \in \mathbb{R}$, we define the operation of **scalar multiplication of vector x by** λ as the multiplication of each element of vector $\mathbf{x}$ by λ, i.e., $\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$.

Properties of scalar multiplication

Proposition 6.1.2. *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ *be vectors in* $\mathbb{R}^n$*, and let* $\alpha, \beta \in \mathbb{R}$*. Then:*

1. *Distributivity of scalar addition over vector multiplication:* $(\alpha+\beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$*;*
2. *Distributivity of scalar multiplication over vector addition:* $\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$*;*
3. *Associativity of scalar multiplication:* $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$*;*
4. *Identity element of scalar multiplication:* $1\mathbf{x} = \mathbf{x}$*;*
5. *Zero element of scalar multiplication:* $0\mathbf{x} = (0, \ldots, 0)$*.*

Proof. Exercise. □

6.1.4 Inner product

The **inner product**, also called the **dot product**, of vectors in $\mathbb{R}^n$ is a real number defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i \,.$$

Example 6.1.3. *Let* $\mathbf{u} = (1, -2, 3), \mathbf{v} = (-3, 2, 5) \in \mathbb{R}^3$. *We have*

$$\mathbf{u} \cdot \mathbf{v} = 1 \times (-3) + (-2) \times 2 + 3 \times 5 = -3 - 4 + 15 = 8 \,.$$

Properties of the inner product

Proposition 6.1.4. *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ *and* $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ *be vectors in* $\mathbb{R}^n$ *and let* $\alpha \in \mathbb{R}$. *Then:*

1. *Commutativity of the inner product:* $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$;
2. *Distributivity of the inner product over vector addition:* $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$;
3. *Compatibility of scalar multiplication with the inner product:* $(\alpha \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\alpha \mathbf{y}) = \alpha(\mathbf{x} \cdot \mathbf{y})$;
4. $\mathbf{x} \cdot \mathbf{x} \geq 0$ *and* $\mathbf{x} \cdot \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = 0$.

Proof.

1. Exercise;
2. We have

$$\begin{aligned} \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= (x_1, x_2, \ldots, x_n).(y_1 + z_1, y_2 + z_2, \ldots, y_n + z_n) \\ &= x_1(y_1 + z_1) + x_2(y_2 + z_2) + \ldots + x_n(y_n + z_n) \\ &= x_1 y_1 + x_1 z_1 + x_2 y_2 + x_2 z_2 + \ldots + x_n y_n + x_n z_n \\ &= (x_1 y_1 + x_2 y_2 + \ldots + x_n y_n) + (x_1 z_1 + x_2 z_2 + \ldots + x_n z_n) \\ &= \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \,; \end{aligned}$$

3. Exercise;
4. We have $\mathbf{x} \cdot \mathbf{x} = x_1^2 + x_2^2 + \ldots + x_n^2$, and since each term x_i^2 is always non-negative, then $\mathbf{x} \cdot \mathbf{x}$ is always non-negative, i.e., $\mathbf{x} \cdot \mathbf{x} \geq 0$. Furthermore, $\mathbf{x} \cdot \mathbf{x}$ is zero if and only if each term x_i^2 is zero, hence $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x}$ is the null vector.

□

Definition 6.1.5. *We say that the vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *are* ***orthogonal*** *if* $\mathbf{x} \cdot \mathbf{y} = 0$.

6.2 Norm and Distance in $\mathbb{R}^n$

Given a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$, the **norm** (or length) of the vector $\mathbf{x}$ is defined, and denoted by $||\mathbf{x}||$, as

$$||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

that is,

$$||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2}\,.$$

6.2.1 Properties of the Norm

Proposition 6.2.1. *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ *be a vector in* $\mathbb{R}^n$ *and let* $\alpha \in \mathbb{R}$. *Then:*

1. $||\mathbf{x}|| \geq 0$;
2. $||\mathbf{x}|| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$;
3. $||\alpha\mathbf{x}|| = |\alpha| \cdot ||\mathbf{x}||$.

Proof. Exercise. □

Given two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, the **distance** between the vectors $\mathbf{x}$ and $\mathbf{y}$ is defined and denoted by $d(\mathbf{x}, \mathbf{y})$, as

$$d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$$

that is,

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,.$$

This distance is usually referred to as the **Euclidean distance**.

Proposition 6.2.2 (Triangle inequality). *Given two vectors in $\mathbb{R}^n$, the norm of their sum is always less than or equal to the sum of their norms, i.e., for two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, we have*

$$||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \,.$$

6.3 Linear Combination of Vectors in $\mathbb{R}^n$

Definition 6.3.1. *Let $\underline{v} = \{v_1, v_2, \ldots, v_p\}$ be a family of p vectors in $\mathbb{R}^n$. Any vector of the form*

$$\alpha_1 v_1 + \cdots + \alpha_p v_p,$$

where the (α_i) are scalars, is called a ***linear combination of the vectors*** *$v_1, \ldots, v_p$.*

6.3.1 Linear Dependence vs. Linear Independence

Definition 6.3.2 (Linear Dependence and Independence). *The m vectors $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_m} \in \mathbb{R}^n$ are said to be* ***linearly dependent*** *if there exist m scalars $c_1, c_2, \ldots, c_m \in \mathbb{R}$, not all zero, such that*

$$c_1\mathbf{a_1} + c_2\mathbf{a_2} + \ldots + c_m\mathbf{a_m} = \mathbf{0}\,. \tag{6.1}$$

If condition (6.1) holds only when $c_1 = c_2 = \cdots = c_m = 0$, then the vectors $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_m}$ are said to be ***linearly independent****.*

Example 6.3.3.

1. *The vectors $\mathbf{a} = (1,2)$ and $\mathbf{b} = (2,4)$ in $\mathbb{R}^2$ are linearly dependent since $2\mathbf{a} + (-1)\mathbf{b} = (0,0)$. In this case, we can simply observe that one vector is a multiple of the other.*

2. *The vectors $\mathbf{a} = (1,2)$ and $\mathbf{b} = (1,3)$ in $\mathbb{R}^2$ are linearly independent. Given two scalars $c_1, c_2 \in \mathbb{R}$ such that $c_1\mathbf{a} + c_2\mathbf{b} = (0,0)$, we have*

$$\begin{aligned} c_1(1,2) + c_2(1,3) = (0,0) &\Leftrightarrow (c_1, 2c_1) + (c_2, 3c_2) = (0,0) \\ &\Leftrightarrow (c_1 + c_2, 2c_1 + 3c_2) = (0,0) \\ &\Leftrightarrow \begin{cases} c_1 + c_2 = 0 \\ 2c_1 + 3c_2 = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} c_1 = 0 \\ c_2 = 0 \end{cases} . \end{aligned}$$

Thus, the only scalars $c_1, c_2 \in \mathbb{R}$ *that satisfy* $c_1\mathbf{a} + c_2\mathbf{b} = (0,0)$ *are* $c_1 = c_2 = 0$. *Hence, by definition, the vectors* $\mathbf{a} = (1,2)$ *and* $\mathbf{b} = (1,3)$ *in* $\mathbb{R}^2$ *are linearly independent.*

Remark 6.3.4. *When we say that a set of vectors is linearly independent (respectively, linearly dependent), it is equivalent to saying that those vectors are linearly independent (respectively, linearly dependent).*

Theorem 6.3.5. *A set of vectors is linearly dependent if and only if one of the vectors is a linear combination of the others.*

Proof. Exercise. □

Remark 6.3.6. *Theorem 6.3.5 can be restated equivalently as: "A set of vectors is linearly independent if and only if none of the vectors is a linear combination of the others".*

Theorem 6.3.7. *Any set of vectors that includes the null vector is linearly dependent.*

Proof. Exercise. □

Theorem 6.3.8. *If in a set of vectors at least two are equal or one is a multiple of another, then the set is linearly dependent.*

Proof. Exercise. □

Example 6.3.9.

1. *The set of vectors*

$$\{(1,2,3,4),(1,2,5,-1),(0,1,0,1),(2,1,8,0)\}$$

is linearly dependent because

$$1(1,2,3,4)+1(1,2,5,-1)+(-3)(0,1,0,1)+(-1)(2,1,8,0) = (0,0,0,0)\,,$$

that is,

$$(2,1,8,0) = 1(1,2,3,4) + 1(1,2,5,-1) + (-3)(0,1,0,1)\,,$$

i.e., the vector $(2,1,8,0)$ *is a linear combination of the vectors* $(1,2,3,4)$, $(1,2,5,-1)$ *and* $(0,1,0,1)$. *This is an application of Theorem 6.3.5, since one of the vectors in the set is a linear combination of the others.*

2. *The set of vectors*

$$\{(1,2,3,4),(1,2,5,-1),(0,0,0,0)\}$$

is linearly dependent because, taking $c_1 = c_2 = 0$ *and* $c_3 \neq 0$*, we have*

$$c_1(1,2,3,4) + c_2(1,2,5,-1) + c_3(0,0,0,0) = (0,0,0,0)\,.$$

This is an application of Theorem 6.3.7, as the set contains the null vector.

3. *The set of vectors*

$$\{(1,2,3,4),(1,2,5,-1),(1,2,3,4)\}$$

is linearly dependent because

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

This is an application of Theorem 6.3.8, since the set contains two equal vectors.

4. *The set of vectors*

$$\{(1,2,3,4),(1,2,5,-1),(2,4,6,8)\}$$

is linearly dependent because

$$2(1,2,3,4) + 0(1,2,5,-1) + (-1)(2,4,6,8) = (0,0,0,0)\,.$$

This is an application of Theorem 6.3.8, since $(2,4,6,8) = 2(1,2,3,4)$*, i.e., one of the vectors is a multiple of another.*

Chapter 7

Matrices

Given two natural numbers m and n, an $m \times n$ **matrix over** $\mathbb{R}$ (also called a **matrix of type** or **matrix of order** $m \times n$) is the table formed by $m \times n$ real numbers arranged in m rows (horizontal) and n columns (vertical), as represented by the following matrix A,

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

where each element (or entry) a_{ij} corresponds to a real number in the (i, j)-position of the matrix A, i.e., it is located in row i and column j, with i ranging from 1 to m and j from 1 to n. The matrix A can also be denoted by $A = [a_{ij}]_{m \times n}$.

The set of all matrices of order $m \times n$ over $\mathbb{R}$ is denoted by $M_{m \times n}(\mathbb{R})$.

Let us now look at some particular cases of matrices:

- If $m = 1$, we have a matrix with only one row, called a **row matrix**, as in the matrix

$$L = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \end{bmatrix} \in M_{1 \times n}(\mathbb{R}).$$

- If $n = 1$, we have a matrix with only one column, called a **column**

matrix, as in the matrix

$$C = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} \in M_{m\times 1}(\mathbb{R}).$$

- If $m = n$, we have a matrix with an equal number of rows and columns, called a **square matrix** of order n, as in the matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} \in M_{n\times n}(\mathbb{R}).$$

The set of all square matrices of order n (i.e., of type $n \times n$) with real entries, denoted by $M_{n\times n}(\mathbb{R})$, can also be written as $M_n(\mathbb{R})$. In such matrices, the **main diagonal** consists of the elements a_{ii} for $i \in \{1, \dots, n\}$.

Let us now consider some particular cases of square matrices:

- If all the elements outside the main diagonal are zero, the matrix is called a **diagonal matrix**, as in the matrix

$$D = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix} \in M_n(\mathbb{R}).$$

- If all the elements below the main diagonal are zero, the matrix is called an **upper triangular matrix**, as in the matrix

$$T_{sup} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ 0 & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix} \in M_n(\mathbb{R}).$$

- If all the elements above the main diagonal are zero, the matrix is called a **lower triangular matrix**, as in the matrix

$$T_{inf} = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ d_{21} & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \in M_n(\mathbb{R}).$$

- A diagonal matrix in which all the elements on the main diagonal are equal to 1 is called the **identity matrix**, as in the matrix

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \in M_n(\mathbb{R}).$$

The identity matrix is usually denoted by I_n, where the subscript n indicates the order of the matrix. If the order of the identity matrix is clear from the context, the subscript may be omitted.

The **null matrix** is the matrix of type $m \times n$ whose elements are all zero, and it is represented by

$$O = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \in M_{m\times n}(\mathbb{R}).$$

Equality of matrices

Two matrices $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{p\times q}$ are said to be **equal** if:

1. they are of the same type, i.e., $m = p$ and $n = q$, and
2. the elements in the same position are equal, i.e., $a_{ij} = b_{ij}$ for each $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$.

7.1 Matrix Operations

7.1.1 Matrix Addition

The operation of **matrix addition** is only defined for matrices of the same type. Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$ be two matrices of the same type. The sum of the matrices A and B is defined as the matrix whose entry in position (i, j) is the sum of the entry in position (i, j) of matrix A with the entry in position (i, j) of matrix B, and is denoted by $A+B = (a_{ij} + b_{ij})_{m\times n}$.

Example 7.1.1. *Given the matrices* $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \end{bmatrix}$ *and* $B = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ *of type* 2×3*, their sum is*

$$A + B = \begin{bmatrix} 1+2 & 0+1 & 1+0 \\ 2+0 & -1+0 & 3+5 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ 2 & -1 & 8 \end{bmatrix}.$$

Properties of Matrix Addition

Proposition 7.1.2. *Let* $A, B, C, O \in M_{m\times n}(\mathbb{R})$*, where* O *is the null matrix of type* $m \times n$*. Then:*

1. *Commutativity:* $A + B = B + A$;
2. *Associativity:* $(A + B) + C = A + (B + C)$;
3. *Identity element:* $A + O = O + A = A$;
4. *Additive inverse:* $A + (-A) = (-A) + A = O$, *where* $-A$ *denotes the matrix obtained from* A *by multiplying each element by* -1.

Proof. Exercise. □

7.1.2 Multiplying a Matrix by a Scalar

The **multiplication of a matrix** $A = [a_{ij}]_{m\times n}$ **by a real number** α is defined as the multiplication of each element of matrix A by the scalar α, i.e., $\alpha A = [\alpha a_{ij}]_{m\times n}$.

Example 7.1.3. *Given the matrix* $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \end{bmatrix}$ *and the scalar* -3*, we have*

$$(-3)A = \begin{bmatrix} (-3)\times 1 & (-3)\times 0 & (-3)\times 1 \\ (-3)\times 2 & (-3)\times(-1) & (-3)\times 3 \end{bmatrix} = \begin{bmatrix} -3 & 0 & -3 \\ -6 & 3 & -9 \end{bmatrix}.$$

Properties of Scalar Multiplication of a Matrix

Proposition 7.1.4. *Let $A, B, O \in M_{m\times n}(\mathbb{R})$, where O is the null matrix of type $m \times n$, and let $\alpha, \beta \in \mathbb{R}$. Then:*

1. *Distributivity of scalar addition over matrix multiplication:* $(\alpha+\beta)A = \alpha A + \beta A$;
2. *Distributivity of scalar multiplication over matrix addition:* $\alpha(A+B) = \alpha A + \alpha B$;
3. *Associativity of scalar multiplication:* $\alpha(\beta A) = (\alpha\beta)A$;
4. *Identity element:* $1A = A$;
5. *Absorbing element:* $0A = O$.

Proof. Exercise. □

7.1.3 Matrix Multiplication

Given two matrices A and B, we can define the **multiplication of matrix** A **by matrix** B . However, for this multiplication to be defined, the number of columns of A must equal the number of rows of B.

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{n\times q}$. The multiplication of matrix A by matrix B, denoted by AB, is defined as the matrix of type $m \times q$ whose entry in position (i, j) is given by

$$\underbrace{\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}}_{\text{row } i \text{ of } A} \cdot \underbrace{\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}}_{\text{column } j \text{ of } B} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj} ,$$

and therefore,

$$AB = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]_{m\times q} .$$

Example 7.1.5. *Given the matrices* $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & 0 & 1 & 1 \\ 0 & 3 & 0 & 2 \end{bmatrix} \in M_{3\times 4}(\mathbb{R})$ *and* $B = \begin{bmatrix} 1 & 2 \\ -1 & -1 \\ 0 & 1 \\ 0 & 3 \end{bmatrix} \in M_{4\times 2}(\mathbb{R})$, *we have that the product AB is defined, and*

$$\begin{aligned} AB &= \begin{bmatrix} 1\times 1+2\times(-1)+0\times 0+1\times 0 & 1\times 2+2\times(-1)+0\times 1+1\times 3 \\ (-1)\times 1+0\times(-1)+1\times 0+1\times 0 & (-1)\times 2+0\times(-1)+1\times 1+1\times 3 \\ 0\times 1+3\times(-1)+0\times 0+2\times 0 & 0\times 2+3\times(-1)+0\times 1+2\times 3 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 3 \\ -1 & 2 \\ -3 & 3 \end{bmatrix} \in M_{3\times 2}(\mathbb{R}) . \end{aligned}$$

Properties of Matrix Multiplication

Proposition 7.1.6. *Let A, B, C, O be matrices of the appropriate types such that the multiplications below are defined, where O is the null matrix. Then:*

1. *Associativity of matrix multiplication:* $(AB)C = A(BC)$;
2. *Left distributivity of matrix multiplication with respect to matrix addition:* $C(A+B) = CA + CB$;
3. *Right distributivity of matrix multiplication with respect to matrix addition:* $(A+B)C = AC + BC$;
4. *Identity element:* $AI = IA = A$;
5. *Absorbing element:* $AO = OA = O$.

Proof. Exercise. □

Remark 7.1.7. *Matrix multiplication is not generally commutative, as shown in the following example. Note that in Example 7.1.5, the multiplication BA is not even defined, since the number of columns of B, which is 2, is different from the number of rows of A, which is 3.*

Example 7.1.8. *Given the square matrices* $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 1 \\ 0 & 3 & 0 \end{bmatrix} \in M_3(\mathbb{R})$ *and* $B = \begin{bmatrix} 1 & 2 & 0 \\ -1 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \in M_3(\mathbb{R})$, *both of order 3, we have that the multiplica-*

tions AB and BA are defined,

$$AB = \begin{bmatrix} -1 & 0 & 2 \\ -1 & -1 & 2 \\ -3 & -3 & 3 \end{bmatrix}$$

and

$$BA = \begin{bmatrix} -1 & 2 & 2 \\ 0 & 1 & -1 \\ -1 & 6 & 1 \end{bmatrix},$$

from which we see that $AB \neq BA$.

Power of a Matrix

Given a matrix A and a natural number k, can we define the k-th power of the matrix A? That is, can we multiply the matrix A by itself k times? Note that the multiplication AA is only defined if the number of columns of A equals the number of rows of A, i.e., if A is a square matrix. Thus, the power of a matrix is only defined for square matrices. Therefore, the **k-th power of a square matrix** A is defined by

$$A^k = \underbrace{A \times A \times \ldots \times A}_{k \text{ times}}, \quad k \in \mathbb{N}.$$

7.1.4 Transpose of a Matrix

Given a matrix $A \in M_{m\times n}(\mathbb{R})$, the **transpose matrix** of A, denoted by A^T, is the matrix whose rows are the columns of A and whose columns are the rows of A. That is, the element in position (i, j) of matrix A becomes the element in position (j, i) of matrix A^T, i.e., if $A = [a_{ij}]_{m\times n}$, then $A^T = [a_{ji}]_{n\times m}$.

Example 7.1.9. *Let*

$$A = \begin{bmatrix} 2 & 3 & 7 \\ -5 & 0 & 8 \\ 2 & -9 & 1 \\ 0 & 4 & -3 \end{bmatrix} \in M_{4\times 3}(\mathbb{R}).$$

Then,

$$A^T = \begin{bmatrix} 2 & -5 & 2 & 0 \\ 3 & 0 & -9 & 4 \\ 7 & 8 & 1 & -3 \end{bmatrix} \in M_{3\times 4}(\mathbb{R}).$$

Definition 7.1.10. *Let $A \in M_n(\mathbb{R})$ be a square matrix of order n. Then A is said to be:*

- *a **symmetric matrix** if $A = A^T$;*
- *a **skew-symmetric matrix** if $A = -A^T$.*

Properties of the Transpose of Matrices

Proposition 7.1.11. *Let A and B be matrices for which the operations below are defined, and let $k \in \mathbb{R}$ and $n \in \mathbb{N}$. Then:*

1. $\left(A^T\right)^T = A$;
2. $(A+B)^T = A^T + B^T$;
3. $(kA)^T = kA^T$;
4. $(AB)^T = B^T A^T$;
5. $(A^n)^T = \left(A^T\right)^n$.

Proof. Exercise. □

7.2 Rank of a Matrix

Given a matrix $A \in M_{m\times n}(\mathbb{R})$, viewing each row of A as a vector in $\mathbb{R}^n$, we can interpret the set of rows of matrix A as a set of vectors in $\mathbb{R}^n$. In this sense, we may ask whether the set of rows of matrix A is linearly independent. If not, we may ask what is the maximum number of rows in A that form a linearly independent set of vectors.

Definition 7.2.1 (**Rank of a Matrix**). *The **rank of a matrix** A is the maximum number of rows of A that form a linearly independent set of vectors, and is denoted by $r(A)$.*

Remark 7.2.2. *If A is the null matrix, then $r(A) = 0$. For any other (non-null) matrix, its rank is a natural number.*

Proposition 7.2.3. $r(A) = r(A^T)$.

Definition 7.2.4 (**Leading(s) of a Matrix**). *Given a non-null row of a matrix, the* ***leading*** *is the leftmost non-zero element of that row. In the case of a null row, we say it has no leading.*

Given a non-null matrix, the ***leadings*** *of the matrix are the leadings of all its non-null rows.*

Definition 7.2.5 (**Row Echelon Form**). *A matrix $A \in M_{m \times n}(\mathbb{R})$ is said to be in* ***row echelon form*** *if A is the null matrix, or if (when not zero) its leadings are in rows $1, \dots, s$, for some $s \in \{1, \dots, m\}$, and in positions $(1, k_1), (2, k_2), \dots, (s, k_s)$, with $1 \leq k_1 < k_2 < \dots < k_s \leq n$, that is, for each leading, the elements to its left (in the same row) and below it (in the same column) are all zero.*

Example 7.2.6. *The matrices, with leadings in bold, are:*

1. $\begin{bmatrix} 0 & \mathbf{1} & -1 & 0 \\ 0 & 0 & 0 & \mathbf{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} \mathbf{1} & 2 & -1 \\ 0 & \mathbf{5} & 3 \\ 0 & 0 & \mathbf{-1} \end{bmatrix}$, $\begin{bmatrix} \mathbf{3} \\ 0 \\ 0 \end{bmatrix}$ *and* $\begin{bmatrix} \mathbf{5} & -1 & 0 & 2 \end{bmatrix}$ *are in row echelon form;*

2. $\begin{bmatrix} \mathbf{1} & 0 & -1 \\ 0 & \mathbf{2} & 5 \\ 0 & \mathbf{3} & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathbf{-1} & 3 & 0 \\ 0 & 0 & \mathbf{6} & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *and* $\begin{bmatrix} \mathbf{1} & 3 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{2} & -2 \end{bmatrix}$ *are not in row echelon form.*

Proposition 7.2.7. *If a matrix $A \in M_{m \times n}(\mathbb{R})$ is in row echelon form, with $s \in \{1, \dots, m\}$ non-null rows, then $r(A) = s$.*

Proof. It suffices to check that the non-null rows of the matrix in row echelon form form a linearly independent set of vectors. □

7.2.1 Elementary operations on matrices

Given a matrix $A \in M_{m \times n}(\mathbb{R})$, an **elementary transformation** (or **elementary operation**) **on the rows** of the matrix A is a transformation of one of the following types:

I) exchange of row i with row j, where $i \neq j$, denoted by $l_i \leftrightarrow l_j$;

II) multiplication of row i by a scalar $\alpha \in \mathbb{R} \setminus \{0\}$, denoted by $l_i \to \alpha l_i$;

III) replacement of row i by its sum with row j multiplied by $\beta \in \mathbb{R}$, where $i \neq j$, denoted by $l_i \to l_i + \beta l_j$.

Example 7.2.8.

$$A = \begin{bmatrix} 3 & 2 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix} \xrightarrow[l_1 \leftrightarrow l_2]{} \begin{bmatrix} 1 & 0 & 2 \\ 3 & 2 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ (elementary transf. of type I)}$$

$$\xrightarrow[l_2 \to l_2 + (-3) l_1]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & -6 \\ 0 & 1 & 0 \end{bmatrix} \text{ (elementary transf. of type III)}$$

$$\xrightarrow[l_2 \to \frac{1}{2} l_2]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 1 & 0 \end{bmatrix} \text{ (elementary transf. of type II)}$$

$$\xrightarrow[l_3 \to l_3 + (-1) l_2]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 3 \end{bmatrix} \text{ (elementary transf. of type III)}.$$

Definition 7.2.9 (Row-equivalent matrix). *A matrix $A \in M_{m \times n}(\mathbb{R})$ is said to be **row-equivalent** to a matrix $B \in M_{m \times n}(\mathbb{R})$ of the same type if B can be obtained from A through a finite number of elementary row transformations.*

Given a matrix $A \in M_{m \times n}(\mathbb{R})$, using elementary row transformations, one can obtain a row-equivalent matrix in row echelon form. This method is called **matrix condensation**, and consists of the following steps:

Step 1 If A is the null matrix or a single-row matrix, then A is already in row echelon form;

Step 2 By swapping rows (elementary transf. of type I), if necessary, obtain a matrix B such that the first row, among all non-null rows of A, has the leftmost leading;

Step 3 Let b_{1j} be the leading of row 1 of matrix B. For each row i of B, with $i \in \{2, \ldots, m\}$, apply an elementary transformation of type III, $l_i \to l_i + \left(-\frac{b_{ij}}{b_{1j}}\right) l_1$, which transforms the element b_{ij} in each row i into 0, thereby obtaining another matrix, call it C;

Step 4 Ignore the first row of matrix C and apply steps 1, 2, and 3 again to the resulting submatrix.

Example 7.2.10. *Given the matrix* $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 9 & 3 & -4 \\ 0 & 0 & 2 & -2 & 3 \\ 0 & 1 & 2 & 1 & -1 \end{bmatrix}$ *and perform-*

ing the following elementary row transformations,

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 9 & 3 & -4 \\ 0 & 0 & 2 & -2 & 3 \\ 0 & 1 & 2 & 1 & -1 \end{bmatrix} \xrightarrow[l_1 \leftrightarrow l_4]{} \begin{bmatrix} 0 & \mathbf{1} & 2 & 1 & -1 \\ 0 & 4 & 9 & 3 & -4 \\ 0 & 0 & 2 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[l_2 \to l_2 + (-\frac{4}{1}) l_1]{} \begin{bmatrix} 0 & \mathbf{1} & 2 & 1 & -1 \\ 0 & 0 & \mathbf{1} & -1 & 0 \\ 0 & 0 & 2 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[l_3 \to l_3 + (-\frac{2}{1}) l_2]{} \begin{bmatrix} 0 & \mathbf{1} & 2 & 1 & -1 \\ 0 & 0 & \mathbf{1} & -1 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = B\,.$$

we obtain the row-equivalent matrix B in row echelon form.

Proposition 7.2.11. *If two matrices $A, B \in M_{m\times n}(\mathbb{R})$ are row-equivalent, then they have the same rank, i.e., $r(A) = r(B)$.*

Example 7.2.12. *In the previous Example 7.2.10, we have $r(B) = 2$, and therefore, by Proposition 7.2.11, we deduce that $r(A) = r(B) = 2$.*

Proposition 7.2.13. *Let $A \in M_{m\times n}(\mathbb{R})$. Then $r(A) \le m$ and $r(A) \le n$, i.e., $r(A) \le \min\{m, n\}$.*

Proof. See [3] or [8]. □

Example 7.2.14.

1. *If $A \in M_{5\times 2}(\mathbb{R})$, then $r(A) \le 2$.*
2. *If $A \in M_{2\times 5}(\mathbb{R})$, then $r(A) \le 2$.*
3. *If $A \in M_5(\mathbb{R})$, then $r(A) \le 5$.*

7.2.2 Inverse of a Matrix

When, for example, two real numbers are multiplied, it is known that, for any non-zero real number, there always exists another real number, called the **algebraic inverse**, such that the product of the two is 1. Recall that 1 is the identity element for real number multiplication.

Does the same happen with matrices? That is, given any matrix, does there exists another matrix such that the product of the two is the identity matrix? The answer to this question is, in general, **NO**. However, there are matrices for which the answer is **YES**. Let us now consider the following definitions and properties.

Definition 7.2.15 (**Invertible Matrix**). *A (square) matrix $A \in M_n(\mathbb{R})$ is said to be* ***invertible*** *(or* ***nonsingular****), or to have an* ***inverse****, if there exists a matrix $B \in M_n(\mathbb{R})$ of the same type such that*

$$AB = BA = I_n\,.$$

If no such matrix B exists, then A is said to be ***singular****.*

Remark 7.2.16. *Note that in Definition 7.2.15, since $AB = BA$, the concept of an invertible matrix only makes sense for square matrices.*

Proposition 7.2.17 (**Uniqueness**). *If $A \in M_n(\mathbb{R})$ is invertible, then there exists one and only one matrix $B \in M_n(\mathbb{R})$ such that $AB = BA = I_n$.*

Proof. Suppose there exist $B_1, B_2 \in M_n(\mathbb{R})$ such that

$$AB_1 = B_1A = I_n \quad \text{and} \quad AB_2 = B_2A = I_n\,.$$

Then we have

$$B_1 = B_1I_n = B_1(AB_2) = (B_1A)B_2 = I_nB_2 = B_2\,,$$

from which it follows that the inverse is unique. □

Remark 7.2.18. *The inverse of the identity matrix is the identity matrix itself, since $I_nI_n = I_n$.*

Definition 7.2.19 (**Inverse Matrix**). *Let $A \in M_n(\mathbb{R})$ be invertible. The unique matrix $B \in M_n(\mathbb{R})$ such that $AB = BA = I_n$ is called the* ***inverse of*** *A and is denoted by A^{-1}.*

Example 7.2.20. *Let $A = \begin{bmatrix} 3 & 5 \\ -1 & -2 \end{bmatrix}$. Since*

$$\begin{bmatrix} 3 & 5 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ -1 & -3 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ -1 & -3 \end{bmatrix}\begin{bmatrix} 3 & 5 \\ -1 & -2 \end{bmatrix} = I_2\,,$$

we have that A is invertible and $A^{-1} = \begin{bmatrix} 2 & 5 \\ -1 & -3 \end{bmatrix}$.

Proposition 7.2.21 (**Inverse and Rank**). *Let $A \in M_n(\mathbb{R})$. The following statements are equivalent:*

1. *A is invertible;*

2. *$r(A) = n$;*

3. *the identity matrix can be obtained from A through elementary row operations.*

Proof. See [3] or [8]. □

Example 7.2.22.

1. *Let $A = \begin{bmatrix} 1 & -2 & 4 \\ 0 & -3 & 5 \\ -2 & 1 & -3 \end{bmatrix}$. Since*

$$\begin{bmatrix} \mathbf{1} & -2 & 4 \\ 0 & -3 & 5 \\ -2 & 1 & -3 \end{bmatrix} \xrightarrow[r_3 \to r_3 + 2r_1]{} \begin{bmatrix} \mathbf{1} & -2 & 4 \\ 0 & \mathbf{-3} & 5 \\ 0 & -3 & 5 \end{bmatrix} \xrightarrow[r_3 \to r_3 - r_2]{} \begin{bmatrix} \mathbf{1} & -2 & 4 \\ 0 & \mathbf{-3} & 5 \\ 0 & 0 & 0 \end{bmatrix},$$

we get $r(A) = 2 < 3 = n$, so, by Proposition 7.2.21, A is singular.

2. *Let $B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ -1 & 0 & 0 \end{bmatrix}$. Since*

$$\begin{bmatrix} \mathbf{1} & 0 & 1 \\ 2 & 2 & 2 \\ -1 & 0 & 0 \end{bmatrix} \xrightarrow[\substack{r_2 \to r_2 - 2r_1 \\ r_3 \to r_3 + r_1}]{} \begin{bmatrix} \mathbf{1} & 0 & 1 \\ 0 & \mathbf{2} & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix},$$

we get $r(B) = 3$, so, by Proposition 7.2.21, B is invertible.

7.2.3 Computing the Inverse Matrix

Given an invertible matrix, how can we find its inverse?

By performing elementary row operations, it is possible to obtain the inverse matrix of a given invertible matrix, as described in Proposition 7.2.21. Here's how:

- Given an invertible matrix $A \in M_n(\mathbb{R})$, perform elementary row operations on A until the identity matrix I_n is obtained. Meanwhile, apply exactly the same sequence of elementary row operations (same operations in the same order) to the identity matrix. In the end, the resulting matrix will be the inverse of A.

In practical terms, a simple way to apply this method of finding the inverse matrix is to place the given matrix $A \in M_n(\mathbb{R})$ side by side with the identity matrix of the same order,

$$[A \mid I_n] \ ,$$

and as elementary row operations are applied to the matrix A, the same operations are also applied to the rows of I_n. When, through these operations, the identity matrix is obtained in place of A, the matrix obtained in place of I_n is the inverse of A,

$$[A \mid I_n] \xrightarrow[\text{elem. row ops.}]{} \cdots \xrightarrow[\text{elem. row ops.}]{} \left[I_n \mid A^{-1}\right] .$$

Consider the following example.

Example 7.2.23. *Let* $B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ -1 & 0 & 0 \end{bmatrix}$. *As seen in Example 7.2.22, the matrix B is invertible. Let us now compute its inverse. We have*

$$[B \mid I_3] = \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow[\substack{l_2 \to l_2 - 2l_1 \\ l_3 \to l_3 + l_1}]{} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[l_2 \to \frac{1}{2} l_2]{} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[l_1 \to l_1 - l_3]{} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right] ,$$

from which we obtain $B^{-1} = \begin{bmatrix} 0 & 0 & -1 \\ -1 & \frac{1}{2} & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Proposition 7.2.24. *Let $A, B \in M_n(\mathbb{R})$ be invertible matrices and let $\alpha \in \mathbb{R} \setminus \{0\}$. Then:*

1. *A^{-1} is invertible and $\left(A^{-1}\right)^{-1} = A$;*
2. *αA is invertible and $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$;*
3. *AB is invertible and $(AB)^{-1} = B^{-1} A^{-1}$;*
4. *for any $k \in \mathbb{N}$, A^k is invertible and $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$;*

5. A^T *is invertible and* $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$;

Proof. Exercise. □

Proposition 7.2.25. *Let* $A_1, \ldots, A_k \in M_n(\mathbb{R})$ *be invertible matrices, with* $k \in \mathbb{N}$. *Then the product* $A_1 \cdot \ldots \cdot A_k$ *is invertible and*

$$(A_1 \cdot \ldots \cdot A_k)^{-1} = A_k^{-1} \cdot \ldots \cdot A_1^{-1}.$$

Proof. Exercise. □

Proposition 7.2.26. *Let* $A, B \in M_n(\mathbb{R})$. *The product* AB *is invertible if and only if both* A *and* B *are invertible.*

Proof. See [3] or [8]. □

Remark 7.2.27. *Note that in item 3 of Proposition 7.2.24, only one direction of the equivalence in Proposition 7.2.26 is mentioned.*

Corolary 7.2.28. *Let* $A, B \in M_n(\mathbb{R})$. *If* $AB = I_n$, *then both* A *and* B *are invertible, and we have* $A^{-1} = B$ *and* $BA = I_n$.

Proof. Exercise. □

Chapter 8

Determinants

Given a square matrix $A \in M_n(\mathbb{R})$, the **determinant** of the matrix A is a real number associated (according to certain rules) with the matrix A, denoted by $det(A)$ or $|A|$.

Let us now see how to associate this real number with each matrix.

8.1 Matrices of order 1 and order 2

Given a matrix of order 1, $A = [a_{11}]$, the determinant of the matrix A is defined as a_{11}, and is written as

$$det(A) = a_{11} \quad \text{or} \quad |A| = a_{11}\,.$$

Given a matrix of order 2, $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the determinant of the matrix A is defined as $a_{11}a_{22} - a_{12}a_{21}$, and is written as

$$det(A) = a_{11}a_{22} - a_{12}a_{21} \quad \text{or} \quad |A| = a_{11}a_{22} - a_{12}a_{21}\,.$$

Example 8.1.1.

1. *If* $A = [-2]$, *then* $|A| = -2$.

2. *If* $A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$, *then* $|A| = 1 \times 3 - 2 \times (-1) = 5$.

8.2 Matrices of order $n \geq 2$ - Laplace's Theorem

In this section, we present the strategy to compute the determinant of a matrix of order n, for any natural number $n \geq 2$.

Let us first consider the following definition.

Definition 8.2.1 (**Algebraic complement**). *Let $A \in M_n(\mathbb{R})$ be a square matrix of order n, with $n \geq 2$. The* ***cofactor*** *or* ***algebraic complement*** *of position ij of A, denoted by $\hat{a}_{ij}$, is the scalar*

$$\hat{a}_{ij} = (-1)^{i+j} \, |A(i|j)| \, ,$$

where $A(i|j)$ is the submatrix obtained from A by removing row i and column j.

Example 8.2.2. *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. *We have*

$$\begin{aligned} \hat{a}_{32} &= (-1)^{3+2} \, |A(3|2)| \\ &= -\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} \\ &= -(6-12) \\ &= 6 \, . \end{aligned}$$

The general strategy for computing the determinant of a matrix of order $n \geq 2$ consists in applying a rule deduced from **Laplace's Theorem**, presented below.

Theorem 8.2.3 (**Laplace's Theorem**). *Let $A \in M_n(\mathbb{R})$ be a square matrix of order n, with $n \geq 2$. The determinant of A is equal to the sum of the products obtained by multiplying the elements of any row of A by the corresponding algebraic complements, i.e.,*

$$|A| = a_{i1}\,\hat{a}_{i1} + a_{i2}\,\hat{a}_{i2} + \ldots + a_{in}\,\hat{a}_{in} \, , \text{ for any row } i \text{ of } A \, .$$

The same result holds if, instead of any row of A, we choose any column, that is,

$$|A| = a_{1j}\,\hat{a}_{1j} + a_{2j}\,\hat{a}_{2j} + \ldots + a_{nj}\,\hat{a}_{nj} \, , \text{ for any column } j \text{ of } A \, .$$

Example 8.2.4. *Given the matrix* $A = \begin{bmatrix} 1 & 0 & 3 \\ -1 & 2 & 4 \\ 3 & 1 & 2 \end{bmatrix}$, *applying Laplace's Theorem (Theorem 8.2.3):*

- *to the first row of A,*

$$\begin{aligned}
|A| &= a_{11}\,\hat{a}_{11} + a_{12}\,\hat{a}_{12} + a_{13}\,\hat{a}_{13} \\
&= 1\,(-1)^{1+1}\ |A\,(1|1)| + 0\,(-1)^{1+2}\ |A\,(1|2)| + 3\,(-1)^{1+3}\ |A\,(1|3)| \\
&= 1 \begin{vmatrix} 2 & 4 \\ 1 & 2 \end{vmatrix} - 0 \begin{vmatrix} -1 & 4 \\ 3 & 2 \end{vmatrix} + 3 \begin{vmatrix} -1 & 2 \\ 3 & 1 \end{vmatrix} \\
&= 1 \times 0 - 0 \times (-14) + 3 \times (-7) \\
&= -21\,.
\end{aligned}$$

- *to the second column of A,*

$$\begin{aligned}
|A| &= a_{12}\,\hat{a}_{12} + a_{22}\,\hat{a}_{22} + a_{32}\,\hat{a}_{32} \\
&= 0\,(-1)^{1+2}\ |A\,(1|2)| + 2\,(-1)^{2+2}\ |A\,(2|2)| + 1\,(-1)^{3+2}\ |A\,(3|2)| \\
&= -0 \begin{vmatrix} -1 & 4 \\ 3 & 2 \end{vmatrix} + 2 \begin{vmatrix} 1 & 3 \\ 3 & 2 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ -1 & 4 \end{vmatrix} \\
&= -0 \times (-14) + 2 \times (-7) - 1 \times 7 \\
&= -21\,.
\end{aligned}$$

Remark 8.2.5. *If the matrix has several zero entries, it is particularly advantageous to apply Laplace's Theorem (Theorem 8.2.3) to the row or column with the most zeros.*

Example 8.2.6. *Let* $A = \begin{bmatrix} 1 & 0 & 3 \\ -1 & 2 & 4 \\ 1 & 0 & -1 \end{bmatrix}$. *Since the second column of A contains two zeros (more than any other row or column of A), it is particularly advantageous to apply Laplace's Theorem (Theorem 8.2.3) to the second column of A, hence*

$$\begin{aligned}
|A| &= a_{12}\,\hat{a}_{12} + a_{22}\,\hat{a}_{22} + a_{32}\,\hat{a}_{32} \\
&= 0 \times (-1)^{1+2} \times |A\,(1|2)| + 2 \times (-1)^{2+2} \times |A\,(2|2)| + 0 \times (-1)^{3+2} \times |A\,(3|2)| \\
&= 2 \begin{vmatrix} 1 & 3 \\ 1 & -1 \end{vmatrix} \\
&= 2 \times (-4) \\
&= -8\,.
\end{aligned}$$

8.3 Properties

Proposition 8.3.1. *Let $A \in M_n(\mathbb{R})$. Then the determinant of A is equal to the determinant of its transpose, i.e., $|A| = \left|A^T\right|$.*

Proof. Recall that the determinant of A can be computed from any row or column (see Theorem 8.2.3 - Laplace's Theorem). Thus, it suffices to observe that, for example, the ith row of A is equal to the ith column of A^T, so computing the determinant of A by applying Laplace's Theorem (Theorem 8.2.3) to the ith row of A is the same as applying Laplace's Theorem to the ith column of A^T. □

Proposition 8.3.2. *A matrix $A \in M_n(\mathbb{R})$ is invertible if and only if $|A| \neq 0$.*

Proof. See [3] or [8]. □

Remark 8.3.3. *Let $A \in M_n(\mathbb{R})$.*

1. *Note that the result stated in Proposition 8.3.2 is equivalent to:*

$$A \text{ is singular if and only if } |A| = 0\,.$$

2. *It is also noteworthy that by Proposition 7.2.21, we can deduce that:*

$$|A| \neq 0 \;\text{ if and only if }\; r(A) = n\,.$$

Proposition 8.3.4. *Let $A, B \in M_n(\mathbb{R})$ and let $\alpha \in \mathbb{R}$.*

1. *If A has a zero row (or column), then $|A| = 0$.*
2. *If A has two rows (or columns) that are multiples of each other, then $|A| = 0$.*
3. *If A is upper triangular (or lower triangular), then $|A| = \prod_{i=1}^{n} a_{ii}$.*
4. $|AB| = |A|\,|B|$.
5. *More generally, if $k \in \mathbb{N}$ and $A_1, \ldots, A_k \in M_n(\mathbb{R})$, then*

$$|A_1 \cdot \ldots \cdot A_k| = \prod_{i=1}^{k} |A_i|\,.$$

6. $|\alpha A| = \alpha^n\,|A|$.

Proof. Exercise. □

Exercise 8.3.5.

1. *If* $D \in M_n(\mathbb{R})$ *is a diagonal matrix, then* $|D| = \prod_{i=1}^{n} d_{ii}$.
2. $|I_n| = 1$.
3. *For any* $A \in M_n(\mathbb{R})$ *and* $m \in \mathbb{N}$, $|A^m| = |A|^m$.
4. *There exist matrices* $A, B \in M_n(\mathbb{R})$ *such that* $|A + B| \neq |A| + |B|$.
5. *Although matrix multiplication is not commutative, for any matrices* $A, B \in M_n(\mathbb{R})$, *we have* $|AB| = |BA|$.

Example 8.3.6.

1. *If* $D = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & 5 \\ 0 & 0 & -3 \end{bmatrix}$, *then* $|D| = 1 \times 2 \times (-3) = -6$.
2. *Let* $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. *We have* $|A| = |B| = 0$ *and* $|A + B| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 5$.
3. *Let* $A, B, C \in M_n(\mathbb{R})$ *such that* $|A| = 2$, $|B| = -5$ *and* $|C| = 4$. *We have:*
   - $\left|AB^TC\right| = |A| \left|B^T\right| |C| = |A| |B| |C| = 2 \times (-5) \times 4 = -40$.
   - $|3B| = 3^n |B| = -5 \times 3^n$.
   - $|B^2C| = |B^2| |C| = |B|^2 |C| = (-5)^2 \times 4 = 100$.

Proposition 8.3.7. *Let* $A \in M_n(\mathbb{R})$ *be an invertible matrix. Then the determinant of its inverse is the algebraic inverse of the determinant of* A, *i.e.,*

$$\left|A^{-1}\right| = \frac{1}{|A|}.$$

Proof. Exercise. □

Example 8.3.8. *Let* $A, B, C \in M_n(\mathbb{R})$ *such that* $|A| = 2$, $|B| = -5$ *and* $|C| = 4$. *Since* $|B| \neq 0$ *and* $|C| \neq 0$, *we have that the matrices* B *and* C *are* *invertible and*

$$\begin{aligned} \left|C^{-1}A^TB^{-1}\right| &= \left|C^{-1}\right|\left|A^T\right|\left|B^{-1}\right| \\ &= |C|^{-1}\,|A|\,|B|^{-1} \\ &= \frac{1}{4}\times 2\times\left(-\frac{1}{5}\right) \\ &= -\frac{1}{10}\,. \end{aligned}$$

8.4 Elementary Operations and the Determinant

Another strategy for calculating the determinant of matrices of particularly high order is through elementary transformations on the rows. Knowing that each elementary transformation on the rows of a matrix affects its determinant in a certain way, one can, by calculating the determinant of row-equivalent matrices, calculate the determinant of a given matrix.

Let us now examine the effect each of the elementary transformations on the rows of a matrix has on its determinant.

Proposition 8.4.1. *Let $A, B \in M_n(\mathbb{R})$ and $\alpha, \beta \in \mathbb{R}$. The following hold:*

I) if $i \neq j$ and $A \xrightarrow[l_i \leftrightarrow l_j]{} B$, then $|A| = -|B|$;

II) if $\alpha \neq 0$ and $A \xrightarrow[l_i \to \alpha l_i]{} B$, then $|A| = \frac{1}{\alpha}|B|$;

III) if $i \neq j$ and $A \xrightarrow[l_i \to l_i + \beta l_j]{} B$, then $|A| = |B|$;

Proof. Exercise. □

Remark 8.4.2. *Note that only the elementary transformation of type III) does not affect the calculation of the determinant.*

How, then, can we calculate the determinant of a matrix A using elementary transformations on its rows? Consider the following procedure:

1. Perform elementary transformations on the rows of A until a matrix B in echelon form (i.e., upper triangular) is obtained;

2. Taking into account the corresponding changes each of these elementary transformations has on the determinant, obtain the relationship between the determinant of A and the determinant of B;

3. Since the determinant of B is easy to calculate (it is simply the product of the elements on its main diagonal), and the relationship between $|A|$ and $|B|$ is known, the value of $|A|$ can be determined.

Example 8.4.3.

$$\begin{aligned}
\begin{vmatrix} 0 & 5 & 10 \\ 1 & 2 & 3 \\ 2 & 6 & 8 \end{vmatrix} &\overset{(l_1 \leftrightarrow l_2)}{=} -\begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & 10 \\ 2 & 6 & 8 \end{vmatrix} \\
&\overset{(l_3 \to l_3 - 2l_1)}{=} -\begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & 10 \\ 0 & 2 & 2 \end{vmatrix} \\
&\overset{(l_2 \to \frac{1}{5} l_2)}{=} -5\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{vmatrix} \\
&\overset{(l_3 \to l_3 - 2l_2)}{=} -5\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{vmatrix} = -5 \times (-2) = 10\,.
\end{aligned}$$

Chapter 9

Systems of linear equations

9.1 Linear Equations and Systems of Equations

An equation of the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$, with $a_1, a_2, \ldots, a_n, b \in \mathbb{R}$, is called a **linear equation** in the variables $x_1, x_2, \ldots, x_n$ over $\mathbb{R}$. The constants $a_1, a_2, \ldots, a_n$ are called the **coefficients** of the equation, and the constant b is called the **constant term**. If $b = 0$, the linear equation is said to be **homogeneous**.

The vector $(\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$ is called a **solution of the linear equation** if substituting each x_i by the corresponding β_i, with $i = 1, \ldots, n$, yields a true statement.

Example 9.1.1. *Given the linear equation $2x + 3y = 5$ in the variables x and y, with coefficients 2 and 3, and constant term 5, it follows that:*

- *the vector $(1, 2)$ is not a solution of the equation, since $2 \times 1 + 3 \times 2 = 8 \neq 5$;*
- *the vector $(1, 1)$ is a solution of the equation, since $2 \times 1 + 3 \times 1 = 5$.*

Does every linear equation have a solution? And if it does have a solution, is it unique?

Remark 9.1.2. *It should be noted that a homogeneous linear equation always has at least one solution, which is the trivial solution, referred to as the* ***trivial solution****. However, this may not be the only solution. Consider the following example.*

Example 9.1.3. *Given the homogeneous linear equation $2x + 3y = 0$ in the variables x and y, with coefficients 2 and 3, and constant term 0, it follows that the vector $(0, 0)$ is a solution of the equation, since $2 \times 0 + 3 \times 0 = 0$. However, note that all vectors of the form $(3c, -2c)$, for any $c \in \mathbb{R}$, are solutions to the given homogeneous linear equation.*

If instead of just one linear equation, there is a finite set of linear equations, all in the same variables, this set is called a **system of linear equations**.

Let $m, n \in \mathbb{N}$ and consider the system (S) of m linear equations, in the variables $x_1, x_2, \ldots, x_n$ over $\mathbb{R}$,

$$(S)\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases},$$

with coefficients $a_{ij} \in \mathbb{R}$ and constant terms $b_i \in \mathbb{R}$, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. If the constant terms $b_1, \ldots, b_m$ are all zero, the system (S) is called a **homogeneous system of linear equations**.

The vector $(\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$ is a **solution of the system** if it is a solution to each of its m linear equations, i.e., if the following statement is true:

$$a_{11}\beta_1 + a_{12}\beta_2 + \ldots + a_{1n}\beta_n = b_1 \wedge \cdots \wedge a_{m1}\beta_1 + a_{m2}\beta_2 + \ldots + a_{mn}\beta_n = b_m ,$$

which can also be written as

$$\begin{cases} a_{11}\beta_1 + a_{12}\beta_2 + \ldots + a_{1n}\beta_n = b_1 \\ \ldots \\ a_{m1}\beta_1 + a_{m2}\beta_2 + \ldots + a_{mn}\beta_n = b_m \end{cases}.$$

Remark 9.1.4. *If the system (S) is homogeneous, then it will always have the trivial solution $(0, 0, \ldots, 0) \in \mathbb{R}^n$, since*

$$\begin{cases} a_{11}0 + a_{12}0 + \ldots + a_{1n}0 = 0 \\ \ldots \\ a_{m1}0 + a_{m2}0 + \ldots + a_{mn}0 = 0 \end{cases},$$

*which is a conjunction of true statements. The null solution $(0, 0, \ldots, 0) \in \mathbb{R}^n$ of a homogeneous system is called the **trivial solution**.*

9.2 Classification of Systems

In general, a system of linear equations may or may not have a solution. And in the case it has a solution, that solution may not be unique. Let us now look at how systems of linear equations are classified in terms of their solutions.

Definition 9.2.1. *A system (S) of linear equations is said to be:*

- ***inconsistent*** *if* (S) *has no solutions;*
- ***consistent*** *if* (S) *has at least one solution, being*
  - ***consistent with a unique solution*** *if that solution is unique;*
  - ***consistent with infinitely many solutions*** *if that solution is not unique, meaning that there are infinitely many solutions.*

Remark 9.2.2. *Every system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.*

Remark 9.2.3. *From what was observed above about the existence of solutions for homogeneous systems, it follows that a homogeneous system is always consistent, and it may be either witha unique solution or with infinitely solutions.*

Example 9.2.4. *The homogeneous system*

1. $\begin{cases} 2x+2y=0 \\ 5x+5y=0 \end{cases}$ *is consistent with infinitely many solutions, which are of the form* $(c,-c)$, *for any* $c \in \mathbb{R}$;
2. $\begin{cases} 2x+2y=0 \\ 2x+3y=0 \end{cases}$ *is consistent with a unique solution, the trivial solution* $(0,0)$.

When dealing with a system of linear equations, one may be interested in classifying the system in terms of the existence or non-existence of a solution, without necessarily calculating its solution set explicitly. Thus, given a system of linear equations, one can:

- **discuss the system**, which consists of classifying the system in terms of the existence or non-existence of a solution without needing to determine its solution set;
- **solve the system**, which consists of determining the set of its solutions.

In this context, what has been studied about matrices will be crucial. Let us consider the following definition and what follows.

Definition 9.2.5. *Given a system of linear equations* (S), *the* ***matrix form of the system*** (S) *is defined by the matrix equation* $AX=B$, *where*

$$\underbrace{A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{\substack{\text{matrix of}\\ \text{coefficients}}}, \quad \underbrace{X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\substack{\text{matrix of}\\ \text{variables}}} \quad \text{and} \quad \underbrace{B=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\substack{\text{matrix of}\\ \text{constant terms}}}.$$

*The matrix of coefficients of the system (S) is called the **coefficient matrix**, and the **augmented matrix** of the system (S) is defined by the matrix*

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right] \in M_{m\times(n+1)}(\mathbb{R}) \ ,$$

which is represented by $[A \mid B]$.

Example 9.2.6. *The system of linear equations in the variables x_1, x_2, x_3 over $\mathbb{R}$:*

1. $\begin{cases} x_1 + x_2 - x_3 = 0 \\ 2x_1 + x_2 = 1 \\ x_1 - x_3 = 1 \\ 3x_1 + x_2 - x_3 = 2 \end{cases}$ *can be written in matrix form as*

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & -1 \\ 3 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix},$$

and its augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & 1 & 0 & 1 \\ 1 & 0 & -1 & 1 \\ 3 & 1 & -1 & 2 \end{array}\right].$$

2. *with the augmented matrix*

$$\left[\begin{array}{ccc|c} 2 & -1 & 3 & 5 \\ -1 & 0 & 4 & -2 \end{array}\right],$$

is $\begin{cases} 2x_1 - x_2 + 3x_3 = 5 \\ -x_1 + 4x_3 = -2 \end{cases}.$

Definition 9.2.7. *Two systems of linear equations are said to be **equivalent** if they have the same set of solutions.*

Proposition 9.2.8. *Let $AX = B$ and $A'X = B'$ be systems of linear equations over $\mathbb{R}$. If the augmented matrices $[A \mid B]$ and $[A' \mid B']$ are equivalent by row* *operations, i.e.,*

$$[A \mid B] \xrightarrow[\substack{\text{elem. transf.}\\ \text{on the rows}}]{} \cdots \xrightarrow[\substack{\text{elem. transf.}\\ \text{on the rows}}]{} [A' \mid B'] \,,$$

then the systems $AX = B$ and $A'X = B'$ are equivalent.

Proof. See [3] or [8]. □

Example 9.2.9. *The systems* $\begin{cases} 2x + 3y + z = 3 \\ -x + y + 2z = 1 \\ -2x + 2y + 4z = 2 \end{cases}$ *and* $\begin{cases} 2x + 3y + z = 3 \\ y + z = 1 \end{cases}$ *are equivalent, because*

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ -1 & 1 & 2 & 1 \\ -2 & 2 & 4 & 2 \end{array}\right] \xrightarrow[l_3 \to l_3 - 2l_2]{} \left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ -1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$
$$\xrightarrow[l_2 \to l_2 + \frac{1}{2}l_1]{} \left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ 0 & \frac{5}{2} & \frac{5}{2} & \frac{5}{2} \\ 0 & 0 & 0 & 0 \end{array}\right]$$
$$\xrightarrow[l_2 \to \frac{2}{5}l_2]{} \left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

9.3 Matrix Rank and System Classification

Given the matrix representation of a given system of linear equations, let's see how, through the rank of the coefficient matrix and the rank of the augmented matrix, we can classify the system in terms of the existence or non-existence of a solution.

Proposition 9.3.1. *Let $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{m\times 1}(\mathbb{R})$. It follows that $r(A) \leq r\left([A \mid B]\right)$. More precisely,*

$$r\left([A \mid B]\right) = r(A) \quad \text{or} \quad r\left([A \mid B]\right) = r(A) + 1\,.$$

Proof. Exercise. □

Thus, relations can be established between the rank of the coefficient matrix and the rank of the augmented matrix of a given system of linear equations and its classification in terms of the existence or non-existence of a solution, as stated in the following theorem.

Theorem 9.3.2. *Let $AX = B$ be a system of linear equations, with $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{m\times 1}(\mathbb{R})$.*

- *if $r(A) < r([A \mid B])$, the system $AX = B$ is inconsistent;*
- *if $r(A) = r([A \mid B])$, the system $AX = B$ is consistent, and:*
  - *consistent with a unique solution if $r(A) = r([A \mid B]) = n$, where n is the number of variables of the system (or the number of columns of A);*
  - *consistent with infinitely many solutions if $r(A) = r([A \mid B]) < n$.*

Proof. See [3] or [8]. □

When dealing with a consistent system with infinitely many solutions, it means that when writing its solution set, one or more of its variables can be expressed in terms of another(s) of its variables. Consider the following definition.

Definition 9.3.3. *Let $AX = B$ be a consistent system of linear equations with infinitely many solutions, where $A \in M_{m\times n}(\mathbb{R})$. The number of free variables, given by $n - r(A)$, is called the* ***degree of freedom*** *of the system.*

9.4 Solving Systems

The process of solving systems of linear equations $AX = B$ is known as the *Gaussian elimination method*, which, in terms of matrices, involves obtaining a row-equivalent matrix in row echelon form from the augmented matrix $[A \mid B]$, and then solving the system by substitution, determining the solution to the equation corresponding to the last non-zero row, then moving to the penultimate one, and so on until reaching the first.

Another process that can also be used to solve systems of linear equations is an extension of the *Gaussian elimination method*, known as the *Gauss-Jordan elimination method*. This method consists of obtaining a row-equivalent matrix in row echelon form from the augmented matrix $[A \mid B]$, but with all pivots equal to 1 and all other elements in the columns of the pivots being zero, and then determining the solution set of the system using this matrix, without the need to perform any substitution. Consider the following example.

Example 9.4.1. *Consider the system of linear equations in the variables x_1, x_2, x_3, x_4 over $\mathbb{R}$,*

$$(S)\begin{cases} x_1 + 2x_2 + x_3 - 3x_4 = -5 \\ 2x_1 + 4x_2 + 4x_3 - 4x_4 = -6 \\ -x_1 - 2x_2 - 3x_3 - x_4 = 3 \end{cases}.$$

Discussion of the system (S) Considering the augmented matrix of system (S), we have

$$[A \mid B] = \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 2 & 4 & 4 & -4 & -6 \\ -1 & -2 & -3 & -1 & 3 \end{array}\right] \xrightarrow[\substack{l_2 \to l_2 - 2l_1 \\ l_3 \to l_3 + l_1}]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 0 & -2 & -4 & -2 \end{array}\right]$$
$$\xrightarrow[l_3 \to l_3 + l_2]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 0 & 0 & -2 & 2 \end{array}\right],$$

where $r([A \mid B]) = r(A) = 3 < 4 = n$ (number of variables). Thus, we conclude that the system (S) is consistent with infinitely many solutions, and degree of freedom $n - r(A) = 4 - 3 = 1$.

Solving the system (S) Considering the augmented matrix of system (S), we have

$$[A \mid B] = \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 2 & 4 & 4 & -4 & -6 \\ -1 & -2 & -3 & -1 & 3 \end{array}\right] \xrightarrow[\substack{l_2 \to l_2 - 2l_1 \\ l_3 \to l_3 + l_1}]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 0 & -2 & -4 & -2 \end{array}\right]$$
$$\xrightarrow[l_3 \to l_3 + l_2]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 0 & 0 & -2 & 2 \end{array}\right]$$
$$\xrightarrow[\substack{l_2 \to \frac{1}{2}l_2 \\ l_3 \to -\frac{1}{2}l_3}]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & -3 & -5 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right]$$
$$\xrightarrow[\substack{l_2 \to l_2 - l_3 \\ l_1 \to l_1 + 3l_3}]{} \left[\begin{array}{rrrr|r} 1 & 2 & 1 & 0 & -8 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right]$$
$$\xrightarrow[l_1 \to l_1 - l_2]{} \left[\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & -11 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right],$$

from which we obtain

$$\begin{cases} x_1 + 2x_2 & = -11 \\ x_3 & = 3 \\ x_4 & = -1 \end{cases},$$

with $x_2 \in \mathbb{R}$ *(free variable). That is, the solution set of the system* (S) *is given by*

$$\begin{aligned} C.S. &= \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 = -11 - 2x_2 \wedge x_3 = 3 \wedge x_4 = -1\} \\ &= \{(-11 - 2x_2, x_2, 3, -1) : x_2 \in \mathbb{R}\}. \end{aligned}$$

9.5 Cramer's Systems

Consider a system of linear equations (S) whose matrix form is $AX = B$.

Definition 9.5.1. *If the matrix* A *is square and invertible, the system* (S) *is called a* ***Cramer's system****.*

Exercise 9.5.2. *Consider a Cramer's system* (S)*, whose matrix form is* $AX = B$.

a) Show that (S) *is a consistent system with a unique solution.*

b) Also, show that the solution to the system (S) *is given by* $X = A^{-1}B$.

Exercise 9.5.3. *Consider the system* (S) $\begin{cases} x + y & = 1 \\ y + z & = 2 \\ x + y + z & = 0 \end{cases}.$

a) Show that (S) *is a Cramer's system.*

b) Calculate the matrix A^{-1} *and show that the unique solution to* (S) *is* $(-2, 3, -1)$.

9.5.1 Cramer's Rule

Given a Cramer's system, there is another way to find its unique solution using determinants. An important feature of this strategy, hereafter referred to as **Cramer's Rule** , is that it allows finding each component of the solution of the system independently of whether or not the other components are calculated.

Proposition 9.5.4 (**Cramer's Rule**). *Let $AX = B$ be the matrix form of a Cramer's system, with $A \in M_n(\mathbb{R})$. For each $j \in \{1, \ldots, n\}$, let $A(j)$ be the matrix obtained by replacing the j-th column of A with the matrix B of the independent terms. Then, the (unique) solution to the system $AX = B$ is the vector in $\mathbb{R}^n$ given by*

$$\left(\frac{|A(1)|}{|A|}, \frac{|A(2)|}{|A|}, \ldots, \frac{|A(n)|}{|A|} \right) .$$

Proof. See [3] or [8]. □

Example 9.5.5. *Returning to the previous exercise (Exercise 9.5.3), note that by applying Cramer's Rule one can compute each component of its solution independently. Now, applying Proposition 9.5.4, we have*

$$x = \frac{|A(1)|}{|A|} = \frac{\begin{vmatrix} \mathbf{1} & 1 & 0 \\ \mathbf{2} & 1 & 1 \\ \mathbf{0} & 1 & 1 \end{vmatrix}}{1} = -2 ,$$

$$y = \frac{|A(2)|}{|A|} = \frac{\begin{vmatrix} 1 & \mathbf{1} & 0 \\ 0 & \mathbf{2} & 1 \\ 1 & \mathbf{0} & 1 \end{vmatrix}}{1} = 3 ,$$

and

$$z = \frac{|A(3)|}{|A|} = \frac{\begin{vmatrix} 1 & 1 & \mathbf{1} \\ 0 & 1 & \mathbf{2} \\ 1 & 1 & \mathbf{0} \end{vmatrix}}{1} = -1 .$$

Bibliography

[1] Tom M. Apostol, *Cálculo*, (Vol.1. Português), Editorial Reverté, 1985.

[2] Tom M Apostol, *Calculus, volume 1*, John Wiley & Sons, 1991.

[3] I. Cabral, C. Perdigão, and C. Saiago, *Álgebra linear: teoria, exercícios resolvidos e exercícios propostos com soluções*, Escolar Editora, 3ª ed., 2012.

[4] Filipa de Carvalho and Telmo Peixe, *Notas de matemática I para as licenciaturas em Economia, Finanças e Gestão*, ISEG, 2017.

[5] ______, *Notas de matemática I para as licenciaturas em Economia, Finanças e Gestão*, ISEG, 2023.

[6] Jaime Campos Ferreira, *Introdução à análise matemática*, Fundação Calouste Gulbenkian, 2008.

[7] Maria do Rosário Grossinho, *Lições de matemática - notas teóricas*, ISEG, 2013.

[8] Serge Lang, *Linear algebra*, Springer Science & Business Media, 1987.

[9] ______, *A first course in calculus*, Springer Science & Business Media, 2012.

[10] Carlos Sarrico, *Análise matemática leituras e exercícios*, Gradiva, 2013.

[11] K. Sydsæter and P.J. Hammond, *Essential mathematics for economic analysis*, FT Prentice Hall, 2008.

Index